Mathematical Methods in Economics 2011

29th International Conference

Mathematical Methods in Economics 2011

Proceedings – part I

We celebrate 20th Anniversary of the Czech and Slovak Operational Research Societies

September 6 – 9, 2011

Janská Dolina, Slovakia

The conference is organised by:

Faculty of Informatics and Statistics, University of Economics, Prague

Faculty of Economic Informatics, University of Economics in Bratislava

Czech Society of Operational Research

Slovak Society of Operational Research

Czech Econometric Society

Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 – part I
Publisher: Professional Publishing
Mikulova 1572/13, 149 00 Praha 4, Czech Republic
Editors: Martin Dlouhý, Veronika Skočdopolová
Pages: 791

ISBN 978-80-7431-058-4

The publication has not passed the language correction.

Programme Committee

Helena Brožová
Faculty of Economics and Management, University of Life Science, Prague

Michal Fendek
Faculty of Economic Informatics, University of Economics in Bratislava

Petr Fiala
Faculty of Informatics and Statistics, University of Economics, Prague

Jana Hančlová
Faculty of Economics, Technical University of Ostrava

Zlatica Ivaničová
Faculty of Economic Informatics, University of Economics in Bratislava

Josef Jablonský
Faculty of Informatics and Statistics, University of Economics, Prague

Ladislav Lukáš
Faculty of Economics, University of West Bohemia, Pilsen

Vladimír Mlynarovič
Faculty of Social and Economic Sciences, Comenius University in Bratislava

Jan Pelikán
Faculty of Informatics and Statistics, University of Economics, Prague

Karel Sladký
Institute of Information Theory and Automation of the Academy of Sciences of CR, Prague

Tomáš Šubrt
Faculty of Economics and Management, University of Life Science, Prague

Milorad Vidović
Faculty of Transport and Traffic Engineering, Uiversity of Belgrade, Serbia

Karel Zimmermann
Faculty of Mathematics and Physics Charles University, Prague

Organization Committee

Martin Dlouhý, Jan Fábry Jana Kalčevová, Martina Kuncová

Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic

Ivan Brezina, Juraj Pekár, Kvetoslava Surmanová, Marian Reiff

Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia

Contents – Part I

Stochastic programming software: a comparison for investment problem

Lukáš Adam[1]

Abstract. Many practical problems involve solving stochastic optimization problems. As the number of scenarios increases, even in the linear case the necessity for specialized algorithms and software arises. This is especially true for the polyhedral case, among which minimization of several risk measures belongs, namely conditional value at risk, mean absolute deviation and others.

In this paper we deal with investment problem minimizing conditional value at risk. This model is solved using several algorithms, namely deterministic equivalent, L–shaped algorithm, both in basic and multicut versions, furthermore we present hot start method. Regarding software, we present the comparison of specialized two stage programming software SLP–IOR with general purpose optimization software GAMS. For large scale problems, the best performance is obtained with the L-shaped method solved in C# using GAMS solver CPLEX.

Keywords: stochastic programming, algorithms, software, CVaR.

1 Introduction

Linear two stage programming is one of the simplest examples of stochastic programming. These problems are well documented in many publications, for example in [2] or [3]. In this paper we provide a practical financial optimization problem. This problem is only one stage problem, but thanks to conditional value at risk in its objective function, it can be laid out into two stage problem.

Conditional value at risk is the expected loss given that loss is greater than or equal to value at risk, which is α–quantile of the loss distribution for given confidence level α. CVaR was devoloped in [6] and its minimizing can be done by solving

$$\begin{aligned} &\min_{x,z}\ c'x + z + \frac{1}{1-\alpha}\,\mathrm{E}[q(\xi)'x - z]^+ \\ &s.t.\ x \in X, z \in \mathbb{R}, \end{aligned} \tag{1}$$

where c is given vector, x are portfolio weights, $L(x,\xi) = q(\xi)'x$ is random loss function, X is polyhedral set and $\alpha \in (0,1)$ is prescribed significance level. In the rest of this paper we will suppose that $q(\xi)$ can attain only finite number of realizations, denote them $q_1, \dots, q_K$ with the corresponding probabilities $p_1, \dots, p_K$. Problem (1) is not linear, but can be linearized by adding slack variables w_k. The problem is

$$\begin{aligned} \min_{x,z,w_k}\ & c'x + z + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k w_k \\ s.t.\ & x \in X, z \in \mathbb{R}, \\ & w_k \geq q_k'x - z,\ k = 1, \dots, K, \\ & w_k \geq 0,\ k = 1, \dots, K. \end{aligned} \tag{2}$$

Because of the expected value in the objective function, problem (1) can be rewritten as a two stage problem with first stage

$$\begin{aligned} &\min_{x,z}\ c'x + z + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k \mathcal{Q}(x,z,\xi_k) \\ &s.t.\ x \in X, z \in \mathbb{R}, \end{aligned} \tag{3}$$

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 18675, Prague 8, Czech republic, e-mail: lukas.adam@yahoo.com.

where $\mathcal{Q}(x, z, \xi_k)$ is optimal value of the second stage problem

$$\begin{aligned} \min_{w_k} \ & w_k \\ s.t. \ & w_k \geq q_k' x - z, \\ & w_k \geq 0. \end{aligned} \tag{4}$$

Problem (2) is called deterministic equivalent and problems (3)–(4) two stage problem. Because of the special structure of the two stage problem, specialized algorithms can be used. The best known is L–shaped algorithm, whose short description will be provided. Because $\phi(x, z) = \sum_{k=1}^{K} p_k \mathcal{Q}(x, z, \xi)$ is polyhedral function, it is equal to maximum of finite number of linear functions, denote them $b_l - c_l' x - d_l z$ for some scalars b_l, d_l, vectors c_l and $l = 1, \ldots, L$. In every iteration at least one of these functions is constructed, these functions are called cuts.

Problem (3) can be rewritten as

$$\begin{aligned} \min_{x,z,\theta} \ & c'x + z + \frac{1}{1-\alpha}\theta \\ s.t. \ & x \in X, z \in \mathbb{R}, \\ & \theta \geq b_l - c_l' x - d_l z, \ l = 1, \ldots, L. \end{aligned} \tag{5}$$

Parameters in cuts are based on the solution of the dual problem to (4):

$$\begin{aligned} \max_{\pi_k} \ & \pi_k (q_k' x - z) \\ s.t. \ & 0 \leq \pi_k \leq 1, \end{aligned} \tag{6}$$

hence the optimal solution is either $\pi_k = 0$ or $\pi_k = 1$.

In L–shaped algorithm in every iteration we construct candidates for optimal solution (x^l, z^l) and if they are not optimal, we add a cut. If we want to find supporting hyperplane to $\phi(x, z)$ at point (x^l, z^l) it can be shown that $b_l = 0$, $c_l = \sum_{k=1}^{K} p_k \pi_k^l (-q_k)$ and $d_l = \sum_{k=1}^{K} p_k \pi_k^l$, where π_k^l are optimal solutions to (6) if we set $x = x^l$ and $z = z^l$. Because either $\pi_k^l = 0$ or $\pi_k^l = 1$, we can write $c_l = -\sum_{k \in \mathcal{K}^l} p_k q_k$, $d_l = \sum_{k \in \mathcal{K}^l} p_k$, where

$$\mathcal{K}_l = \{k; \ \pi_k^l = 1\} = \{k; \ q_k' x^l - z^l > 0\}. \tag{7}$$

Because of this, it is not necessary to solve a subproblem for every k, but it is sufficient to find the sign of $q_k' x^l - z$.

By master problem we will understand

$$\begin{aligned} \min_{x,z,\theta} \ & c'x + z + \frac{1}{1-\alpha}\theta \\ s.t. \ & x \in X, z \in \mathbb{R}, \\ & \theta \geq x' \sum_{k \in \mathcal{K}^l} p_k q_k - z \sum_{k \in \mathcal{K}^l} p_k, \ l = 1, \ldots, \nu, \end{aligned} \tag{8}$$

where ν is the number of current iteration. It is a modification of problem (5), but among constraints there are only already generated cuts.

L–shaped method was described for the first time in [7] and its application to (3)–(4) is simple. Different approach with the same result can be found in [5], this approach is more concentrated on properties of sets $\mathcal{K}^l$. Now we will summarize the whole algorithm.

1. Set $\nu = 0$, $\mathcal{K}^0 = \{1, \ldots, K\}$ and to master problem (8) add cut $\theta \geq x' \sum_{k=1}^{K} p_k q_k - z$.
2. Increase ν by one, solve (8) and denote its optimal solution $(x^\nu, z^\nu, \theta^\nu)$.
3. According to (7) compute set $\mathcal{K}^\nu$ and

   $$\omega^\nu = (x^\nu)' \sum_{k \in \mathcal{K}^\nu} p_k q_k - z^\nu \sum_{k \in \mathcal{K}^\nu} p_k.$$

   If $\omega^\nu \leq \theta^\nu$, stop the algorithm, (x^ν, z^ν) is optimal solution to (2). If preceding inequality is not fulfilled, add cut to master problem and return to the first step.

One important question arises: which algorithm should be used and in which situation. Firstly we will compare algorithms from theoretical point of view. In deterministic equivalent we solve one big problem. When using L–shaped algorithm, this problem is split into one master problem and K subproblems. In every iteration all of these problems are solved. In linear programming, mainly the number of constraints, not variables is important. This is because it is necessary to invert base matrix, whose dimension is equal to number of constraints.

That implies that L–shaped method should be able to solve larger problem than deterministic equivalent. Furthermore, for larger problem it should provide results faster. On the other hand, we have to hope that the cuts will be chosen properly. For example when solving master problem (8) the number of added constraints can be as high as 2^K, which is an immense number even for small K. As we will see later, the algorithm is fortunately well behaved and the size of master problem is kept small. However it is not difficult to find an example, in which it is necessary to add all possible cuts. In this case L–shaped algorithm would be more time consuming than solving the deterministic equivalent.

We will meet one of these examples later in this paper, let us shortly mention it. There are many variants of L–shaped algorithm, one of them is multicut version. In base version we estimate from below $\sum_{k=1}^{K} p_k \mathcal{Q}(x, z, \xi_k)$. Only one parameter θ is needed. In multicut version we need K parameters $\theta_1, \ldots, \theta_K$ and all the functions $p_k \mathcal{Q}(x, z, \xi_k)$ are estimated separately. The idea of this variant is that in every iteration more information about $\phi(x, z)$ is gathered, so less iterations will be needed. On the other hand, the master problem will be larger.

Because (6) may have only two optimal solutions for any (x, z), for any fixed k the function $\mathcal{Q}(x, z, \xi_k)$ is piecewise linear with only two pieces, on one being equal to zero. Thus multicut version will be virtually identical to solving deterministic equivalent. This would not probably happen for any problem in which second stages are more complex.

In next section we will compare several types of software. This is of course not the first attempt to compare software or solvers, good descriptions of different software can be found for example in [4] or [8]. However, in most comparisons only small problems were taken into account. In my opinion, this is not the correct way to address the problem. For small problems if the difference should become significant, lots of problems would have to be queued. For example in SLP–IOR it is not possible to write source code, thus it is not possible to automatically queue several problems.

To test software problem described in [1] was used. We solved problem (1) with q discrete and depending on seven independent random interest and exchange rates. Together there were $K \in \{128, 2187, 16384, 78125, 311040\}$ scenarios. Loss function with opposite sign indicated profitability and each instrument depended only on two to four from seven random variables. The polyhedral set X depended on parameter $h \in [0, 1]$ in such a way that

$$h_1 < h_2 \Rightarrow X(h_1) \supset X(h_2). \tag{9}$$

2 SLP–IOR

First of the compared software is SLP–IOR, which is specialized on linear stochastic programming. Unfortunately it is not able to solve any nonlinear problems except for two types of problems which can be linearized. First complication appears when loading problem, SLP–IOR does not support user's programming, so the problem must be loaded in SMPS format, which must be generated first. Doing so caused several difficulties, mainly the length of generation and the size of generated file. Concrete numbers can be seen in table 1.

K	Time [s]	Size [MB]	Loading time [s]
128	0.101	0.763	2
2187	1.543	14.821	2
16384	12.368	114.874	55
78125	56.759	553.945	4:30
311040	4:06.130	2214.273	?

Table 1: Time of generation, size of file and loading time of SMPS format

In the first column we can see the number of scenarios, the second one refers to the time in seconds which was taken by generating the SMPS file. The largest problem needed four minutes only to generate. In the Size column, there is the size of generated files, the largest file was bigger than 2GB.

The problem is the core design of SMPS format. Besides other things we have to specify all the random variables for every scenario. As we have already said, generally profitability may depend on seven indexes, but one particular profitability of given financial instrument will depend only on four or less of them. So in programmable environment, it is possible to create and store several smaller tables and use them on the fly during the generation of problem, whilst when using SMPS, all this information has to be computed and saved ahead. In SMPS format there will be plenty of recurrent information, every value will be repeated at least hundred times on different places in the file.

After generating, the file was loaded into SLP–IOR, the times needed for SLP–IOR to load the file can be seen again in table 1 in the last column. The times are not precise, because it was possible to measure them only by stopwatch. The interrogation mark means that the times differed significantly, the average time was around twenty minutes, but in one case the SLP–IOR froze and had to be terminated.

Furthermore, SLP–IOR managed to solve the problem only by generating and solving deterministic equivalent. But for this way we do not need specialized software, we will discuss it later on. For the record, QDECOM, solver using quadratic decomposition, which is based on the multicut version of L–shaped algorithm, was not able to find any feasible solution.

3 GAMS

In the rest of this paper we will consider another optimization software GAMS, which is general purpose software. That means that we cannot use decomposition algorithms directly. One possibility is solving deterministic equivalent, which becomes progressively larger and for large K GAMS is not capable of solving the problem. Precise times can be seen in table 2 in two columns corresponding to header CPLEX, which is GAMS core solver for linear programming. The first column denotes the number of iterations, which were carried out by GAMS. The second number represents the total time in which the problem was solved.

Another possibility is to implement L–shaped algorithm described earlier in this paper. The results are stated in two L–shaped columns. The first one is again the time needed to solve the problem and the other one the number of iterations, which has a slightly different meaning than in previous case. Here it is the number of iterations in L–shaped algorithm, not the number of iterations performed by a solver. It may be surprising that for smaller problems the deterministic equivalent provides better solution times. This is probably due to the quality of CPLEX.

It is important to realize that the size of master problem (8) does not depend on number of scenarios K, but only on number of iterations ν. Because the number of iteration is kept small even for large K, the master problem will be small as well. This means that the master problem is solved in negligible time, while the time consuming part is the generation and summation over $\mathcal{K}^\nu$, whose dimension can be as great as K. Because GAMS is not good at working with large arrays, it turned out that it is better to work with some object oriented language. I chose C#, in which the majority of optimization is written, mainly the computation of necessary variables and values, generation of source code for GAMS and searching through $\mathcal{K}^\nu$. GAMS is only called to solve the master problem.

The results are in table 2 in seventh and eighth columns. For smaller problems the solution times are not as good, this is caused by the communication between C# and GAMS, this time is again independent of K. For greater K this method is superior. The number of iterations is clearly identical to solving in GAMS, because the algorithms are the same, only the realizations differ.

Another improvement can be achieved by using cuts generated in previous solving. Indeed, for $h \in \{0, 0.1, \ldots, 0.5\}$ denote the corresponding feasible sets $X(h)$. Because h influences only the constraints, the objective function remains the same. This means that the cuts generated in problem with h_1 remains lower estimation of $\phi(x, z)$ for problem with $h_2 > h_1$. Because the solution time is not so influenced by the size of master problems (and thus by the number of cuts), but mainly by the number of iterations, this shows to be a good way to save some time when solving nested problems.

On the other hand, if we had long sequence of $h_1 < \cdots < h_N$, it could happen that we will store huge

h	K	CPLEX		L–shaped		C#		Hot start	
0	128	0.407	7	1.727	10	2.852	10	2.902	10
0.1	128	0.356	11	2.291	14	4.319	14	2.929	10
0.2	128	0.349	11	2.205	14	4.344	14	0.334	2
0.3	128	0.355	11	2.152	14	4.189	14	0.657	3
0.4	128	0.414	11	2.186	14	4.131	14	0.339	2
0.5	128	0.372	11	2.014	13	3.950	13	1.003	4
0	2187	1.540	358	18.138	45	18.571	45	19.117	45
0.1	2187	1.600	373	14.608	37	15.472	37	11.890	26
0.2	2187	1.742	435	19.338	49	20.391	49	16.125	34
0.3	2187	1.640	470	19.593	48	19.820	48	14.341	28
0.4	2187	1.841	464	15.866	42	17.440	42	13.491	26
0.5	2187	1.850	473	16.384	42	17.369	42	12.693	23
0	16384	17.667	2663	1:14.060	42	38.611	42	36.707	42
0.1	16384	26.101	3494	1:50.894	64	56.817	64	55.032	54
0.2	16384	19.666	2984	2:03.248	67	1:01.387	67	54.135	55
0.3	16384	30.704	3546	1:45.881	58	52.483	58	48.682	47
0.4	16384	28.848	3390	1:06.133	36	31.965	36	24.709	23
0.5	16384	35.904	3836	1:24.989	47	43.168	47	34.178	31
0	78125	4:32.370	13948	6:25.737	52	2:34.848	52	2:29.557	52
0.1	78125	5:41.063	15382	11:42.975	93	4:39.386	93	4:44.238	94
0.2	78125	5:30.011	14932	6:25.400	53	3:01.012	53	2:33.667	45
0.3	78125	8:12.615	17788	6:59.819	55	2:44.145	55	2:14.903	42
0.4	78125	5:33.441	16218	7:14.018	60	3:01.343	60	2:32.247	45
0.5	78125	8:24.959	18247	6:50.092	58	2:57.148	58	2:26.726	43
0	311040	x	x	57:10.192	111	22:42.622	111	25:26.102	111
0.1	311040	x	x	54:36.342	111	23:06.111	111	8:43.769	38
0.2	311040	x	x	32:24.348	67	13:36.060	67	12:28.154	55
0.3	311040	x	x	31:01.003	62	13:34.051	62	12:44.721	55
0.4	311040	x	x	31:33.420	63	12:41.594	63	11:34.290	50
0.5	311040	x	x	37:07.353	64	13:09.851	64	11:44.770	50

Table 2: Software comparison

number of redundant cuts and solving master problem would take too long. Fortunately, this is not the case for $N = 6$ because no more than 100 cuts were generated in majority of problems, so the master problem was be kept small. For larger N a technique deciding which cuts to store and which to throw away would have to be developed. Because in every solving we do not start from scratch, but with some cuts obtained earlier, this method is called hot start.

Numerical results are again in table 2 in the last two columns. Solution time was slightly improved. There is another method: to consider decreasing sequence of $h_1 > \cdots > h_N$ and apply the hot start method again, this time the previously generated cuts will not be only lower bounds, but also supporting hyperplanes for $\phi(x, z)$. Despite this, the solution time remains approximately the same.

In the end we return to the multicut version. We have already said that in some situations it may provide similar results to the deterministic equivalent. This is the case of our problem. Because CVaR denotes the expected loss in $(1 - \alpha)100\%$ cases, $|\mathcal{K}^\nu| \geq (1 - \alpha)K$ must hold true. This means that at least $(1 - \alpha)K$ cuts must be added to master problem. Unfortunately, this is only a theoretical bound, in reality it will be much greater.

As we can see in table 3, in most cases only two iterations were necessary to find optimal solution, in some cases it was necessary to complete three iterations. These cases are marked with / in $\mathcal{K}^2$ column.

h	K	Time [s]	$\lvert\mathcal{K}^2\rvert$	h	K	Time [t]	$\lvert\mathcal{K}^2\rvert$
0	128	0.441	64	0	16384	10.184	4236/4780
0.1	128	0.407	88	0.1	16384	11.594	4499/4503
0.2	128	0.367	96	0.2	16384	12.677	4800/4805
0.3	128	0.370	98	0.3	16384	7.519	5147
0.4	128	0.373	100	0.4	16384	9.994	5462
0.5	128	0.361	128	0.5	16384	10.611	5595
0	2187	1.884	810/856	0	78125	2:29.663	24235/25847
0.1	2187	1.084	819	0.1	78125	2:26.571	24818/25000
0.2	2187	1.242	833	0.2	78125	3:00.984	25443/25446
0.3	2187	1.258	832	0.3	78125	1:49.108	26075
0.4	2187	1.307	843	0.4	78125	1:55.736	26800
0.5	2187	1.246	838	0.5	78125	2:40.379	26922

Table 3: Multicut version

The theoretical sufficient size of $|\mathcal{K}^\nu| = \frac{1}{20}K$ is not achieved, the actual size corresponds approximately to $\frac{1}{3}K$. Because of this, the multicut version was not able to optimize the largest problems. On the other hand, in small problems, the solution times are on a par with the best tested methods.

4 Conclusion

In this paper we compared several types of software and algorithms, where the emphasis was put mainly on GAMS. In my opinion the best idea is to use deterministic equivalent because of its simplicity. If deterministic equivalent is too large to be solved, some decomposition algorithm has to be used. Concerning whether to use only GAMS or connect GAMS with any object oriented language, I would recommend relying only on GAMS for simple problems. When the problem becomes complex, for example with necessity of code repetition, object oriented languages should be used.

References

[1] Adam, L., Lendel, G., and Sůva, P: Krizové scénáře vývoje ekonomiky. Seminar thesis MFF UK, 2011.

[2] Birge, J. R., and Louveaux, F.: *Stochastic linear programming. Models, theory, and computation.* Springer Series in Operations Research. New York, NY: Springer, 1997.

[3] Kall, P., and Mayer, J.: *Stochastic linear programming. Models, theory, and computation. 2nd ed..* International Series in Operations Research & Management Science. New York, NY: Springer, 2010.

[4] Kopa, M., ed.: *On selected software for stochastic programming.* Matfyzpress, Praha, 2008.

[5] Künzi-Bay, A., and Mayer, J.: Computational aspects of minimizing conditional value-at-risk, *Computational Management Science* **3:1** (2006), 3–27.

[6] Rockafellar, R. T., and Uryasev, S.: Optimization of Conditional Value-at-Risk, *Journal of Risk* **2:3**, (2000), 1–26.

[7] Van Slyke, R. M., and Wets, R.: L–shaped linear programs with applications to optimal control and stochastic programming, *SIAM Journal on Applied Mathematics* **17** (1969), 638–663.

[8] Wallace, S. W., and Ziemba, W. T., eds.: *Applications of stochastic programming.* MPS/SIAM Series on Optimization 5. Philadelphia, 2005.

Perspective Mathematical Method for Employees' and Organizational Units' Comparison on Projects

Vítězslav Antoš[1], Roman Kvasnička[2]

Abstract. The aim of this paper is to introduce a new approach to be used for comparison of employees or organizational units on different hierarchical levels of project organizational structure. Compliance to the project organizational structure is covered by use of AHP (Analytical Hierarchy Process). Within the AHP, the comparison is performed in order to derive reward coefficients for every organizational unit (on higher levels of the hierarchy) and for every specific role/employee (on lower levels of the hierarchy). These reward coefficients are used within the rewarding system for calculation of the personal rewards, based on the overall profit of every specific project. The system is being developed by the authors of this contribution and it is called CRAMS (Clear Rewarding and Motivating System). The first part provides a brief description of the system and issues related to the original approach. The second part contains the description of methods used within the new approach. The third part of this paper describes the new approach to rewards coefficients setup itself. The last part of this article contains a brief conclusion.

Keywords: rewarding, project, AHP, score method, WBS, OBS, responsibility matrix.

JEL Classification: C44, M12
AMS Classification: 90B50

1 Introduction – The purpose of comparisons

The purpose of employees' and organizational units' (OUs) comparison is to set up rewarding coefficients (RCs) for every company's OU, for every person involved in a project. These coefficients are used for the calculation of personal reward for specific employee within the financial rewarding system for project oriented companies, called **CRAMS** (Clear Rewarding and Motivating System). The system is being developed by the authors of this article. [1]

1.1 Brief description of the system

The purpose of the system depicted bellow (see Fig. 1) is "Continuous stimulation of employees involved in the project leading (and execution) to a correct decision making and maximization of the company's profit."[1].

The every single project can be seen as an analogy to *self-governing workshops* as they were used by Tomas Bata in his managerial system. [2]

To commence the stimulation effect on employees, they are acquainted with the *value of reward* that the every specific employee can expect (and will gain) if the project is performed *exactly as planned*, with reaching *exactly the planned costs and selling price* or income of the project.

This sets a *goal* for involved employees and it also clearly sets the *personal reward* linked to reaching the goal. The element of *interpersonal motivation* is brought in by having one overall goal linked to rewards, which is "Profit of the project equal to (defined value)", instead of specific SMART goals for everyone. This also initiates *creativity* and *pro-activity* of involved employees, because the higher profit of the project means higher reward, bigger benefit, for everyone, including the company itself.

To keep the stimulation effect, these rewards should respect the individual's *contribution* to the result of a project, or the individual *influence* on the result and they must be derived from the real project profit. The Saaty's [3] Analytic Hierarchy Process (AHP) and some comparison methods within the AHP are used to set up the individual rewards coefficients - shares on the profit of project actually.

[1] Czech University of Life Sciences Prague/FEM, Department of Systems Engineering, Kamýcká 129, antosv@pef.czu.cz.
[2] Czech University of Life Sciences Prague/FEM, Department of Systems Engineering, Kamýcká 129, kvasnicka@pef.czu.cz.

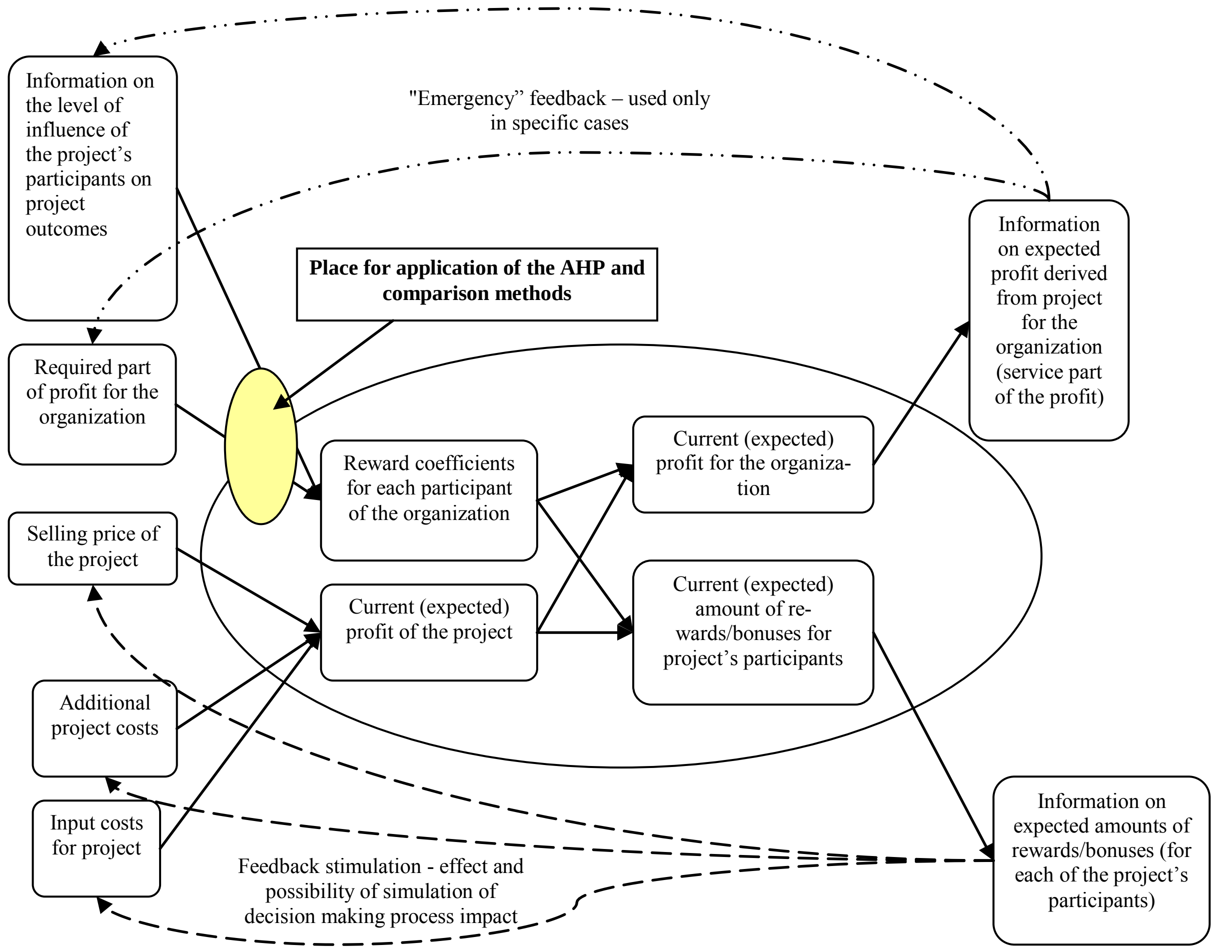


Figure 1 System diagram [1]

1.2 AHP application

The goal of the AHP application is to set reward coefficients, to distribute rewards in accordance to expected contribution of involved OUs and individual employees. The AHP is applied in order to follow the hierarchy of every project – its project organizational structure (POS). Compared to POS, there is the only difference in the hierarchy of AHP – on the top level there is the profit of the project itself. It is distributed into two parts on the second level of the hierarchy: **rewards** and **profit for the company**. Then the branch for rewards follows POS: the reward coefficient for project manager (or project management team), reward coefficients for involved OUs and on the lowest levels, RCs for every individual employee involved in the project. On every level of POS (and related AHP levels) the superior is responsible for definition of *preference information* for subordinated OUs or individuals. The rewards coefficient on the higher level is sum of RCs on the lower level of the same branch of hierarchy. The preference information expresses the expected contribution of specific OU or employee, compared to the expected contribution of the rest of involved subordinates. See Fig. 2 for example. [1]

Different levels of the hierarchy require different approaches and methods of comparison. On the second level, the decision maker is top management, or its representative. He has to consider what part of profit can be used to cover rewards. This level actually does not need any method for comparison. The decision should be based on business and financial needs, owner's expectations, the total amount of employees involved and other relevant factors, which can vary company from company. On the third level of hierarchy, there must be set rewards coefficients for project manager as a person responsible for the complex project and for the rest of project team. This should be also performed, or approved at least, by representative of top management. On the fourth level, the decision maker is the project manager. He should use appropriate method to set up reward coefficients according to expected contribution and influence of every involved OU (OU can be department, business unit, work package team). The similar situation is on the levels bellow, where decision makers are managers of departments, business units, work package managers or team leaders and objects of decision making are some other OUs, or more often individual subordinates.

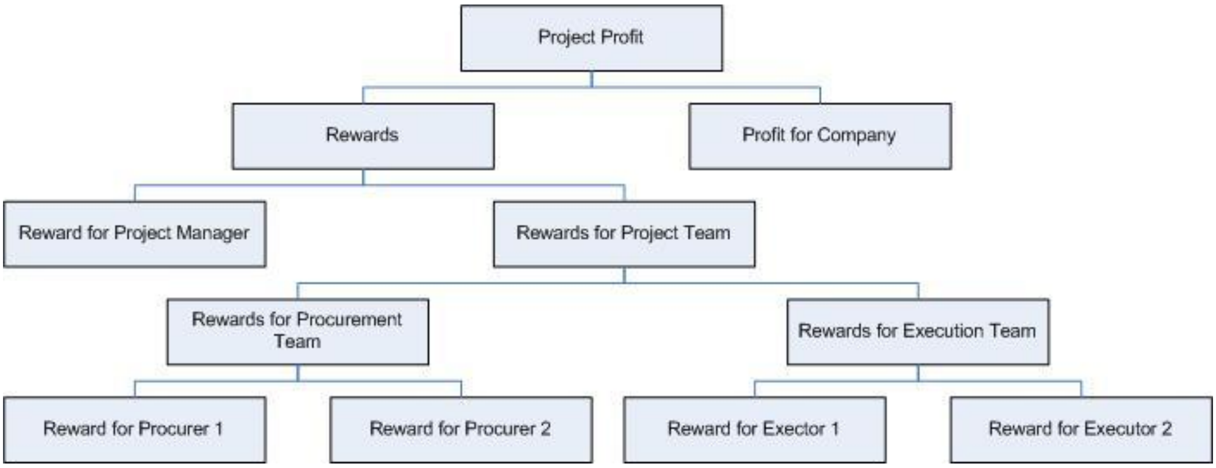


Figure 2 Example project profit distribution hierarchy [4]

Authors originally suggested using the pairwise comparison method on every level of the hierarchy [1]. It has allowed calculation of reward coefficients in compliance with the hierarchy of AHP, but there were also identified two major issues of this method's application:

1. The consistency of the pairwise comparison matrix for more than five compared objects [4].
2. High subjectivity of preference information inserted into the process.

These two points led authors to a deeper research of a different, more acceptable way of the reward coefficients setup.

2 Material and Method

In the new approach to reward coefficients setup, there are used following methods and terms: **WBS**, **OBS**, **responsibility matrix**, **AHP**, **score method**. There can be also used the **team Delphi method** as a supportive method for group definitions of score for WBS parts and roles.

2.1 WBS, OBS and responsibility matrix

- **WBS** (*Works Breakdown Structure*) is product-oriented hierarchical breakdown of the project objective. It is performed by decomposition of the project objective into work packages. On the top level of this hierarchy, there is the project objective, on the lowest level there are work packages, or partial activities if this level of WBS is necessary.
- **OBS** (*Organization Breakdown Structure*) is temporary hierarchical organizational structure of the project, consisting of elements (people) who are in charge of managing and performing project works and responsible for deliveries.
- **Responsibility matrix** is clear and specific definition of competencies and responsibilities of project team members related to defined project activities and work packages. It is the combination of WBS and OBS. [5]

2.2 AHP

Thomas L. Saaty, the author of the AHP [3], defines hierarchy as "a powerful way of classification used by the mind to order information gained from experience or from your own thinking to understand the complexity of the world around us according to the order and distribution of influences that make certain outcomes happen". It can be used for many assignments, e.g. for multi criteria analysis, for human resource allocation [6] etc.

In the system the AHP is used to define a hierarchy for rewards distribution for the project team. Now, the AHP is being applied in the compliance with the responsibility matrix of every project. It allows us to respect all the relations of different employees to different parts of WBS and to respect their roles (e.g. Project Manager manages the whole project, but also is responsible for approvals of parts of WBS; Head of design manages the whole design, but he also executes a part of design and he consults the overall management of the project). For the example see Fig 3.

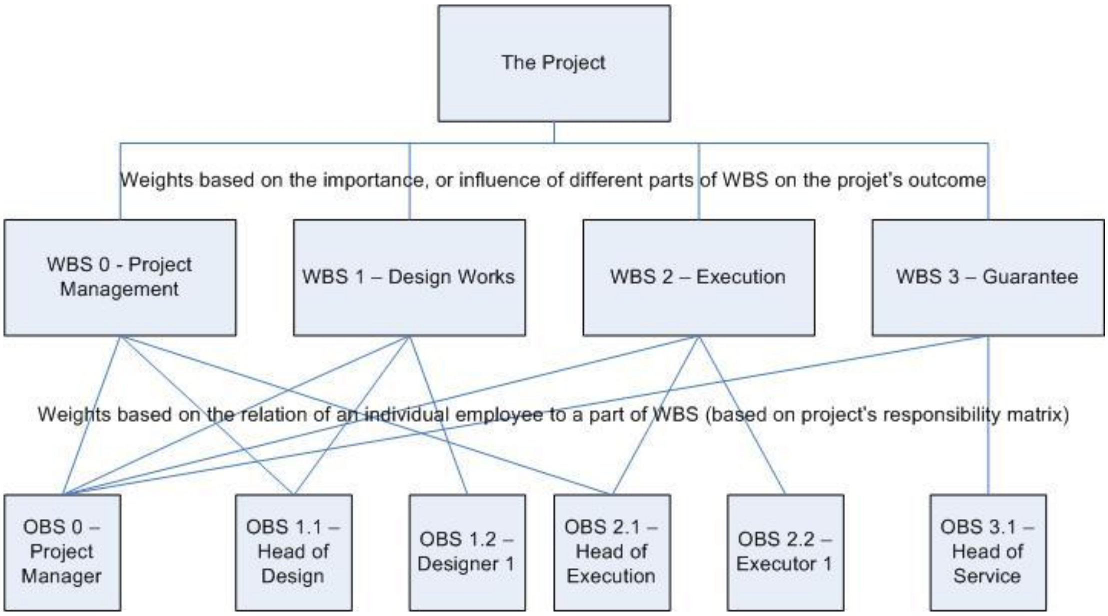


Figure 3 Example of AHP in compliance with the responsibility matrix of a project

2.3 Score method

Score method is method used to estimate the weights, in this case for estimate the reward coefficients. The score method is simple and to the coefficients estimation is necessary going through two parts of method.

1. Scoring

Score from a predetermined interval assigned for elements of the WBS. This score expresses the *influence* or *importance* of every *WBS element* on the result of the whole project. The other scores are assigned to different classes of *responsibilities*. This kind of score expresses the measure of *OBS element's influence* based on the class of responsibility.

2. Weight estimation

Weight (wi) estimation is based on the assigned scores. The intrinsic value is calculated according to a simple formula [7]:

$$w_i = \frac{p_i}{\sum_{i=1}^{k} p_i} \tag{1}$$

where p_i is the own score of the WBS or OBS element and k is the number of WBS elements or OBS elements related to the same WBS element.

2.4 Team Delphi method

Team Delphi is a method for group decision making. It can be used for definition of the scores for WBS and responsibility classes by the group of company's top managers. Landeta [8] defines Delphi method as the "popular technique for forecasting and an aid in decision-making based on the opinions of experts, which has been in existence for over half a century". Originally, this method was created in the RAND Corporation (Santa Monica, California), who used the experience of experts to determine the decision, more in [8,9].

Landeta [8] summarizes the main characteristics to the next few points:

- It is a repetitive process. The experts must be consulted at least twice on the same question, so that they can reconsider their answer, aided by the information they receive from the rest of the experts.
- It maintains the anonymity of the participants or at least of their answers, as these go directly to the group coordinator. This means a group working process can be developed with experts who do not coincide in time or space and also aims to avoid the negative influence that could be exercised by factors in the individual answers in terms of the personality and status of the participating experts.
- Controlled feedback. The exchange of information between the experts is not free but is carried out by means of a study group coordinator, so that all irrelevant information is eliminated.

- Group statistical response. All the opinions form part of the final answer. The questions are formulated so that the answers can be processed quantitatively and statistically. [8]

3 The new approach to the coefficient setup

To be able to use the new approach, it is necessary to have *WBS* and *OBS* of the project defined and combined into the *responsibility matrix*. The WBS defines partial outcomes (in the terms of performed work) of the project. The OBS defines people who will perform a project. The responsibility matrix defines key responsibilities of every involved person, related to partial project outcomes or works (see Tab. 1).

		PMT	Design dept.		Execution dept.		Service dept.
		Project manager	Head of deign	Designer 1	Head of exeuti	Executor 1	Head of service
Deliveries / Tasks / WBS	0 Management of the project itself	M, A	C		C		
	1 Design works	A	M, E1	E2	C		C
	2 Exectuion	A			M	E1	
	3 Service under warranty period	A					M, E1

Responsibilities	
A	Is responsible for approvals
M	Manages
E1	Executes (senior level)
E2	Executes (junior level)
C	Consults

Table 1 Project responsibility matrix (based on WBS and OBS)

In the next step the *score method* is applied to define *weights* for different parts of WBS and to define the *score* related to influence and contribution of different responsibilities. The weights for WBS are always unique for every different project and should be defined by top management in cooperation with the project manager. If there is more than one decision maker at this point, the *Team Delphi* method can be used to setup these weights consensually, while taking into account opinions of every involved decision maker. The score related to different responsibilities can be defined as one overall business rule of a company (see Tab. 2).

Weights for WBS				PMT	Design dept.		Execution dept.		Service dept.
Score	Weights			Project manager	Head of deign	Designer 1	Head of exeuti	Executor 1	Head of service
10	0,323	Deliveries / Tasks / WBS	0 Management of the project itself	M, A	C		C		
8	0,258		1 Design works	A	M, E1	E2	C		C
8	0,258		2 Exectuion	A			M	E1	
5	0,161		3 Service under warranty period	A					M, E1

Responsibilities		Score
A	Is responsible for approvals	6
M	Manages	10
E1	Executes (senior level)	8
E2	Executes (junior level)	6
C	Consults	4

Table 2 Responsibility matrix with WBS weights calculation and role score setup

When the weights for WBS and score for responsibilities are set, it is time to replace the responsibility symbols by adequate score. When one person has more than one responsibility related to one element of WBS, the final score is the sum of partial responsibility scores (see Tab. 3).

Weights for WBS				PMT	Design dept.		Execution dept.		Service dept.
Score	Weights			Project manager	Head of deign	Designer 1	Head of exeuti	Executor 1	Head of service
10	0,323	Deliveries / Tasks / WBS	0 Management of the project itself	16	4		4		
8	0,258		1 Design works	6	18	6	4		4
8	0,258		2 Exectuion	6			10	8	
5	0,161		3 Service under warranty period	6					18

Responsibilities		Score
A	Is responsible for approvals	6
M	Manages	10
E1	Executes (senior level)	8
E2	Executes (junior level)	6
C	Consults	4

Table 3 Responsibility matrix with responsibilities replaced by score

Then the responsibility score is transformed into standardized weights for every element of WBS - the sum of weights in one line is always equal to 1 (see Tab. 4).

Weights for WBS				PMT	Design dept.		Execution dept.		Service dept.
Score	Weights			Project manager	Head of deign	Designer 1	Head of exeuti	Executor 1	Head of service
10	0,323	Deliveries / Tasks / WBS	0 Management of the project itself	0,667	0,167	0,000	0,167	0,000	0,000
8	0,258		1 Design works	0,158	0,474	0,158	0,105	0,000	0,105
8	0,258		2 Exectuion	0,250	0,000	0,000	0,417	0,333	0,000
5	0,161		3 Service under warranty period	0,250	0,000	0,000	0,000	0,000	0,750

Table 4 Responsibility matrix with responsibilities scores replaced by standardized weights

The table 5 shows final overall reward coefficients for every involved employee. They are calculated as the sum of weights related to different WBS elements. This table also shows possibility to link the rewarding to different elements of WBS using partial weights for different persons and relevant WBS, but this option would cancel the effect of the rewards linked to a project in its wholeness.

Weights for WBS				PMT	Design dept.		Execution dept.		Service dept.
Score	Weights			Project manager	Head of deign	Designer 1	Head of exeuti	Executor 1	Head of service
10	0,323	Deliveries / Tasks / WBS	0 Management of the project itself	0,215	0,054	0,000	0,054	0,000	0,000
8	0,258		1 Design works	0,041	0,122	0,041	0,027	0,000	0,027
8	0,258		2 Exectuion	0,065	0,000	0,000	0,108	0,086	0,000
5	0,161		3 Service under warranty period	0,040	0,000	0,000	0,000	0,000	0,121
			Coefficient for the whole project	**0,361**	**0,176**	**0,041**	**0,188**	**0,086**	**0,148**

Table 5 Responsibility matrix with final reward coefficients

The exact value of the **expected personal reward** from the project is simple multiplication of the amount of money dedicated for rewards and the personal reward coefficient.

4 Conclusion

The new approach is in compliance with the logic of structures in project oriented organizations. The new approach to the coefficients setup requires a project plan, especially the WBS, OBS and the responsibility matrix as the inputs for reward coefficients setup. The advantage of the new approach is that it does not face the issue with the consistency of the pairwise comparison. It also enables more objective setup and brighter communication of reward coefficients, because they are clearly related to the responsibility classes of individuals and the scored importance and influence of different WBS elements.

Acknowledgements

This article and related research are performed with the support of Internal Grant Agency (IGA) of Czech University of Life Sciences Prague in project No. 201111160040.

References

[1] Antoš, V., Kvasnička, R.: *AHP in System for Project Staff Rewarding and Motivation*. In Mathematical Methods in Economics 2010. Vyd. 1. České Budějovice : University of South Bohemia in České Budějovice, 2010. ISBN 978-80-7394-218-2.

[2] Zelený, M.: *Cesty k úspěchu trvalé hodnoty soustavy Baťa*. ČINTÁMANI, Brno, 2005. ISBN 80-239-4969-1.

[3] Saaty, T.: *Decision-making with the AHP: Why is the principal eigenvector necessary*. In European Journal of Operational Research 145 (2003), 85-91.

[4] Antoš, V., Kvasnička, R.: The Consistency of Pairwise Comparison within the Reward System. In *Think Together 2011,* ISBN: 978-80-213-2169-4, online: http://www.thinktogether.cz.

[5] Doležal J., Máchal P., Lacko B. and col.*: Projektový management podle IPMA*. Grada Publishing, Praha 2010. ISBN 978-80-247-2848-3.

[6] Saaty, T. et al.: The analytic hierarchy process and human resource allocation: Half the story, *Mathematical* and Computer Modelling 46 (2007), 1041-1053.

[7] Jablonský, J.: *Operační výzkum, kvantitativní metody pro ekonomické rozhodování (in czech)*. Praha, Professional Publishing, 2002, ISBN: 80-86419-42-8.

[8] N. Dalkey and O. Helmer, An experimental application of the Delphi method to the use of experts, *Manage. Sci. 9* (1963) (3), pp. 458–467.

[9] Landeta, J.: Current validity of the Delphi method in social sciences , *Technological Forecasting and Social Change*, Volume 73, Issue 5, 2006, Pages 467-482.

Modification of the Three-point PERT Estimate for Practical Use

Jan Bartoška[1], Tomáš Šubrt[2]

Abstract. The PERT method is widely used as a tool for the estimation of activity duration in project management. The traditional PERT approach is usually used to estimate the mean and variance of the activity duration. The PERT uses the beta distribution that is a unimodal probability distribution defined on a closed interval. At real projects the three-point PERT estimate is often adapted to suit specific requirements and project needs. The peak of the beta distribution is possible to shift to the left or right and get a different mean estimate of the activity duration. The shift of the distribution peak depends on the human behaviour while working on the task. The paper proposes the modification of the three-point PERT estimate using parameterization of the weighted average. It proposes to shift the beta distribution peak, which is dependent on a human work effort and which is reachable by adjusting the PERT three-point estimate. Such a modification shifts the beta distribution peak where the work effort of an allocated source occurs. We demonstrate how to use a suitable parameterization of the three-point PERT estimate to reduce the risk of violating important project terms and due dates.

Keywords. PERT; Project management; Beta distribution; Work effort; Three-point PERT estimate; Activity duration; Risk of violating terms.

JEL Classification: C61.
AMS Classification: 90B99.

1 Introduction

Project management accentuates project time analysis whose output for every project is the determination of project duration and stating critical activities or overall tasks according to time reserves. A key and initial step during the project time analysis is stating the duration of individual activities. These individual activities are used for the determination of a project critical path and for the calculation of time reserves in the project. Duration periods are usually estimated. The determination or estimate of the activity duration period is usually a non-trivial task based on the experience or professional knowledge of a project manager. In everyday practice, to estimate the duration period the so-called three-point estimate is applied, originally developed for the PERT method (Program Evaluation and Review Technique). The PERT method is widely known among experts as well as professional public and its principles and means of application are described in [10], [2], [7] or [5].

Deficits and possible modifications of a commonly used approach for the calculation of a middle value and time estimate dispersion in the PERT method are dealt with for example in [6] and [12]. The author [6] focuses on the use of an oblong beta distribution and its possible advantages during the application. A different approach to the modification of contemporary means of calculation in the PERT method is offered in [12]. The author proposes his own manner of middle value approximation and time estimate dispersion and compares it with the existing. Both articles work only with mathematical apparatus and propose either the beta distribution modification or the modification of the manner of the middle value calculation and dispersion. However, the mentioned articles do not directly offer the inclusion of a human factor to the determination of duration period estimate. The manner of application and possible modifications of the PERT method are further discussed for instance in [1], [4] or [11].

The following text of the article tries to determine the modification of the middle time estimate value. It takes into account not only the human factor of an allocated source per activity but also the possibility of a different project manager's approach. The objective of the article is to consider a delay risk while creating the time estimate and for different types of an allocated source.

[1] Ing. Jan Bartoška, Ph.D., bartoska@pef.czu.cz, Department of Systems Engineering,
Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague 6, Czech Republic
[2] Doc. Ing. Tomáš Šubrt, Ph.D., subrt@pef.czu.cz, Department of Systems Engineering,
Faculty of Economics and Management, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Prague 6, Czech Republic

2 Theoretical Background

2.1 Determining activity duration for the PERT method

For the PERT method distribution it is supposed that an arbitrary quantity describing the activity duration will resemble the beta distribution. For the beta distribution the following density of probability is used (Malcolm et al., 1959):

$$\int_a^b f(t)dt = 1 \qquad f(t) = \begin{cases} 0 & \text{if} \quad t < a \\ c(t-a)(b-a) & \text{if} \quad a \le t \le b \\ 0 & \text{if} \quad t > b \end{cases} \tag{1}$$

More details concerning mathematical basics of the PERT method, especially the density of the beta distribution, can be found in [10] or [8]. The reason why it is possible to presuppose the beta distribution with a significant right-handed asymmetry for the activity duration consists in the fact that people responsible for the activity realisation do not state real duration of their activity based on their true judgement. On the contrary, they state the time period grossed up with a particular time reserve. When questioned, prior to the activity realisation, the time estimates of activity duration periods will be overestimated. Therefore, after the activity has been realised, it often shows that, in reality, the activities were finished earlier or they could be finished earlier. For every activity in the project the PERT method implements three parameters: an optimistic estimate of a duration period (a), a pessimistic estimate of a duration period (b), and the estimate of the most frequent activity duration period (m). The PERT method defines the formulas used in practice as follows:

$$\mu = \frac{a + 4m + b}{6} \tag{2}$$

$$\sigma_t^2 = \frac{(b-a)^2}{36} \tag{3}$$

The aim of the calculations based on the formulas (2) and (3) is to determine middle values (2) and dispersions (3) of all deadlines as soon as possible and allowable at the latest for all activities and a consequent determination of a critical path. As the three-point estimate we regard the formula (2), where three parameters are used and which is applied to estimate the activity duration period in the project (see Figure 1):

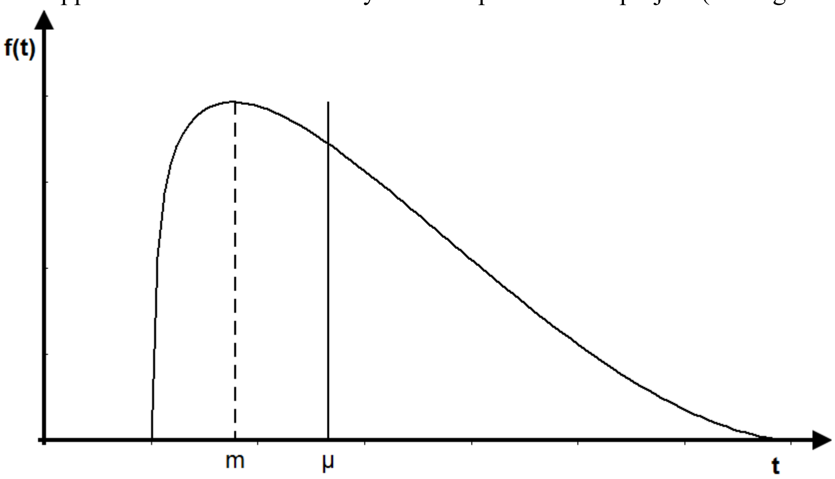


Figure 1 Beta distribution for values: a=1; m=2; b=7.

2.2 The impact of human factor

If there is a deadline determined for the completion of a task and a human factor is a source, during the activity realisation the source labours unevenly and with a variable intensity. Delay during activities realisation with human source participation leads to stress or to tension aimed at the source or the tension of the source. During the development and growth of the tension the human factor in an allocated source evokes the increase in work effort. Figure 2 below demonstrates possible behaviour of the human source known as the "Student Syndrome". The "Student Syndrome" phenomenon and its course are dealt with for example in [3]. The phenomenon is common in both fiction and non-fiction. It is mentioned for example in [9]. It emerged on the basis of the observation of students during their work on assigned tasks. The phenomenon has its equivalent in psychology – there it is known as „procrastination". Having been assigned a task, the human factor is under the influence of his or her natural behaviour and after a short time-interval s/he tries to create reserves on behalf of work effort. At the moment of discovery that the task has not been fulfilled in the time framework, there occurs a breakthrough in the source's behaviour. An increasing stress of the source leads to tension and a rapid growth of work effort, and to enormous values.

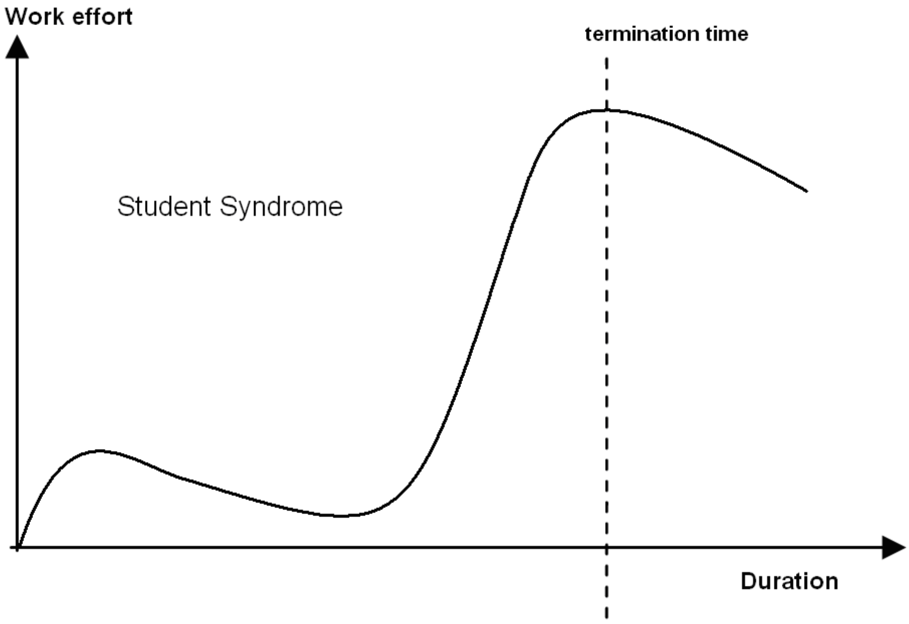


Figure 2 The variability of work effort during the "Student Syndrome".

3 Results and Discussion

3.1 Three-point estimate modification

The behaviour of an allocated source described in Material and Methods leads to a shift in a modal value of activity duration to the right, towards a pessimistic estimate. The shift towards the pessimistic estimate is provided by the variability of source work effort during activity realisation. The more the source succumbs to the "Student Syndrome", the more the distribution peak of the activity duration shifts towards the pessimistic estimate. The source's succumbing to the "Student Syndrome" is conditioned by the source's behaviour which does not change in a short period of time. The relevance of mode estimate, i.e. the position of the distribution peak, will depend on the source's consciousness, which the PERT method does not take into account. It is possible to understand the consciousness of the source - as the level of the estimate quality of the most frequent duration time. If the source consciously or unconsciously determines a modal parameter in a different way from reality or experience, middle values of duration time will not be estimated accurately. This may endanger the whole project. The estimate credibility of the most frequent activity duration determines the risk of activity delay and thus the whole project. The verification of the estimate quality of duration may be an arduous or impossible task. Respecting this the three-point estimate modification of activity duration may be proposed without a modal value as follows:

$$\mu = \frac{\gamma b + a}{1 + \gamma} \quad (4)$$

Where parameters b and a in the formula (4) are pessimistic and optimistic estimates of activity duration, and parameter γ is the level of the influence of the „Student Syndrome“ on an allocated human source in the activity in the range of (0; +∞). The level of γ influence should be set according to project manager's own experience, project type, his or her knowledge of specific resource behaviours and skills. A very low value of parameter γ expresses the state in which the "Student Syndrome" has almost no effect and a middle value of activity duration approaches an optimistic estimate. An extremely high value of parameter γ expresses an absolute impact of the syndrome where the middle value of activity duration approaches a pessimistic estimate. The more the parameter value γ approaches 0, the more conscious is the source in his/her work effort, and on the contrary. The proposed parameter γ can be further interpreted as an estimate coefficient for expected maximum source workload during activity realisation.

3.2 Influencing delay risk: Concealed time reserve

If the nature of the activity and allocated source prompts that the source is conscious in his or her work effort and does not succumb to the influence of the "Student Syndrome", and if it is obvious that the source tries to create concealed time reserve, it is possible to consider for example the value 0.5 for parameter γ:

$$\mu = \frac{0.5b + a}{1 + 0.5} \quad (5)$$

As concealed time reserve, it is possible to indicate such a part of the pessimistic estimate which unanimously expresses an unused time reserve of an allocated source. The time reserve prevents the source from delay and therefore the source will not consider it redundant. The duration period distribution is then similar to the beta distribution from the PERT method (see Figure 3).

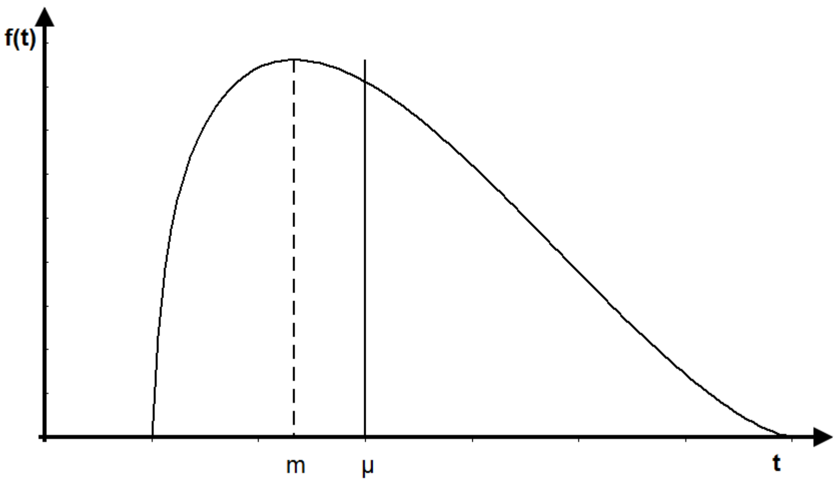


Figure 3 Beta distribution for modified three-point estimate (a=1; b=7; γ=0.5).

Analytically, it can be assumed that parameter γ value in the range of (0; 1) will express the existence of bigger or smaller concealed time reserve of an allocated source. The determination of parameter γ value and consequently of activity duration distribution should be the task of a project manager and it should also depend on the experience with a particular source.

3.3 Influencing delay risk: Work at the last moment

If the nature of the activity and an allocated source prompts that the source is not conscious in his or her work effort and succumbs to the influence of the “Student Syndrome”, or if it is obvious that the source does not create any concealed time reserve and overestimates his or her abilities, it is possible to consider value 2 for parameter γ:

$$\mu = \frac{2b + a}{1 + 2} \tag{6}$$

If an allocated source is inclined to carry out work at the last moment, it is possible to consider the following formula for the Beta duration distribution (see Figure 4):

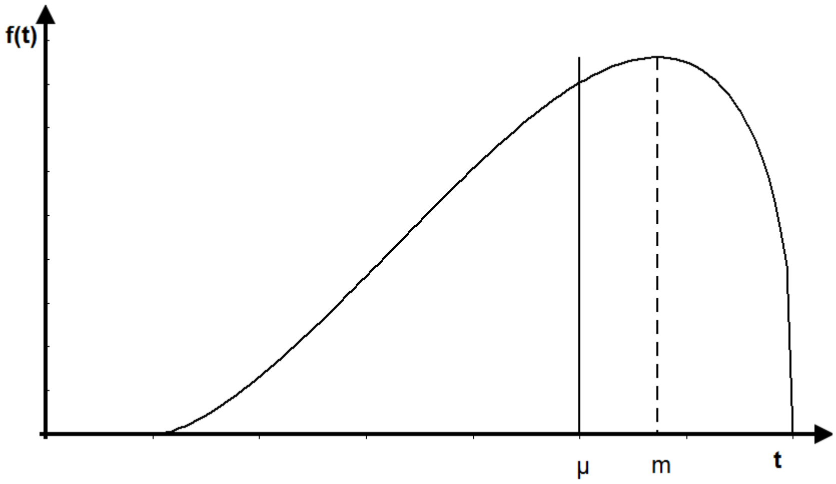


Figure 4 Beta distribution for modified three-point estimate (a=1; b=7; γ=2).

Again we can analytically consider the range in parameter γ values in the interval (1; +∞). The determination of parameter γ value and thereby a specific activity duration distribution should be the task of a project manager and it should depend on the experience with the source. We can expect the distribution (Figure 4) and the use of formula (6) for the calculation of a middle value of duration time especially for technically or technologically demanding activities and projects. Let us mention as an example software engineering and computer code programming.

3.4 Source participation on delay risk

If the project manager or responsible project team member does not select a proposed modification for time analysis and three-point estimate, and in the meantime, questions the allocated source about the mode estimate of duration time, the allocated source thus directly influences the level of delay risk. The more quality and true the source estimate is, the lower the risk of activity delays. On the contrary, a wrong estimate of the allocated source may lead to endangering the whole project. If it is possible to estimate the development of allocated source work effort during activity realisation, we can rather recommend the use of a proposed modification of the three-point estimate and thus decreasing the influence of the allocated source on delay risk. The reason for this conclusion may be the fact that the level of risk influence on the side of the source may not be known and the source behaviour may not always be conscious. The decision whether to use the modified three-point estimate or not should depend on the fact how known and predictable is the allocated source behaviour.

4 Conclusion

The article deals with the modification of the three-point estimate of activity duration. The concept of the three-point estimate which is used for the calculation for the middle value in the PERT method is common and applied especially to project management. The topic of the article indirectly follows the works of other authors and it brings its own contribution to the issue. The three-point estimate modification proposed in the article stems from the influence of the human factor in project management and consists in the shift in the distribution peak of activity duration towards either pessimistic or optimistic estimate. The shift in the distribution peak of activity duration is proposed by authors ‘own calculation of a middle value. The proposal is based on the elimination of a mode value, i.e. the estimate value of the most frequent activity duration. The shift direction is determined by a newly introduced parameter γ with the range of values $(0; +\infty)$. The newly introduced parameter expresses the influence of the human factor in the form of the "Student Syndrome" on activity duration distribution. The proposal presupposes that during variable work effort, even the position of the duration time distribution peak will be variable. The proposed three-point estimate modification can be used during any project management. Its use will thus take into consideration the particularity of the activity or allocated source in a duration time estimate. The proposed modification may lead to the decrease in delay risk during the management of the project.

References

[1] AZARON, A. et al.: A multi-objective resource allocation problem in PERT networks. *European Journal of Operational Research* **172** (2006), 838-854.

[2] CLARK, C. E.: The Greatest of a Finite of Random Variables. *Operations Research* **9** (1961), 145-162.

[3] COOK, S. C.: *Applying Critical Chain to Improve the Management of Uncertainty in Projects.* Massachusetts Institute of Technology, 1998. Thesis (Online).

[4] CRUZ, S. et al.: Stochastic models alternative to the classical PERT for the treatment of the risk: mesokurtic and of constant variance. *Central European Journal of Operational Research* **7** (1999), 60-62.

[5] EISNER, H.: A Generalized Network Approach to the Planning and Scheduling of a Research Project. *Operations Research* **10** (1962), 115-125.

[6] HANH, E. D.: Mixture densities for project management activity times: A robust approach to PERT. *European Journal of Operational Research* **188** (2008), 450-459.

[7] HEALY, L. Th.: Activity Subdivision and PERT Probability Statements. *Operations Research* **9** (1961), 341-348.

[8] KLUSOŇ, V.: *Kritická cesta a PERT v řídící praxi*. SNTL n.p., Praha, 1968.

[9] LEACH, L. P.: Critical Chain Project Management Improves Project Performance. *Project Management Journal* **30** (1999), 39-51.

[10] MALCOLM, D. G. et al.: Application of a technique for research and development program evaluation. *Operations Research* **7** (1959), 646-669.

[11] PLEGUEZUELO, R. H. et al.: A note on the reasonableness of PERT hypotheses. *Operations Research Letters* **31** (2003), 60-62.

[12] PREMACHANDRA, I. M.: An approximation of the activity duration distribution in PERT. *Computers & Operations Research* **28** (2001), 443-452.

Modeling multivariate volatility using wavelet-based realized covariance estimator

Jozef Barunik[1], Lukas Vacha[2]

Abstract. Our work brings complete theory for the realized covariation estimation generalizing current knowledge and bringing the estimation to the time-frequency domain for the first time. The results generalize the popular realized variance framework by bringing the robustness to noise as well as jumps and ability to measure the realized covariance not only in time but also in frequency domain. Noticeable contribution is brought also by the application of the presented theory. Our time-frequency estimators bring not only more efficient estimates, but decomposes the realized covariation into arbitrarily chosen investment horizons.

Keywords: multivariate realized volatility, covariation, jumps, wavelets

JEL classification: C22,C51,C58,G17

1 Introduction

High frequency data becoming increasingly available for wide range of securities allowed the shift from parametric conditional covariance estimation based on daily data towards the model-free measurement of so-called "realized quantities" on intraday data. Using a seminal result in semimartingale process theory, [4] shows that realized variance becomes a consistent estimator of the integrated volatility with increasing sampling frequency under the assumption of zero microstructure noise. [10] generalize the idea to multivariate setting of so-called "realized covariation" and provide asymptotic distribution theory for covariance (and correlation) analysis - again with the assumption of zero microstructure noise.

Although the theory is very appealing and intuitive, it assumes that observed high-frequency data are true underlying process. But the real-world data are contamined with microstructure noise and jumps, which makes statistical inference difficult. Realized measures suffer from the large bias and inconsistency with the presence of the noise and jumps in the observed data. The first approach to deal with noise actually throws away large amount of data. While it may not seem to be a logical step, the reason can be found quickly when looking at the data at various sampling frequencies. The higher the frequency of the data we use (i.e. 1 second, 1 tick), the more microstructure noise they contain and the more biased the estimator is. Thus lot of researchers use lower frequencies (i.e. 5 minutes), which results in throwing away the very large amount of the data directly. This is not an appropriate solution a statistician should use. In the recent literature, number of ways have been proposed to restore the consistency through subsampling, i.e. [17], [16], [9]. While inference under the noise and jumps in realized variation theory is widely studied in recent contributions, its generalization to covariation theory only emerges in literature. Together with important contributions by [16] and [9], [13] and [1] deal with an microstructure noise and non-synchronous trading and propose a consistent and efficient estimator of realized covariance, [5] proposes a forecasting model for realized correlations. This research is becoming very active and stands at the frontier of current research in financial econometrics.

In this paper, we contribute to current literature and provide generalization to our wavelet-based realized variation theory [11]. Due to the very limited space of these proceedings, we provide a limited notion of the proposed estimators (complete theory including mathematical proofs of the theory can be found in [11]) in the first part, while second part illustrates how the estimators can be used to estimation and decomposition of the dynamics of the stock market dependencies.

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Institute of Economic Studies, Charles University, Prague, e-mail: barunik@utia.cas.cz

[2]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Institute of Economic Studies, Charles University, Prague, e-mail: vachal@utia.cas.cz

2 General multivariate framework

The pioneering contribution extending univariate framework for realized measures has been made by [10]. Authors bring unified framework for modeling multivariate high frequency financial data using realized covariation and they provide asymptotic distribution theory for standard methods such as regression, correlation analysis, and covariance. Following [10], we introduce multivariate setting for the estimators.

Consider an m-dimensional logarithmic asset price process defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The return process evolving in continuous time over the $[t-h, t]$ time interval, $0 \leq h \leq t \leq T$ is $\mathbf{r}_{t,h} = \mathbf{p}_t - \mathbf{p}_{t-h}$, where $\mathbf{p}_t = (p_{(1)t}, \ldots, p_{(m)t})'$ denote $m \times 1$ vector of log prices at time t, and is used as notation for multivariate price process. We further consider the natural information filtration, an increasing family of σ-fields $(\mathcal{F}_t)_{t \in [0,T]} \subseteq \mathcal{F}$, which satisfies the usual conditions. Information set $\mathcal{F}_t$ contains the full history up to time t of the realized values of asset price and other relevant state variables. A fundamental result of stochastic integration theory states that ssuch process can be uniquely decomposed. For any m-dimensional, square-integrable, continuous sample path, logarithmic price process $(\mathbf{p}_t)_{t \in [0,T]}$, with continuous sample path and a full rank of the associated $m \times m$ quadratic variation process, $[r, r]_t$, there exists a representation such that for all $0 \leq t \leq T$

$$\mathbf{r_{t,h}} = \int_{t-h}^{t} \mu_s ds + \int_{t-h}^{t} \Theta_s dW_s, \tag{1}$$

where μ_s is integrable, predictabe, and finite-variation $m \times 1$ vector, and Θ represents multivariate stochastic volatility process with càdlàg elements and vector W_t is a $m \times 1$ standard Brownian motion. The l-th row of the matrix Θ_t is $(\sigma_{(l,1)t}, \sigma_{(l,2)t}, \ldots, \sigma_{(l,m)t})$. Then the spot (or instantaneous) covariance is defined as $\Sigma_t = \Theta_t \Theta_t'$, satisfying for all $t < \infty$: $\int_{t-h}^{t} \Sigma_{(l,l)u} du < \infty, l = 1, \ldots, m$, where $\Sigma_{(l,q)t}$ is the (l, q) element of the Σ_t process. The requirement that $m \times m$ matrix $[r, r]_t$ is of full rank for all t implies that no asset is redundant at any time[1]. The object of interest, quadratic covariation between l-th and the q-th price processes over $[t-h, t]$, for $0 \leq h \leq t \leq T$ can be expressed as

$$CV_{(l,q)t,h} = \int_{t-h}^{t} \Sigma_{(l,q)s} ds = \int_{t-h}^{t} \sum_{i=1}^{m} \sigma_{(l)i,s} \sigma_{(q)i,s} ds. \tag{2}$$

$CV_{(l,q)t,h}$ consistently estimates the integral of the conditional covariance of the increments of the local martingale component of $\mathbf{r}_{t,h}$ over the assumed interval. In a special case of $\mathbf{r}_{t,h}$ being univariate $(m = 1)$, $CV_{(l,q)t,h}$ represents quadratic variation of $p_{t,h}$.

2.1 Estimation of realized covariation measure

[4] suggest to estimate the quadratic covariation matrix analogously to realized volatility by taking the outer-product of the observed high-frequency return over the period. The realized covariance over $[t-h, t]$, for $0 \leq h \leq t \leq T$, is then defined by

$$\widehat{RC}_{t,h} = \sum_{i=1}^{n} \mathbf{r}_{t-h+\left(\frac{i}{n}\right)h} \mathbf{r}'_{t-h+\left(\frac{i}{n}\right)h}, \tag{3}$$

where n is number of observations in the $[t-h, t]$. [4] and [10] show that *ex-post* realized covariance $\widehat{RC}_{t,h}$ is an unbiased estimator of *ex-ante* expected covariation $RC_{t,h}$. With increasing sampling frequency, realized covariance is moreover consistent estimator of the covariation over any fixed time interval $h > 0$, as $n \to \infty$.

3 Jumps and noise

Consistency of the realized covariance implies a simple alternative empirical return-volatility measurement. But there are two issues which complicate the practical usage of the nice convergence results. As realized covariance is consistent estimator with increasing sampling frequency, $n \to \infty$, continuum of instantaneous return observations must be used in order the realized volatility estimate will converge to the

[1] This condition is not redundant, see [4].

realized covariance. In practice, we can observe only discrete prices, and thus an inevitable discretization error is present. On the other hand, market microstructure effects such as price discreteness, bid-ask spread and bid-ask bounce contaminate the return observations. Thus in practice, return process should not be sampled too often, regardless the number of observations available, to avoid the large bias from market microstructure. Literature has extensively studied the noise-to-signal ratio, constructed optimal sampling schemes, which ranges from 5 to 30 minutes, for instance. The main literature is nicely surveyed by [14, 7, 15, 2]. Recent notable contributions to this literature include [17, 6, 8].

Another issue bringing bias to the estimates of variance-covariance matrix is presence of jumps in the data. In the literature, jump diffusion model is assumed to extend the presented theory. An example of general bivariate jump-diffusion process we will use in this work is $dp_{(q)t} = \mu_{(q)t}dt + \sigma_{(q)t}dW_{(q)t} + \xi_{(q)t}dz_{(q)t}, q = \{1, 2\}$, where z is a constant-intensity Poisson process with the magnitude of jump controlled by $\xi_t \sim N(\bar{\xi}, \sigma_\xi^2)$, $W_{1,t}$ is a standard Brownian motion, and for all $t \in [0, T]$, $dW_{(2)t} = \rho_t dW_{(1)t} + \sqrt{1 - q_t^2}dW_{(3)t}$, where $W_{(3)t}$ is an independent standard Brownian motion, and ρ is a stochastic process with càdlàg paths. Finally, $z_{(1)t}$ and $z_{(2)t}$ are possibly correlated pure jump processes. Quadratic covariation of this process over the $[t - h, t]$ time interval, $0 \leq h \leq t \leq T$, is then

$$CV_{(1,2)t,h} = \underbrace{\int_{t-h}^{t} \rho\sigma_{(1)s}\sigma_{(2)s}ds}_{IC_{t,h}} + \underbrace{\sum_{t-h \leq s \leq t} J_{(1)s}J_{(2)s}}_{\text{Jump Covariation}}, \tag{4}$$

where $J_{(q)t} = \xi_{(q)t}dz_{(q)t}$ and is non-zero only if we have co-jumps. Thus quadratic covariation will compose of the *Integrated Covariance* and covariance of common jumps.

4 Wavelet-based covariation theory

Due to the very limited space of the proceedings, we introduce the main results of the wavelet-based covariation theory extending our results from the univariate part and the results of [12], who first bring the wavelet-based realized volatility estimator to the literature. In our work, we generalize the results of [12] in several ways. Instead of using Discrete Wavelet Transform we use Maximum Overlap Discrete Wavelet Transform (MODWT), which is more efficient estimator and it is not restricted to the sample size limited by the power of two. We also use the Daubechies family of wavelets instead of the Haar type in our work. Our biggest contribution is in providing complete theoretical framework for estimation of multivariate wavelet realized covariation, as it can not be found in current literature. The theory improves efficiency of the estimated covariances. We build on the results and define also new measures of correlation and realized beta based on our wavelet-based estimators. We use wavelets to decompose the realized covariance to obtain information about behavior (energy contribution) at every scale and thus gain deeper knowledge about dependence between the two examined processes. The realized wavelet covariance (using MODWT) is a scale by scale decomposition of the realized covariance defined by the Equation 6.

Definition 1. Wavelet-based realized covariance estimator
The realized wavelet covariation of l-th and q-th asset return from the m-dimensional vector $\mathbf{r}_{t,h}$ over $[t - h, t]$, for $0 \leq h \leq t \leq T$ can be defined as

$$\widehat{RC}^{(WRC)}_{(l,q)t,h} = \sum_{j=1}^{J_s+1} \sum_{k=1}^{n} \widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h} \widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h} \tag{5}$$

where n is number of intraday observations over $[t - h, t]$ and J_s is number of scales considered, $\widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}$ are MODWT coefficients on $j = 1, ..., J_s + 1$ scales, where $J_s \leq \log_2 n$.

In [11] we show that the estimator is unbiased and consistent estimator of the realized covariance. Having defined the basic tool for our analysis, wavelet-based realized covariation estimator which is able to consistently estimate the integrated covariation, we generalize the jump detection test presented in univariate setting to the multivariate setting in the following section.

5 Disentangling jumps from co-jumps

[12] first proposed the usage of wavelets in estimating jumps in high-frequency data. We generalize this concept to multivariate framework. We detect all jumps in m assets separately using wavelet decomposition, and then we estimate co-jumps. Let us define the procedure. Let $\widetilde{\mathcal{W}}_{(q)1,k}$ be a 1^{st} level wavelet coefficients of $(y_{(q)t})_{t\in[0,T]}$. Then for $q = 1,\ldots,m$ assets, we estimate jumps as $|\widetilde{\mathcal{W}}_{(q)1,k}| > \frac{median\{|\widetilde{\mathcal{W}}_{(q)1,k}|, k=1,\ldots,n\}}{0.6745}\sqrt{2\log n}$, where $\hat{\tau}_{(q)l} = \{k\}$ is estimated jump location with size of $\bar{y}_{(q)\hat{\tau}_{l+}} - \bar{y}_{(q)\hat{\tau}_{l-}}$, the averages over $[\hat{\tau}_{(q)l}, \hat{\tau}_{(q)l} + \delta_n]$ and $[\hat{\tau}_{(q)l}, \hat{\tau}_{(q)l} - \delta_n]$ respectively, with $\delta_n > 0$ being a small neighborhood of estimated jump location $\hat{\tau}_{(q)l} \pm \delta_n$. The jump variation of q-th asset is then estimated by sum of squares of its all estimated jump sizes: $\widehat{MWJC}_{(q)} = \sum_{l=1}^{N_t}(\bar{y}_{(q)\hat{\tau}_{l+}} - \bar{y}_{(q)\hat{\tau}_{l-}})^2$. Following the theory in [12], we can say that $\widehat{MWJC}_{(j)}$ will be consistent estimator of jumps for all q assets in the $\mathbf{p}_t$. This result is shown in [11]. Once we have estimated all independent jumps in the studied $\mathbf{p}_t$ vector, we can propose an analysis of co-jumping in the series. The idea is to compare all the jump locations, and those which will be same across all $q = 1,\ldots,m$ assets in some small neighborhood will be co-jumps. Let $\hat{\tau}_{(q)l}$ be estimated jump locations of $(y_{(q)t})_{t\in[0,T]}$ for all $q = 1,\ldots,m$. Then co-jump location $\hat{\tau}^*_l = \{k\}$ can be estimated as: $\hat{\tau}_{(q)l} - \delta_n < \hat{\tau}^*_l < \hat{\tau}_{(q)l} + \delta_n, \quad$ for all $\quad q = 1,\ldots,m$. Co-jumps are important particularly in the portfolio theory. For the well diversified large portfolio in the sense of the Arbitrage Pricing Theory, idiosyncratic jumps are diversified away, but common jumps, or co-jumps remain the problem. Thus in the following subsection, we illustrate our technique on the portfolio multivariate extension.

6 Wavelet-based realized covariance estimator robust to jumps and noise

After the basic introduction of the wavelet-based estimation of realized covariances, we propose an estimator of covariance which is robust to noise and also is able to deal with jumps in the data. Moreover, we will be able to decompose the integrated covariance into J_s components using our estimator.

Definition 2. Jump wavelet TSCV (J-WTSCV) estimator
Let $\widehat{RC}^{(estimator,J)}_{(l,q)t,h}$ denote an estimator of realized covariance between l-th and q-th asset return on the jump-adjusted observed data, $\mathbf{y_{t,h}}^{(J)} = \mathbf{y}_{t,h} - \widehat{\mathbf{MWJC}}$. Jump-adjusted wavelet two scale realized covariance estimator is defined as:

$$\widehat{RC}^{(J-WTSCV)}_{(l,q)t,h} = c_N\left(\widehat{RC}^{(WRC,J)}_{(l,q)t,h} - \frac{\bar{n_G}}{n_S}\widehat{RC}^{(S,J)}_{(l,q)t,h}\right), \tag{6}$$

where $\widehat{RC}^{(WRC,J)}_{(l,q)t,h} = \frac{1}{G}\sum_{g=1}^{G}\sum_{j=1}^{J_s+1}\sum_{k=1}^{n}\widetilde{\mathcal{W}}_{(l)j,t-h+\frac{k}{n}h}\widetilde{\mathcal{W}}_{(q)j,t-h+\frac{k}{n}h}$ obtained from wavelet coefficient estimates using MODWT on the grid of size $\bar{n} = n/G$ on the jump-adjusted observed data, $\mathbf{y}^{(J)}_{t,h} = \mathbf{y}_{t,h} - \widehat{\mathbf{MWJC}}$ and c_N is a constant that can be tuned for small sample performance.

In [11] we show that the estimator 6 is unbiased and consistent estimator of the realized covariance. Thus it converges in probability to the *true* integrated covariance of the $\mathbf{p_{t,h}}$ which is of primary interest in this analysis. Thus we have defined a new wavelet-based covariation theory which is able to estimate realized covariation consistently in the presence of noise and jumps. In the next section, we utilize this theory to propose estimator of covariance and realized beta which are important for the practitioners in the finance.

6.1 Wavelet-based realized correlation

A simple transformation of the realized covariation matrix which is an important tool for practitioners, is correlation measure. Basic results are introduced by [3], while [10] provides also asymptotic theory for the estimators. using pour estimators, we can simply extend these results and propose wavelet-based realized correlation as $\hat{\rho}^{(JWR)}_{(l,q)t,h} = \frac{\widehat{RC}^{(J-WTSCV)}_{(l,q)t,h}}{\sqrt{\widehat{RV}^{(J-WTSRV)}_{(l)t,h}\widehat{RV}^{(J-WTSRV)}_{(q)t,h}}}$, where $\widehat{RC}^{(J-WTSCV)}_{(l,q)t,h}$ is defined by Eq. 6 and estimates covariation between l-th and q-th series, $\widehat{RV}^{(J-WTSRV)}_{(l)t,h}$ and $\widehat{RV}^{(J-WTSRV)}_{(q)t,h}$ is special case of Eq. 6 and estimates variation of l-th and q-th series respectively. To obtain the contribution of every

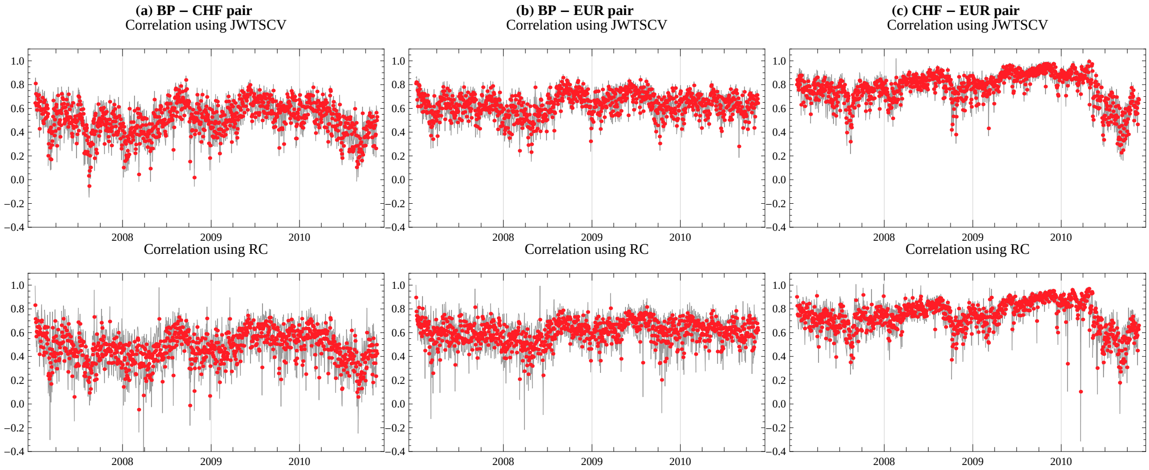


Figure 1: Correlations with 95% CI. $WRCorr$ in the first row and using RC in second row.

particular level j to the overall correlation, we have to weight the wavelet correlations at each level with their energy contribution.

7 Dynamics of realized correlations

In this chapter, we utilize our theoretical results on the real-world data. J-WTSCV proved to be the most efficient estimator of integrated covariance and correlation, thus we expect it to improve also understanding of the true relationship between assets. In addition, we also utilize the power of wavelets to decompose the realized measures to several investment horizons.

Foreign exchange future contracts are traded on the Chicago Mercantile Exchange (CME) on a 24-hour clock basis. We will estimate the realized covariance of British pond (BP), Swiss Franc (CHF) and Euro futures (EUR), while we will focus on the pairs: BP-CHF futures pair, BP-EUR pair and CHF-EUR pair. After estimating the covariance, we will study also correlations between the currencies. All contracts are quoted in the unit value of the foreign currency in US dollars, which makes them comparable. The cleaned data are available from Tick Data, Inc. It is important to understand the trading system before we begin the study. In August 2003, CME started to offer Globex trading platform, which brought large increase in the liquidity of currency futures. For the first time ever in, a single month trading volume on the electronic trading platform exceeded 1 million contracts every day. Beginning Monday, December 18, 2006, CME Globex(R) electronic platform starts offering trading within 23 hours a day. Weekly trading cycle begins on Sunday, 5:00pm and ends on Friday at 4:00pm, while every day the trading is interrupted for one hour from 4:00pm until 5:00pm. These changes in the trading system had crucial impact on the trading activity. For this reason, we restrict ourselves to use the sample period extending from January 5, 2007 through November 17, 2010 containing the most recent financial crisis. The futures contracts we use are automatically rolled over to provide continuous price records, thus we do not have to deal with different maturities. The tick-by-tick transactions are recorded in Chicago Time referred to as Central Standard Time (CST), therefore in a given day, trading activity starts at 5:00pm CST at Asia and continues in Europe followed by North America and finally closing at 4:00pm in Australia. We redefine the day in accordance with the electronic trading system. Moreover, Saturdays and Sundays, US federal holidays, December 24 to December 26, and December 31 to January are eliminated because of the very small activity which would bias the estimates. Finally, we are left with 944 days in the sample.

7.1 Dynamics of decomposed dependencies

Figure 1 brings comparison of the correlations dynamics computed using two estimators: simple realized correlation and our jump-adjusted wavelet correlation ($\hat{\rho}^{(JWR)}_{(l,q)t,h}$) estimator. It is noticeable that our WR-Corr estimator provides estimate with lower variance (basically due to jumps), and confidence intervals.

Precisely, BP - CHF futures pair, BP - EUR futures pair and CHF - EUR futures pair has average estimated WRCorr correlation with 95% confidence interval in parentheses 0.506 (±0.069), 0.629 (±0.053) and 0.769 (±0.051) respectively. Average correlation for the same pairs estimated using standard RC method is 0.47 (±0.1), 0.602 (±0.086) and 0.738 (±0.062) respectively. Even though the correlations changes significantly over time, average correlation estimated using our method is approximately 0.03 larger than using simple RC. This result is economically significant and can have direct impact on portfolio diversification. Moreover, our method provides much narrower confidence intervals for the estimates.

8 Conclusion

It this short conference proceedings paper we present our wavelet-based realized covariance estimators and apply them to estimation of correlation dynamics on the stock markets. Main finding is that our estimator provides more precise correlation estimate as it is robust to noise and jumps in the data.

Acknowledgements

The support from the Czech Science Foundation under Grants 402/09/0965, 402/09/H045, and P402/10/1610 and from the Department of Education MSMT 0021620841 is gratefully acknowledged.

References

[1] Y. Aït-Sahalia, J. Fan, and D. Xiu. High frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105:1504–1517, 2010.

[2] T.G. Andersen and L. Benzoni. Realized volatility. In T.G. Andersen, R.A. Davis, J.P. Kreiss, and T. Mikosch, editors, *Handbook of Financial Time Series*. Springer Verlag, 2007.

[3] T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, (96):42–55, 2001.

[4] T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, (71):579–625, 2003.

[5] F. Audrino and F. Corsi. Modeling tick-by-tick realized correlations. *Computational Statistics and Data Analysis*, 54:2372–2382, 2010.

[6] F.M. Bandi and J. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, (79):655–692, 2006.

[7] F.M. Bandi and J. Russell. Volatility. In J. Birge and V. Linetsky, editors, *Handbook of Financial Engineering*. Elsevier, 2006.

[8] O.E. Barndorff-Nielsen, P. Hansen, A. Lunde, and N. Shephard. Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Working Paper, University of Aarhus, Denmark; Stanford University; Nuffield College, Oxford, UK*, 2007.

[9] O.E. Barndorff-Nielsen, P.R. Hansen, A. Lunde, and N. Shephard. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149–169, 2011.

[10] O.E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3):885–925, 2004.

[11] J. Barunik. Wavelet-based realized variation and covariation theory. Unpublished dissertation. Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, 2011.

[12] J. Fan and Y. Wang. Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, (102):1349–1362, 2007.

[13] J.E. Griffin and R.C.A. Oomen. Covariance measurement in the presence of non-synchronous trading and market microstructure noise. *Journal of Econometrics*, 160(1):58–68, 2011.

[14] P.R. Hansen and A. Lunde. Realized variance and market microstructure noise. *Journal of Business and Economic Statistics*, (24):127–161.

[15] Medeiros MC McAleer M. Realized volatility: A review. *Econometric Reviews*, (27):10–45, 2008.

[16] L. Zhang. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47, 2011.

[17] L. Zhang, P.A. Mykland, and Y. Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association*, (100):1394–1411, 2005.

Some comparative macroeconomic analysis of Visegrad countries using differential geometry approach

Jiří Beck[1], Ladislav Lukáš[2]

Abstract. Paper presents a comparative macroeconomic analysis based upon GDP, inflation and stock exchange index time series of Visegrad countries within the period 1997Q1 till 2010Q4. Parametric representation of data in 3-D state space enables to tackle the macroeconomic development as a curve. Based upon differential geometry approach Frenet frames are constructed at selected set of discrete points along the curve. Investigation of frame translations provides both incremental and accumulated curve lengths, whereas frame rotations may generate traces on unit sphere. Numerical results together with corresponding macroeconomic interpretations of these quantities generated are discussed in detail, and computer implementation is presented, too.

Keywords: Macroeconomic data analysis, GDP, stock exchange index, inflation, time series analysis.

JEL Classification: C02, C32, C82, E23, E27, E31
AMS Classification: 90C15, 91B51

1 Introduction

Macroeconomic analysis is a very attractive field of research. Sure, there is also a lot of literarure and references devoted to the topic and based upon various techniques and methods. Some of them are focused on summarizing data and monitoring facts, other ones concern theoretical modelling and are based upon various theoretical reasonings, e.g. the New Keynesian approach being the most popular recently, see e.g. [4], and there are also works and studies which try to apply new sights and different methods. Our paper belongs to this third category.

We are concerned with a comparative macroeconomic analysis based upon GDP, inflation and stock exchange index time series of Visegrad countries within the period 1997Q1 till 2010Q4. A particular analysis of an evolution of the monetary policy transmission mechanisms of V4 countries is presented in [1].

Reasoning about the length of investigated period was motivated with three intuitive aspects – first, to eliminate prospective transient development of economy after 1989, second, to include period of relative stable and progressive economic development, and third, to involve periods which were impacted by global financial crisis, too. Parametric representation of data in 3-D state space enables to tackle the macroeconomic developments as curves in the state space. We do believe that our non-traditional approach provides new insights to macroeconomic analysis and is able to build new characteristics.

2 Macroeconomic data processing

We have used the data accumulated and presented in [5], which had been downloaded from Eurostat, OECD databases and public databases of Stock exchanges of individual V4 countries, i.e. Czech republic, Hungary, Poland and Slovak republic, in abbreviation CZ, HU, PL, and SK, respectively. As we have focused our analysis upon three important macroeconomic quantities – GDP, inflation and national stock exchange index, so in total, we have got four sets of three time series with 52 entries each.

Let ${}_k\boldsymbol{x}$ denotes a 3-D vector time series for k-th country, k=1,2,3,4, where the prefix k values are induced by lexicographical ordering of countries, i.e. k=1 ~ CZ, k=2 ~ HU, k=3 ~ PL, k=4 ~ SK. Each ${}_k\boldsymbol{x}$ has three 1-D time series components denoted ${}_kx_n$, n =1,2,3, where n=1 ~ stock exchange index, n=2 ~ GDP, n=3 ~ inflation rate. Since each of them has 52 entries, we need even more subtle notation to identify any particular entry.

[1] Univ.of West Bohemia, Faculty of Econ., Dept. of Econ. and Quantitative Methods, Husova 11, 30614 Pilsen, beck@kem.zcu.cz.
[2] Univ.of West Bohemia, Faculty of Econ., Dept. of Stat. and Quantitative Methods, Husova 11, 30614 Pilsen, lukasl@kem.zcu.cz.

Let ${}_kx_{n,m}$, m =1,2,…,52 denote individual entries. However, for some purposes we need to identify a particulatr year and quarter, too. Hence, we adopt the following notation ${}_kx_{n,m(i,j)}$, where the index function is $m(i,j)=4(i-1)+j$, where i=1,2,..,13 denotes sequence of years 1997, 1998,.., 2010, and j=1,2,3,4 stands for quarters of year Q1,..,Q4.

. The raw downloaded data for V4 countries are denoted ${}_k\boldsymbol{y}$ and their entries ${}_ky_{n,m}$, m=1,2,…,52, respectively. However, such data are given in their natural units, which are rather different in their numerical values. In order to make them suitable for our analysis based on differential geometry the raw data ${}_k\boldsymbol{y}$ have been scaled and transfered into ${}_k\boldsymbol{x}$ by procedure given by (1) and (2).

$$ {}_k r_n = \sum_{j=1}^{4} {}_k y_{n,m(9,j)} \ , \ k=1,2,3,4, \text{ and } n=1,2,3, \tag{1} $$

where index function $m(9,j)$, j =1,2,3,4, i=9, gives values {33,34,35,36}.

Quantities ${}_k r_n$ represent the scaling factors and give averaged values for corresponding V4 countries macroeconomic time series. In a role of reference year we have selected the year 2005, hence i=9, since we simply accepted this year as a break-year between 'stable development' and period accused by global financial crisis. Of course, it could be a matter of discussion, but the scaling procedure is rather flexible and could be applied for any other selected basis year, a quarter, or another period averaged.

Vector time series ${}_k\boldsymbol{x}$ with their components are given by (2), where $^{\mathrm{T}}$ stands for transposition.

$$ {}_kx_{n,m}, = {}_ky_{n,m} / {}_k r_n, \quad m=1,2,\ldots,52, \text{ and } k=1,2,3,4, \ n=1,2,3, $$
$$ {}_k\boldsymbol{x} = ({}_kx_{1,m}, {}_kx_{2,m}, {}_kx_{3,m})^{\mathrm{T}}. \tag{2} $$

3 Differential geometry approach and results

We refer [2] and [3] for more details on differential geometry as being short and individual reference selection only. In general, differential geometry deals with curves, surfaces or more general manifolds in analytic way. Hence in order to get a curve, our first step is to define interpolants of ${}_k\boldsymbol{x}$. First, we choose piecewise linear interpolation, and the second one is cubic spline interpolation, which both provide continuous functions.

Let ${}_k\boldsymbol{p}(t)$, ${}_k\boldsymbol{q}(t)$ denote piecewise linear and cubic spline interpolants of vector time series ${}_k\boldsymbol{x}$ defined on interval [1,52] given by (3a) and (3b), respectively. Let d be a time series with m entries, then $s_1(t;d)$ and $s_3(t;d)$ stand for scalar piecewise linear and cubic spline interpolants, respectively, defined on interval [1,m], and built by well-known procedures.

$$ {}_k\boldsymbol{p}(t) = (s_1(t;{}_kx_{1,m}), s_1(t;{}_kx_{2,m}), s_1(t;{}_kx_{3,m}))^{\mathrm{T}}, \tag{3a} $$

$$ {}_k\boldsymbol{q}(t) = (s_3(t;{}_kx_{1,m}), s_3(t;{}_kx_{2,m}), s_3(t;{}_kx_{3,m}))^{\mathrm{T}}. \tag{3b} $$

.

3.1 First result

On upper parts of Figures 1-4 we present curves in 3-D space, on the left ${}_k\boldsymbol{p}(t)$, on the right ${}_k\boldsymbol{q}(t)$, representing pathes of macroeconomic development of V4 countries (CZ, HU, PL, and SK) using data ${}_k\boldsymbol{x}$. The points on the curves are located at the following quarters 1997Q1, 1999Q1, 2001Q1, 2003Q1, 2005Q1, 2007Q1, 2009Q1 and 2010Q4, thus giving better insight to movement along the whole time span [1997Q1,2010Q4]. We may conclude that there is not a great difference between piecewise linear and cubic spline interpolants in all V4 countries, at the first sight.

3.2 Second result

Figures 5 and 6 present plots of all four curves in 3-D space. The piecewise linear interpolants ${}_k\boldsymbol{p}(t)$, k=1,2,3,4 of macroeconomic development of V4 countries is given on the Fig. 5, whereas the Fig. 6 brings a bunch of ${}_k\boldsymbol{q}(t)$. Since we have selected the averaged macroeconomic data of the year 2005 as a reference ones, we add corresponding points at 2005Q1,Q2,Q3 and Q4 to plotted curves, too. Here, we are already ready to see the differences, and these have also motivated us to further analysis, which is presented as the third result. Anyway, a straight-sight and common "U-shape" development of all V4 countries is rather interesting, too. The paper limited space prevent us to give more graphical outputs under various rotations angles.

All computations were performed by sw Mathematica 7.0, WolframResearch Inc., and outputs were in color.

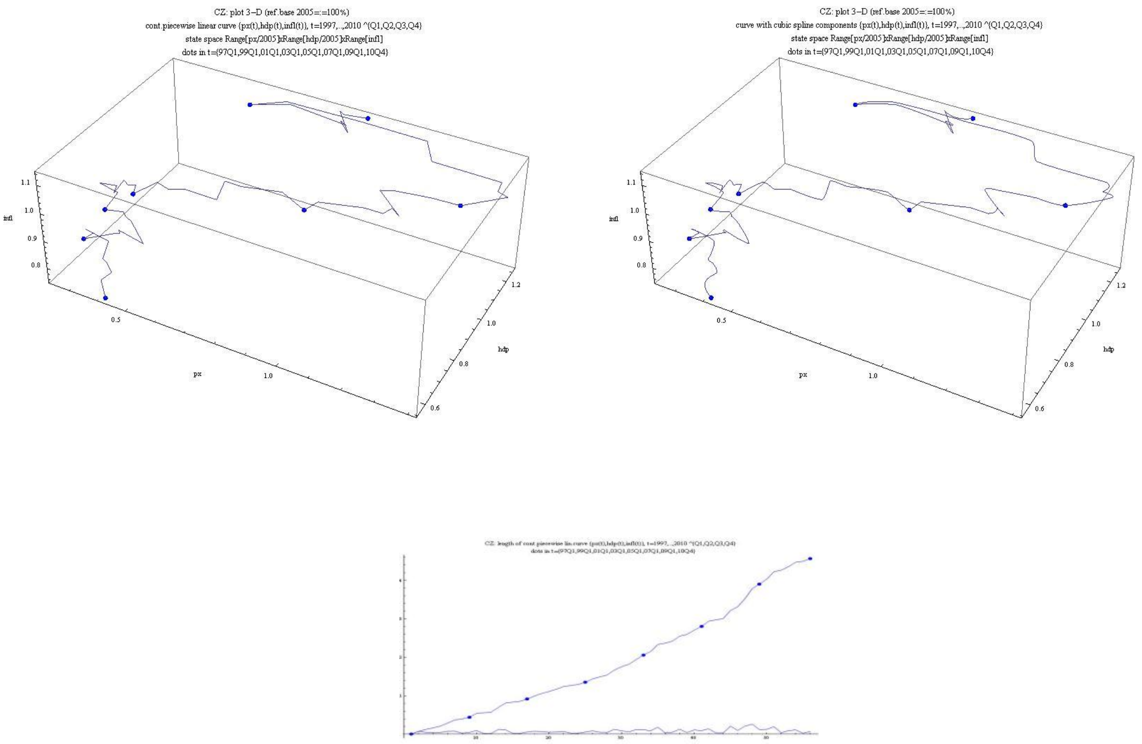


Fig. 1 CZ – piecewise linear ${}_1\boldsymbol{p}(t)$, cubic spline ${}_1\boldsymbol{q}(t)$, length of ${}_1\boldsymbol{p}(t)$.

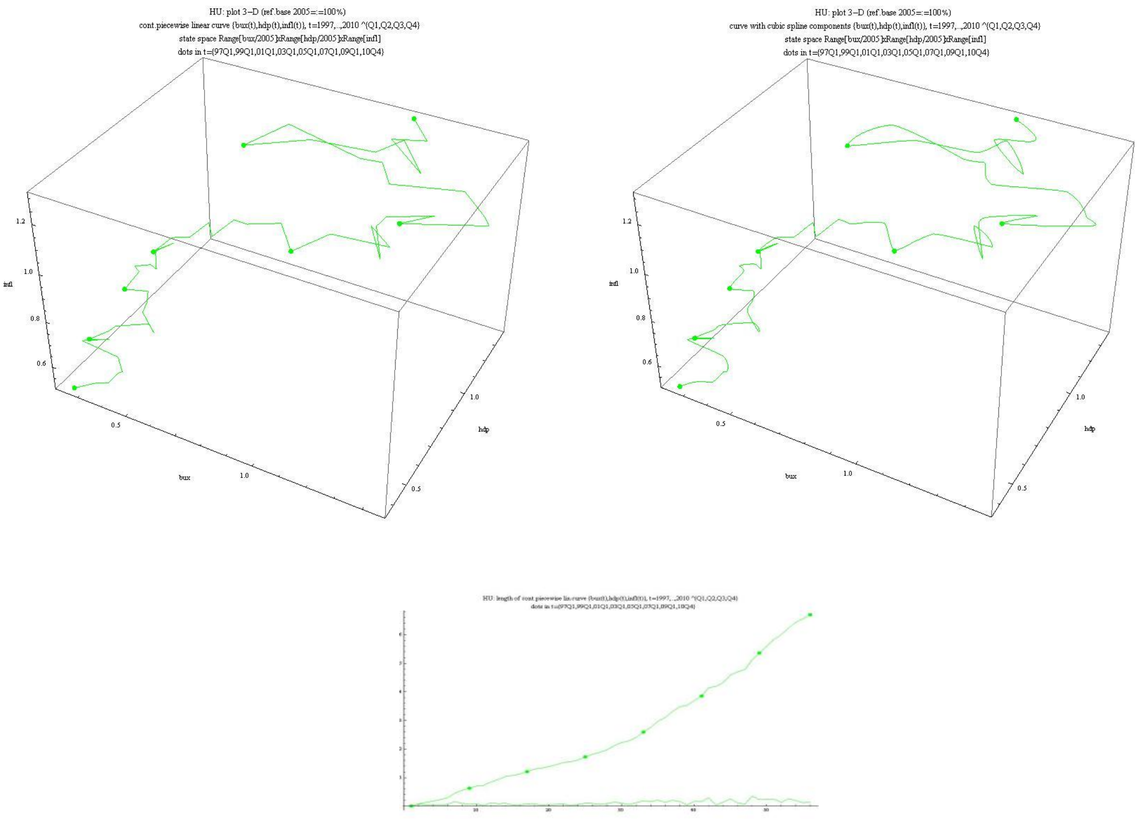


Fig. 2 HU – piecewise linear ${}_2\boldsymbol{p}(t)$, cubic spline ${}_2\boldsymbol{q}(t)$, length of ${}_2\boldsymbol{p}(t)$.

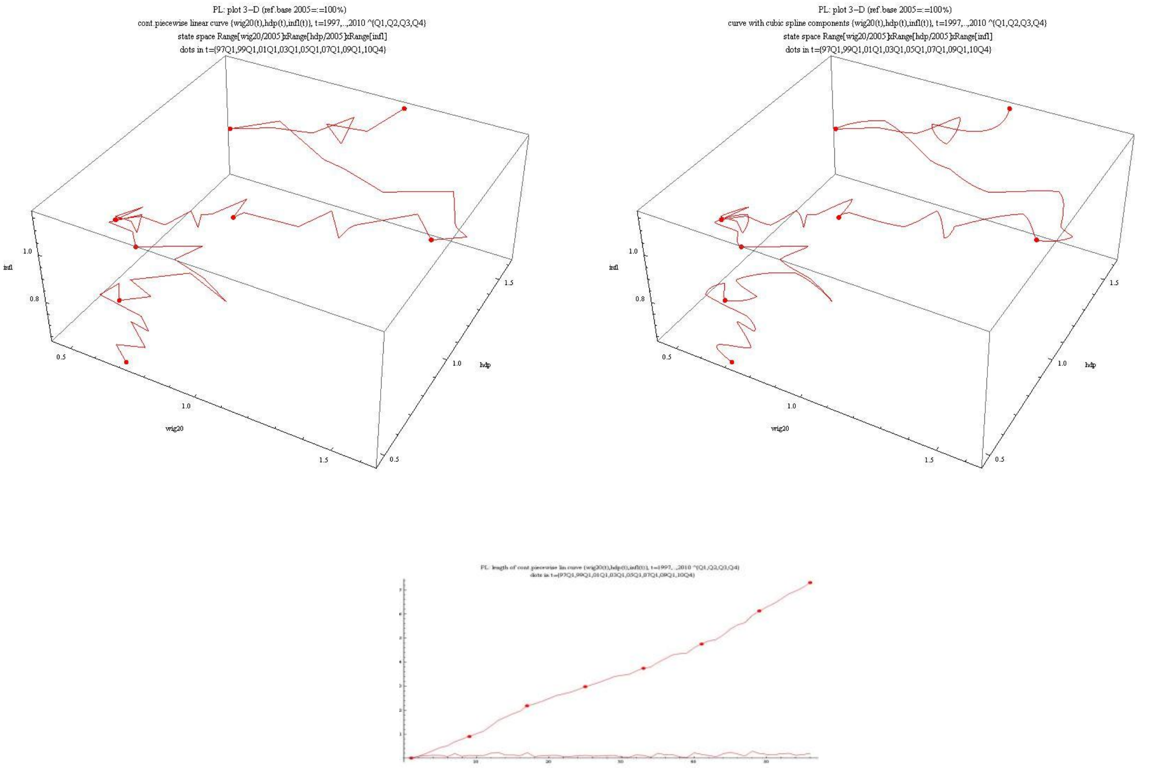


Fig. 3 PL – piecewise linear ${}_3\boldsymbol{p}(t)$, cubic spline ${}_3\boldsymbol{q}(t)$, length of ${}_3\boldsymbol{p}(t)$.

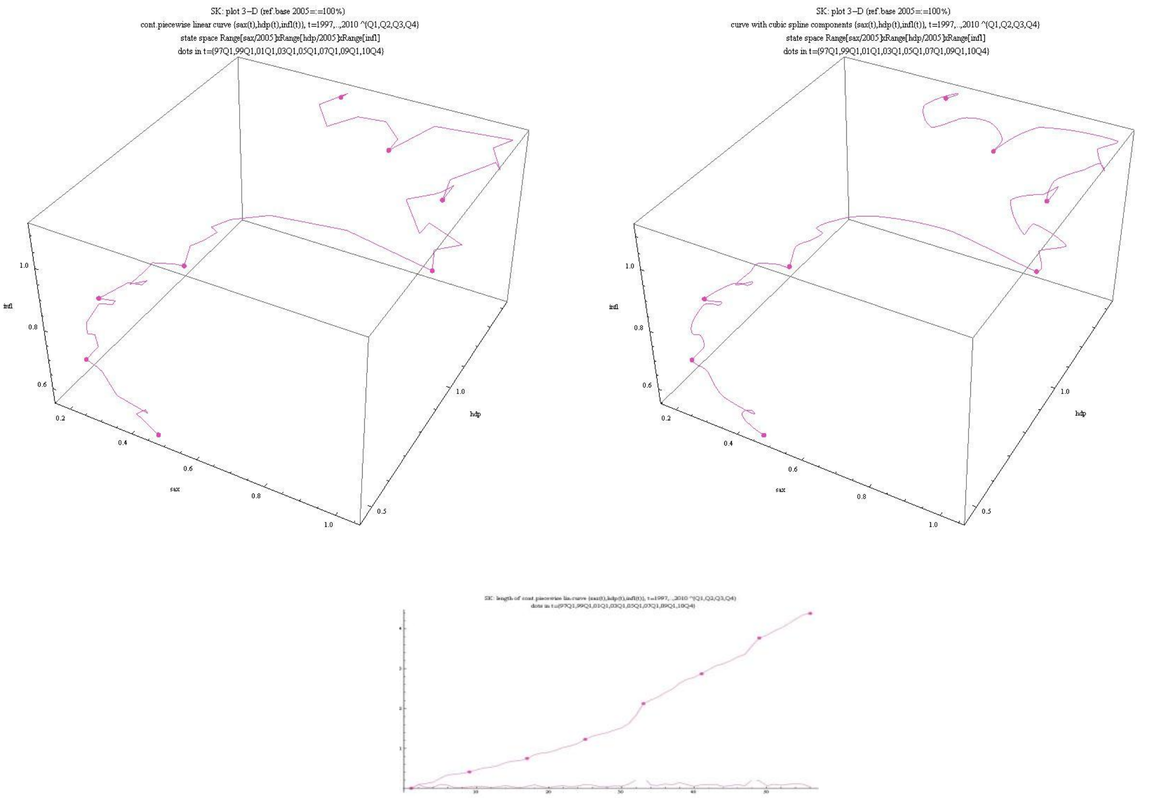


Fig. 4 SK – piecewise linear ${}_4\boldsymbol{p}(t)$, cubic spline ${}_4\boldsymbol{q}(t)$, length of ${}_4\boldsymbol{p}(t)$.

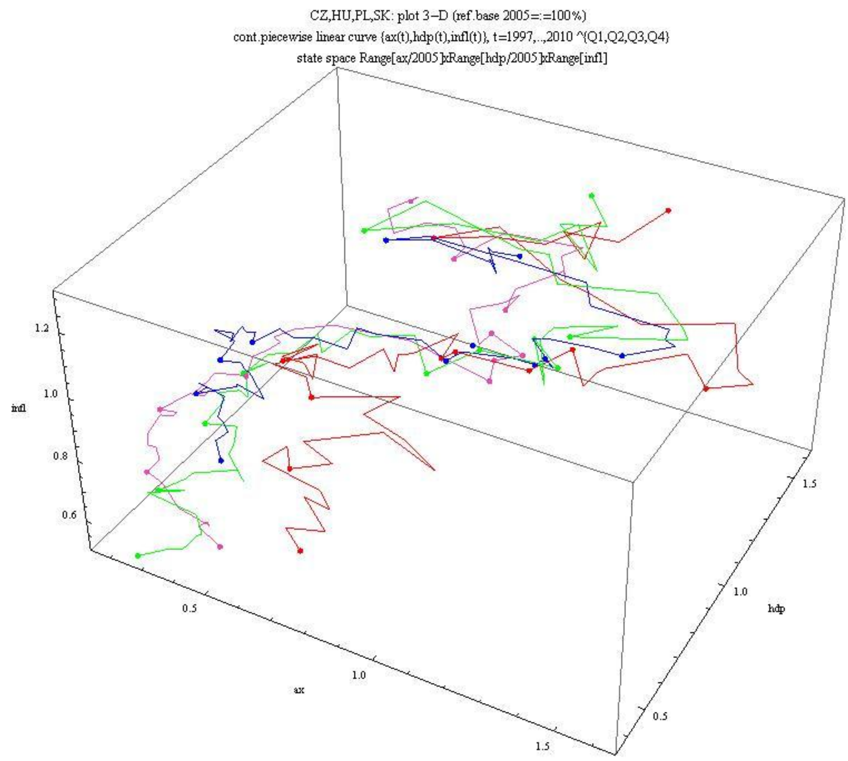


Fig. 5 CZ-HU-PL-SK – $\{{}_1\boldsymbol{p}(t), {}_2\boldsymbol{p}(t), {}_3\boldsymbol{p}(t), {}_4\boldsymbol{p}(t)\}$ with 2005 Q1,2,3,4 dots.

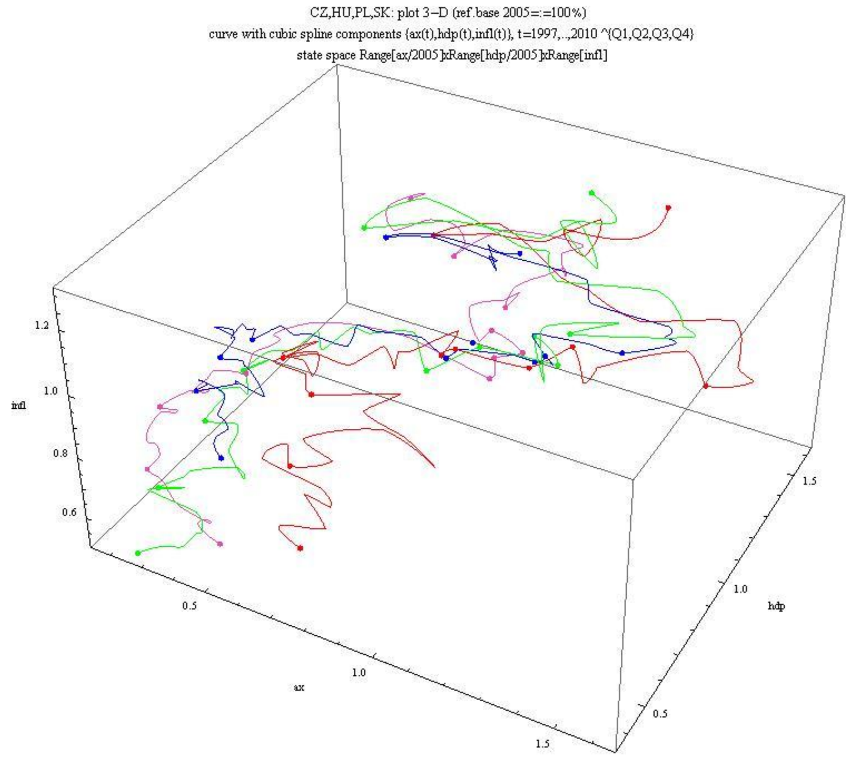


Fig. 6 CZ-HU-PL-SK – $\{{}_1\boldsymbol{q}(t), {}_2\boldsymbol{q}(t), {}_3\boldsymbol{q}(t), {}_4\boldsymbol{q}(t)\}$ with 2005 Q1,2,3,4 dots

3.3 Third result

Lower parts of Figures 1-4 present lengths of curves ${}_k\boldsymbol{p}(t)$, and Fig. 7 shows all lengths of {CZ, HU, PL, SK} macroeconomic development curves ${}_k\boldsymbol{p}(t)$, k=1,2,3,4. The longest curve belongs to HU, the second one as to the lenght has PL, and finally CZ an SK curves are very similar at all. However, inspecting precisely the global lengths by elevation of the 2010Q4 points we identify the shortest path belongs to SK and the length of CZ curve is a little bit longer. Just above time axis t the length increments between adjacent quarter data are depicted. The total length of ${}_k\boldsymbol{p}(t)$ is calculated simply by addition of corresponding increments.

Differential geometry introduces moving Frenet frame along a curve and provides definition of curve length in 3-D. Let $\mathbf{z}(t)=(z_1(t), z_2(t), z_3(t))^{\mathrm{T}}$ be a smooth curve defined over an interval $[a,b]$. Its length $l(a,b)$ is given by (4).

$$l(a,b) = \int_a^b \|\dot{\mathbf{z}}(t)\| dt\,,\quad \dot{\mathbf{z}}(t) = (\,\dot{z}_1(t)\,, \dot{z}_2(t)\,,\ \dot{z}_3(t)\,)^{\mathrm{T}}\,,\ \dot{z}_n(t) = \frac{dz_n(t)}{dt}\,, n=1,2,3. \qquad (4)$$

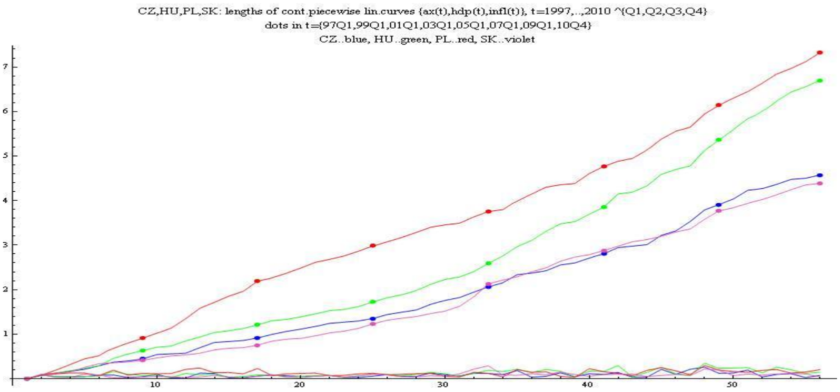


Fig. 7 CZ-HU-PL-SK – lengths of $\{{}_1\boldsymbol{p}(t), {}_2\boldsymbol{p}(t), {}_3\boldsymbol{p}(t), {}_4\boldsymbol{p}(t)\}$.

4 Conclusion

We have analyzed some macroeconomic development curves of V4 countries and their lengths by differential geometry approach. We consider the analysis interesting and promising, and further research is ongoing.

Acknowledgements

Authors acknowledge gratefully funding by the project LC06075 of the Ministry of Education, Czech Republic.

References

[1] Bartóková, L.: The Evolution of the Monetary Policy Transmission Mechanisms of V4 Countries. *E+M Ekonomie a Management* (2/2010), 6-18.

[2] Chern, S-S., Chen, W., Lam, K.S.: *Lectures on differential geometry*. World Scientific Publ., Singapore, 2000, ISBN 981-0234-945.

[3] Gray, A., Abbena, E., Salamon, S.: *Modern Differential Geometry of Curves and Surfaces in Mathematica*. 3.e., , Chapman and Hall CRC, 2006, pp.1016, ISBN 1584884487.

[4] Skála, M.: New Keynesian Macroeconomics – A New Perspective on Labour Market and Macroeconomic Stability (in czech). *E+M Ekonomie a Management* (4/2010), 6-16.

[5] Telínová, L.: Comparison of stock exchange of the Visegrad group (in czech). Plzeň: ZČU/FEK, Master thesis, 2011, pp.80.

Importance of the interactions between Czech and Slovak economy

Veronika Beláková[1]

Abstract. The paper tries to answer the question to what extent there are interactions between Czech and Slovak economy. For this purpose, a DSGE model of two small open economies with foreign sector is employed. Both small open economies consist of four types of agents: households, firms, importers and monetary authority. Compared to the model framework with one small open economy, the model allows for trade between the small open economies which implies that there are two types of importers according to the origin of the imported goods. The model framework assumes monopolistically competitive firms, complete international financial markets, perfectly competitive labour market and allows for deviations from the law of one price. Four variants of model are estimated using Bayesian method (incorporated in Dynare toolbox for Matlab). The variants differ in the assumptions about interactions between Czech and Slovak economy. The variant that best explains data is used to assess the importance of interactions between these two economies. To complete the assessment of interactions between these two economies, the paper presents forecast error variance decomposition of shocks and shock decomposition of selected variables. The contribution of the paper is in the recommendation for economic policy makers as to whether they should take into account the business cycle of the neighbour economy.

Keywords: Bayes factor, interactions, Czech economy, real and nominal rigidities, Slovak economy, two-country model with exogenous foreign sector.

JEL Classification: C51, C52, F41, F42

1 Introduction

This article is a follow-up to Beláková, Vašíček [2] which analyzed behaviour of Czech and Slovak economy. The model used in [2] assumes that representative household and firms in both economies have identical preferences. This assumption means that structural parameters connected with these types of agents were equal. But the territorial structure of foreign trade turnover indicates that influences of neighbour economy to monitoring economy are probably different.[2] For Slovak economy, an average of foreign turnover come from the Czech Republic during 1999-2010 is 16%. Considering only the foreign turnover between the Czech economy and group of economies of Euroarea-12, average share of turnover coming from the Czech economy is approximately 25%. On the other hand, on average, 7 percent of Czech foreign trade turnover come from Slovakia. If we consider distribution of foreign trade turnover between Slovak and Euroarea-12 economies, the foreign trade turnover coming from Slovakia accounts for 10%. These figures show that the influence of Czech economy to Slovak one's could be different as in the inverse case.

The aim of the paper is to quantify the interactions between Czech and Slovak economy. We also try to answer the question if policy makers of one economy should take into account the business cycle of the second economy. For this purpose we estimate a DSGE two-country model with exogenous foreign sector. Macroeconomic time series of Slovak, Czech and Euroarea-12 economies enter the estimation. To achieve the aim of the paper, the estimated values of selected structural parameters, shock decompositions and variance decompositions are discussed. The assessment of importance of interactions between Czech and Slovak economy is based on the Bayes factor of four variants of baseline model.

[1] Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, 60200 Brno, Czech Republic, ciganova.veronika@gmail.com.

[2] In this paper, the monitoring economies are the Czech and Slovak ones. The neighbour economy for Slovakia is the Czech economy and the Slovak economy is the neighbour economy for the Czech Republic.

2 The baseline model

In this section we present main characteristics of the used model and the main differences against the model used in [2]. The model contains two open economies which can influence each other but are not able to affect exogenous foreign sector. The open economies have similar structure, but their structural parameters may differ.

Let us describe the features of the Slovak economy. There are four types of agents: a representative household, home producers, importers and monetary authority.[3] The representative household maximizes its utility function with respect to budget constraint. Its utility is increasing in consumption above the previous consumption – the model contains habit formation - and decreasing in hours of labour. The consumption basket consists of goods produced in Slovakia and goods imported to Slovakia either from the Czech Republic or from Eurozone-12. From the decisions-making of household we get relations for intratemporal and intertemporal decisions and demand functions.

Both types of firms – home producers and importers - that trade on Slovak goods market are monopolistically competitive. They have monopoly power to set their price optimally. The calculation of the optimal price stems from maximization of net present value of their profits subject to their demand function. Only a part of the firms reset their price in any period. The rest of firms either adjust their prices according to indexation rule (which depends on inflation in previous period) or they leave their price without any change. This type of price setting is called Calvo price settings with indexation. The home producers set their prices as a mark-up to real marginal costs. The existence of importers is connected with deviation from the law of one price which means that the price of Czech goods in the Czech Republic expressed in units of Slovak economy is different from the price which Slovak buyer pays for it. The importers also set the prices as mark-ups to their real marginal costs which are the deviations from the law of one price. We assume that the deviations from the law of one price differ for importers from the Czech Republic and Euroarea-12.

Monetary policy is conducted according to Taylor rule. Policy responds to values of inflation and output. Fiscal policy is a zero debt policy which is not able to affect the macroeconomic time series. Other assumptions of the model are a complete international financial market and a perfect competitive labour market. The foreign sector is represented by the group of countries Euroarea-12 and introduced into the model by the AR(1) processes. This type of foreign sector structure enter into the model assumption that foreign sector is not affected by Czech and Slovak economy. This also means that the foreign sector is asymptotically closed and its terms of trade are asymptotically one.

There are eleven shocks in the model. Three shocks come from the AR(1) processes of foreign sector. Four shocks stem in the Slovak economy: preference shock, technological shock, monetary policy shock and shock in the real exchange rate. The equal source of shocks is in the Czech economy. The monetary policy shocks are i.i.d., other shocks are first-order autoregressive processes.

The difference between the model used in this article and the model used in [2] is in the structure of import sector and in the fact that we distinguish between Czech and Slovak structural parameters. With respect to large number of endogenous variables (which indicates the same quantity of equations) we do not introduce the log-linearized form of the model.

3 Estimation of baseline model

Real macroeconomic time series of the Czech Republic, Slovakia and the group of countries EA-12 enter the estimation. Quarterly data from 1Q1999 to 3Q2010 of these three countries on output, inflation, interest rate and real exchange rate are used. All data enter the estimation as deviations from the long run trend which is computed by HP filter and they are not annualized. All data are downloaded from the Eurostat database and are seasonally adjusted by X-12 ARIMA incorporated in IRIS toolbox for Matlab. The following list presents more detailed information about data:

- y_t^{CZ} , y_t^{SK} and y_t^{EA} are output gap of seasonally adjusted real GDP of the Czech Republic, Slovakia and Euroarea-12;
- π_t^{CZ} , π_t^{SK} and π_t^{EA} are seasonally adjusted q-o-q HCPI inflation gaps of the Czech Republic, Slovakia and Euroarea-12;
- i_t^{CZ} , i_t^{SK} and i_t^{EA} are 3-month nominal interest rate gaps of PRIBOR, BRIBOR and EURIBOR;[4]

---

[3] Fiscal policy is specified as zero debt policy with non-distortionary taxes.

[4] Because Slovakia is member of the eurozone from 2009Q1, data of Slovak nominal interest rate is equal to EURIBOR from this period on.

- q_t^{CZ} and q_t^{SK} are expressed in log-deviations from the long time trend of the real exchange rates of the Czech and Slovak Republic.[5]

Model is estimated using Bayesian method, specifically by Metropolis Hastings algorithm, which is incorporate in Dynare toolbox for Matlab. Sources of used prior settings of structural parameters are in [8], [7], [9], [11] and [10]. The prior setting and estimated values of structural parameters and standard deviation of shocks can be found in Table 3 and Table 4 in Appendix A.

In this section we only mention estimated values of parameters which are connected to the aim of the paper or are new compared to [2].The most important parameter is the parameter that represents the share of imports coming from neighbour small open economy in imported consumption of monitoring small open economy. For Slovakia, the estimated value of this parameter is 0.242 which means that 24.2 per cent of Slovak imports come from Czech economy. The ratio of Czech imports which origin in Slovakia is only 2.7 per cent. These results indicate comparatively higher linkage of Slovak economy with Czech one than linkage of Czech economy with Slovak one. A new parameter is a parameter of price stickiness of imports coming from neighbour economy. A Slovak Calvo parameter is 0.776 and compared to the Czech one (0.703), the price contracts last one quarter longer. Parameters of indexation of Slovak importers who import from Czech economy and Czech importers who imports from Slovakia are the same. As the posterior distributions copy the prior counterparts there is not enough information about these parameters in data.

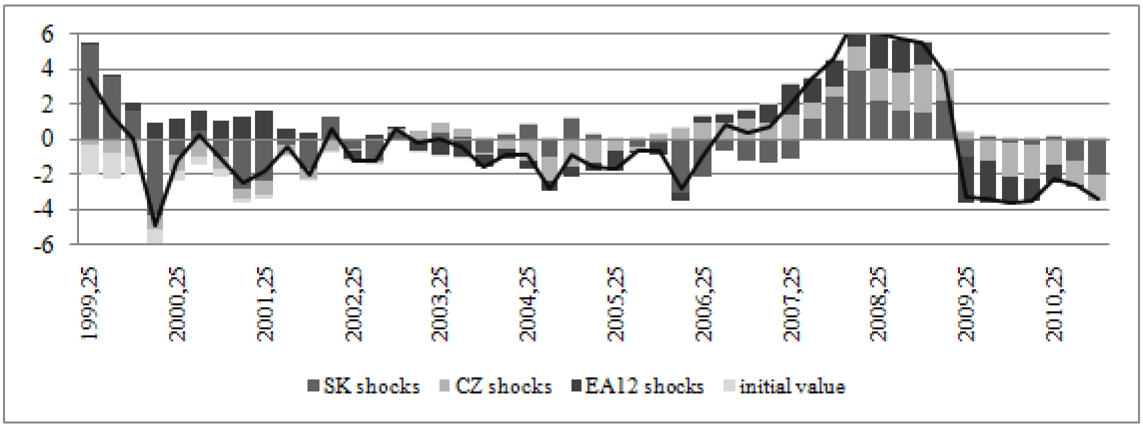


Figure 1 Shock decomposition of Slovak output gap

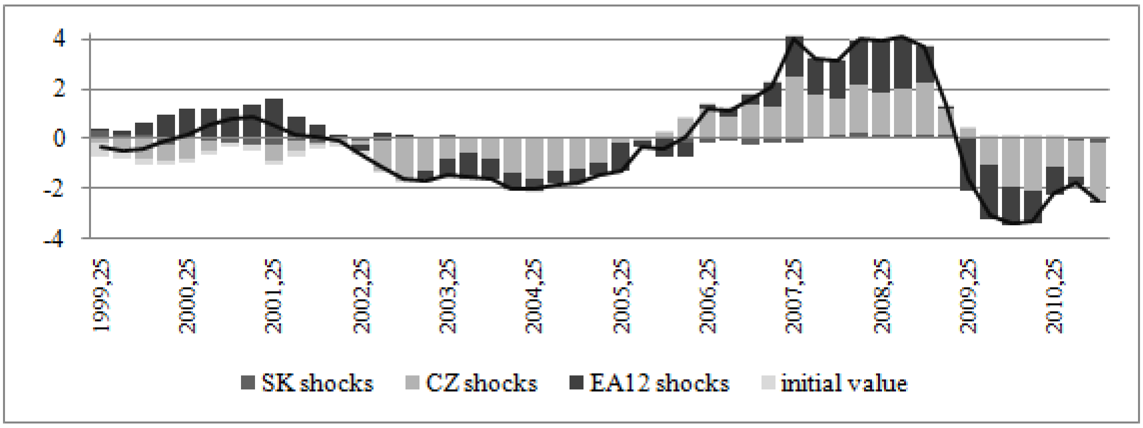


Figure 2 Shock decomposition of Czech output gap

The next step of our investigation of importance of interactions between the two economies is shock decomposition. In this paper you can only see shock decompositions of Slovak (Figure 1) and Czech (Figure 2) output gap. It is useful for our purpose to see from which economy shocks originate. In the figures there are contributions of shocks coming from Slovakia, the Czech Republic and the group of countries EA-12.[6] The difference between these two figures is in the influence of the neighbour country's shocks to the output gap of monitoring country. In Slovakia, the contribution of the Czech shocks is more evident. The share of contributions of Czech

[5] Conversion rate substitute flexible exchange rate from 2009Q1 for Slovakia.
[6] The contribution of country's shocks is sum of these shocks.

shocks on contributions of all shocks in period from 1999Q1-2010Q3 is more than 20 per cent. On the other hand, the contribution of Slovak shocks to explaining of Czech output gap is only one per cent.[7]

Another way to show the importance of neighbour economy's shocks is forecast error variance decomposition (FEVD). You can see the results of this indicator in Table 1 for Czech and Slovak output gap. FEVD is simulated from the posterior means of structural parameters and shocks. It is computed at the first, fourth and eighth quarter horizons and at long-horizon. Table 1 reports the variance shares summed according to origin of shocks. The variance shares also indicate the relative higher influence of Czech shocks on explanation of Slovak output gap. The variance share of Czech shocks in Slovak output at first horizon is 6.3 per cent and this share is increasing with the length of horizon. At the long-horizon variance it is two times larger. Slovak shocks account for less than three per cent of Czech output gap variance in the short run. At the long-horizon, this variance share is only 1.5 per cent.

The results from the estimation of the baseline model indicate relatively larger importance of Czech macroeconomic time series in explaining Slovak data. The next section investigates this fact by comparison among four models.

horizon	SK output gap			CZ output gap		
	SK shocks	CZ shocks	EA12 shocks	SK shocks	CZ shocks	EA12 shocks
1	0.848	0.063	0.089	0.029	0.714	0.258
4	0.800	0.115	0.085	0.019	0.789	0.192
8	0.782	0.131	0.088	0.016	0.807	0.177
inf	0.778	0.133	0.089	0.015	0.812	0.173

Table 1 Forecast error variance decompositions of Slovak and Czech output gap

4 Are the interactions between Czech and Slovak economy important?

The linkage between Czech and Slovak economy in our model is represented by parameters of the share of imports coming from neighbour economy in imported consumption of the monitoring economy. If the value of this parameter is zero, then the linkage between small open economies will be partly disrupted. In case that both parameters are zero, then the model framework is a combination of two models for small open economies. By settings of these parameters we get four types of model. The first model, which is our baseline model, is labelled "SK↔CZ←EA12**"** and represents a two-country model of small open economies with foreign sector. The second model with switched off both parameters labels "SK←EA12; CZ←EA12". The model labelled "EA12→CZ←SK; SK←EA12" represents the situation when the Czech economy is affected by Slovak and Euroarea-12 economies, but the Slovak economy is affected only by the exogenous foreign sector. The last version of the model describes the case when only Slovak economy is affected by the second small open economy. This model has label "EA12→SK←CZ; CZ←EA12".

The models are compared by using Bayes ratio and results are in Table 2. The reference model is our baseline model whose results of estimation are in previous sections. The Bayes ratio lower than one means that the baseline model has higher conditional probability. Results show that only in comparison with "EA12→SK←CZ; CZ←EA12" Bayes ratio is higher than one, attaining the value of almost 163. According to the guidance for interpreting Bayes factor in [5], there is decisive evidence in favour of this model. We can say that the model "EA12→SK←CZ; CZ←EA12" is the best model among chosen models with respect to used data.

Model	**SK↔CZ←EA12**	**SK←EA12; CZ←EA12**	**EA12→CZ←SK; SK←EA12**	**EA12→SK←CZ; CZ←EA12**
Log Marginal Density	-540.949	-541.295	-542.039	-535.855
Bayes Ratio	1.000	0.707	0.336	162.954

Table 2 Bayes Ratio

[7] The percentage of shock contribution is calculated as the portion of the squared contributions of this shock on sum of squared contributions of all shocks.

5 Conclusions

This paper tries to answer the question as to what extent are important the interactions between Czech and Slovak economy. The results of the Bayesian estimation of the baseline model, i.e. two-country model of small open economies with a foreign sector, imply that the proportion of Slovak imports coming from the Czech Republic is almost one-fourth. On the other hand, the share of Czech imports which origin in Slovakia is only 3 per cent. These results are also supported by shock decomposition and forecast error variance decomposition of output gaps. The main conclusion of this paper is that the Slovak policy makers should take into account the business cycle of the Czech economy besides business cycle of Euroarea-12. For that reason, we conclude that the interactions between Czech and Slovak economy are important for Slovakia. The last statement follows from the results obtained from the comparison of four variants of model.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot, S.: *Dynare: Reference Manual, Version 4*, *Dynare Working Papers*, 1, CEPREMAP, 2010. http://www.dynare.org

[2] Beláková, V., Vašíček, O.: *Analysis of interactions between Czech and Slovak economies*. In Mathematical Methods in Economics 2010, České Budejovice: University of South Bohemia in České Budejovice, 2010. ISBN 978-80-7394-218-2.

[3] Calvo, G.: *Staggered Contracts in a Utility-Maximising Framework*, Journal of Monetary Economics,12, 1983, pages 383-398.

[4] Cigáňová, V., Vašíček, O.: *A Small New Keynesian Model of Slovakia*. In Mathematical Methods in Economics 2009. Praha : Czech University of Life Sciences Prague, 2009. ISBN 978-80-213-1963-9.

[5] DeJong, D., Dave, Ch.: *Structural macroeconometrics*. Princeton : Princeton University Press, 2007. 338 s. ISBN 9780691126487

[6] Galí, J.: *Monetary policy, inflation, and the business cycle : an introduction to the new Keynesian framework*. Princeton : Princeton University Press, 2008. 203 p. ISBN 9780691133164.

[7] Justiniano, A., Preston, B.: *Small open economy DSGE models: Specification, estimation and model fit*, Department of Economic, Columbia University, Working Paper, 2004

[8] Liu, P.: *A small New Keynesian model of the New Zealand economy*, Reserve Bank of New Zealand, Reserve Bank of New Zealand Discussion Paper Series DP2006/03, 2006.

[9] Remo, A.: *Monetary policy and stability of Czech Economy: Optimal Commitment Policy in NOEM DSGE Framework*, CVKS MU, working paper No 28/2008, Brno 2008

[10] Senaj, M., Výškrabka, M., Zeman, J.: *MUSE: Monetary Union and Slovak economy model*, NBS Working paper 1/2010, ISSN 1337-5830.

[11] Vašíček, O., Musil, K.: *Behavior of the Czech Economy: New Open Economy Macroeconomics DSGE Model*. Brno : CVKSČE MU, 2006. 113 p. WP 23/2006.

[12] Wickens, M.: *Macroeconomic theory: a dynamic general equilibrium approach*. Princeton : Princeton University Press, 2008. 477 p. ISBN 9780691116402.

Appendix A

parameters	Prior settings			Posterior estimation					
	distr.	mean	std	mean	conf. interval		A	B	C
the share of imports coming from CZ in SK imports	B	0.20	0.1	0.242	0.074	0.396	0.00	0.00	0.34
degree of indexation: SK producers	B	0.60	0.1	0.471	0.293	0.641	0.46	0.45	0.48
Calvo parameter: SK producers	B	0.75	0.1	0.540	0.395	0.673	0.60	0.61	0.50
degree of indexation: SK importers importing from CZ	B	0.60	0.1	0.601	0.434	0.762	0.60	0.60	0.60
Calvo parameter: SK importers from CZ	B	0.75	0.1	0.776	0.637	0.918	0.75	0.76	0.78
degree of indexation: SK importers importing from EA12	B	0.60	0.1	0.527	0.347	0.700	0.52	0.51	0.54
Calvo parameter: SK importers from EA12	B	0.75	0.1	0.471	0.304	0.648	0.47	0.42	0.52

parameters	Prior settings distr.	mean	std	Posterior estimation mean	conf. interval		A	B	C
inverse elasticity of labour supply in SK	G	1.00	0.5	1.107	0.352	1.824	1.47	1.43	1.05
inverse elasticity of intertemporal subst. in consumption in SK	G	1.00	0.5	0.449	0.177	0.697	0.50	0.44	0.49
habit formation parameter in SK	B	0.64	0.1	0.552	0.398	0.702	0.58	0.59	0.52
smoothing of interest rate in SK	B	0.75	0.1	0.902	0.871	0.934	0.91	0.91	0.90
weight of monetary policy on inflation in SK	G	1.50	0.2	1.439	1.143	1.734	1.47	1.48	1.43
weight of monetary policy on output gap in SK	G	0.13	0.04	0.135	0.071	0.197	0.14	0.14	0.13
inertia of technology shock in SK	B	0.50	0.2	0.224	0.034	0.402	0.23	0.23	0.24
inertia of preference shock in SK	B	0.50	0.15	0.493	0.241	0.743	0.56	0.61	0.43
elasticity of subst. between SK produced and imported goods	G	1.00	0.5	0.920	0.235	1.566	0.84	0.85	1.09
elasticity of subst. between CZ goods and EA12 goods in SK	G	1.00	0.5	0.979	0.225	1.710	1.01	0.79	0.99
the share of imports coming from SK in CZ imports	B	0.20	0.1	0.027	0.007	0.047	0.00	0.05	0.00
degree of indexation: CZ producers	B	0.60	0.1	0.482	0.321	0.640	0.49	0.49	0.49
Calvo parameter: CZ producers	B	0.70	0.1	0.730	0.642	0.815	0.72	0.70	0.71
degree of indexation: CZ importers importing from SK	B	0.60	0.1	0.603	0.435	0.762	0.60	0.60	0.60
Calvo parameter: CZ importers from SK	B	0.70	0.1	0.703	0.551	0.861	0.69	0.68	0.70
Degree of indexation: CZ importers importing from EA12	B	0.60	0.1	0.566	0.396	0.737	0.56	0.57	0.56
Calvo parameter: CZ importers from EA12	B	0.70	0.1	0.722	0.613	0.838	0.71	0.75	0.71
inverse elasticity of labour supply in CZ	G	1.00	0.5	1.286	0.465	2.074	1.28	1.33	1.25
inverse elasticity of intertemporal subst. in consumption in CZ	G	1.00	0.5	0.519	0.185	0.842	0.49	0.53	0.45
habit formation parameter in CZ	B	0.80	0.1	0.844	0.759	0.933	0.85	0.83	0.85
smoothing of interest rate in CZ	B	0.70	0.1	0.910	0.881	0.939	0.91	0.91	0.91
weight of monetary policy on inflation in CZ	G	1.50	0.2	1.409	1.126	1.684	1.42	1.42	1.41
weight of monetary policy on output gap in CZ	G	0.13	0.04	0.157	0.090	0.223	0.15	0.15	0.17
inertia of technology shock in CZ	B	0.50	0.2	0.262	0.067	0.458	0.28	0.28	0.27
inertia of preference shock in CZ	B	0.50	0.15	0.585	0.381	0.791	0.61	0.51	0.64
elasticity of subst. between CZ produced and imported goods	G	1.00	0.5	0.627	0.218	1.020	0.59	0.63	0.65
elasticity of subst. between SK goods and EA12 goods in CZ	G	1.00	0.5	0.901	0.228	1.538	1.00	0.99	0.85
inertia of EA12 output gap	B	0.80	0.1	0.843	0.765	0.923	0.84	0.85	0.84
inertia of EA12 interest rate	B	0.80	0.1	0.781	0.693	0.869	0.78	0.77	0.79
inertia of EA12 inflation	B	0.80	0.1	0.481	0.348	0.611	0.48	0.48	0.47
inverse elasticity of intertemporal subst. in consumption in EA12	G	1.00	0.5	0.797	0.212	1.352	0.88	0.93	0.72
habit formation parameter in EA12	B	0.70	0.1	0.623	0.470	0.780	0.61	0.62	0.63
elasticity of substitution among imported goods in EA12	G	1.00	0.5	0.127	0.040	0.209	0.10	0.13	0.11
inertia of real exchange rate shock in SK	B	0.60	0.15	0.651	0.512	0.787	0.64	0.64	0.66
inertia of real exchange rate shock in CZ	B	0.60	0.15	0.646	0.513	0.790	0.65	0.64	0.66

Table 3 Prior settings and the estimated values of structural parameters. There are also estimated values of structural parameters for next three variants of model - columns A, B and C. Column A correspond to model "SK←EA12; CZ←EA12", B = "EA12→CZ←SK; SK←EA12" and C ="EA12→SK←CZ; CZ←EA12".

standard deviation of shocks	Prior settings distr.	mean	std	Posterior estimation mean	CI		A	B	C
technological shock in SK	INVG	1.00	Inf	7.064	2.875	10.878	8.37	8.13	6.14
preference shock in SK	INVG	1.00	Inf	9.747	4.663	14.791	11.66	11.46	9.26
monetary policy shock in SK	INVG	1.00	Inf	0.291	0.234	0.348	0.29	0.29	0.29
technological shock in CZ	INVG	1.00	Inf	6.695	2.247	11.094	6.31	5.27	5.89
preference shock in CZ	INVG	1.00	Inf	8.872	4.235	12.861	8.18	7.44	7.92
monetary policy shock in CZ	INVG	1.00	Inf	0.150	0.124	0.175	0.15	0.15	0.15
EA12 output gap shock	INVG	1.00	Inf	0.681	0.571	0.793	0.69	0.69	0.68
EA12 interest rate shock	INVG	1.00	Inf	0.161	0.132	0.187	0.16	0.16	0.16
EA12 inflation shock	INVG	1.00	Inf	0.340	0.283	0.397	0.34	0.34	0.34
real exchange rate shock in SK	INVG	1.00	Inf	0.931	0.571	1.283	0.97	0.96	0.91
real exchange rate shock in CZ	INVG	1.00	Inf	0.795	0.469	1.116	0.78	0.83	0.74

Table 4 Prior settings and the estimated values of standard deviations of shocks. There are also estimated standard deviations of shocks for next three variants of model - columns A, B and C. Column A correspond to model "SK←EA12; CZ←EA12", B = "EA12→CZ←SK; SK←EA12" and C ="EA12→SK←CZ; CZ←EA12".

Multi-criteria models with trend evaluation in decision making support

Agáta Bodnárová [1], Martin Gavalec [2], Karel Mls[3]

Abstract. Decision making support by multi-criteria models with evaluation of trends is considered. Multi-criteria models are designed to satisfy the expectations of customers, and data based on customer preferences are used for modification of the parameters in the multi-criteria model. Various segments of customers are found and observed by cluster analysis. Prevailing trends in customers' behavior are reflected in trends for individual clusters, which enables predicting the future customers' preferences in individual segments. The recently developed method of cognitive hierarchy process (CHP) is applied to find the optimal offer of goods with respect to the customer needs and expectations.

Keywords: decision making, multi-criteria models, trend evaluation
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Computer support of managers' decision making, typically on operational level, is broadly studied and there are many real applications as well. In this paper, the role of clustering of data sets and the fuzzy trend identification and its contribution to better prediction of awaiting data in the decision process is presented. The main tool used by the method is the evaluation of fuzzy cognitive maps described in [5]. The basic idea of the CHP method is similar to the well-known analytical hierarchy process (AHP) method, invented by T. L. Saaty [18], but CHP uses individual evaluations for sets of criteria in every alternative.

2 Multi-criteria models

The method of cognitive hierarchy process and trend evaluation is applied in this paper to optimizing a general business portfolio. The model describes a standard decisional situation, in which the manager is preparing the new commercial offer for the coming period. We assume that in the previous period, the CHP method was used to help customers to find their most suitable destination according to their individual preferences. The records of the decisional process accumulated from all customers during the history, contain useful information which can be used by the managers for optimizing the next version of the offer.

The goals of the decision-making process in this situation are:

- to optimize the future profit under incomplete information
- to improve the business portfolio
- to fit offers to individual customers' preferences
- to address expected future customers' demands

The managers consider the characteristic properties of available alternatives involved in the company's portfolio. Data related to any particular offer are given and their position in the decisional model can be taken as constant. Based on the records of customers' responses from the previous period, the alternatives

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: agata.bodnarova@uhk.cz
[2]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[3]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: karel.mls@uhk.cz

are evaluated. Then the worst ones will be left out, and new alternatives with the characteristic properties of the most successful ones will be added to the portfolio.

In the above approach, the information about the preferences of different types of customers is suppressed. Hence, some minor segments of customers might be neglected and possibly lost. To avoid this situation, the evaluation and optimization of the portfolio should also consider the preferences of significant groups of customers. The relevant groups of customers are identified from the available records by the methods of cluster analysis.

3 Fuzzy Cognitive Maps

Cognitive map (CM) as a modeling tool was introduced by Axelrod [1] as a system composed of the set of concepts and the set of causal relationships. Each particular concept influences other related concepts via causal relationships in positive or negative sense, and there are no interactions between independent concepts. A cognitive map can be represented by a directed graph, where concepts of the CM correspond to nodes of the graph and causal relationships correspond to arcs oriented from the cause concepts to the effect concepts. Causality strength is expressed by signs $+$ and $-$ as positive or negative dependence between concepts.

The notion of a fuzzy cognitive maps (FCM) was proposed by Kosko [10], [11] and later enriched by several authors [9], [14], [17]. Involving fuzzy logic helps to solve problems with respect to human-like way of thinking. FCM works on the principle that the causal relationships and concepts are accompanied by a number within the real unit interval $\langle 0, 1\rangle$. By this evaluation, fine differences in causal relationships can be expressed and partial activation of concepts can be used, in contrast to the binary activation in CM.

In this paper, a fuzzy cognitive map is formally defined as an ordered pair $M = (C, A)$, where C is a finite set of cardinality $|C| = n$ whose elements are called *concepts*, and A is a matrix of type $n \times n$ with values in the real interval $\langle 0, 1\rangle$ (alternatively, in $\langle -1, 1\rangle$). The entries of matrix A are interpreted as the levels of causal relations between pairs of concepts in C. Further, we shall consider an evaluation vector of the fuzzy cognitive map M, which is defined as a mapping $e : C \to \langle 0, 1\rangle$ and its values are interpreted as activation levels of concepts in C. Decision support, and model behavior prediction as well, represent the most often cited domains of FCMs utilization, see [2], [8], [13], [16], [19].

4 Cognitive hierarchy process

A standard approach in decision making is based on dividing the decision problem into smaller parts (alternatives, criteria, goals). Evaluations of the importance degree of various objectives and preferences for alternative solutions are then used to find the final decision. The well-known analytical hierarchy process (AHP) methodology uses a relative normalization approach, in which the total sum of the weights of all subcriteria for a given parent criterion is equal to 1, see [3], [7].

Fuzzy cognitive maps as supporting tool for decision making process were considered in [4]. The cognitive hierarchy process (CHP), described in [5], merges AHP and evaluation of complex systems of objectives described by fuzzy cognitive maps. In contrast to the AHP method, which works with fixed evaluation of relative weights of different alternatives (cases), CHP uses a specific system of weights for every individual alternative.

CHP works with a system of individual FCM's unified by a common template. We say that a fuzzy cognitive map $M^* = (C^*, A^*)$ is a template for a system

$$\mathcal{M} = \big(M_s; s \in \mathcal{I}\big)$$

of individual FCM's $M_s = (C_s, A_s)$, if $C_s = C^*$ and $A_s \leq A^*$ holds true for every individual $s \in \mathcal{I}$.

For the purpose of CHP, a tree structure of the template $M^* = (C^*, A^*)$ is assumed. C^* denotes the set of template nodes, A^* is the set of weighted template edges (zero-weighted edges are not considered) and M^* is a root-tree with the root $c_0 \in C^*$. Any node $c \in C^*$, $c \neq c_0$ has the unique predecessor denoted by $p(c)$, and for any node $c \in C^*$, $S(c)$ denotes the set of all successors of c.

For every fuzzy cognitive map $M_s \in \mathcal{M}$, an individual evaluation mapping $e_s : C \to \langle 0,1 \rangle$ is defined recursively, according to the tree structure of template M^*. The detailed formulas are given below in Section 5.

5 Mathematical model

The CHP method can by applied in various conditions. In this section we describe a mathematical model for the situation, when the deciding manager has at disposal a set of data on customer preferences, and the data should be processed and considered in designing the optimal business portfolio of alternatives.

The customer preferences to available alternatives consist of several criteria, which need not be independent. The criteria are considered as cognitive concepts, and for description of their mutual interdependence the optimization model uses cognitive maps related to a fixed template $M^* = (C^*, A^*)$.

An example of the structure of a template cognitive map of dimension $n = 16$ is shown in Figure 1.

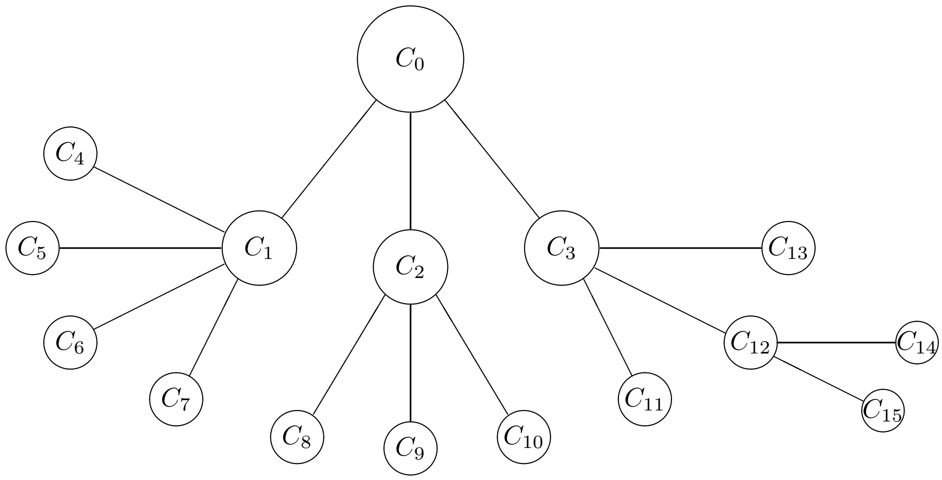


Figure 1: Structure of a template cognitive map

We assume that M^* possesses a tree structure with the root $c_0 \in C^*$. Moreover, we assume that the nodes are ordered in a sequence $(c_0, c_1, \ldots, c_{n-1})$, in which the predecessor $p(c)$ of every node $c \in C^*$ occurs before the node c itself. Hence, the first node in the sequence must be the root node c_0, which is the only one with no predecessor. Any node $c \in C^*$ with $S(c) = \emptyset$, i.e. any node without successors, is called a *leaf* in the tree M^*. The set of all leaves is denoted by L.

The current portfolio contains r alternatives $\mathcal{D} = \{D_1, D_2, \ldots, D_r\}$ and the characteristic properties of each alternative $D \in \mathcal{D}$ are given as fuzzy values $h(D, c)$ for all leaves $c \in L$. The value $h(D, c)$ describes the degree of correspondence of the particular alternative D to the leaf concept c.

The data from the previous activity period describe the priorities of s customers $\mathcal{B} = \{B_1, B_2, \ldots, B_s\}$. The priorities of any particular customer $B \in \mathcal{B}$ are fuzzy values $q(B, c)$ for all non-root nodes $c \neq c_0$. Every such value is interpreted as the degree of relative importance of the concept c, in relation to its parent concept $p(c)$, from the subjective point of view of the customer B.

Different segments of customers can be found by methods of fuzzy cluster analysis (FCA) according to the customers' preferences. We get a set of t customer groups (clusters) $\mathcal{G} = \{G_1, \ldots, G_t\}$. To each cluster $G \in \mathcal{G}$, there are assigned its own priorities $\bar{q}(G, c)$ and a membership function $\mu_G : \mathcal{B} \to \langle 0,1 \rangle$ expressing the degree of membership for each particular customer $B \in \mathcal{B}$ in the cluster G.

In the evaluation procedure, a system $\mathcal{M} = \big(M_B; B \in \mathcal{B}\big)$ of individual fuzzy cognitive maps $M_B = (C_B, A_B)$ related to the template system M^* is defined by putting $C_B = C^*$ and by putting $A_B(p(c), c) = q(B, c) \cdot A^*(p(c), c)$ for every $c \in C^*$.

For each customer $B \in \mathcal{B}$ and for each alternative $D \in \mathcal{D}$, the evaluation mapping $e = e_{BD} : C^* \to \langle 0,1 \rangle$ is defined by a backward recursion, from leaves to the root node:

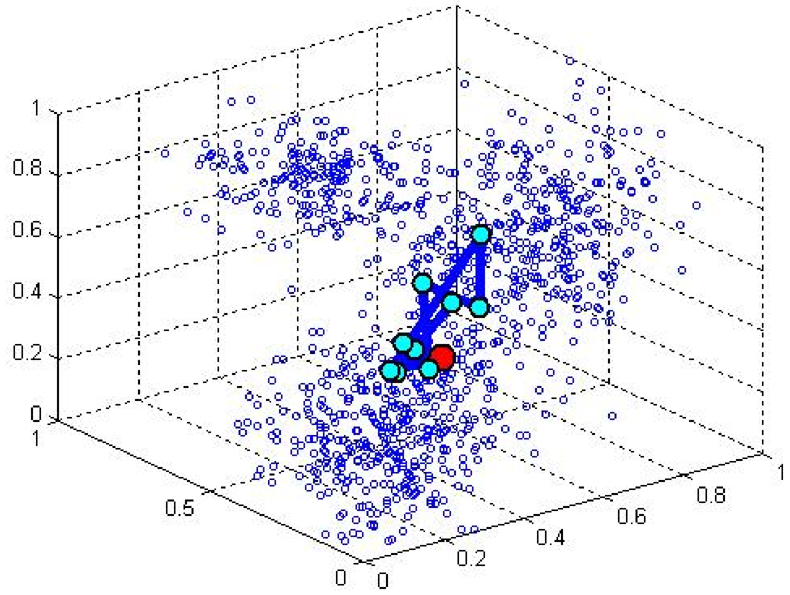


Figure 2: Evolution of population center in 3D projection of 16D space

if $c \in L$ is a leaf in the template tree M^*, i.e. if $S(c) = \emptyset$, then we put

$$e(c) = h(D, c) \tag{1}$$

for every non-leaf node we have $S(c) \neq \emptyset$ and we put

$$e(c) = \frac{1}{|S(c)|} \sum_{k \in S(c)} \sqrt{e(k) \cdot A_B(p(c), c)} \tag{2}$$

For every concept $c \in C$, the value $e(c) = e_{BD}(c)$ means the evaluation of the concept c by the customer B in relation to the alternative D. The evaluation values of the leaf concepts $c \in L$ are given by the customer, according to formula (1), as well as his subjective priorities $A_B(p(c), c)$ for the predecessors of non-leaf concepts $c \notin L$. In the example in Figure 1, the non-leaf concepts are c_0, c_1, c_2, c_3 and c_{12}. For these concepts, the values $e(c) = e_{BD}(c)$ are computed by formula (2).

The root-value of the above evaluation function e_{BD}, computed from the subjective priorities of the customer B, can be interpreted as the aggregated assessment $e_B(D) = e_{BD}(c_0)$ for every alternative D. In the case when the customer B has to decide, he/she chooses the one of the alternatives with the highest assessment $e_B(D)$. In formula (2), a simple aggregation function combining arithmetical and geometrical means is used. However, more sophisticated aggregation functions, e.g. those using weights of individual variables, are also possible.

The manager, willing to optimize the portfolio of alternatives, uses the set of assessments in a more complex way. One possible approach would be assigning weights to every alternative according to the ordering of assessments from every individual customer. The total weighted sum of assessments from all customers, in notation $e(D)$, then shows the aggregated global evaluation of any alternative D.

In deciding how many of the old alternatives should be deleted and which of the new available alternatives should be included into the portfolio for the next season, the aggregated customers priorities $e(D)$ are then used. The newly considered alternatives are added to the set of all previous ones, and the common ordering of the old destinations and new ones gives the answer to the problem.

Different segments of customers are addressed when the evaluation and the assessment of the alternatives is computed separately for different clusters $G \in \mathcal{G}$. The assessments according to cluster priorities $\bar{q}(G, c)$ are used in combination with the weighted mean of individual assessments by customers, weighted by the membership function μ_G. The cluster analysis is also useful in investigation of trends as it is shown in the next section.

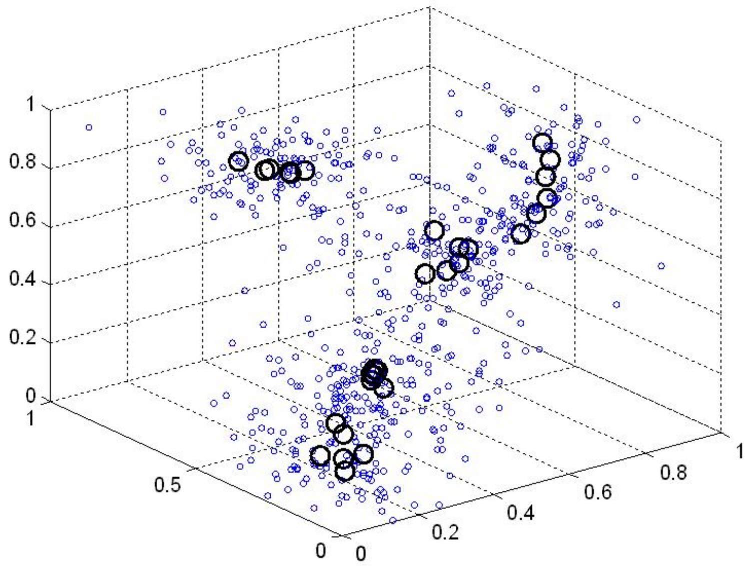


Figure 3: Evolution of multiple cluster centers in 3D projection of 16D space

6 Trend evaluation

The evaluation procedure described above will be more efficient, if time dependence of the available data is considered. The preferences and priorities of customers are changing in time and the observable trends in the evaluation of alternatives can be involved into the decision process. By dividing the data into several time intervals, we get a time series of optimal alternatives according to customers preferences. By standard method of processing the time series we get a (linear or non-linear) estimation of the trend in series of optimal alternatives. The trend (or fuzzy trend) can be further used for prediction of the future optimal alternative [15]. The efficiency of the trend prediction depends on the size of the computed value in comparison with the stochastic component of the available data.

With sufficient amount of data, extended in a reasonable time period, the trends can be computed not only for the group of all customers, but also for different segments of the population. This combination of trend calculation and the cluster analysis turns out to be considerably more efficient.

Two figures demonstrate the advantage of using the cluster analysis on the given data set before computing the trends. The semi-random data were computed in a simulation model, with 10 time steps. In Figure 2, the successive position of the center of the population are connected by lines, so that the stochastic character of the changes without observable trend is clearly visible. (For simplicity, 3D projections with respect to chosen three axis, of the complete 16-dimensional data set, are shown in figures 2 and 3).

The fuzzy cluster analysis revealed five groups of customers with typical preferences for each group. The FCA method produced also membership values for all individual customers, in each of the clusters. The membership functions were used in computing the evaluation of alternatives by customer clusters. Figure 3 shows the evolution of five cluster centers found in the same data set during first 6 steps. Here the present trends of individual clusters can be immediately seen. In similar situation, using the individual trends to predict the future preferences of customers in each cluster, gives significant improvement of the optimal results.

7 Conclusions

A new model using the CHP method with trends and cluster analysis was suggested. Known customer priorities are involved into the managerial decision model for optimizing the portfolio of alternatives. While the global view, inferred from the former ratings of customers, helps to identify optimal alternatives, there are differences in priorities representing the opinion of particular groups of customers. This analysis is combined with trend evaluation which makes prediction of the future customers preferences possible, with stronger impact to the future customers.

Acknowledgements

The support of Czech Science Foundation # 402/09/0405 and MSM0021620838 is kindly acknowledged.

References

[1] Axelrod, R.: Structure of Decision: The Cognitive Maps of Political Elites. Princeton University Press, New Jersey (1976)

[2] Chiu, C., Hsu, P., Norcio, A.: A fuzzy reasoning approach to support knowledge acquisition. International Journal of General Systems, 1999 **28**(2-3): 227–241

[3] Dryer, R.F., Forman, E.A.: An Analytical Approach to Marketing Decisions. Prentice Hall, Englewood Cliffs New Jersey (1991)

[4] Gavalec, M., Mls, K.: Fuzzy cognitive maps and decision making support. Proc. of the 21st Conf. Mathem. Methods in Economics, Czech University of Agriculture in Prague, Prague (2003) 87–93

[5] Gavalec, M., Mls, K.: Cognitive hierarchy process - an approach to decision making support. Proc. of the 24th Conf. Mathem. Methods in Economics, University of Pilsen, Pilsen (2006) 187–193

[6] Gavalec, M., Mls, K.: On-line fuzzy optimization based on cognitive hierarchy process. Abstracts of the 9th Inter. Conf. Fuzzy Sets Theory and its Applications, Liptovský Ján (2008) 46–46

[7] Golden, B.L., Harker, P.T., Wasil, E.A.: The analytical hierarchy process - applications and studies. Springer Verlag, New York (1989)

[8] Khan, M.S., Chong, A., Gedeon, T.: A Methodology for Developing Adaptive Fuzzy Cognitive Maps for Decision Support. Journal of Advanced Computational Intelligence (2000) **4**(6)

[9] Kim, H., Lee, K.: Fuzzy implications of fuzzy cognitive map with emphasis on fuzzy causal relationship and fuzzy partially causal relationship. Fuzzy Sets and Systems (1998) **97**(3): 303–313

[10] Kosko, B.: Fuzzy cognitive maps. International Journal of Man-Machine Studies (1986) **24**(1): 65–75

[11] Kosko, B.: Fuzzy engineering. Prentice-Hall, Upper Saddle River (1997) 549

[12] Krapka, M., Zima, M.: NEFRIT - a tool for multi-criteria fuzzy evaluation (in Czech). In: Inteligentní systémy pro praxi - Fuzzy logika, neuronové sítě, pravidlové expertní systémy. Luhačovice: AD&M Ostrava - Poruba (2000)

[13] Liu, Z., Satur, R.: Contextual fuzzy cognitive map for decision support in geographic information systems. IEEE Transactions on Fuzzy Systems (1999) **7**(5): 495–507

[14] Park, K., Kim, S.: Fuzzy Cognitive Maps Considering Time Relationships. International Journal of Human-computer Studies (1995) **42**(2): 157–168

[15] Rossiter, J.M.: Humanist Computing for Knowledge Discovery from Ordered Datasets. Dissertation, University of Bristol (2000)

[16] Taber, W.R., Siegel, M.: Estimation of Expert Weights with Fuzzy Cognitive Maps. In: 1st IEEE International Conference on Neural Networks (ICNN-87), San Diego (1987)

[17] Tsadiras, A.K., Margaritis, K.G.: Cognitive Mapping and Certainty Neuron Fuzzy Cognitive Maps. Information Sciences (1997) **101**: p. 109–130

[18] Saaty, T.L.: The analytic hierarchy process. McGraw-Hill, New York (1980)

[19] Zeleznikow, J.: Building Decision Support Systems in discretionary legal domains. International Review of Law, Computers&Technology, (2000) **14**(3): p. 341–356

Investment Instrument for Pension Protection

Adam Borovička, Petr Mynařík, Petr Fiala[1]

Abstract. The paper deals with very actual problem of financial funding in a pension age. A conception of the state pension system is very vague for a long time in Czech Republic, so people still focus more and more on creating of personal financial resources, from which they will draw in a retirement age. Starting frame of article is comparison of investment instruments for solution a pension protection. This contribution shows and compares different tools, which can be used to ensure incomes during retirement. We stated five categories of investment instruments – retirement income insurance, capital life insurance, investment life insurance, building savings and special investment programs. The solution procedure has two steps. In first step we choose the options that will be further analyzed. We apply multiple criteria decision making methods in terms of each group to choose effective products which will participate in potential client portfolio. Second step is devoted to create a suitable structure of investment instruments. We outline possible portfolio scenarios within the scope of time period representing approximately productive investor's age. A portfolio composition is primarily influenced by claims on costs, rate of profit and risk rate. The goal of the paper is introduction of principles of dynamic and stochastic programming with practical application in the field of pension protection scheme.

Keywords: pension protection, portfolio, multiple criteria, dynamic and stochastic programming.

JEL Classification: C61, G11
AMS Classification: 93E20, 91B28

1 Introduction

We solve the situation of financial protection in retirement, which is today very topical and which is not entirely clear, how the pension system will function in the future in our country. Continuous savings for retirement is an essential part of rational behavior of everybody (not only young people) today.

So we decided to look at investment possibilities in this direction. We select a few investment instruments in terms of multiple criteria decision making method that will contribute to the future structure of the portfolio. The task is conceived as a dynamic and stochastic optimization problem. The aim is to create an optimal portfolio in each reporting period with regard to the required limits and objective function.

2 Dynamic and stochastic programming

Dynamic programming can be described as a mathematical method that is able to solve some optimization process, which consists of several stages (see [2]). In the decision making processes we determinate the optimal decision in any given instance, that successive. In our case, we consider the phase periods. Each stage is characterized by a particular state, which is determined by the values of factors and variables. Method of dynamic programming can solve the problem in stages, but in accordance with the requirements of the optimization process as a whole (see [8]).

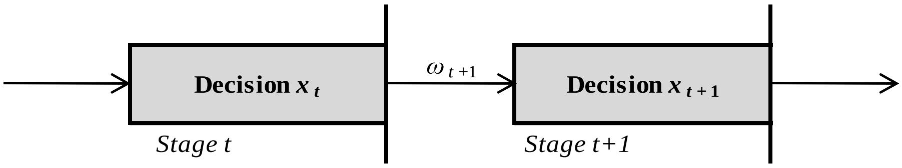
As mentioned earlier, each stage will be seen as a period of time. If it is a countable set of time steps, we are talking about discrete decision-making processes. In addition to discrete dynamic programming, there is still continuous dynamic programming, which addresses the processes that must be decided at any moment of a certain time interval (see [9]).

A classification of dynamic processes can be performed on the basis of the nature of the variables that occur in the model. In the case of deterministic models, we consider all factors and variables in the model predetermined, explicitly given. If we do not want to come to a great simplification in many cases we abandon this as-

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 130 67 Prague 3, adam.borovicka@vse.cz, petr.mynarik@vse.cz, pfiala@vse.cz.

sumption and include the nature of the stochastic model. The presence of random variables in the process gives the resulting solution of probabilistic nature. The optimal solution in terms of searching for extreme values of the objective function is determined by the realizations of random variables in the model (see [5]).

In practice, we consider the random variable of yield rate has a continuous probability distribution, namely the normal distribution with two parameters, mean and standard deviation, it is $N(\mu,\sigma)$. Yield rate of specific product can therefore take an infinite number of values within each stage. In this case, we will proceed so that in each period to generate a limited number of continuous distribution of values according to the spirit of *Monte Carlo* method (see [3]), which will form the basis of the possible scenarios dynamic structure of the investment portfolio. To create a portfolio, it should be kept in mind that the decisions in the individual sections are linked, respectively the actual decision is based on previous decisions. We can graphically display the *sequential model* for two adjacent periods, where x_t (or x_{t+1}) denotes unknown (decision) variable in time period t (or t+1) and ω_{t+1} marks stochastic nature's decision in stage t+1 (see figure 1).



$$x_{t+1} = f\,(x_t, \omega_{t+1})$$

Figure 1

Taking into account the fact described relations between the unknown variables in the dynamic model we can describe the solving procedure in every time t (t = 1, 2, ..., T) as

$$\begin{aligned} &f(x_{1t}, x_{2t}, \ldots, x_{nt}) \to \min(\max) \\ &g_i(x_{1t}, x_{2t}, \ldots, x_{nt})\, R_i\, b_i \quad i = 1,2,\ldots,m, \end{aligned}$$

where x_{1t}, x_{2t}, …, x_{nt} are unknown variables for each stage t, f is the objective function, which is minimized or maximized, g_i is the left side, R_i is a relational sign, and b_i is the right side of the *i-th* limit, m is the number of constraints in the model. At the same time must hold

$$\begin{aligned} x_{j1} &= f(\omega_1) \quad j = 1,2,\ldots,n \\ x_{j2} &= f(x_{j1}, \omega_1, \omega_2) \quad j = 1,2,\ldots,n \\ x_{j3} &= f(x_{j1}, x_{j2}, \omega_1, \omega_2, \omega_3) \quad j = 1,2,\ldots,n \\ &\vdots \\ x_{jT} &= f(x_{j1}, x_{2j}, \ldots, x_{nT-1}, \omega_1, \omega_2, \ldots, \omega_T) \quad j = 1,2,\ldots,n, \end{aligned}$$

whereas $x_{j1}, x_{j2}, \ldots, x_{jT}$ are unknown (decision) variables for indexes j = 1, 2, ..., n in each stage (t = 1, 2, ..., T) and $\omega_1, \omega_2, \ldots, \omega_T$ are nature's decision for each time period (t = 1, 2, ..., T). It's obvious that ω_t makes model stochastic. From the practical point of view ω_t is represented by product yield rates which embody a stochastic character in each time period.

3 Pension protection in practice

In this section we present a practical application, set in the area of financial protection in retirement. We consider the situation of a man aged 35, who is graduated from high school and for long time working in a well-paid position. Because thinking about the future, decide to put some of their money available to certain investment instruments related to financial protection in retirement.

3.1 Products

A potential investor (client) selects from five groups of products - investment life insurance, capital life insurance, retirement income insurance, special investment programs, and building savings.

As we can see in [1], *investment life insurance* combines insurance protection with the possibility of a very interesting yield rate, but that is not guaranteed. On the market there are several investment funds, which differ in the expected appreciation of money and the level of risk. The investment risk is all on the client side. In the case of *capital life insurance* we speak of insurance against death or survival. Compared investment life insurance offers a guaranteed yield of invested money. Next we come to the *retirement income insurance*, which we define as a form of savings backed by the state. Finally, we see *building savings products* with the state aid and opportunity to negotiate special home loans. *Special investment programs* usually represent investments in mutual funds with an optimized investment strategy that is tailored to the longer-range investment horizon (see [7]).

Within each group, we have included a number of investment products from different companies providing financial services in the Czech Republic. List of banks tracked, and their products are shown in the following two tables.

Retirement income insurance	**Investment life insurance**		**Capital life insurance**	
Pension fund	***Company***	***Product name***	***Company***	***Product name***
Allianz PF	AVIVA	Benefit	Allianz	Allianz kapitálové poj.
AXA PF	AXA	Comfort Plus	Amcico	Život Plus
PF České spořitelny	ČP	Dynamik Plus	AXA	Smíšené ŽP
PF České pojišťovny	ČSOB	Maximal	ČP	KŽP
PF Komerční banky	Generali	Clever Invest	ČPP	Spoření s Filipem
ČSOB Progres PF	ING	Investor Plus	ČSOB	Kvarteto
Generali PF	Kooperativa	Perspektiva	ING	KŽP
ČSOB Stabilita PF	PČS	Flexi	Kooperativa	Horizont
ING PF	Uniqa	IŽP	Uniqa	Akord
			Wüstenrot	Smíšené ŽP

Table 1 List of investment products (retirement income insurance, investment life insurance and capital life insurance)

Special investment programs		**Building savings**
Company	***Product name***	***Product name***
Conseq	Horizont Invest	Českomoravská stavební spořitelna
Conseq	Active Invest	Modrá pyramida stavební spořitelna
Atlantik	Target Click Fund	Raiffeisen stavební spořitelna
Česká spořitelna	Fondy životního cyklu	Stavební spořitelna České spořitelny
ČP Invest	Partner Invest	Wüstenrot stavební spořitelna
Pioneer	Rentier Invest	

Table 2 List of investment products (special investment programs and building savings)

3.2 Multiple criteria decision making

From each group of products we choose several investment alternatives by the help of multiple criteria decision making method which they can partake in final form of rising portfolio.

Before an application of some of methodical approaches we must determine evaluative criteria and their weights as a representation of decision maker's preferences. We stated characteristics of efficiency, costs and riskiness, minimal payment connected with product and interest rate associated with credit from building savings as well. Whereas we reflect rather more conservative investor, the biggest weight is allotted to yield rate for longer time period and risk resulting from investment in particular product which is measured as a standard deviation of yield rates. A cost intensity of investment obtains the smallest weight according to investor preferences.

We apply *ELECTRE I* method to choose a few investment alternatives from each group of products. In [4] or [6], this method evaluates alternatives in accordance with preferential relation. We don't need an ordering of alternatives. The method provides a division of alternates into two classes – efficient and inefficient. ELECTRE I scoops efficient alternatives from each group after subjective election of two threshold values – preference and dispreference threshold. The investor sets these values as 0.5. From each of five groups two alternatives were sorted out apart from the building savings with only one (see table 3).

	Company	**Product name**
Investment life insurance	AVIVA	Benefit
	ČSOB	Maximal
Capital life insurance	AXA	Smíšené ŽP
	Wüstenrot	Smíšené ŽP
Retirement income insurance	Alllianz	Alllianz PF
	Generali	Generali PF
Special investment programs	Atlantik	Target Click Fund
	Česká spořitelna	Fondy životního cyklu
Building savings	Modrá pyramida stavební spořitelna	SS Modrá pyramida

Table 3 Choice products as efficient by ELECTRE I method

3.3 Economic model

In previous chapter, we can see that the algorithm of ELECTRE I method chose nine efficient investment products. A presence of four time stages with duration of six years denotes dynamic character of the solving problem.

The investor specifies several conditions for future portfolio ensuring a positive financial certainty in pensionable age. Firstly he requires monthly payment in an amount from 3500 to 5000 CZK. The annual yield rate of portfolio should be at least 5 %, on the contrary costs should not over limits 4 % from monthly payment. Further the investor does not want to insert his free financial resources to more than one product of each group, generally then to more than three products within the scope of each time period. The risk for a whole time is stated as a minimizing optimization criterion.

The risk of investment in concrete product is described by standard deviation of its yield rate. For simplification we introduce a possible investment in building savings only 1667 CZK per month. This variant is the most attractive possibility regarding a state dotation and other specific conditions of building savings. The yield rate is a random variable with normal distribution. Lastly we assume that relative costs and risk corresponding to investment in particular product will be constant for a whole watched time period.

3.4 Mathematical model

In the mathematical model we apply two-dimensional variable x_{ij} showing amount of monthly investment in product *i* in stage *j*. The next variable y_{ij} takes the value 1 if the investor will invest in product *i* in stage *j*, failing which is 0. Symbol r_i marks a riskiness of *i-th* product, p_j indicates a general monthly payment in *j-th* stage. Costs associated with product *i* are expressed by c_i and yield rate of *i-th* product in *j-th* stage by e_{ij}. MIN_i represents the minimal monthly investment in *i-th* product.

Accept anticipated constraints we have to handle another condition. If the investor decides to invest in products *investment* and *capital life insurance* and *retirement income insurance*, he will never get out these products. This situation results from product character. The final model located in oncoming page has the following form:

$$z = \sum_{i=1}^{9}\sum_{j=1}^{4} r_i x_{ij} \Big/ \sum_{i=1}^{9}\sum_{j=1}^{4} x_{ij} \to MIN$$

$$\sum_{i=1}^{9} x_{ij} = p_j \quad j = 1, 2, \ldots, 4$$

$$3500 \le p_j \ge 5000 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=1}^{9} e_{ij} x_{ij} \Big/ \sum_{i=1}^{9} x_{ij} \ge 0{,}05 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=1}^{9} c_i x_{ij} \Big/ \sum_{i=1}^{9} x_{ij} \le 0{,}04 \quad j = 1, 2, \ldots, 4$$

$$MIN_i\, y_{ij} \le x_{ij} \le p_j y_{ij} \quad i = 1, 2, \ldots, 9 \quad j = 1, 2, \ldots, 4$$

$$x_{ij} = 1667 y_{ij} \quad i = 9 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=1}^{9} y_{ij} \le 3 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=1}^{2} y_{ij} \le 1 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=3}^{4} y_{ij} \le 1 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=5}^{6} y_{ij} \le 1 \quad j = 1, 2, \ldots, 4$$

$$\sum_{i=7}^{8} y_{ij} \le 1 \quad j = 1, 2, \ldots, 4$$

$$if\ \ y_{ij} = 1, then\ y_{ij+1} = 1 \quad i = 1,2,\ldots,6 \quad j = 1,2,3$$

$$x_{ij} \ge 0 \quad i = 1, 2, \ldots, 9 \quad j = 1, 2, \ldots, 4$$

$$y_{ij} \in \{0, 1\} \quad i = 1, 2, \ldots, 9 \quad j = 1, 2, \ldots, 4$$

3.5 Evaluation

We generated two values of yield rate for each product in each of four sexennial stages by means of *Monte Carlo* method, hence we get 16 scenarios how the potential portfolio should seem. Thus the probability of each scenario is 1/16.

We obtain six portfolio scenarios consisting of investments in capital life insurance AXA, retirement income insurance Allianz, building savings Modrá pyramida and special investment program Česká spořitelna, the others do not contain investment in special investment program. The detached scenarios differ in the size of embedded financial resources into these products in particular stages, watched expenses and efficiency as well. The following tables display two mentioned groups of acquired portfolios (investment portfolio I and II).

Product	Stage 1	Stage 2	Stage 3	Stage 4
Capital life insurance AXA	*125 CZK*	*3 753 CZK*	*125 CZK*	*125 CZK*
Retirement income insurance Allianz	*3 208 CZK*	*1 147 CZK*	*3 208 CZK*	*3 208 CZK*
Building savings Modrá pyramida	*1 667 CZK*	*0 CZK*	*1 667 CZK*	*1 667 CZK*
Special investment program ČS	*0 CZK*	*100 CZK*	*0 CZK*	*0 CZK*

Table 4 Investment portfolio I

Product	Stage 1	Stage 2	Stage 3	Stage 4
Capital life insurance AXA	*125 CZK*	*125 CZK*	*125 CZK*	*125 CZK*
Retirement income insurance Allianz	*3 208 CZK*	*3 208 CZK*	*3 208 CZK*	*3 208 CZK*
Building savings Modrá pyramida	*1 667 CZK*	*1 667 CZK*	*1 667 CZK*	*1 667 CZK*

Table 5 Investment portfolio II

In the first case the objective function is 0.0081, in the second 0,009. It means that the portfolio I globally implicates riskiness 0.81 % as a standard deviation of yield rate. The portfolio II embodies a little bit higher value 0.9 %. The first mentioned portfolio generally signals lower values of average efficiency and higher values of average costs for the whole time than the second one. In the quantitative expression, approximately 5.5 % contra 6.2 % in case of yield rate and 2.3 % contra 1.9 % in costs. The advantage of the portfolio II could be the fact that a potential investor has not to change structure during the whole period.

In virtue of described facts we suggest to choose investment portfolio II. This one has slightly higher risk, but better values in costs and efficiency and guarantees invariable structure. It is essential to emphasize that the recommended portfolio disposes probability of occurrence 1/16 considering selected scope of generated values of random variable in each stage.

4 Conclusion

An ambition of the article was to minister to easier decision making, what kinds of products investor should choice for financial protection in pensionary age. In terms of subjective stated constraints and objective function as the best alternative sizes up investment in **capital life insurance AXA**, **retirement income insurance Allianz** and **building savings Modrá pyramida** in thereinbefore structure for 24 years. We apply the dynamic model with stochastic (!) character for real expression of studied situation.

The subject of deeper research should be check on normality of probability distribution of yield rates, an application of concept of stochastic dominance (see [10]) as well. For instance, this approach should serve in multiple criteria decision making process. Another theme is an inclusion of longer watched time period with shorter length of stages with the view of more detailed problem solving, eventually the larger number of possible scenarios. In these cases, the model would become more complicated whose solving would not be lightly attainable.

Acknowledgements

The research project was supported by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Capital life insurance, investment life insurance, retirement income insurance and building savings, accessible from: *http://www.finance.cz/.*
[2] Denardo, E.: *Dynamic programming – models and application*. Prentice-Hall Inc., New Jersey, 1982.
[3] Dlouhý, M., Fábry, J., Kuncová, M., Hladík, T.: *Simulace podnikových procesů*. Computer Press, Brno, 2007.
[4] Fiala, P.: *Modely a metody rozhodování*. Oeconomica, Praha, 2008.
[5] Hadley, G.: *Nonlinear and Dynamic programming*. Addison-Wesley Publishing Company Inc., USA, 1964.
[6] Hwang, C. L., Yoon, K.: *Multiple Attribute Decision Making. Methods and Applications*. Springer-Verlag, Berlin, 1981.
[7] Lifecycle fund, accessible from: *http://www.iscs.cz/web/fondy/fondy_zivotniho_cyklu/*.
[8] Samek, J., Nečasová, Z., Kodera, J.: *Dynamické programování v ekonomických procesech*. Státní pedagogické nakladatelství, Praha, 1983.
[9] White, D. J., *Dynamic Programming*. Oliver & Boyd, Edinburgh; Holden-Day, San Francisco, 1969.
[10] Zelený, M.: *Human Systems Management - Integrating Knowledge, Management and Systems*. World Scientific Publishing, New Jersey, 2005.

Comparison of the Selected Departments in Hospitals Using DEA Models

Jana Borůvková[1], Martina Kuncová[2], Josef Jablonský[3]

Abstract. DEA (Data Envelopment Analysis) models are usually used to find the relative efficiency among homogenous units according to selected criteria (inputs and outputs). In this article we try to compare the not-for-profit hospitals that are situated in Vysocina Region in these towns: Jihlava, Havlickuv Brod, Trebic and Nove Mesto na Morave. In this article we compare only the Departments of Dermatology in the hospitals named above. All the criteria for the comparison were selected in cooperation with the Section of Health Care of the Vysocina Region which also has given all the data. This article describes results of DEA models and suggestions how to improve the efficiency of the departments.

Keywords: DEA models, hospitals, Vysocina Region, department of dermatology, efficiency

JEL Classification: C44, C67
AMS Classification: 90B50, 90B90

1 Introduction

In this article we are trying to answer the question how to evaluate the efficiency of the dermatology departments. We aim at the dermatology departments in the hospitals of the Vysocina Region. In this region we may find 5 regional hospitals in towns Havlickuv Brod, Jihlava, Nove Mesto na Morave, Pelhrimov and Trebic, but in Pelhrimov there is no dermatology department, so we compare only four hospitals. The research raised from the interest of the Section of Health Care of the Vysocina Region that wanted to know the analysis of the dermatology departments and so they gave us appropriate data for the models.

For the evaluation of the efficiency of hospitals it is possible to use various kinds of techniques and methods [5]. In this first phase we have decided for the methods of data envelopment analysis (DEA). DEA belongs to the operational research methods, especially to the linear programming models, that have been used many times in private or public sector to find the efficiency of the homogenous units (countries, regions, enterprises, schools, hospitals, insurance companies, military units etc.). These units must have identical inputs and outputs to measure the efficiency from the same parameters.

DEA models are widely used in health care especially for the evaluation of hospitals – for example Novosadova and Dlouhy [8] created model for the evaluation of acute hospitals. The authors chose these inputs for the evaluation: number of doctors, number of beds and number of nurses. The outputs were defined as: number of in-beds patients and number of day treatment. Another example is the work of the authors Clement et al. [2] that also compared the hospitals according to the number of registered nurses, number of licensed nurses, number of other staff and number of beds (as inputs) and number of babies born, number of out-patient cases, number of first-aid treatments, number of out-patients and total case mix (as outputs). Giokas [6] mentions only one input: total cost and the outputs in his case are: in-patients days, out-patients visits and ancillary services. Pi-Fang and Hui-Chen [11] defined different inputs: number of beds, number of physicians, number of other medical personnel and number of nurses. As outputs they chose payment for in-patients, payment for out-patients, number of in-patients, number of out-patients, number of visits to emergency service and number of surgical service.

Dexter and O´Neill [9] also compared the departments of the hospitals but contrary to our case their units were 8 different kinds of departments (cardiology, neurology, orthopedy, gynecology, pulmonary, urology, vascular and general surgery). The selected inputs were: number of beds, level of technical equipment, number of doctors, case-mix of the states and regions that have all these departments. To the outputs belong treated illness-

[1] College of Polytechnics Jihlava, Department of Mathematics, Tolstého 16, 58601 Jihlava, Czech Republic, boruvkova@vspj.cz
[2] College of Polytechnics Jihlava, Department of Economic Studies, Tolstého 16, 58601 Jihlava, Czech Republic, kuncova@vspj.cz
[3] University of Economics Prague, Department of Econometrics, W.Churchill Sq.4, 13067 Prague 3, Czech Republic, jablon@vse.cz

es by all the selected departments. So you can see that in each case there might be different ideas about the inputs and outputs to measure the efficiency of the units.

2 Data and methods

For the evaluation of our units (Departments of Dermatology) the DEA analysis was used. The basic idea of DEA models consists in estimation of an efficient frontier that defines production possibility set of the problem. Based on the set of available decision making units (DMUs) DEA estimates so-called efficient frontier, and projects all DMUs onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. DEA also provides efficiency scores and reference units for inefficient DMUs. Reference units are hypothetical units on the efficient frontier, which can be regarded as target units for inefficient units. DEA models can be oriented to inputs or outputs. In the case of input oriented models we assume fixed level of outputs (CCR-I), the output oriented model assumes fixed level of inputs and maximize level of outputs with respect to given inputs (CCR-O) [3]. 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. The review and detailed information about DEA models can be found in [1] and [10]. The basic idea for the efficiency calculation is to maximize the rate of weighted sum of outputs divided by weighted sum of inputs. For example the model transformed (Charles-Cooper transformation) into the linear programming form can be defined as follows (CCR-I):

$$\text{Maximize } z = \sum_{i=1}^{r} u_i y_{iq}$$

$$\text{Subject to: } \sum_{i=1}^{r} u_i y_{ik} \leq \sum_{j=1}^{m} v_j x_{jk}, \; k=1,2,\ldots,n \tag{1}$$

$$\sum_{j=1}^{m} v_j x_{jq} = 1$$

$$u_i \geq 0, \; i=1,2,\ldots,m$$

$$v_j \geq 0, \; j=1,2,\ldots,r,$$

where q represents the DMU, y_{rj} are known outputs, x_{ij} are known inputs of the jth DMU, u_r and v_j are the variable weights to be determined by the solution of this problem. 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 than 1 (in CCR-I model) [3].

The aim of DEA is to separate the DMUs into efficient and inefficient ones according to the used inputs and “produced” outputs. DEA models are based on the fact that there exists a set of possible production possibilities formed by all possible combinations of inputs and outputs – and the set is given by the efficient frontier. The units lying on the frontier are considered as efficient and the remaining ones as inefficient. Their efficiency score is measured as a distance from the efficient frontier [4]. The number of DMUs should be high enough because if we have few units and a lot of inputs and outputs, all the units are considered to be efficient. So it is very important to chose appropriate criteria for the comparison. On the other hand it is possible to use one input and one output to describe in a graph the efficient frontier. As we use the graphical representation for the model with two outputs and one input, the Figure 1 shows the shape of the efficient frontier in this situation.

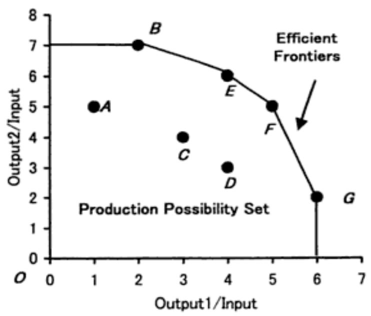


Figure 1 Efficient frontier for 2 outputs and 1 input [4]

In our case we should compare four hospitals, especially their dermatology departments. The problem is that they are only four so it is not possible to use a lot of inputs and outputs as mentioned above. According to the data and with cooperation with Section of Health Care of the Vysocina Region we have defined these inputs and outputs:

Inputs:

- Costs for out-patients
- Costs for beds
- Total costs
- Number of nurses
- Number of doctors
- Medical staff

Outputs:

- Number of in-patients
- Number of out-patients
- Number of in-patient days

The data (from the year 2010) that we could use for the comparison in order to find the efficient and inefficient department are in the Table 1. As you can see, in the dermatology department of Havlickuv Brod there is zero in the number of in-patients – it is because of the fact that this hospital does not have beds on this department. That is why the total cost are lower than in other departments and also the number of doctors and nurses is smaller.

Parameters	Type	Jihlava	Havl. Brod	Trebic	Nove Mesto
Number of in-patients	output	434	0	480	359
Number of out-patients	output	25769	10348	34957	28414
Number of in-patient days	output	5095	0	6661	4106
Costs for out-patients	input	8 394 714	1 550 065	2 618 984	3 131 087
Costs for beds	input	14 214 214	0	10 165 622	7 067 260
Total costs	input	22 608 928	1 550 065	12 784 606	10 198 347
Number of nurses	input	9	1	11	9
Number of doctors	input	3,9	1	4	4
Medical staff	input	12,9	2	15	13

Table 1 Data for the analysis of the hospitals – Department of Dermatology

3 Results

The analysis was made using two software products– Frontier Analyst from Banxia Software (www.banxia.cz). In our analysis we have to select among the given inputs and outputs because of the fact that we have only four DMUs (dermatology departments). We have selected these five variables:

Inputs:

- Total costs
- Number of nurses
- Number of doctors

Outputs:

- Number of in-patients
- Number of out-patients

In the analysis two basic DEA models were used – BCC-I and CCR-I (as we suppose the hospital can influence especially inputs). The results of the models are in Table 2:

Model	Jihlava	Havl.Brod	Trebic	Nove Mesto
BCC-I	1.000	1.000	1.000	0..976
CCR-I	1.000	1.000	1.000	0.971

Table 2 Results of the DEA models using 3 inputs and 2 outputs.

All models have given similar information – the only inefficient dermatology department is in Nove Mesto. The potential improvement of the variables of the dermatology in Nove Mesto is little bit different according to the models, but the virtual (improved) numbers of the inputs and outputs are in Table 3.

	Inputs			Outputs	
Real data	10,198,347	9	4	359	28,414
Model/ Virtual numbers	**Total costs**	**Nurses**	**Doctors**	**In-patients**	**Out-patients**
BCC-I	9,952,565	8.479	3.244	359	28,754
CCR-I	9,901,713	8.446	3.211	359	28,414

Table 3 Possible improvement of the variables of Nove Mesto dermatology

As the hospital or the department can influence mainly the inputs, it is evident that they should decrease the total costs for about 3% (about 300 thousand Czech crowns) and also decrease a little bit the number of medical staff (or the other way round – as if they decrease the staff, they decrease the costs).

We have tried more DEA models for:

1 input and 1 output

1 input and 2 outputs

2 inputs and 1 output

2 inputs and 2 outputs

In most cases all the units were efficient but for example in the model with 2 outputs (number of in-patients, number of out-patient) and 1 input (number of nurses) we can see (Figure 2) that not only Nove Mesto but the dermatology in Trebic is also a little bit far from the efficient frontier.

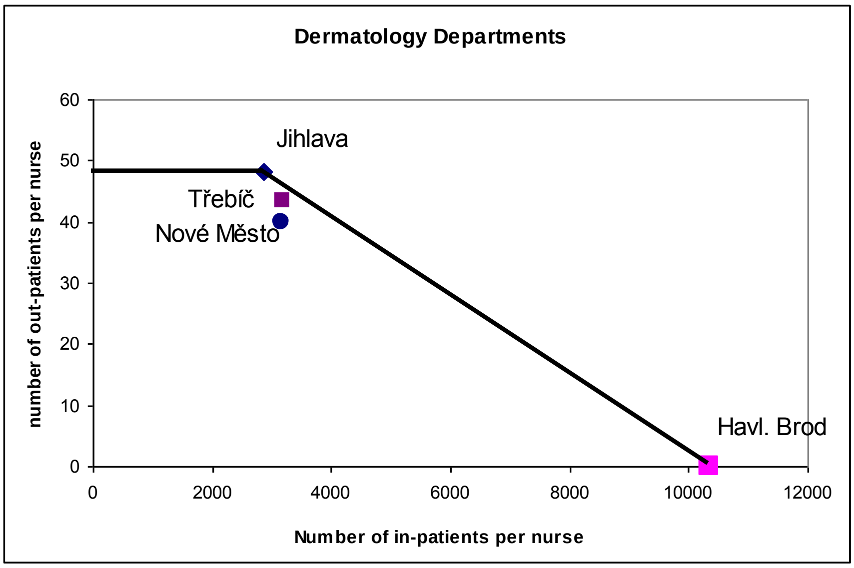


Figure 2 Efficient frontier for 2 outputs and 1 input – comparison of the dermatologies

4 Results

The task of this contribution was to compare the dermatology departments in the Vysocina Region. Because the number of the compared departments is only four, it is possible to use a limited number of inputs and outputs. That is why 5 variables to judge the efficiency of the dermatology departments in Vysocina Region were selected – 3 inputs and 2 outputs. Nearly all models marked the dermatology in Nove Mesto as inefficient, one model also showed that Trebic dermatology could be inefficient. To improve the efficiency the hospital should decrease number of doctors and nurses at the department (which lead to the decrease of the costs) – for example in a way that is typical for Havlickuv Brod – it means without hospital beds. So there could exist only two (Jihlava, Havlickuv Brod) or three (Trebic) hospitals serving for the in-patients in Vysocina Region.

We think about the enlargement of the model by addition of other variables – but for this purpose it is necessary to increase the number of DMUs. As the number of hospitals in Vysocina Region is given, we may add data from some previous years (for example to include data from the year 2009 and 2010) – but by this time the data are not available. Then we have 8 DMUs and we can also watch the possible changes of the efficiency of each dermatology through the years. Another possibility could be the creation of the model with the weights restriction as in [1].

Acknowledgements

The research was supported partly by the Grant Agency of the Czech Republic GAČR No. 402/09/0231 and also by the Internal Grant Agency of the College of Polytechnics Jihlava – grant no. 12002: “Comparison of the Hospitals in Vysocina Region”.

References

[1] Allen, R., Athanassopoulos, A., Dyson, A.G., Thanassoulis, E.: Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Annals of Operations Research* **73** (1997), 13–34.

[2] Clement, J. P., et al.: Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion. *Health Care Management Science.* **11** (2008), 67–77.

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

[4] Cooper, W.W., Seiford, L.M., Tone, K.: *Introduction to Data Envelopment Analysis and Its Uses*. Springer Science + Business Media Inc., New York, 2006.

[5] Dlouhy, M.: Efficiency and Productivity Analysis in Health Services. *Prague Economic Papers*. **18** (2009): 176-184.

[6] Giokas, D.: The use of goal programming, regression analysis and data envelopment analysis for estimating efficient marginal costs of hospital services. *Journal of Multicriteria Decision Analysis* **11** (2002), 261–268.

[7] Hosseinzadeh Lotfi, F., Noora, A.A., Jahanshahloo, G.R., Jablonský, J., Mozaffari, M.R., Gerami, J.: An MOLP based procedure for finding efficient units in DEA models, *CEJOR* **17** (2009), 1–11.

[8] Novosadova, I., Dlouhy. M.: Evaluation of technical efficiency of acute hospitals and its relation to wages of health personnel. *Ekonomicky casopis* **55** (2007), 783–792.

[9] O´Neill, Liam; Dexter, F.: Results with an Applicaiton to Hospital Inpatient Surgery. *Health Care Management Science* **8** (2005), 291–298.

[10] Pedraja-Chaparro, F., Salinas-Jiménez J., Smith P.: On the Role of Weight Restrictions in Data Envelopment Analysis. *Journal of Productivity Analysis* **8** (1997), 215–230.

[11] Pi-Fang, H.; Hui-Chen, H. The development and application of a modified data envelopment analysis for assessing the efficiency of different kinds of Hospitals. *International Journal of Management* **24** (2007), 318–330.

Third-degree stochastic dominance and DEA efficiency – relations and numerical comparison

Martin Branda [1]

Abstract. We propose efficiency tests which are related to the third-degree stochastic dominance (TSD). The tests are based on necessary conditions for TSD and on related mean-risk models. We test pairwise efficiency as well as portfolio efficiency with respect to full diversification of available assets.

We apply the proposed tests to 25 world financial indexes and we select the efficient ones. The test data set is divided into the periods – before financial crises and during it, and it is also considered at once.

Keywords: third-degree stochastic dominance, stochastic dominance efficiency, mean-risk efficiency, DEA efficiency.

JEL classification: C44
AMS classification: 90C15

1 Introduction

Dealing with uncertainty on financial markets is very difficult task. The investor's decision is highly dependent on the selected criteria which should help him to select the best among available investment opportunities. Harry Markowitz, [12], introduced his mean-risk model more than 50 years ago where variance was used as the risk measure. Many other risk measures has been proposed since then. The axiomatic definition of coherent risk measures is accepted by theorists as well as by practitioners, cf. [1]. The purpose of the mean-risk models is to maximize the mean return and to minimize the risk at the same time under given constraints on portfolio composition leading to biobjective optimization problem.

Another possible method how to find the best investment opportunity is to use an utility function, cf. [13]. To compare two possible outcomes, it is necessary to choose a particular nondecreasing function which corresponds to the investor's aversion to risk and serves as the utility function, and then to find an investment opportunity with the highest expected utility.

Stochastic dominance, introduced by [6, 7], is very closely related to the utility functions. It is defined over a whole set of utility functions with desired properties and compares the portfolios with respect to the whole class. Note that the stochastically dominating random variables are also optimal with respect to particular classes of risk measures, cf. [5, 14]. Note that mean-variance efficiency does not imply stochastic dominance efficiency, see [10]. Third-degree stochastic dominance (TSD) was introduced in [16] as a natural extension of stochastic dominances of lower orders. It is suitable for investors with decreasing absolute risk aversion. Recently, qualitative stability of an investment model with TSD constraint was investigated in [2].

Data Envelopment Analysis (DEA) was introduced by [4] as a tool of selection efficient units among units with the same structure of inputs and outputs. We will formulate models which help us to select efficient investment opportunities where historical rates of return are used as the inputs and the mean and risk as the outputs. By an appropriate choice of the risk measure we can obtain an efficiency test which is consistent with TSD. Efficiency tests for dominances of lower orders were proposed in [8, 9].

The paper is organized as follows. In Section 2, the third-degree stochastic dominance is defined and the basic properties are summarized. We propose various efficiency tests in Section 3. The tests are then used to find efficient world financial indeces in Section 4.

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Prague 8, Czech Republic, e-mail: martin.branda@seznam.cz

2 Third-degree stochastic dominance

Let $\mathcal{X}$ be a set of available investment opportunities with finite second moments. We prefer higher values to lower, i.e. we deal with profits, rates of return etc. Possible choices of the set will be discussed in the next section. We will propose two equivalent definitions of the third-degree stochastic dominance. The first was given originally by [16].

Definition 1. Let $\mathcal{U}_3$ be a class of real-valued differentiable functions with $u' > 0$, $u'' \leq 0$, and $u''' \geq 0$. The relation $X \succeq_{TSD} Y$ is equivalent to the condition that $\mathbb{E}u(X) \geq \mathbb{E}u(Y)$ holds for all utility functions $u \in \mathcal{U}_3$ for which both expectations are finite, and the strict dominance, $X \succ_{TSD} Y$, holds iff moreover there exists $u \in \mathcal{U}_3$ such that $\mathbb{E}u(X) > \mathbb{E}u(Y)$.

Below we will give two arguments why to consider the third-degree stochastic dominance instead of dominances of lower orders. The index of absolute risk aversion is usually defined as $ara(x) = -\frac{u''(x)}{u'(x)}$. It can be interpreted as the normalized relative change in marginal utility due to a change in wealth and it relates to instantaneous aversion to risk. The utility functions for which $ara'(x) < 0$ are usually referred as decreasing absolute risk aversion (DARA) utility functions. For example, with constant absolute risk aversion, our risk-taking behavior is the same regardless of the size of the wealth. The necessary but not sufficient condition for decreasing absolute risk aversion is that $u''' > 0$, because it holds

$$ara'(x) = \frac{-u'''(x)u'(x) + (u''(x))^2}{(u'(x))^2} < 0,$$

The second heuristic motivation why to consider the condition on the third derivative of the utility functions can be found in [10] and says: if we denote w the initial wealth and $X \in \mathcal{X}$ a random variable with finite third moment $\mathbb{E}X^3$, we can expand the utility function $u(w + X)$ into Taylor series at the point $w + \mathbb{E}X$, compute its expected value and we approximately obtain

$$\mathbb{E}[u(w + X)] \cong u(w + \mathbb{E}X) + \frac{u''(w + \mathbb{E}X)}{2!}\sigma_X^2 + \frac{u'''(w + \mathbb{E}X)}{3!}\nu_X^3,$$

where σ_X^2 is the variance of X and ν_X^3 its third central moment. If the other factors are held constant, then the higher σ_X^2, the lower the expected utility of an investor is, and the higher the skewness, the higher the expected utility. Hence, under our assumptions on $\mathcal{U}_3$ the investor dislikes variance and likes positive skewness. Fortunately, in practical applications of the third-degree stochastic dominance we do not need the restrictive condition $\mathbb{E}X^3 < \infty$.

Now we propose an alternative definition which we will use in the next section. We consider the cumulative distribution functions which are derived from the distribution function $F_X^{(1)} = F_X$ of $X \in \mathcal{X}$:

$$F_X^{(k)}(\eta) = \int_{-\infty}^{\eta} F_X^{(k-1)}(\xi)d\xi, \ \forall \eta \in \mathbb{R}, \ k = 2, 3,$$

Definition 2. Let $X, Y \in \mathcal{X}$ be two random variables on $(\Omega, \mathcal{F}, P)$ with the distribution functions F_X, F_Y. We say that X dominates Y in the sense of third-degree stochastic dominance, denoted by $X \succeq_3 Y$, if and only if the following two conditions hold:

$$\begin{aligned} F_X^{(3)}(\eta) &\leq F_Y^{(3)}(\eta), \ \forall \eta \in \mathbb{R}, \\ \mathbb{E}X &\geq \mathbb{E}Y. \end{aligned}$$

We say that X strictly dominates Y in the sense of third-degree stochastic dominance, denoted by $X \succ_3 Y$, if and only if $X \succeq_3 Y$ and $Y \succeq_3 X$ does not holds.

Note that the condition which compares the expectations is not necessary if the supports of the compared random variables are unbounded, see [15].

3 Efficiency tests

In this section we propose several tests which should help us to identify efficient investment opportunities. Pairwise efficiency as well as portfolio efficiency allowing full diversification across the assets are taken into consideration. Using pairwise comparisons, an asset is classified as efficient if there is no other asset that strictly dominates the asset with respect to the criteria. These efficiency may be more useful for financial indeces. Since investors may combine the assets, tests for portfolio efficiency allowing full diversification across the assets are of interest too.

We consider n assets and denote R_i the rate of return of i-th asset. The following two choices of the set of investment opportunities will be used:

1. $\mathcal{X}^P = \{R_i, i = 1, \ldots, n\}$, which corresponds to investment into one single asset, and enables us to test pairwise efficiency,
2. $\mathcal{X}^{FD} = \{\sum_{i=1}^n R_i x_i : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$, which enables diversification of our portfolio across all assets, hence we will use it to test efficiency with respect to full diversification.

Another choices of the set are also possible, e.g. allowing short sales, and will be aimed in future research. We will show how the efficiency tests can be constructed in general or based on discretely distributed returns. Let r_i^t, $t = 1, \ldots, T$, be the t-th realizations of the i-th asset return R_i. It can be computed as: $r_i^t = \frac{P_i^t}{P_i^{t-1}} - 1$ where P_i^t, P_i^{t-1} is the price of the i-th asset at the end of the t-th, $(t-1)$-st time period, respectively.

3.1 Mean-risk efficiency

Let $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ denote a risk measure which is a function of available investment opportunities and which quantifies the corresponding risk as a real number.

Definition 3. We say that $X \in \mathcal{X}$ strictly dominates $Y \in \mathcal{X}$ in the sense of mean-risk criterion, denoted $X \succ_{\mathbb{E},\mathcal{R}} Y$, if $\mathbb{E}X \geq \mathbb{E}Y$ and $\mathcal{R}(X) \leq \mathcal{R}(Y)$ with at least one strict inequality.

Definition 4. We say that $X \in \mathcal{X}$ is mean-risk efficient if there exists no $Y \in \mathcal{X}$ such that $Y \succ_{\mathbb{E},\mathcal{R}} X$.

By an appropriate choice of the risk measure we can get various efficiency tests. Our tests are based on Theorem 1 in [14], which states necessary conditions for $X \succeq_{FSD} Y$: $\mathbb{E}X \geq \mathbb{E}Y$ and $\mathbb{E}X - lsd(X) \geq \mathbb{E}Y - lsd(Y)$, where lsd denotes the lower semideviation. For $X \in \mathcal{X}$, it is defined as

$$lsd(X) = \left(\mathbb{E}[X - \mathbb{E}X]_-^2\right)^{1/2},$$

where $[\cdot]_-^2 = (\min\{0, \cdot\})^2$. If we consider discretely distributed returns, we get for the i-th asset

$$lsd(R_i) = \left(\frac{1}{T}\sum_{t=1}^{T}[r_i^t - \overline{r}_i]_-^2\right)^{1/2},$$

where $\overline{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_i^t$.

If at least one of the following conditions holds with strict inequality, then $Y \succ_{\mathbb{E},lsd} X$:

$$\mathbb{E}Y \geq \mathbb{E}X, \quad lsd(X) \geq lsd(Y). \tag{1}$$

Using the following program for $\alpha \in (0, 1]$ we can obtain portfolios which are mean-lsd efficient and consistent with TSD, cf. [14]:

$$\begin{aligned} \max \sum_{i=1}^{n} \overline{r}_i x_i + \mu \frac{1}{T} \sum_{t=1}^{T} z_t^2 & & \\ \sum_{i=1}^{n} x_i(\overline{r}_i - r_t^t) &\leq z_t, \\ \sum_{i=1}^{n} x_i &= 1, \\ x_i, z_t &\geq 0, \end{aligned} \tag{2}$$

where z_t are auxiliary decision variables which help us to model the positive parts.

3.2 TSD efficiency

Definition 5. We say that $X \in \mathcal{X}$ is efficient with respect to TSD if there exists no $Y \in \mathcal{X}$ such that $Y \succ_{TSD} X$.

To compare two random variables we will use the alternative expression of the integrated distribution function, see [5, 14]:

$$F_X^{(3)}(\eta) = \frac{1}{2}\int_{-\infty}^{\eta} [\eta - \xi]_+^2 dF_X(\xi).$$

where $[\cdot]_+^2 = (\max\{0, \cdot\})^2$. We can test $Y \succ_k X$ by investigating if the following conditions hold with at least one strict inequality in

$$\begin{aligned} F_X^{(3)}(\eta) - F_Y^{(3)}(\eta) &\geq 0, \forall \eta, \\ \mathbb{E}Y - \mathbb{E}X &\geq 0. \end{aligned} \tag{3}$$

3.3 DEA efficiency test

In this section, we will propose new DEA portfolio efficiency test with respect to the third-degree stochastic dominance. Any asset is compared with all portfolios which can be mixed from all considered assets, i.e. full diversification is enabled. The test for a benchmark $b \in \{1, \ldots, n\}$ can be formulated in general as follows:

$$\begin{aligned} &\max \delta^m + \delta^r \\ \sum_{i=1}^{n} x_i \mathbb{E}R_i &= \mathbb{E}R_b + \delta^m, \\ lsd^2\Big(\sum_{i=1}^{n} x_i R_i\Big) &\leq lsd^2(R_b) - \delta^r, \\ \sum_{i=1}^{n} x_i &= 1, \\ x_i, \delta_m, \delta_r &\geq 0, \end{aligned}$$

where the mean and lower semideviation of the benchmark are compared with portfolio mean and risk. If the optimal value is equal to 0, then the benchmark asset is said to be efficient, otherwise it is not efficient. Similar three tests were proposed and compared in [11]. Note that the proposed test as well as our test state only necessary condition for TSD-efficiency.

For discretely distributed random returns, we obtain the following quadratic programming problem which can be easily solved by standard solvers:

$$\begin{aligned} &\max \delta^m + \delta^r \\ \sum_{i=1}^{n} \overline{r}_i x_i &= \overline{r}_b + \delta^m, \\ \sum_{i=1}^{n} x_i(\overline{r}_i - r_t^i) &\leq z_t, \\ \frac{1}{T}\sum_{t=1}^{T} z_t^2 &\leq lsd_b^2 - \delta^r, \\ \sum_{i=1}^{n} x_i &= 1, \\ x_i, z_t, \delta_m, \delta_s &\geq 0. \end{aligned} \tag{4}$$

4 Stock indices efficiency – empirical study

We consider the following 25 world financial (stock) indices which are listed on Yahoo Finance:

- **America** (5): MERVAL BUENOS AIRES, IBOVESPA, S&P TSX Composite index, S&P 500 INDEX RTH, IPC,
- **Asia/Pacific** (11): ALL ORDINARIES, SSE Composite Index, HANG SENG INDEX, BSE SENSEX, Jakarta Composite Index, FTSE Bursa Malaysia KLCI, NIKKEI 225, NZX 50 INDEX GROSS, STRAITS TIMES INDEX, KOSPI Composite Index, TSEC weighted index,
- **Europe** (8): ATX, CAC 4, DAX, AEX, SMSI, OMX Stockholm PI, SMI, FTSE 100,
- **Middle East** (1): TEL AVIV TA-100 IND.

In our analysis we describe each index by its weekly rates of returns. We divided the returns into three datasets:

- before crises (B): September 11, 2006 - September 15, 2008
- during crises (D): September 16, 2008 - September 20, 2010
- whole period (W).

We choose September 16, 2008 to divide the data because all financial indices strongly fell down in week starting with this day. The descriptive statistics of the returns can be found in [3], where the same dataset was analyzed using different techniques. It can be observed that almost all returns are negatively skewed. Moreover, comparing the before crises data with during crises data we found that the during crises returns usually have higher standard deviation and kurtosis.

Table 1 shows efficient indeces according to the tests introduced in the previous section: pairwise tsd (1), pairwise mean-lsd (2), full diversification mean-lsd (3), and DEA (4). The pairwise comparison using the integrated distribution functions $F^{(3)}$ was implemented in Matlab using optimization toolbox. The quadratic programming tests were solved using the modelling system GAMS 23.0 and the solver Cplex 12.0.

The pairwise mean-lsd test selects most of efficient indeces and all the indeces selected by another tests are among them. The efficient indeces selected by mean-lsd model and DEA test are the same. It can be also seen that the returns observed during crises influence the tests based on the whole period more than the returns obtained before crises.

	P-TSD			P-ML			F-ML / DEA		
	B	D	W	B	D	W	B	D	W
IBOVESPA				X			X		
S&PTSX Composite index				X					
S&P 500 INDEX,RTH	X			X					
IPC				X		X			
BSE SENSEX				X		X			
Jakarta Composite Index					X	X		X	X
FTSE Bursa Malaysia KLCI		X	X		X	X			
NZX 50 INDEX GROSS	X			X		X			
TSEC weighted index					X				

Table 1: Efficient indeces (B - before crises, D - during crises, W - whole period)

Acknowledgements

The research was supported by the Czech Science Foundation (grant P402/10/1610).

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. *Mathematical Finance* **9** (1999), 203-228.

[2] Branda, M. (2009). Qualitative stability of stochastic programs with third-degree stochastic dominance constraint induced by mixed-integer linear recourse. *Proceedings of Mathematical Methods in Economics* 2009 (H. Brožová, R. Kvasnička, eds.), Czech University of Life Sciences, Prague (2009), 24-28.

[3] Branda, M., Kopa, M.: DEA-risk efficiency and stochastic dominance efficiency of stock indices. Submitted to *Czech Journal of Economics and Finance* (2011).

[4] Charnes, A., Cooper, W., Rhodes, E.: Measuring the efficiency of decision-making units. *European Journal of Operations Research* **2** (1978), 429-444.

[5] Gotoh, J.-Y., Konno, H.: Third-degree stochastic dominance and mean-risk analysis. *Management Science* **46** (2000), No. 2, 289-301.

[6] Hadar, J., Russell W.R.: Rules for ordering uncertain prospects. *American Economic Review* **9** (1969), 25-34.

[7] Hanoch, G., Levy H.: The efficient analysis of choices involving risk. *Review of Economic Studies* (1969), 335-346.

[8] Kopa, M., Chovanec, P.: A second-order stochastic dominance portfolio efficiency measure. *Kybernetika* **44** (2008), No. 2, 243-258.

[9] Kopa, M., Post, T.: A portfolio efficiency test based on FSD optimality. *Journal of Financial and Quantitative Analysis*, **44** (2009), No. 5, 1103-1124.

[10] Levy, H.: *Stochastic dominance: Investment decision making under uncertainty.* Second edition, Springer, New York, 2006.

[11] Lozano, S., Gutierrez, E.: TSD-consistent performance assessment of mutual funds. *Journal of the Operational Research Society* **59** (2008), 1352-1362.

[12] Markowitz, H. M.: Portfolio selection. *The Journal of Finance* **7** (1952), No. 1, 77-91.

[13] von Neumann, J., Morgenstern, O.: *Theory of games and economic behavior.* Princeton University Press, 1944.

[14] Ogryczak, W., Ruszczynski, A.: On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programmming*, Ser. B **89** (2001), 217-232.

[15] Schmid F.: A note on third-degree stochastic dominance. *OR Spectrum* **27** (2005), 653-655.

[16] Whitmore, G. A.: Third-degree stochastic dominance. *The American Economic Review* **60** (1970), No. 3, 457-459.

Analysis of Decision-Making Problems with Key State of Nature

Helena Brožová, Martin Flégl[1]

Abstract. Decision-making is the most important managerial activities of each enterprise. Every day decisions are facing uncertainty of information during decision-making processes. The problem is how to express or use obtained knowledge and information, and how to work with uncertainty. Crossing the border between certainty and uncertainty, it moves our decisions on higher level and gives us key competitive advantage. This is important mainly if the managers have to take into account specific type of event – crisis situation, situation which can distort or improve results of their decisions or situation with unknown probability of realization. This type of situation can be called Key state of nature. This contribution demonstrates the decision-making process in which this special type of state of nature is included. The aim is to show possible approach of how to analyze and support solving of decision-making problems with key state of nature.

Keywords: Decision-making, key state of nature, EMV, probability analysis.

JEL Classification: C79
AMS Classification: 91B06

1 Introduction

Contemporary economic decision-making is very complex and difficult process which has an important impact on firm competitiveness. Interests of the most managers are mostly focusing on reaching positive economic results and as well as on extending of the approaches how to reach these economic results. The decision-making problems can be described as a situation when we have to choose one decision alternative from the list of possible alternatives. Decision model helps to choose the best decision according to the quantified data. It is based on the theory of games against nature.

The decision maker selects one of his strategies – alternatives that are available. Their effect depends on possible future states of nature. Payoffs are associated with each combination alternative – state of nature. The future result of selected alternative depends also on the probabilities of realization of states of nature. If the promising state of nature has high probability, the result of alternative will be better and vice versa. Estimation of probability of the states of nature is very complicated task. Therefore some approaches exist for refinement of these data. Commonly known is application of so called additional information. This approach is again based on using of crisp estimation and do not allow some sensitivity analysis.

In this paper we want to suggest how to select the best decision based on the changes of states nature probabilities in view of the probability of specific state of nature. This specific type of event can be called **Key state of nature**. This state of nature can be

- crisis situation, situation which can distort results of decisions,
- advantageous situation, situation which can improve results of decision, or
- situation, with unknown probability of realization.

2 Decision-making under risk with the key state of nature

The general format of a decision model under risk ([1], [4], [5]) is in a decision table (**Table 1**).

		States of nature			
		S_1	S_2	.....	S_n
	A_1	v_{11}	v_{12}	.....	v_{1n}
Alternatives	A_2	v_{21}	v_{22}	.....	v_{2n}
	.....	.....	.....	.....	.....
	A_m	v_{m1}	v_{m2}	.....	v_{m3}
Probabilities		p_1	p_2	.....	p_n

Table 1 Decision table

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

where $\boldsymbol{A}_i$ is the i-th alternative, i = 1, ..., m,

$\boldsymbol{S}_j$ is the j-th state of nature, j = 1, ..., n,

v_{ij} is the payoff of alternative $\boldsymbol{A}_i$ and state of nature $\boldsymbol{S}_j$ combination, and

p_j is probability of state of nature $\boldsymbol{S}_j$.

The appropriated alternative is specified according to decision criterion as maximising output or minimising input, next we will suppose that the best alternative maximises the payoff value. Commonly used criterion is the Expected monetary value criterion (*EMV*), the alternative is selected if it has the maximal mean value of payoff (1)

$$
\begin{aligned}
&EMV_i = \sum_{j=1}^{n} p_j v_{ij} \quad i = 1,\dots,m \\
&A_I\colon \quad EMV_I = \max_{i=1,\dots,m} EMV_i
\end{aligned}
\tag{1}
$$

Suppose now, that the key state of nature is state $\boldsymbol{S}_n$. Analysis of the best alternative now has to be made according to the change of probability p_n. If this probability increases, probabilities of other states of nature have to decrease and vice versa. This relation has to be proportional, because in the other cases it would not be possible to make analysis for the all possible values of probability p_n, which are from the interval $\langle 0, 1 \rangle$.

For clearness we start with $p_n = 0$ and the best alternative $\boldsymbol{A}_I$ has to be choosing using formula (2), where the *EMV* is calculated from the first n-1 pay-offs.

$$
\begin{aligned}
&EMV_i = \sum_{j=1}^{n-1} p_j v_{ij} \quad i = 1,\dots,m \\
&A_I : EMV_I = \max_{i=1,\dots,m} EMV_i
\end{aligned}
\tag{2}
$$

If now the probability p_n will be increasing, its value represents the expectation of realization of key state of nature $\boldsymbol{S}_n$. The *EMV* in relation to the key state of nature EMV^{KSN} will be calculated as follows (3) and again the best alternative $\boldsymbol{A}_I$ must maximize this value.

$$
\begin{aligned}
&EMV_i^{KSN} = (1 - p_n) \sum_{j=1}^{n-1} p_j v_{ij} + p_n v_{in}, \quad i = 1,\dots,m \\
&A_I : EMV_I^{KSN} = \max_{i=1,\dots,m} EMV_i^{KSN}{}_i
\end{aligned}
\tag{3}
$$

This formula for calculation of the expected monetary value is based on an anticipation of the non-realisation and the realisation of the key state of nature.

Graphical representation of values of the EMV^{KSN} in relation to the probability of the key state of nature serves for practical application, because it shows the switching of the best alternatives. In such graph the axe $\boldsymbol{x}$ contains probability of the key state of nature and the axe $\boldsymbol{y}$ describes the corresponding EMV^{KSN} value. In such representation the best alternative is showed by the line above all other lines. The best alternative is changed in all breakdown points of the EMV^{KSN}.

2.1 Analysis of the best decision according to the key state of nature

Example 1

The simple example of decision making model is in the following **Table 2** and **Figure 1**. Because the key state of nature $\boldsymbol{K}$ is really crisis situation (very worth pay-offs of both alternatives), with its higher probabilities the difference between the alternatives is shrinking and their results become very bad. The best alternative is $\boldsymbol{A}_2$ for all values of probability of $\boldsymbol{K}$.

	S_1	S_2	S_3	S_4	K
A_1	5	15	-1	4	-10
A_2	4	8	3	10	-10
Probabilities	0,1	0,15	0,45	0,3	0

Table 2 Example 1 - Decision table

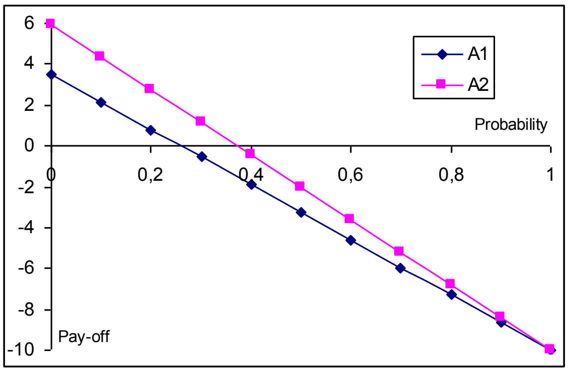


Figure 1 Example 1 – The dependency of EMV^{KSN} on probability of key state of nature

Example 2

The second example of decision making model is in the **Table 3**. The key state of nature ***K*** is again situation with worth pay-offs. If the probability of ***K*** is very low, the best alternative is $\boldsymbol{A_2}$. With its higher probability the best alternative is changed from alternative $\boldsymbol{A_2}$ to alternative $\boldsymbol{A_1}$. This change occurs with probability $p_5 = 0{,}23445$ (**Figure 2**).

	S_1	S_2	S_3	S_4	K
A_1	5	15	-1	4	-2
A_2	4	8	3	10	-10
Probabilities	0,1	0,15	0,45	0,3	0

Table 3 Example 2 - Decision table

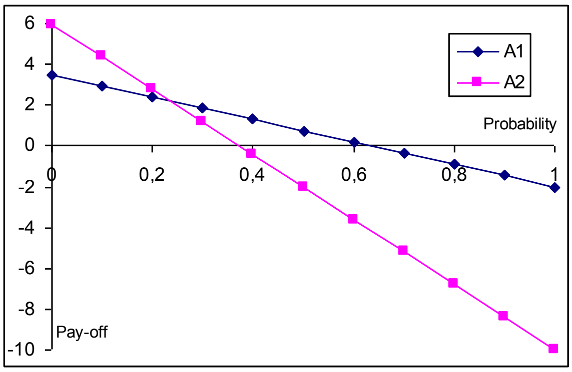


Figure 2 Example 2 – The dependency of EMV^{KSN} on probability of key state of nature

3 ISO decision problem analysis according to the key state of nature

Company XYZ provides complex spectrum of services in the field of system integration ([2], [3]). Currently the company is looking for new opportunities how to increase their competitiveness. First possible way is establish Quality management system based on ISO 9001 or establish Environmental management system based on ISO 14001 ([6], [7]). It is going to be decision-making under the risk because there is a lot of information which has to be included into forecasts of the future trends on the market demand.

Analysis of solved decision problem leads to the following definition of elements of this decision problem. Because establishing ISO 14001 follows the ISO 9001 the possible strategies are:

$\boldsymbol{A_1}$ = nothing establish
$\boldsymbol{A_2}$ = establish only ISO 9001
$\boldsymbol{A_3}$ = establish ISO 9001 and ISO 14001

The possible future situation is divided into following five states of nature:

S_1 = low preference of ISO norms by customers
S_2 = lower preference, customers begin oriented onto ISO 9001
S_3 = average preference, customers require ISO 9001
S_4 = higher preference, customers require ISO 9001 and tend also onto ISO 14001
S_5 = high preference, customers require both ISO 9001 and ISO 14001

Profit in the sector of system integration was evaluated around 500 million Czech crown. This company has market share 25 %. If we consider company's turnover in the amount 125 million Czech crowns and estimation of expenses necessary for implementation managerial approaches following ISO norms and potential profit growth, the pay-offs are estimated and the decision model will be construct as it is shown below (**Table 4**).

	S_1	S_2	S_3	S_4	S_5
A_1	125	120	115	110	100
A_2	131	140	136	137	135
A_3	120	126	132	144	150
Probabilities	0,1	0,3	0,4	0,1	0,1

Table 4 Decision table of ISO problem

The best alternative according to the EMV criterion is the second variant A_2 with the highest value (**Table 5**). The alternative A_3 yields also a sufficient value of EMV.

EMV_1	115,5
EMV_2	**136,7**
EMV_3	132

Table 5 EMV criterion values

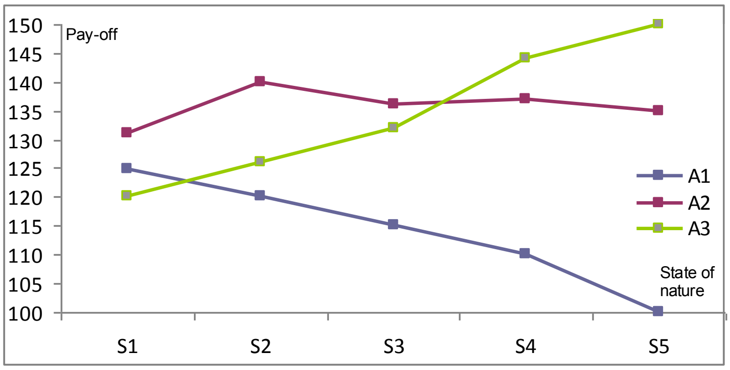


Figure 3 ISO problem – Graphical description of pay-offs

The graphical representation of dominance according to the states of nature in ISO decision problem shows, that there is not just one best alternative (**Figure 3**). If the market is not significantly oriented onto ISO 9001 and ISO 14001 certification (**S_1** - low preference of ISO norms by customers, **S_2** - lower preference, customers begin oriented onto ISO 9001, and **S_3** - average preference, customers require ISO 9001); the best alternative is alternative **A_2** (establish only ISO 9001). In the case of high preference of both ISO certifications, the best alternative becomes **A_3** (establish ISO 9001 and ISO 14001).

3.1 Analysis of the best decision according to the key state of nature S_1

Now we will analyse the best decision alternative according to the key state of nature **S_1** (low preference of ISO norms by customers**)**. This situation describes the hypothetic tendency to non implement ISO norms. First of all the probabilities of the other four states of nature have to be proportionally converted to receive their new values with a sum equal to 1 (**Table 6**).

	S_1	S_2	S_3	S_4	S_5
Probabilities	0	0,333333	0,444444	0,111111	0,111111

Table 6 New values of states of nature probabilities

The following graph (**Figure 4**) describe this situation, calculation was made using formulas in (3). The best alternative is always the second variant A_2 (establish only ISO 9001).

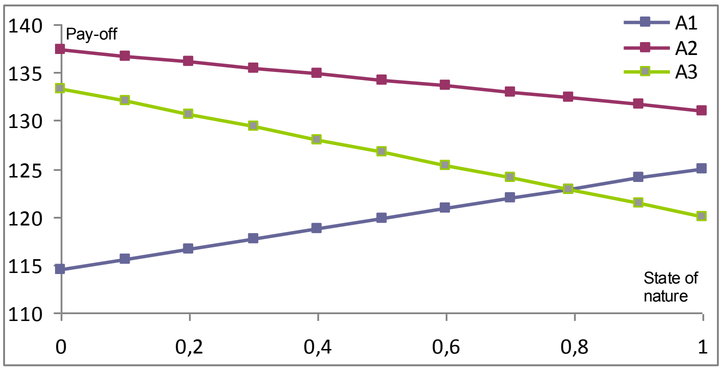


Figure 4 ISO problem – EMV^{KSN} dependency on probability of key state of nature S_1

3.2 Analysis of the best decision according to the key state of nature S_5

Tendency to implement ISO norms is described if the key state of nature is the state of nature S_5 (high preference of both ISO 9001 and ISO 14001 by customers). The different result is obtained (**Table 7** and **Figure 5**).

	S_1	S_2	S_3	S_4	S_5
Probabilities	0,111111	0,333333	0,444444	0,111111	0

Table 7 New values of states of nature probabilities

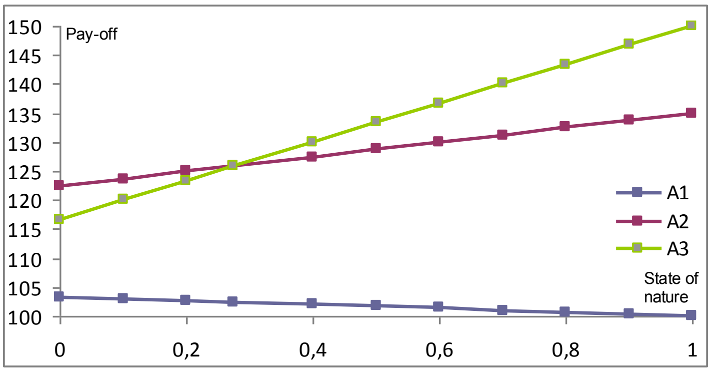


Figure 5 ISO problem – EMV dependency on probability of key state of nature S_5

The best alternative is the second variant A_2 (establish only ISO 9001) only for low probability of the key state of nature S_5 (high preference of both ISO 9001 and ISO 14001 by customers). If its probability will be greater ($p_5 = 0.2742$ is the breakdown point) the best alternative will become A_3 (establish ISO 9001 and ISO 14001).

4 Conclusion

This contribution shows analysis of decision model in relation of probability of the key state of nature. This analysis is based on calculation of EMV^{KSN}. This form of the expected monetary value includes a compromise between the impossibility and the certainty of realisation of the key state of nature.

Graphically it is very easy to make the analysis of the best alternatives according to the breakdown points of their EMV^{KSN} lines. This graph shows not only the best alternatives but also corresponding intervals of probability of key state of nature.

Practical example of ISO problem shows that implementation (certification) of ISO 9001 and ISO 14001 is getting to be very important and companies without these certifications probably will not be successful.

Acknowledgements

The paper is supported also by the grant project of the Ministry of Education, Youth and Sports, the Czech Republic, No. MSM6046070904 “Informační a znalostní podpora strategického řízení” (“Information and knowledge support of strategic management”).

References

[1] Bonini, Ch. P., Hausmann, W. H., Bierman, H.: *Quantitative analysis for management*, Irwin, Boston, 1997.

[2] Flégl, M., Brožová, H.: Decision making models with fuzzy values, *In: Mathematical Methods in Economics 2010*, České Budějovice, 2010, ISBN 987-80-7394-218-2, pp. 138-143.

[3] Flégl, M., Brožová, H.: Fuzzy Decision-Making for Implementing ISO 9001 and/or ISO 14001, *In: WSEAS Proceedings: Mathematics and Computers in Biology, Business and Acoustics* , pp. 33-38, ISBN: 978-960-474-293-6

[4] Raiffa, H.: *Decision Analysis: Introductory Readings on Choices Under Uncertainty*. McGraw Hill. 1997. ISBN 0-07-052579-X

[5] Shelton, R. B.: *Gaming the Market: Applying Game Theory to Create Winning Trading Strategies*, Wiley Finance, John Wiley and Sons, 1997.

[6] *Státní politika životního prostředí České republiky* (2004 - 2010). Praha: Ministerstvo životního prostředí, 2004. 56 s. ISBN 80-7212-283-5.

[7] Veber, J. et al.: *Management kvality, environmentu a bezpečnosti práce : Legislativa, metody, systémy, praxe*. 2. aktualizované. Praha: Management Press, 2010. 359 s. ISBN 978-80-7261-210-9.

Stock Market Co-Movements in Europe

Vít Bubák [1], Evžen Kočenda [2], Filip Žikeš[3]

Abstract. We analyze daily as well as intraday stock market indices for the stock markets in Budapest (BUX), Prague (PX 50), Warsaw (WIG 20), Frankfurt (DAX 30), Paris (CAC 40) and London (UKX). The sample begins on June 3, 2003 and runs till December 31, 2010. We first employ a VAR model to model their joint behaviour. The residuals from the VAR model are then used as a basis for the analysis of dynamic correlations between the six markets under research. To model the dynamics of the conditional variance of the innovation process we employ the Dynamic Conditional Correlation (DCC) model. Pattern on daily level shows that European markets become well integrated as dynamic correlation coefficients are quite high. On the intraday level we find strong co-movements between Western European markets but not among the Central European markets. We do not detect excessive changes in co-movement patterns due to the recent crisis period.

Keywords: Stock markets, European Union, co-movements, conditional correlation.

JEL Classification: G17
AMS Classification: 91B28

1 Introduction

The ongoing process of globalization and European integration has entailed large cross-border capital flows and resulted in stronger real economic linkages between old and new EU member states. Portfolio capital flows accompanied by a deepening of the financial systems in Central and Eastern Europe (CEE) may also have promoted financial market integration in Europe. Greater real and financial integration may imply higher synchronization between developed and emerging European stock markets, as well as among the CEE markets as a group. The hypotheses of higher synchronization are an issue we aim to address in this paper.

We analyze co-movements of the European stock markets. We use data for three Western European stock markets (Paris, Frankfurt, and London) and for three CEE markets (Budapest, Prague, and Warsaw). We employ the Dynamic Conditional Correlation GARCH (DCC-GARCH) model introduced in [7] to analyze co-movements among the markets. First, we analyze daily co-movements and then intra-day co-movements.

The present article offers an integrated approach to studying the co-movements between the Western and Central European equity markets, one that differs from the previous studies by focusing both on the day-to-day as well as intraday interactions between these markets.

In the day-to-day analysis, we aim to explain what drives the changes in co-movements between the Western and CEE markets over time. Motivated by an extant literature on the financial market co-movements that shows the importance of correlated changes in fundamental/macroeconomic variables for international asset allocation and portfolio diversification (see, e.g., [6], [16], [9]), we set out in the same direction and explain the fluctuations in conditional correlation between the Western and CEE markets using a set of indicators that help us proxy for the state of the Western developed markets (note: we effectively treat the CEE markets as endogenous). These include, among others, local VIX indices that measure the volatility in the relevant markets and are otherwise used as indicators of investors' confidence in the market, and liquidity variables such as ted-spread (difference between unsecured and secured loans) and percentage bid/ask spread. Other variables, such as measure of aggregate business conditions in [2] are also considered.

A solid understanding of international stock market co-movements is not possible without a joint analysis of the conditional behavior of equity returns and the corresponding exchange rates. This subject is again of key importance for international investors – mainly hedge funds and mutual funds – contemplating the diversification benefits of allocating parts of their portfolios to Central European assets. As noted in [11], developed-country-domiciled mutual and hedge fund holdings already account for about 14–19% of the free-float adjusted market capitalization in Central Europe. As international stock market co-movements tend to be stronger in peri-

[1] FitchSolutions, 30 North Colonnade London E14 5GN, vit.bubak@fitchsolutions.com.
[2] Charles University and Academy of Sciences CERGE-EI, Politickych veznu 7 P.O. Box 882 111 21 Prague Czech Republic, evzen.kocenda@cerge-ei.cz.
[3] Imperial College London Business School, Exhibition Road London SW7 2AZ United Kingdom, f.zikes06@imperial.ac.uk.

ods of distress, an increase in foreign exchange volatility may further amplify the variability of internationally allocated portfolios for investors whose consumption is denominated in a developed-country currency as shown in [3]. The associated rise in the cost of hedging foreign exchange risk then plays an important role in the investment decision-making process and is likely to have an equally important effect on asset market co-movements.

The paper is organized as follows. In the next section we describe our data in detail. In section 3 we introduce the methodology employed. In se section 4 we present our results and briefly conclude.

2 Data

Our dataset consists of daily as well as intraday data available from Bloomberg for the stock markets in Budapest (BUX), Prague (PX 50), Warsaw (WIG 20), Frankfurt (DAX 30), Paris (CAC 40) and London (UKX). Thus, in our analysis we consider three emerging EU markets (Hungary, the Czech Republic and Poland) and three developed EU markets (Germany, France and the United Kingdom). The sample begins on June 3, 2003 and runs till December 31, 2010. The time difference between the markets is accounted for by using Central European Daylight Time (CEDT) for all indices, which eliminates the time difference between London and continental Europe. The index returns are computed as log first differences.

In Figure 1 we present daily levels and returns of all six indices. Levels follow a similar pattern of the upward trend that is replaced by a structural break due to the world financial and economic crises. Post-crisis behavior indicates that a revival is slow and does not exhibit a uniformly increasing pattern.

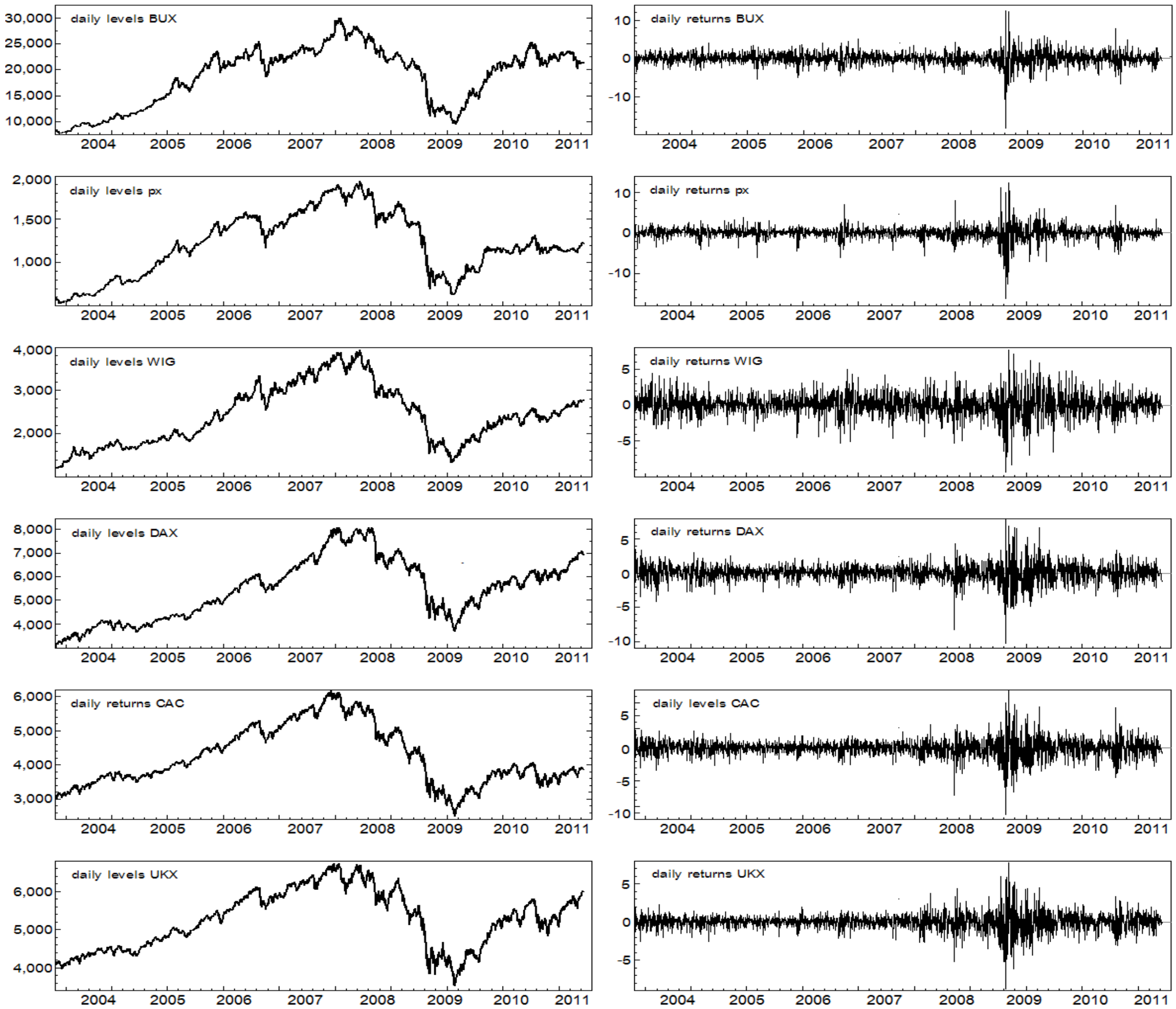


Figure 1 Plots of daily levels (left) and daily returns (right) for the Central European equity indices (top three rows), and the Western European equity indices (bottom three rows). The sample runs from June 3, 2003 to December 31, 2010.

We compute daily returns and obtain 1,885 observations whose descriptive statistics are presented in Table 1. Further, we compute two types of the intraday returns and obtain 135,720 5-minute returns and 5,655 two-hour (120-minute) returns whose descriptive statistics are presented in Tables 2 and 3, respectively. All three types of returns exhibit abnormal extent of kurtosis that is notably excessive in case of the 5-minute returns. In accord with expectations standard deviation of returns increases as the frequency of returns gets lower.

The reason for analyzing two-hour returns – in addition to 5-minute ones - follows from the analysis of the volatility signature plots. It is well know that intraday data sampled at very high frequencies tend to be contaminated by the so-called microstructure noise. The noise arises from a number of frictions inherent to the process of trading and posting bid and ask quotes. See [15] for an overview of the theory of market microstructure and [10] for the implications of the presence of noise for estimating volatility from high-frequency data.

The vast majority of papers in the literature circumvent the problem of noise by sampling sparsely, that is, by sampling at frequencies at which the bias is small. To this end, the so-called volatility signature plot that shows the average daily realized volatility calculated at different sampling frequencies was introduced in [1]. In the absence of noise, this plot should be at zero. If, on the other hand, the noise is present, the signature plot will reveal the frequency at which the bias induced by it kicks in.

Figure 2 shows the volatility signature plots for the realized variances of three Central European markets considered in this study. In addition, Figure 3 shows the signature plots for the realized covariances.

Index	Mean	Std Dev	Skew	Kurt	Min	Max
BUX	0,05	1,73	-0,64	15,91	-19,36	12,42
PX	0,04	1,63	-0,63	17,70	-16,19	12,36
WIG	0,04	1,60	-0,29	5,76	-9,25	7,66
DAX	0,04	1,34	-0,23	8,63	-10,25	7,94
CAC	0,01	1,34	-0,16	8,88	-10,23	8,94
UKX	0,02	1,13	-0,22	9,53	-8,70	7,78

Table 1 Descriptive statistics for daily returns. There are 1,885 observations in the sample that runs from June 3, 2003 to December 31, 2010.

Index	Mean	Std Dev	Skew	Kurt	Min	Max
BUX	-0.001	0.122	0.132	80.03	-5.198	5.272
PX	-0.000	0.085	-0.142	33.56	-1.953	1.948
WIG	-0.000	0.145	0.199	18.98	-2.132	3.007
DAX	-0.000	0.104	-0.194	28.47	-2.181	2.763
CAC	-0.000	0.103	-0.070	26.08	-1.865	2.317
UKX	-0.000	0.087	-0.038	23.32	-1.601	1.903

Table 2 Descriptive statistics for intraday returns. There are 135,720 5-minute returns in the sample that runs from June 3, 2003 to December 31, 2010.

Index	Mean	Std Dev	Skew	Kurt	Min	Max
BUX	-0.035	0.644	-0.800	13.37	-6.706	5.695
PX	-0.009	0.557	-0.807	17.89	-7.347	5.031
WIG	-0.003	0.706	-0.109	10.13	-7.505	6.089
DAX	-0.001	0.483	-0.037	13.02	-4.553	5.246
CAC	-0.004	0.481	0.084	13.97	-4.642	5.170
UKX	-0.008	0.422	-0.050	15.02	-4.127	4.549

Table 3 Descriptive statistics for intraday returns. There are 5,655 2-hour (120-minute) returns in the sample that runs from June 3, 2003 to December 31, 2010.

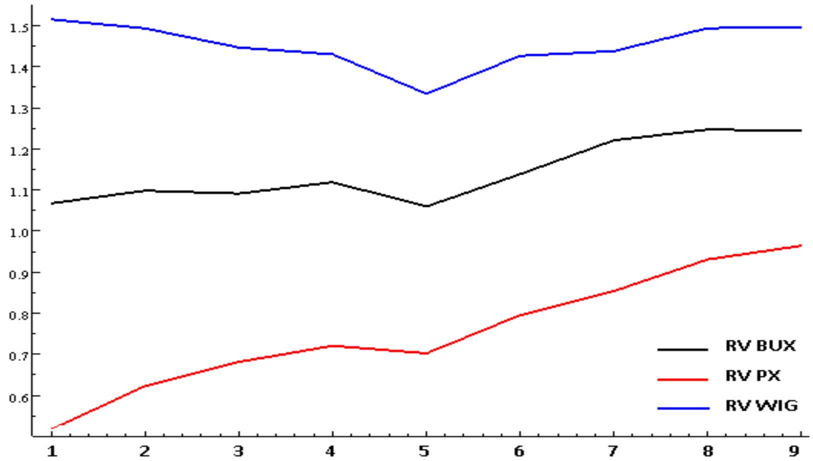


Figure 2 Plots of realized variance based on 5, 10, 15, 20, 25, 30, 60, 90, and 120 minute intraday returns (represented on the x-axis by numbers 1, 2, …, 9, respectively) for the three CEE markets. The sample runs from June 3, 2003 to December 31, 2010.

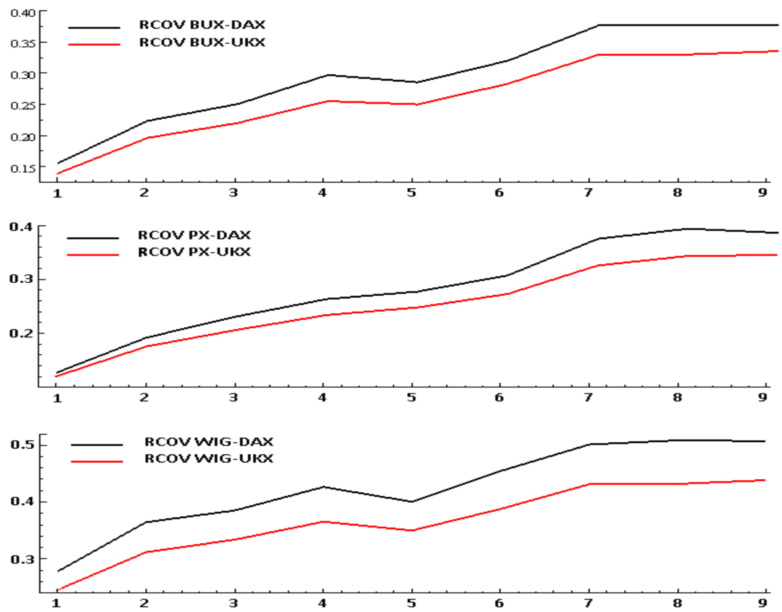


Figure 3 Plots of realized covariance based on 5, 10, 15, 20, 25, 30, 60, 90, and 120 minute intraday returns (represented on the x-axis by numbers 1, 2, …, 9, respectively) for the CEE vs. Western markets. The sample runs from June 3, 2003 to December 31, 2010.

3 Methodology

Given the time series nature of index returns, we first employ a VAR model to model their joint behavior. The residuals from the VAR model are then used as a basis for the analysis of dynamic correlations between the six markets under research. To model the dynamics of the conditional variance of the innovation process (further denoted by ε_t) we employ the DCC model introduced in [7]. In this model, the variance covariance matrix evolves according to

$$H_t = D_t R_t D_t$$

where

$$D_t = diag\left(h_{11,t}^{\frac{1}{2}}, \cdots, h_{KK,t}^{\frac{1}{2}} \right)$$

and $h_{ii,t}$ represents any univariate (G)ARCH(p,q) process, $i = 1,\ldots, k$. The particular version of the dynamic conditional correlation model that we use is due to [8] and [7]. In this model, the correlation matrix is given by the transformation

$$R_t = diag\left(q_{11,t}^{-\frac{1}{2}}, \cdots, q_{KK,t}^{-\frac{1}{2}}\right) Q_t diag\left(q_{11,t}^{-\frac{1}{2}}, \cdots, q_{KK,t}^{-\frac{1}{2}}\right)$$

where $Q_t = (q_{ij,t})$ in turn follows

$$Q_t = (1-\alpha-\beta)\overline{Q} + \alpha\eta_{t-1}\eta'_{t-1} + \beta Q_{t-1}$$

where

$$\eta_t = \frac{\varepsilon_{i,t}}{\sqrt{h_{ii,t}}}$$

are standardized residuals,

$$\overline{Q}_t = T^{-1}\sum_{t=1}^{T}\eta_t\eta'_t$$

is a k x k unconditional variance matrix of η_t, and α and β are non-negative scalars satisfying the condition that

$$\alpha + \beta < 1$$

Recall that it is an ARMA representation of the conditional correlations matrix that guarantees the positive definiteness of Q_t and hence of R_t.

To estimate the DCC-MGARCH model, we proceed as follows. First, we find a suitable specification for each of the four equations of the volatility transmission system as discussed earlier in this section. We continue in the usual way by iteratively removing from each equation the least significant variables until all the variables are significant. The DCC model is then fitted to the series of residuals, where the estimation is performed by optimizing the likelihood function using the Feasible Sequential Quadratic Programming (FSQP) algorithm shown in [14].[4] We estimate the model efficiently in one step to obtain valid standard errors for the DCC estimates.[5]

4 Results

Pattern on daily level shows that CEE markets become well integrated with the developed European markets as dynamic correlation coefficients are quite high (Figure 4). This finding is in line with fragmented evidence in the literature. Further, it conforms to the process of deepening in the CEE capital markets and increased degree of the CEE markets' economic integration with Western Europe as a result of the European integration process and advancements of the CEE countries towards the Eurozone shown in [12] and [13].

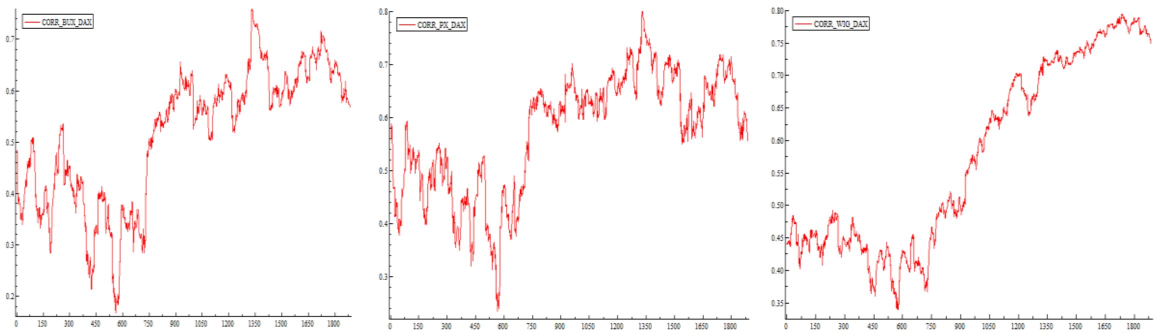


Figure 4 Plots of dynamic daily correlations between the CEE and Frankfurt stock markets. The sample runs from June 3, 2003 to December 31, 2010.

[4] The FSQP algorithm is based on a sequential programming technique to maximize a non-linear function subject to non-linear constraints. See [4] for more details on these two optimization functions.

[5] It is well known that the volatility and the correlation parts of the DCC-MGARCH system can be estimated consistently in two steps. However, the estimators obtained from two-step estimation are limited information estimators (see [8]) and hence are not fully efficient. In our estimation, we used the two-step estimation procedure to obtain accurate starting values for the one-step estimation. Note, however, that we performed both the one-step and the two-step estimations and the corresponding estimates were nearly identical.

Along with the day-to-day analysis, we also investigate the behavior of the markets on intraday level. The intraday analysis poses several challenges, such as the effect of non-synchronous trading on the market data. To this end, we take a great care in correctly specifying the mean equation first and model the contemporaneous correlation using a DCC model.

On the intraday level we find strong co-movements between Western European markets but we find only a little systematic co-movements between the Western European stock markets and the three CEE counterparts. Little co-movements are also found among the CEE markets themselves (figures not presented). These findings fully correspond to those in [5] for an earlier period. The low correlation on the instantaneous level is also tested for a delayed response. We study this effect by computing the empirical correlations of the lagged effects on the markets. Our results show that correlations somewhat increase when we allow for 5 to 10 minutes lags but the effect dissipates very quickly afterwards. Finally we do not detect excessive changes in co-movement patterns due to the recent crisis period that involved dramatic increases in volatility spillovers on financial markets as evidence in [3].

5 Conclusions

The intraday analysis helps us understand the nature of information flow between the equity markets, including the adjustment of the CEE markets to the shocks occurring in the Western developed markets. Our study shows the relevance of its finding for portfolio application, with the objective of comparing the conditional (or modeled) and unconditional behavior the markets in question. Finally, we show that CEE markets integrations progresses and co-movements were not excessively affected by the recent crisis.

References

[1] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P.: Modeling and forecasting realized volatility. *Econometrica* **71** (2003), 579-625.

[2] Aruoba, S. B., Diebold, F. X., and Scotti, C.: Real-Time Measurement of Business Conditions. *Journal of Business & Economic Statistics* **27** (2009), 417 – 427.

[3] Bubák, V., Kočenda, E., and Žikeš, F.: Volatility Transmission in Emerging European Foreign Exchange Markets. *CESifo Working Paper* **3063** (2010).

[4] Doornik, J. A.: *Ox 5.0 - An Object-oriented Matrix Programming Language*. Timberlake Consultants Ltd., London, 2007.

[5] Égert, B., and Kočenda, E.: Time-Varying Synchronization of the European Stock Markets. *Empirical Economics* **40** (2011), 393-407.

[6] Eichengreen, B., and Andrew, K. R.: *Contagious currency crises: Channels of conveyance, in Changes in Exchange Rates in Rapidly Developing Countries*. Ed.: T. Ito and A. Krueger, 1998.

[7] Engle, R. F.: Dynamic Conditional Correlation – A Simple Class of Multivariate GARCH Models. *Journal of Business and Economic Statistics* **20** (2002), 339 – 350.

[8] Engle, R. F., and Sheppard, K.: Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. *National Bureau of Economic Research Working paper* **8554** (2001).

[9] Glick, R., and Rose, A.: Contagion and trade: Why are currency crises regional? *Journal of International Money and Finance* **18** (1999), 603 – 617.

[10] Hansen, P. H., and Lunde, A.: Realized variance and market microstructure noise. *Journal of Business & Economic Statistics* **24** (2006), 127 – 161.

[11] Jotikasthira, C., Lundblad, C., and Ramadorai, T.: Asset Fire Sales and Purchases and the International Transmission of Financial Shocks. *University of Oxford Working Paper* **7595** (2009).

[12] Kočenda, E., Kutan, A. M., and Yigit, T. M.: Pilgrims to the Eurozone: How Far, How Fast? *Economic Systems* **30** (2006), 311-327.

[13] Kočenda, E., Kutan, A. M., and Yigit, T. M.: Fiscal Convergence in the European Union**.** *North-American Journal of Economics and Finance* **19** (2008), 319-330.

[14] Lawrence, C. T., and Tits, A. L.: A Computationally Efficient Feasible Sequential Quadratic Programming Algorithm. *SIAM Journal of Optimization* **11** (2001), 1092–1118.

[15] O'Hara, M.: *Market Microstructure Theory*. Blackwell Publishing Limited, 1995.

[16] Rigobon, R.: Informational speculative attacks: Good news is no news. *MIT Working paper* January (1998).

Optimizing the Parallel Logistic Processes with Shared Regeneration Plants

Robert Bucki[1]

Abstract. The article highlights the problem of mathematical modelling of the independent logistic production systems which share the same regeneration plants. Each of the production systems consists of production units which are arranged in series. Each unit consists of production stands which are made active in case of necessity. A production stand is equipped with a tool which is used during the production process. If the tool is completely worn out, it needs to be replaced or regenerated. The problem consists in determining the best order realization sequence in order to minimize the lost flow capacity of production stands. The idea of the time scaling method is proposed in order to maximize the total production output. Heuristic algorithms are used to control the choice of the order vector elements and determine the production route. The criterion of production output maximization is presented.

Keywords: mathematical modelling, optimization, heuristic algorithms, computer simulation, production systems, regeneration procedure.

JEL Classification: C02
AMS Classification: 68R99

1 Introduction

Increasingly competitive markets bring new challenges and customers' demands are constantly changing. As a result, manufacturers need to become more responsive, more quickly, more efficiently, yet often within tighter budgets and timescales. Simulators are helping them to visualize, analyze and optimize their processes to achieve these business performance improvements [5]. Fundamental principles for modelling complex, real-world systems with the aim to simulate efficiently in order to closely resemble reality remain the basic goal of all optimization techniques [4]. The ability to model and simulate abstract concepts in the computer is truly a remarkable tool and it is worth exploring its usage to augment the design of highly-complex production systems. Optimization must be based upon business objectives to find the schedule that produces the product with the least cost or the shortest production lead time [2]. It all means that the planner ought to be presented with the best algorithm enabling him to manufacture the best sequence of products. Knowing that there is an order to be realized, we can easily find out in advance what kind of charge will be necessary and what amount of it will be needed [3].

Order realizing requires complying with specific features of the system in which this order is made. Each production system uses specific tools which get worn out throughout the order realization period. There are tools that can be subjected to regeneration and those which must be replaced with the identical new ones. Tools can be regenerated a certain number of times. After exceeding this number the tool is excluded from the production process and is subject to utilization. Then it is replaced by a new identical tool in order to restore the required number of tools. The problem of regenerating tools is described in detail in [6] along with the procedures implemented. Simulations carried out with the use of those procedures prove that there is a need to seek for a better solution. The bigger the number of simulations, the bigger chance to obtain a satisfactory solution.

The problem of mathematical modelling of the independent logistic production systems which share the same regeneration plants is highlighted hereby. The best order realization sequence in order to minimize the total production time is sought for. Heuristic algorithms control the production process. Maximization of the production output and minimization of the lost flow capacity of the system as well as minimization of the tool replacement time remain the priority criterion [7]. This approach in elastic production systems is carried out by means of discrete optimization [8,12]. Moreover, logistic methods play an important role in supporting complexes of operations while seeking for the satisfactory solution to the specified optimization task [11].

Manufacturing systems are indispensably connected with e-commerce systems which form a standard environment and support business activities characterized by a wide range of parameters. These parameters describe individual e-commerce system components, are important indicators of e-commerce system status and are of great importance for business process management. All e-commerce system parameters must be

[1] The College of Informatics and Management, ul. Legionów 81, 43-300 Bielsko-Biała, rbucki@wsi.edu.pl

continuously monitored and more precisely measured as well as categorized [13]. New modelling and system design techniques are required for information technologies that can support the enterprise in achieving and sustaining the necessary flexibility [10]. Systems of linear discrete equations employed to support optimizing the output of the economic system require determining corresponding initial data which generate adequate bounded solutions [1]. A strict control of the procedures through each step is required in such a case [9].

2 General assumptions

Let us introduce a sample logistic structure which requires the adequate control approach. The problem itself consists in determining the sequence of elements of the order vector which are to be realized subsequently. The proposed heuristic algorithms choose the required element on which certain operations are carried out. The state of orders decreases after each production decision which influences the state of the whole logistic system at each stage *k*, $k = 1,\ldots,K$. The given criteria are used on condition that each of them is associated by adequate bounds. Assuming that the results of calculations which are made for a chosen heuristic algorithm do not deliver a satisfactory solution, there arises a need to test other algorithms.

Let us introduce the vector of charges $W = [w_l]$ where w_l - the *l*th charge material, $l = 1,\ldots,L$ and the vector of orders $Z = [z_n]$ where z_n - the *n*th production order (given in units), $n = 1,\ldots,N$.

Now, we propose the assignment matrix of products to charges $\Omega = \lfloor \omega_{n,l} \rfloor$ where $\omega_{n,l}$ - the assignement of the *n*th product to the *l*th charge material, $n = 1,\ldots,N$, $l = 1,\ldots,L$. Elements of the assignment matrix take the following values:

$$\omega_{n,l} = \begin{cases} 1 & \text{if the } n\text{th product is realized from the } l\text{th charge,} \\ 0 & \text{otherwise.} \end{cases}$$

We also assume that used charge vector elements are immediately supplemented which means that we treat them as the constant source of charge material. However, for simplicity reasons, we assume that each *n*th product is made from the universal charge which enables realization of the given *n*th product from any *l*th charge material element, $l = 1,\ldots,L$.

The logistic system presented hereby consists of *I* parallel manufacturing subsystems. Each *i*th subsystem, consists of *J* production stands arranged in series, $i = 1,\ldots,I$. Realized products are passed subsequently through each stand in the *i*th subsystem. There are no buffer stores between stands. Each *j*th production stand located in the *i*th row of the logistic system carries out an operation on the *n*th product. It is assumed that each production stand placed in the *j*th column realizes the same operation with the use of the identical tool. Moreover, it is also assumed that $I < N$.

Let us introduce the vector of regeneration plants $R = \lfloor r_j \rfloor$, $j = 1,\ldots,J$ (the *j*th regeneration plant regenerates tools which are used in each manufacturing stand placed in the *j*th column of the discussed logistic system) and at the same time $r_j = \begin{cases} 1 & if \quad active, \\ 0 & otherwise. \end{cases}$

The problem highlighted in the paper consists in maximizing the total production output by means of minimizing the regeneration time of tools as well as finding the optimal sequence of production decisions which are meant to send totally worn out tools to the *j*th regeneration plant - r_j. The given tool can be regenerated only a defined number of times. If this number is exceeded, the tool is excluded from the production process and must be replaced by a new one. However, there are also tools which cannot be regenerated and must be replaced ($r_j = 0$).

Let us introduce the structure matrix of the logistic system at the *k*th stage:

$$E^k = \left[\underset{r_j \lrcorner}{e^{k}_{i,j}} \right],$$

where: $\underset{r_j \lrcorner}{e^{k}_{i,j}} = \begin{cases} 1 & if \quad active, \\ 0 & otherwise. \end{cases}$

The flow of the order elements in the logistic system is formed on the basis of the discussed assumptions is shown below:

$$[w_{\omega,l}] \Rightarrow \left\{ \begin{array}{ccccc} e^{k}_{1,1} & \to \cdots \to & e^{k}_{1,j} & \to \cdots \to & e^{k}_{1,J} \\ r_1 & & r_j & & r_J \\ \cdots & & \cdots & & \cdots \\ e^{k}_{i,1} & \to \cdots \to & e^{k}_{i,j} & \to \cdots \to & e^{k}_{i,J} \quad \cdots \\ r_1 & & r_j & & r_J \\ \cdots & & \cdots & & \cdots \\ e^{k}_{I,1} & \to \cdots \to & e^{k}_{I,j} & \to \cdots \to & e^{k}_{I,J} \\ r_1 & & r_j & & r_J \end{array} \right\} \Rightarrow \begin{bmatrix} z_1 \\ \cdots \\ z_n \\ \cdots \\ z_N \end{bmatrix}$$

The life vector of the logistic system for a brand new set of tools is defined: $G = \lfloor g_j \rfloor,\ j = 1,\ldots,J$, where g_j is the number of the nth product units which can be realized in any production stand in the jth column before its tool is completely worn out and requires an immediate replacement with either a regenerated or new tool.

We assume that every decision about production, replacement or regeneration is made at the stage $k-1$.

Let $S_n^{k-1} = \left[s(n)_{i,j}^{k-1} \right]$ be the matrix of state of the logistic system for the nth product realization at the stage $k-1$ where $s(n)_{i,j}^{k-1}$ is the number of units of the nth product already realized in the stand in the ith row of the jth column with the use of the installed tool.

Let $P_n^{k-1} = \left[p(n)_{i,j}^{k-1} \right]$ be the matrix of the flow capacity of the logistic system for the nth product realization at the stage $k-1$ where $p(n)_{i,j}^{k-1}$ is the number of units of the nth product which still can be realized in the stand in the ith row of the jth column. If $\underset{1 \le i \le I}{\forall} p_{i,1}^{k} = 0$, then the nth order awaits for completing the regeneration process and installing a new tool to enter the production system.

On the basis of the above assumptions we can determine the flow capacity of the production stand in the ith row of the jth column for the nth element of the order vector Z at the stage $k-1$: $p(n)_{i,j}^{k-1} = g_j - s(n)_{i,j}^{k-1}$

The manufacturing procedure consists in realizing orders in parallel production routes in sequence. It is assumed that manufacturing another order element in a route can begin when the previously realized one leaves the route. Its disadvantage consists in the need of waiting for completing the manufacturing process of a certain product in this route before resuming it again for the next one. This results in not using the available flow capacity of the whole production system. Moreover, during the production course tools must be replaced. The state of the system has to be recalculated when any decision is made in the system.

Let us define the matrix of production times for the nth product, $n = 1,\ldots,N$ in the production stand in the jth column $T^{pr} = \lfloor \tau_{n,j}^{pr} \rfloor$. If the nth product is not realized in the production stand in the jth column, then $\tau_{n,j}^{pr} = 0$. If $\tau_{n,j}^{pr} \le \tau_{n,j+1}^{pr}$, then the jth production stand becomes blocked while manufacturing the nth product.

Let us define the vector of replacement times for the tools in the logistic system $T^{repl} = \left[\tau_j^{repl} \right]$ where τ_j^{repl} - the replacement time of the tool in the production stand in the jth column.

Let us introduce the production rate vector $V = \lfloor v_n \rfloor$. Its element v_n is the number of units of the nth product made in the time unit in any ith production line, $i = 1,\ldots,I$.

In order to calculate the total manufacturing time of all elements from vector Z it is necessary to take into account the production time, the replacement time and, finally, the regeneration time. The order realization time can be optimized by either employing more production lines at the same time to realize the nth element or replacing tools only then, when they are fully worn or optimizing the regeneration process so that the tool after regeneration is available on demand.

The total order realization time T is calculated beginning with the moment when any nth element enters the logistic system till the moment when the last element of the order vector leaves any stand in the Jth column.

3 Equations of state

The state of the discussed parallel logistic system changes in case of manufacturing the *n*th product as follows:

$$S_n^0 \to S_n^1 \to \ldots \to S_n^k \to \ldots \to S_n^K$$

The state of the production stand in case of production the *n*th product changes consequently:

$$s(n)_{i,j}^0 \to s(n)_{i,j}^1 \to \ldots \to s(n)_{i,j}^k \to \ldots \to s(n)_{i,j}^K$$

which can be written in the following form:

$$s(n)_{i,j}^k = \begin{cases} s(n)_{i,j}^{k-1} & \text{if no } n\text{th product is realized in the } i\text{th stand of the } j\text{th column at the } k\text{-1 stage,} \\ s(n)_{i,j}^{k-1} + x_n^k & \text{otherwise.} \end{cases}$$

The state of the production stand in case of replacement of the tool changes in the way shown below:

$$s(n)_{i,j}^k = \begin{cases} s(n)_{i,j}^{k-1} & \text{if the tool is not replaced in the } i\text{th stand of the } j\text{th column at the stage } k\text{-1,} \\ 0 & \text{otherwise.} \end{cases}$$

The order vector Z changes after every production decision:

$$Z^0 \to Z^1 \to \ldots \to Z^k \to \ldots \to Z^K$$

The order vector is modified after every decision about production:

$$z_n^k = \begin{cases} z_n^{k-1} - x_n^k & \text{if the number of units } x_n^k \text{ of the } n\text{th order is realized at the } k\text{th stage,} \\ z_n^{k-1} & \text{otherwise.} \end{cases}$$

4 Order heuristics

In order to control the choice of the order vector elements we need to implement heuristics which determine elements from the vector Z for the production process. The following control algorithms for production are put forward:

- the algorithm of the maximal order,
- the algorithm of the minimal order,
- the algorithm of the relative order.

4.1 The algorithm of the maximal order

This algorithm chooses the biggest order vector element characterized by the biggest coefficient γ_n^{k-1} in the state S_n^{k-1}. To produce the element a, $1 \le a \le N$ the condition $(q_n^k = a) \Leftrightarrow \left[\gamma_a^{k-1} = \max_{1\le n\le N} \gamma_n^{k-1}\right]$, $\gamma_n^{k-1} = z_n^{k-1}$ must be met. This approach is justified by avoiding excessive bringing the production line to a standstill in order to change an element to be manufactured. If only minimal orders were chosen in state S_n^{k-1}, in consequence the number of orders might be reduced. Such control is favorable because the *n*th production line is blocked and must be stopped only in order to replace the tools in certain stands (on condition that the replacement process disturbs the flow of the material).

4.2 The algorithm of the minimal order

This algorithm chooses the smallest order vector element characterized by the smallest coefficient γ_n^{k-1} in the state S_n^{k-1}. To produce the element a, $1 \le a \le N$ the condition $(q_n^k = a) \Leftrightarrow \left[\gamma_a^{k-1} = \min_{1\le n\le N} \gamma_n^{k-1}\right]$, $\gamma_n^{k-1} = z_n^{k-1}$ must be met. This approach is justified by the need to eliminate the elements of the order vector Z which could be sent to the customer just after the *n*th product leaves the production line on condition that the customer sets such a requirement.

4.3 The algorithm of the relative order

This algorithm chooses the order element characterized by the biggest relative order coefficient γ_n^{k-1} in the state S_n^{k-1}. To produce the element a, $1 \le a \le N$ the condition $(q_n^k = a) \Leftrightarrow \left[\gamma_a^{k-1} = \max_{1 \le n \le N} \gamma_n^{k-1}\right]$, $\gamma_n^{k-1} = \frac{z_n^{k-1}}{z_n^0}$ must be met. It is assumed that the orders are realized one after another that is to say each order element z_n in the state S_n^{k-1} is reduced partly. Such control is advantageous when some parts of the order are needed earlier.

5 Sub-line heuristics

In order to control the choice of the line we need to implement heuristics which determine the subsystem for producing the order on the basis of the flow capacity of the routes (subsystems) defined in the vector $P_i^{k-1} = \left\lfloor p(n)_i^{k-1} \right\rfloor$ where $p(n)_i^{k-1}$ - the flow capacity of the ith subsystem. The algorithms of the maximal and minimal capacity of the subsystem are put forward.

5.1 The algorithm of the maximal flow capacity of the sub-system

This algorithm chooses the route characterized by the maximal flow capacity of the subsystem i.e. the maximal coefficient ρ_i^{k-1}. To choose the route b, $1 \le b \le I$ the condition $(q_i^k = b) \Leftrightarrow \left[\rho_b^{k-1} = \max_{1 \le i \le I} \rho_i^{k-1}\right]$, $\rho_i^{k-1} = p(n)_i^{k-1}$ must be met. It is assumed that the nth order is manufactured in the minimal number of subsystems which may not lead to splitting the order.

5.2 The algorithm of the minimal flow capacity of the sub-system

This algorithm chooses the route characterized by the minimal flow capacity of the subsystem i.e. the maximal coefficient ρ_i^{k-1}. To choose the route b, $1 \le b \le I$ the condition $(q_i^k = b) \Leftrightarrow \left[\rho_b^{k-1} = \min_{1 \le i \le I} \rho_i^{k-1}\right]$, $\rho_i^{k-1} = p(n)_i^{k-1}$ must be met. It is assumed that the nth order is manufactured in the bigger number of subsystems which may not lead to faster manufacturing the order.

6 Regeneration procedures

Let us introduce the matrix $\Delta = \left\lfloor \delta_{i,j} \right\rfloor$ which elements represent the number of allowable regeneration processes of the tool for the stand located in the jth column. If $r_j = 0 \Rightarrow \forall_i \delta_{i,j} = 0$, then tools from the stands in the jth column cannot be regenerated and must be replaced by a new one. If $\forall_i \delta_{i,j} \ne 0$ and at the same time the number of already carried out regeneration procedures has been exceeded, the tool cannot be regenerated anymore and must be replaced with a new one. We assume that the replaced tools are available at once.

Each active regeneration plant r_j uses the FIFO procedure consisting in regenerating the worn out elements which are in the queue as the first ones. Then they are returned to the adequate production stand which has been the longest period of time in the standstill mode. Machines in the production stands use tools which either can be regenerated or must be replaced by a new one. Each tool subjected to regeneration is indexed. The matrix of allowable regeneration procedures is given in the vector $R^{reg} = \left\lfloor \varsigma_j^{reg} \right\rfloor$ where $\varsigma_j^{reg} = \varphi$, $\varphi = 0,1,\ldots,\Phi$. If $\varsigma_j^{reg} = 0$, then tools from stands in the jth column cannot be regenerated and must be replaced.

7 The production maximization criterion

The production maximization criterion takes the form $Q_1 = \sum_{k=1}^{K} q_1^k = \sum_{k=1}^{K} \sum_{n=1}^{N} x_n^k \to \max$ where x_n^k is the number of units of the nth element realized at the kth stage. This criterion requires the use of the tool replacement bound $\sum_{i=1}^{I} \sum_{j=1}^{J} y_{i,j}^k \tau_j^{repl} \le c$ where c - the maximal allowable tool replacement time, τ_j^{repl} - the replacement time of the used tool in the stand in the *j*th column and if $y_{i,j}^k = 1$, the replacement procedure of the tool in the stand $e_{i,j}$ is carried out (otherwise $y_{i,j}^k = 0$) as well as the flow capacity bound $y_{i,j}^k \sum_{i=1}^{I} p(n)_{i,j}^k \leq g_j$. The production maximizing criterion is reduced to the replacement time of tools and flow capacity bounds.

8 Final remarks

It would be advisable to verify another regeneration approach i.e. the LIFO procedure. Moreover, implementing the regeneration heuristic consisting in choosing the tool with the highest number ς_j^{reg} may result in minimizing the total order realization time. The number of production lines forming the discussed logistic system should be minimized as their bigger number leads to generating unnecessary costs due to the fact that there must be an operator employed. In conclusion, the need for carrying out a computer simulation must be met in order to project the logistic production system which will be able to realize the order in the lowest possible time at the lowest possible costs. To do this, there must be a multi-criterion model with adequate bounds created. Another idea accelerating the order realization process would be beginning realization of the next element from the order vector without having to wait for completing manufacturing the previous units of the *n*th element which means an immediate employing the stand with no manufacturing duty. However, it requires separate control for each *i*th production line. There is also the need to examine the production time minimization criterion $Q_T \to \min$ and the minimal lost flow capacity criterion $Q_P \to \min$. The latter criterion should be implemented in case there remains an extra time to realize the order. These criteria are to be implemented along with adequate bounds.

Acknowledgements

The work was supported by the grant MEB051031 (VŠE Praha – WSIZ Bielsko-Biała) financed by MŠMT ČR.

References

[1] Bastinec, J. Diblik, J.: On Initial Data Generating Bounded Solutions of Linear Discrete Equations. *Journal of Applied Mathematics*. Vol. 1, 2008, No. 2, pp. 111-116.

[2] Bucki, R.: Managing Flexibility of the Production System with Interoperation Stores. In: *Mezinárodní vědecký seminář - Nové trendy v ekonometrii a operačním výzkumu*. Katedra ekonometrie Fakulty informatiky a statistiky VŠE v Prahe a Katedra operačného výskumu a ekonometrie Fakulty hospodárskej informatiky EU v Bratislave, Zbornik 15. – 17. december 2010, Ekonomická univerzita v Bratislave, Vydavateľstvo EKONÓM, Bratislava 2010.

[3] Bucki, R.: Managing a Flexible Economic System. *Acta academica karviniensia*, Slezská univerzita v Opavě, Obchodně podnikatelská fakulta v Karviné, č. 2, roč. 2010, s. 51-58.

[4] Bucki, R.: Thorough Analysis of the Technological Case Control. *Management & Informatics*, Network Integrators Associates, Parkland, Florida, Vol. 1, No. 1, 2007, pp. 68-112.

[5] Bucki, R. and Antonyová, A. and Suchánek, P.: Economic Control of the Distribution System by Means of Heuristic Approach. *Information Technologies and Systems*. Vol. 12, No. 2. National Academy of Sciences of Ukraine, State Department on Communication and Informatization of Ukraine, State Scientific and Research Institute of Information Infrastructure, 2009, pp. 57-66.

[6] Bucki, R. and Marecki, F.: *Digital Simulation of Discrete Processes*. Network Integrators Associates, Parkland, Florida, 2006, p. 162.

[7] Chramcov, B., Daníček, L.: Simulation Study of the Short Barrel of the Gun Manufacture. *In: Proceedings of 23rd European Conference on Modelling and Simulation.* Madrid: European Council for Modelling and Simulation, 2009. s. 275-280.

[8] Dlouhý, M., Fabrý, J., Kuncová, M., and Hladík, T.: *Simulace podnikových procesů*, Praha, Computer Press, 2007, 201 p.

[9] FIALA, P.: *Modelování a analýza produkčních systemů*. Praha: Professional Publishing, 2002, 259 p.

[10] Hunka, F., Hucka, M., Kasik, J., and Vymětal, D.: REA Value Chain Applied on Production Planning Model. *The 2nd International Conference on Logistics and Transport & The 1st International Conference on Business and Economics*. Rydges Lakeland Resort Queenstown, Queenstown, New Zealand. 16-18 December 2010.

[11] Marecki, J., Łużny, D., Marecki F.: *Logistyka Informatyczna Kompleksów Operacji*. Wyższa Szkoła Informatyki i Zarządzania, 2010. 166 p.

[12] Sawik, T.: *Optymalizacja dyskretna w elastycznych systemach produkcyjnych*, Wydawnictwa Naukowo-Techniczne, Warszawa, 1992, str. 14-31, 162-173.

[13] Suchánek, P.: Systemization of Measurable Parameters of E-commerce Systems. *In Vedecký monografický zborník Informačné technológie v manažmente a marketingu*. Žilina: Žilinská univerzita – Fakulta řízení a informatiky, 2010. s. 97-102.

Modelling and Simulation of Complexes of Operations in Logistic Systems

Robert Bucki[1], Zbigniew Frąckiewicz[2], Franciszek Marecki[3]

Abstract. The paper highlights the problem of modelling and simulation of complexes of operations concerning objects servicing in the logistic system. As an example of the logistic system, the set of two container ports is analyzed. Each port has a serial-parallel structure (loading berths and discharging wharfs). The complex of operations consisting in servicing ships in these ports (unloading and loading) is modelled by means of multi-stage programming. Modelling and digital simulation of servicing objects in the logistic system let us determine the effectiveness of heuristics at different stages of the system. This knowledge is used to control complexes of operations in logistic systems.

Keywords: mathematical modelling, complexes of operations, heuristic control, logistic system.

JEL Classification: C02
AMS Classification: 68R99

1 Introduction

Logistics originates from operations research [18,19], military uses as well as mass service systems. The basics of contemporary logistics lie in information technology (computers, information nets and mobile telephony). Logistics deals with examining complexes of operations [6,9]. Complex logistic systems are analyzed by means of modelling and simulation methods [1,2]. Modelling and simulation are traditionally used during projecting systems whereas logistics manages and controls realization of the complex of operations. From this point of view, the use of modelling and simulation is not a conventional optimization method. However, operators who implement heuristics resulting from their knowledge - natural intelligence, make decisions in logistic systems. Artificial intelligence allows us to verify the effectiveness of these heuristics [5,10,11,12]. Moreover, digital simulation lets us gather knowledge on effects of heuristic approach usage in logistic systems. The article hereby presents modelling and simulation of complexes of operations in logistic systems based on the above remarks.

Modelling and simulation can be used to either optimize operative control or create time-scales for logistic processes. Operative control is implemented in random environment [6,13,14] whereas time-scaling is used in deterministic environment [3,4]. Practically, there are mixed situations in which short-time scaling (e.g. per a working shift) and operative control (the decision about each operation) must be used.

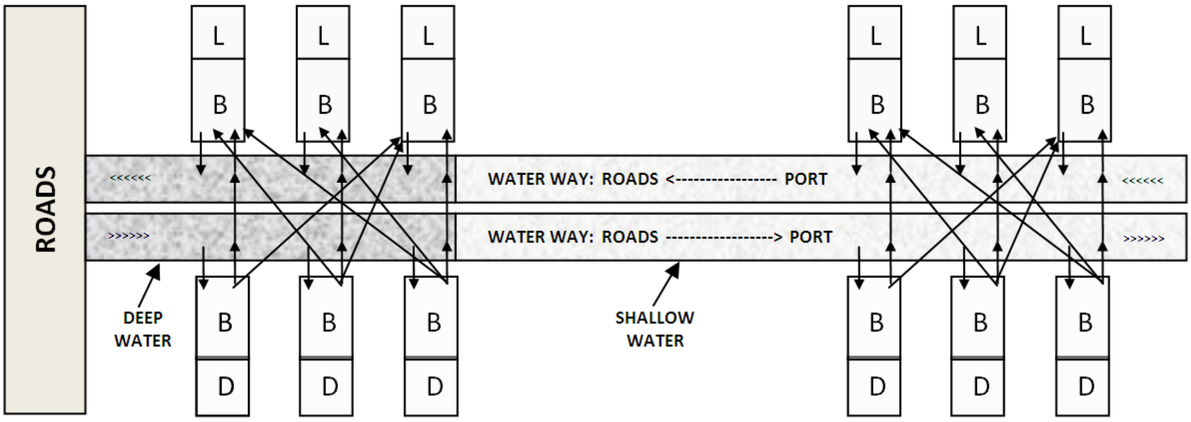
Logistic systems usually have a certain structure (i.e. serial, parallel, tree-like, etc.). These systems realize tasks of complexes of operations: servicing, transporting, storing, etc. These operations are characterized by logical, time and location (space) bounds. Complexes of operations result from the need to service streams of objects in the systems which have a certain structure. From this point of view control of the complexes of operations is not classical optimization. Streams of objects are analyzed in time intervals (e.g. days) as there are no data about future objects. The article is an example of the logistic system of two container ports. Ships which are unloaded, loaded and transported are determined objects to be serviced. To model servicing ships in ports the method of multi-stage programming is used [6,9]. Other logistic systems can be modelled analogously.

2 Logistic problem formulating

Let us consider the logistic system modelled on the basis of the port complex consisting of two container ports. Such a situation is common in case of ports situated at a river mouth. Traditionally, the basic port is built inland and an additional port is at the mouth of the river. The ports of Szczecin (the basic port) and Świnoujście (the additional port) or Hamburg (the basic port) and Cuxhaven (the additional port) are examples of such structures. Building of the port complex is justified, among others, by the constant depth of water ways. New ships are bigger and bigger and their extreme draught increases. Due to this fact, big and loaded ships cannot reach the basic port directly from the open sea roads because of the draught which is bigger than the depth of the port canal. Bigger ships are unloaded in an additional port to decrease their draught. Only after the operation of discharging the load partly, they are tugged to the basic port. The analogue situation happens during loading big

[1] The College of Informatics and Management, ul. Legionów 81, 43-300 Bielsko-Biała, rbucki@wsi.edu.pl
[2] The College of Informatics and Management, ul. Legionów 81, 43-300 Bielsko-Biała, z.frackiewicz@wp.pl
[3] The College of Informatics and Management, ul. Legionów 81, 43-300 Bielsko-Biała, fmarecki@gmail.com

ships. They are loaded only partly in the basic port. After tugging them to the additional port, the loading process is completed. There are cases of big ships which were launched in the basic port of Hamburg. Then they left Hamburg with no load for the managing ship owner, never to return to the basic port of Hamburg as loaded ones. Port canals are constantly silted up by the sea sand which makes them shallow. Moreover, the depth of port canals is changeable depending on the season of the year as the state of rivers varies. There are also tides and ebbs caused by the Moon's gravity. The availability of the port canal depends also on weather. The general structure of the port as the logistic system is shown in Fig. 1. There are roads, the basic port, the additional port, the water ways in the port canals and harbour basins.



B – harbour basin; **L** - loading berth; **D** - discharging wharf; >>> route of the ship; → route of the tugged ship

Figure 1 The structure of the logistic system: the system of two ports

The ships await for servicing in the port system at the roads. Let us assume that the ship is to be discharged and then loaded. In this case, the route runs from the roads to the additional port where the ship is partly discharged and tugged from the additional port to the basic port. There, the ship is completely unloaded and partly loaded. After this operation, the ship is tugged back to the additional port where an extra loading process is carried out. Then, the ship goes to the roads. Discharging and loading operations of the ship are carried out at the wharfs equipped with the use of proper cranes. Ships are tugged from the roads to the ports. Moreover, ships await for servicing (discharging, loading or transport) in the harbour basins which are shown in Fig. 1.

The logistic problem consists in such servicing of ships in the port system so that economic effects of ports are maximal. The economic effects result from the freight account. If the ship is serviced in the port system before the declared term, then the port staff are awarded a bonus. Otherwise, the port team pays a fine for the delay. The problem consists in the fact that time-scales are set for a short period of time (e.g. for a week) and economic effects can be maximized in a year's time. It is known that a weekly multiple optimization process delivers a worse result than one-time yearly optimization. However, yearly optimization cannot be carried out directly as ships which will be serviced in the port system in the future are not known [9].

Optimization of ships in the port can be discussed as the classical static problem. First of all, there are not data about future ships in the analyzed time horizon. Secondly, times of servicing and transporting ships are not deterministic due to atmospheric disturbances. Moreover, transport of ships in the port system is realized by tugs (an extra resource). Tugs are treated as multi-agent operatively controlled system on the basis of calls to transport ships [8].

Because of the above, optimization of ships in the port system is carried out by means of mathematical modelling and digital simulation. Control of the port system as a logistic system is treated as two-level. Time-scales of servicing ships (discharging and loading) are determined at the higher level. This case requires decisions made on the basis of heuristics. The mathematical model of servicing ships includes heuristics which allow us to generate allowable time-scales. These time-scales are determined by means of the digital simulation method at the given initial state of the port system and are supported by heuristic algorithms. Ships are discharged and loaded according to the time-scales. When either discharging or loading is completed, the tug is required. The transport time of the ship is given but the tug may not be available. This is the reason why operative control of tugs disturbs discharging and loading ships.

3 Complex of logistic operations

The port system consists of two ports (basic and additional). The water area of the harbour and the port territory are distinguished in these ports. The water area of the harbour comprises basins and port canals. Container stores are situated in the port territory. The contact area of the water area of the harbour and the port territory is formed by the quay. Discharged and loaded ships are placed adequately at the discharging wharf and at the loading berth. The quays are classified as discharging and loading ones. The port system can be decomposed into: discharging wharfs of the basic port, loading berths of the basic port, discharging wharfs of the additional port and loading berths of the additional port.

Let us assume there are *M* quays at which *N* ships are to be serviced (discharged or loaded). The matrixes of operation times of servicing ships are introduced:

$$\Xi = \lfloor \upsilon_{m,n} \rfloor,\ m = 1,\ldots,M\ , n = 1,\ldots,N \tag{1}$$

where: $\upsilon_{m,n}$ - the time of servicing the *n*th ship at the *m*th quay.

In the general case, the number of the discharging wharfs and the number of the loading berths differ in the basic and additional ports. Moreover, the number of discharged ships and the number of loaded ships at these quays are different. So the dimensions of the matrix (1) are different for various quays. However, operations of servicing ships at quays are always described by matrixes (1).

Let us assume that there are given transport times of ships between:

- the roadstead and discharging wharfs of the additional ports,
- the discharging wharfs of the additional port and the discharging wharfs of the basic port,
- the discharging wharfs and the loading berths of the basic port,
- the loading berths of the basic port and the loading berths of the additional port,
- the loading berths of the additional port and the roadstead.

Let us assume that there are *I* initial points and *J* target points for transporting ships. The matrixes of transporting ships are given:

$$\mathrm{T} = \lfloor \tau_{i,j} \rfloor,\ i = 1,\ldots,I\ , j = 1,\ldots,J \tag{2}$$

where: $\tau_{i,j}$ - the time of transporting the ship from the *i*th to the *j*th point.

In a general case, the number of initial points and the number of target points for transporting ships differ in various areas of the port: the roadstead – the additional port – the basic port. Moreover, the transport time depends on the ship and the tug. Transporting some ships requires more than one tug, etc. However, transport operations are always described by matrixes (2). There exist certain logical and localization (space) relations between operations of servicing ships at quays. Unloading operations precede loading operations. Unloading operations at an extra quay precede loading operations at the basic quay. Loading operations at the basic quay precede loading operations at an extra quay. Moreover, it is important to determine the sequence and place of unloading and loading a container on the ship.

There are also time bounds in the complex of operations of servicing ships, e.g. loading operations can begin after the determined time due to availability of containers in stores. Containers which are unloaded from a certain ship can be loaded on another ship which means that the sequence of servicing ships is important as well. From the logistic point of view, the number of weeks during which the ship awaits for servicing is of a high importance. To comply with this requirement, ships are given priorities equaling the number of weeks during which they have to await for servicing. Ships with higher priorities are serviced as the first ones.

4 Modelling complexes of operations

Modelling the process of servicing ships in the port should deliver a possibility to simulate subsequent operations of unloading, loading and transport. In practice, realization of these operations depends on the port dispatcher. The port dispatcher uses heuristics while choosing a ship and a quay. This is justified by the fact that there are not all available data, the data are not deterministic and there is no time for classical optimization of the combinatory problem. There are heuristics in the mathematical model of servicing ships in the port system. Thanks to such a model, a time-scale of servicing ships in the port system must be determined. The time-scale depends on implemented decisive rules. The multi-stage programming method is used to model the logistic process. This method requires defining: the state of the logistic system, the procedure of generating trajectories of states, the value of states and the rules of eliminating non-perspective states. The logistic system of the port system can be decomposed into separate groups of discharging wharfs and loading berths, the basic port or the additional port.

Let us assume that there are *M* quays at which *N* ships are to be serviced. The state X of such a system can be defined in the matrix form composed of two columns and *N* rows. The rows are connected with ships which are serviced and columns are defined as follows:

$$\mathrm{X} = \lfloor x_{n,j} \rfloor, \; n = 1,\ldots,N \; , \; j = 1,2 \tag{3}$$

where: $x_{n,1} = m$ if the *n*th ship was serviced at the *m*th quay,

$x_{n,1} = 0$ if the *n*th ship was not serviced,

$x_{n,2} = t_{m,n}$ if the *n*th ship was serviced at the *m*th quay at the moment $t_{m,n}$,

$x_{n,2} = 0$ if the *n*th ship was not serviced.

The initial state marked by X^0 is given (e.g. it is the matrix with zero elements). Each final state marked by X^N represents allowable solutions to the problem (the allowable time-scale). If the time of servicing ships is not limited, the final state is the matrix with positive elements. If there are time limits (e.g. *C* represents the port operating time limit of a week), the number of positive elements of the state matrix is maximal but none of them is not bigger than *C*.

The allowable solution is obtained as a result of generating the trajectory of states:

$$\mathrm{X}^0 \to \mathrm{X}^1 \to \ldots \to \mathrm{X}^{e-1} \to \mathrm{X}^e \to \mathrm{X}^{e+1} \to \ldots \to \mathrm{X}^E$$

where: e – the stage number of generating states, *E* – the number of stages.

The procedure of generating states assumes that the state X^{e-1} as well as the heuristic decision $h(m,n)$ are given but the subsequent state X^e must be determined. This procedure takes the following form:

$$\forall_n \forall_m \; \left\{ \left(x_{n,1}^{e-1} = 0 \right) \wedge \left\lfloor \; \max(T_m^{e-1}; \varphi_n) + \upsilon_{m,n} \le C \; \right\rfloor \right\} \Rightarrow \left\{ X^e = F \; \left\lfloor \; X^{e-1}, h(m,n) \; \right\rfloor \right\} \tag{4}$$

at the same time: φ_n - the availability time of the *n*th ship.

The availability term of the *m*th quay is determined from the state X^{e-1} as follows:

$$T_m^{e-1} = \max x_{i,2}^{e-1} \tag{5}$$

and for the ships serviced at the *m*th quay, i.e.: $$x_{i,1}^{e-1} = m \tag{6}$$

In the new state X^e we receive: $$x_{i,1}^{e} = m \quad \text{and} \quad x_{i,2}^{e} = t_{m,n} \tag{7}$$

Moreover, the moment $t_{m,n}$ is determined as $$t_{m,n} = \max(T_m^{e-1}; \varphi_n) + \upsilon_{m,n} \tag{8}$$

Other elements of the matrixes X^{e-1} and X^e are identical. Analogously, state trajectories for every subsystem of the port system can be generated. There are a lot of logical conditions which are specific for the port, however, they are not mentioned above.

5 Simulation of complexes of operations

The process of servicing ships in the port can be simulated on the basis of the mathematical model by determining allowable time-scales. The exemplary computer simulators, which are based on modelling by means of the multi-stage programming, are shown among others in the works [15, 16]. In order to carry out the simulation of complexes of operations, the chronological order of generating states of trajectories must be introduced and certain heuristics for the simulation process must be chosen.

Introducing the chronological order consists in directing the *n*th ship to the *m*th quay on condition that it is able to service ships as soon as possible in the state X^{e-1}. As a consequence, the chosen quay must meet the following condition:

$$T_m^{e-1} = \max_{1 \le j \le M} T_j^{e-1} \tag{9}$$

Using the chronological order during generating allowable trajectories of states enables us to determine the allowable time-scale for servicing ships according to the chosen heuristic rule. This is the way of simulating the dispatcher's operations who uses his own heuristic rule during operative control. Operative control consists in making decisions after announcing that the quay is available. This method is justified when data are disturbed (e.g. operating the fire engines, ambulances, etc. is treated as operative control, not time-scaling). In case of servicing ships, data can be disturbed (e.g. by atmospheric conditions). Then the time-scale differs from the real process and is to be determined again. Determining the time-scale is necessary because the quay must be prepared for servicing ships (e.g. by moving cranes to the determined place, gathering containers, etc.). Preparing quays for servicing ships increases the effectiveness of servicing procedures.

The time-scale can be determined for a longer period of time (e.g. a week) by means of heuristics. This time-scale is realized up till the moment when deviations from the real process are too big. At this moment a new time-scale has to be determined with the use of the same method. This procedure requires using terms of the step and the horizon of time-scaling. The horizon of time-scaling usually comprises only one working shift. The dispatcher's orders are given for this period of time (not only for one servicing operation). The time-scaling horizon usually comprises one week (from Monday to Friday). After this time, the subsequent weekly time-scale is determined.

The case discussed hereby enables us to use the following heuristics:

1) The choice of the *n*th ship which is the fastest available ship.
   This rule is effective if the availability terms vary. If they were the same, the ship could not be chosen.
2) The choice of the *n*th ship which has to be serviced the fastest possible in the port.
   The choice of the ship which is to be served the fastest minimizes the delay in servicing of ships. If these terms are the same, then the ship cannot be chosen.
3) The choice of the *n*th ship which gives the biggest bonus.
   This rule is implemented when each ship is dealt with before the term generating the profit for the port.
4) The choice of the *n*th ship which demands the biggest fine.
   This rule is implemented when each ship is serviced with a delay generating the loss for the port.
5) The choice of the *n*th ship which is characterized by the longest service time.
   This rule is used when servicing the ship requires the replacement of servicing devices (e.g. cranes) at the wharf bringing a relatively short delay in servicing the ship.

In a general case the system uses the heuristic rule which enables the effective selection of ships. The choice of the heuristic to make a decision about servicing a ship at the quay depends on the state of the port. The effectiveness of heuristics depends on the data so knowledge about the effectiveness of heuristics can be gathered by means of the simulation method. Time-scales can be determined for the same data with the use of different heuristics. Effectiveness coefficients of heuristics can be determined by repeating simulation experiments for random data. These coefficients are probabilities of heuristics which will give the best time-scale. The effectiveness weight of heuristics depends on the state of the port. Its real state is not important but its characteristics (the class of states) are vital. The number of real states is immense and this is the reason for analyzing the characteristic state. Heuristics are correlated with characteristic states. Recognizing states by means of artificial intelligence methods is a specific feature of logistic control systems.

6 The control system

The logistic system takes the form of the port system consisting of the basic port and the additional port. Ships come to the roadstead continuously and are serviced (unloaded, loaded and transported) in the ports. The ship which was serviced is directed to the roadstead and goes to the next destination. After coming back to the roadstead, the economic effect is calculated (bonus or fine). The control consists in maximization of economic effects. The problem consists in determining optimal weekly time-scales and the economic effect is calculated in a longer period of time (e.g. yearly). Unfortunately, yearly time-scales cannot be determined because ships which will arrive at the port are not known. The specification of the control system of servicing ships in the port distinguishes modules for ships with different routes:

1) The ship is discharged only in the additional port.
2) The ship is loaded only in the additional port.
3) The ship is discharged and subsequently loaded only in the additional port.
4) The ship is not serviced in the additional port but only in the basic one.
5) The ship is discharged only in the additional and basic ports.
6) The ship is loaded only in the additional and basic ports.
7) The ship is discharged in the additional port and subsequently loaded in the basic port.
8) The ship is discharged in the additional and basic ports and loaded in the basic port.
9) The ship is discharged in the additional and basic ports and loaded in the additional port.
10) The ship is discharged in the additional and basic ports and loaded in the basic and additional ports.

Ships going between the basic and additional ports in order to transport containers inside the system are neglected.

The control system consists of the following states of modules: discharging wharfs of the additional port (X), loading berths of the additional port (Y), discharging wharfs of the basic port (V) and loading berths of the basic port (W). Modelling and simulation of servicing ships in the port system integrates different states depending on the route of the ship, which is specified hereby. Controlling servicing ships consists in making decisions about choosing a ship for servicing at the earliest available quay. This decision is made on the basis of the decision table [7]. There are *K* columns corresponding with the heuristics and *L* rows representing characteristic states. Recognizing the real state consists in adjusting it to a certain characteristic state. The heuristic is determined for the characteristic state which helps to make a decision about servicing the ship. The heuristic is chosen by means of the mixed strategy at random (with defined probabilities) from the set of the given heuristics. Each subsequent decision is made analogously.

7 Final remarks

Modelling and simulation of complexes of operations in logistic systems is the problem which requires the logistic system specification, the control system project, modelling the process of servicing objects, creating computer simulators, carrying out simulation experiments in order to evaluate heuristics and preparing the decision table for the control system. The article presents these stages on the basis of the port system servicing container ships. The goal of modelling and simulation of complexes of operations in the logistic system is determining the time-scale of servicing ships. These time-scales specify discharging and loading of ships with the use of forecast transport operation times. The control of transport of ships is realized operatively i.e. the tug is required only after completing discharging or loading the ship. As the time-scale is given, it can be used by the control system. Advanced information about transport operations of ships can increase the usage effectiveness of tugs in the port system.

Acknowledgements

The work was supported by the grant MEB051031 (VŠE Praha – WSIZ Bielsko-Biała) financed by MŠMT ČR.

References

[1] Bucki, R., and Marecki, F.: *Modelling and Simulation*. Network Integrators Associates, Parkland, Florida, 2005.

[2] Bucki, R., and Marecki, F.: *Digital Simulation of Discrete Processes*. Network Integrators Associates, Parkland, Florida, 2006.

[3] Frąckiewicz, Z., and Marecki, J.: *Logistical Ships Scheduling Model in the Parallel Port*. Zeszyty Naukowe Wyższej Szkoły Morskiej, Szczecin, 2003, nr 70, 73-84.

[4] Frąckiewicz, Z., and Marecki, J.: Logistical Ships Scheduling in Serial Structure Port. *International Conference on "Marine Traffic Engineering"*, Świnoujście, Szczecin Meritime University, 2003, 94-104.

[5] Frąckiewicz, Z., and Marecki J.: Zastosowanie metod sztucznej inteligencji do sterowania ruchem statków w porcie. *Rozdział w monografii „Efektywność zastosowań systemów informatycznych" (red. J. K. Grabara)*, Wydawnictwa naukowo-Techniczne, Warszawa, 2004, 453-466.

[6] Marecki, J., Łużny, D., and Marecki F.: *Logistyka informatyczna kompleksów operacji*. WSIZ, Bielsko-Biała, 2010.

[7] Marecki, F.: Expert Control of Production and Transport Systems. *Information Technologies and Systems*, Vol. 10, No. 1, 2007, 191-207.

[8] Marecki, J.: *Planning with Continuous Resources in Single/Multiagent Systems*. Network Integrators Associates, Parkland, Florida, 2006.

[9] Marecki, J., and Marecki F.: *Logistyka informatyczna obsługi statków w portach*. WSIZ, Bielsko-Biała, 2007.

[10] Marecki, J.: *Metody sztucznej inteligencji*. Monografia. Wydawnictwo Pracowni Komputerowej J. Skalmierski, Gliwice, 2001, ss.115.

[11] Marecki, J.: *Semantic Networks and Intelligence Agents*. Network Integrators Associates, Parkland, Florida, 2003.

[12] Marecki, J.: *Advances in Artificial Intelligence Planning Techniques*. Network Integrators Associates, Parkland, Florida, 2009.

[13] Marecki, J.: *Allocation of Continuous Resources in Agent Systems*. VDM Verlag, Saarbrucken, 2009.

[14] Marecki, J.: *Cooperative and Adversarial Deployed Multiagent Systems*. Network Integrators Associates, Parkland, Florida, 2010.

[15] Olbrzymek, M., and Marecki, J.: Symulator komputerowy obsługi statków w porcie z pojedynczym nabrzeżem. *Workshop on „Methods and Applications of Artificial Intelligence"*, WSIZ, Bielsko-Biała, 2009, pp. 93-106.

[16] Olbrzymek, M., and Marecki, J.: Symulator obsługi statków w porcie o strukturze równoległej. *Project: Information Logistics of Transport, Production and Storing Systems, Id. No. MEB051024*. Faculty of Applied Informatics, Tomas Bata University in Zlín, 10-12 November 2010.

[17] Olbrzymek, M. : *Symulatory komputerowe obsługi statków w portach*. WSIZ, Bielsko-Biała, 2011.

[18] Sawik, T.: *Badania operacyjne dla inżynierów zarządzania*. Akademia Górniczo-Hutnicza, Kraków, 1998.

[19] Wagner, H. M.: *Badania operacyjne*. Państwowe Wydawnictwo Naukowe, Warszawa, 1980.

Changes in the behavior of economies in a DSGE model in the course of time

Jan Čapek[1]

Abstract. The contribution uses a New Keynesian Dynamic Stochastic General Equilibrium model as a tool for analysis of model behavior. The used macroeconomic model is a Small Open Economy model derived from microeconomic foundations and presumes four types of agents.

Aim of the contribution is to identify changes in behavior of modeled economies during dramatic economic events and explain observed changes in the mechanics of the model.

The method proceeds from Bayesian estimates of parameters and impulse response functions. In order to analyze changes in parameter estimates and impulse response functions in time, a number of estimation procedures are carried out. Each estimation involves the same baseline model with the same prior setting and varies only in the time-frame of the data set. Resulting changes in the estimates of parameters and different shapes of impulse response functions may point to changes in behavior of the system caused only by using different data span. Such model results may be explained in the context of changed behavior of economic agents.

Keywords: DSGE model, economic recession, change in model behavior, Bayesian estimates, impulse response function

JEL classification: C32, C52, E32, E43, E52, F41, F43
AMS classification: 91B51, 91B64

This paper presents some results of a continuous research on changes in behavior of economies represented by a DSGE model. Although some other time-frames may be also of interest, alotted space allows for single example – an actual one being contemporary recession.

1 Model – Introduction

This section introduces New Keynesian (NK) Dynamic Stochastic General Equilibrium (DSGE) model that is used for analysis in this paper. The model is derived from microeconomic behavior of particular economic agents. These include domestic and foreign households, domestic and foreign producers, domestic importers and domestic and foreign monetary authority.

The model is in small open economy (SOE) setting, so that it presumes two countries – small open economy that is influenced by a big closed economy. The small open economy is home economy, big large economy is the foreign economy.[1]

Most of the model assumptions are adopted from Lubik and Schorfheide [5]. Due to the lack of space, the model is not derived in this paper – similar models with details of derivation can be found in textbook Galí [2]. The model used in this paper differs from frequently cited basic models like in Liu [4], Justiniano and Preston [3] or already mentioned Galí's textbook [2] by households' utility function, incorporation of the world-wide technology shock and related transformations of marginal utility of income.[2]

When applicable, variables with H subscript (e.g. $X_{H,t}$) denote goods produced and activities associated with Home economy and variables with F subscript (e.g. $X_{F,t}$) denote goods produced and activities associated with

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno, e-mail: capek@econ.muni.cz

[1]Strictly speaking, since the big foreign economy is modeled structurally, uncovered interest parity makes it possible for leaks from small economy to big economy. However, model results indicate that it does not occur.

[2]The difference mentioned last can be also found in Adolfson et al. [1, pp. 15–16].

Foreign economy. The location of economic activities is denoted with a star superscript (e.g. X_t^*) for the foreign economy and is left without extra notation for home economy.

2 Model – Log-linearized form

This section summarizes log-linearized model form; equations themselves are not written in full due to lack of space.

Starting with households, the system contains equations for evolution of marginal utility of income, the law of motion of habit stock, Euler equation, and the definition of inflation from domestic and imported inflation:

Behavior of producers yield the New-Keynesian Phillips Curve. Importers' optimization is analogous to that of producers and for the log-linearized system is utilized importers' Phillips curve.

Incorporation of open economy setting contributes to the system with a number of equations, namely definition of the depreciation rate of nominal exchange rate, differenced definition of terms of trade and combined definition of real exchange rate and LOP gap. The model also includes some equilibria equations that must hold. These include equation regarding international risk-sharing, uncovered interest parity (UIP) condition and log-linearized market clearing equation.

Foreign economy is modeled structurally so that there exist foreign households and producers that also show optimizing behavior. Optimizing behavior of foreign agents contribute to the log-linearized system with equations analogous to domestic case.

The model is closed by specifying monetary policy. Towards this end, standard Taylor-type rule is used. This formulation of monetary policy assumes that central banks respond to deviations of inflation from steady state, growth rate of output from steady state growth rate γ and possibly to deviations of nominal exchange rate depreciation from steady state.

The model is supplemented with AR(1) processes describing evolution of government expenditures (acting as a demand or market clearing shock) g_t, country-specific technology shock to production function (acting as a supply shock) and the evolution of z_t, which is growth rate of world-wide non-stationary technology shock.

3 Data

The data set used for estimation is depicted in Figure 1. There are four observed variables for domestic Czech economy in the first row and three observed variables for foreign EuroArea12 economy in the second row. Fourth panel is depreciation of nominal exchange rate – Czech Koruna.

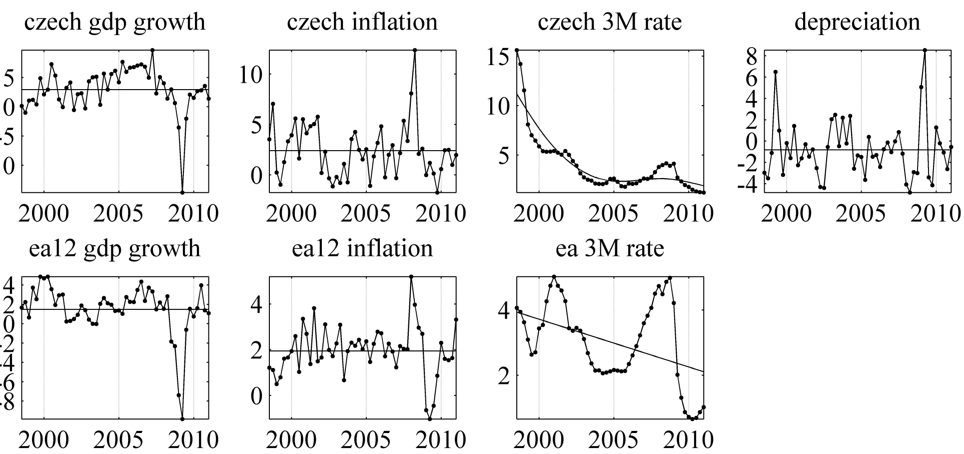


Figure 1: Data set

The data are quarterly and time is denoted at the end of period so that e.g. 2001.25 is first quarter of 2001 and 2008 is fourth quarter of 2007.

All variables are per cent, inflations and output growths are annualized, interest rates are per annum and depreciation rate is not annualized. Dotted line denotes original data and solid line denotes trend that is deducted prior to estimation. All variable except for interest rates use constant trend (mean) as a trend. Although this is standard procedure, it is not appropriate for domestic and foreign interest rates since these series show noticeable non-constant trends. Linear trend is sufficient for foreign interest rate but domestic interest rates exhibits even more curvature. After some experiments with polynomial trends and various filters, Hodrick-Prescott filter seems to be the best option.[3]

4 Parameter estimates with different time-frames

Since the research question at hand addresses changes in economy's behavior during recession, it is imperative to estimate the model on different time frames, namely on a time frame without recession and on a time frame that includes the recession.[4] Then, it is possible to compare the estimates and draw results as to the nature of the recession.

This intuitive idea of two model estimates and their comparison is clouded with two problems. One of the problems is that two estimates is not enough, because the choice of time frames would necessarily be arbitrary and moreover, the information acquired from comparison of just two results would be limited. Another problem, which is illustrated by Figure 2 stems from the choice of first observation, which may affect the estimate of the parameter to the extent that the choice of first observation may be far more important than the choice of the last observation. In another words, the choice of few oldest data could be in the estimate more important than inclusion or exclusion of crisis data.

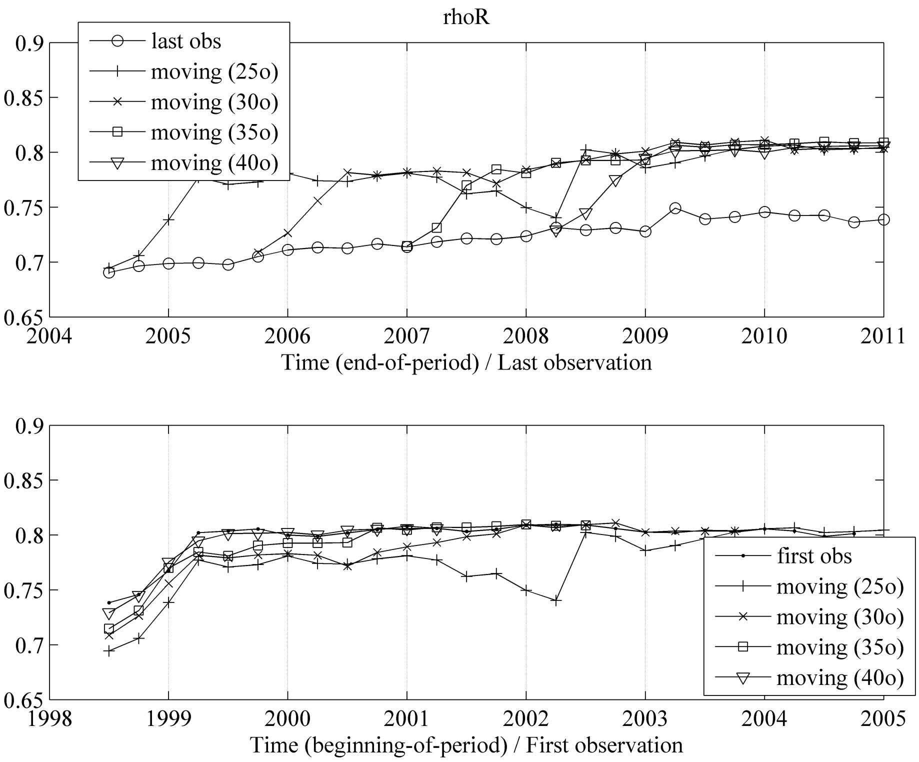


Figure 2: Estimate of backward-looking parameter in domestic Taylor rule ρ_R with different time-frames

These problems call for a more elaborate tool, which would minimize or eliminate the issues. In order to investigate sensitivity of the estimate to (a) the choice of the last observation, (b) the choice of the first observation and (c) the length of the time frame, a complex tool that addresses all of these variants is used.

[3]Sensitivity analysis of the determination of trends was conducted and model results does not change significantly when different method of detrending is used.

[4]Each time frame is defined by the choice of first observation and the last observation, the frequency always stays the same – all estimates are on quarterly data.

Due to lack of space, the methodology is not described in detail. On the other hand, the working of the procedure is explained on an example, which follows.

To illustrate the merits of the methodology, the example is focused on rather less intuitive case, which was very important for the remainder of the analyses in this paper. The chosen example is a case where the important information is in the beginning of the data series.

Figure 2 contains estimates of backward-looking parameter in domestic Taylor rule ρ_R and has the following structure – upper panel plots estimates where last observation varies (“last obs”) and four different moving-window estimates with three different lengths of time frame. The length of the time frame is in brackets (“25o”, “30o”, “35o” and “40o”.). The horizontal axis is time at the end of period. The last values of estimates are for 2011, which is the fourth quarter of 2010. The circle, plus, cross, square and triangle at time 2011 have the following meaning: “last obs” estimate at 2011 is an estimate of the parameter on time frame from the beginning (which is 1998.5) to the time denoted by the horizontal axis (2011). “moving (25o)” estimate uses 2011 as the last observation as well, but it uses just 25 observations, the beginning of the time frame is in this case 2005. “moving (30o)” begins at time 2003.75, “moving (35o)” begins at 2002.5 and “moving (40o)” begins at 2001.25 – all of the estimate have the end of the time frame at 2011.

The lower panel draws estimates with varying first observation (and fixed last observation) and again four different lengths of moving-window estimates. The difference also lies in the horizontal axis which is now the time at the beginning of the time frame. First value of estimates is for 1998.5, which is second quarter of 1998. Following explanation of the meaning of the dot, plus, cross, square and triangle address values at 1998.5. The “first obs” estimate naturally starts at 1998.5 and ends with the latest observation, which is 2011. “moving (25o)”, “moving (30o)”, “moving (35o)” and “moving (40o)” all start in 1998.5 and end in 2004.5, 2005.75, 2007, and 2008.25 respectively.

Now that it is clear what Figure 2 shows, this paragraph proceeds to interpretation of the contents of the Figure. Upper panel displays a gradual rise in “last obs” estimate but moving-window estimates all jump up at some point and end up in higher value of the parameter – “last obs” estimate ends at 2011 by value approximately 0.71 and all moving-window estimates end up with parameter value 0.8. This result indicates, that there might be some important information somewhere in the beginning of the time series. The lower panel confirms the suspicion because all the displayed estimates exhibit a rise of the parameter estimate by 0.1 in the first 4 observations. It means that no matter the length of the time frame, the exclusion of observations from 1998.5 to 1999 lead unambiguously to a rise in the point estimate of the parameter. In this context, the gradual rise in “last obs” estimate is illusory. New added information in the “last obs” estimate just dilute the strong information in the first three observations.

Similar situation as in the case of parameter ρ_R occurred with some other parameters (namely Calvo parameters $\theta_{\cdot}$), which indicates, that data in time frame from 1998.5 to 1999 belong to different historic era, which was not under investigation and the data in question were therefore discarded from the set for further analyses.

Further analyses are therefore focused on a truncated data set that should represent homogeneous structure.

5 Changes in parameter estimates during recession

After discarding of the data that represented different economic era, the remaining interval from first quarter of 1999 to the last quarter of 2010 was used for the analyses. Since there is an assumption of homogeneous structure (supported by previous analysis), there is no need to carry out “first obs” and moving-window estimates. Figure 3 therefore depicts “last obs” estimates.

The two depicted parameter estimates show foreign central bank’s reaction to output growth ψ_2^* (`psi2star`) in the left panel and AR1 persistence parameter in the world-wide technology shock ρ_Z (`rhoZ`) in the right panel. The Figure also depicts two widths of confidence bands. The wider confidence band is 90% band and the narrower band varies and is stated in the title of the panel in question. The width of the narrower confidence band is chosen so that such confidence bands do not overlap for at least two different periods. This can be a measure of how much the point estimate changes realtive to the amount of uncertainty associated with the estimate.

Left panel shows that the reaction of foreign central bank to output growth diminishes from fourth quarter of 2007 and stabilizes after “crisis quarter” 2009.25. This can be explained by looking at the data – foreign output drops, but the interest rates start to go down not before fourth quarter of 2008 and then it drops in first quarter of 2009. This can be interpreted as central bank’s negligence towards output growth. However, this conclusion is not

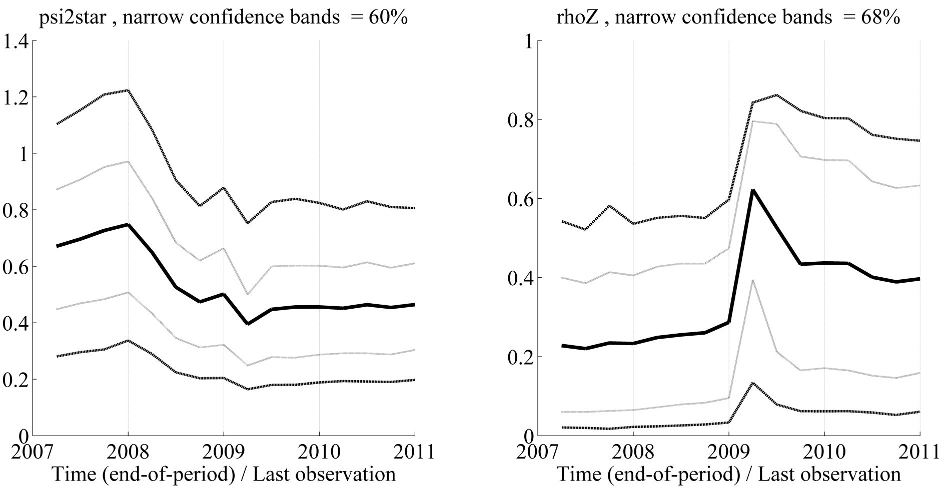


Figure 3: Parameter estimates with different time-frames. Left panel: foreign central bank's reaction to output growth ψ_2^*, right panel: AR1 persistence parameter in the world-wide technology shock ρ_Z, wide confidence bands = 90%

surprising considering the evolution of inflation in first quarters of 2008 – the inflation was quite high and inflation is primary objective of foreign central bank. It therefore had to disregard from output decline and focus on the danger of inflation.

The width of the narrower confidence bands can also be interpreted in a way that 60% confidence bands mean that the confidence bands do not overlap with significance level 0.4. With this significance level, it is possible to say that the reaction of the central bank's changed as a result of recession.

Another example is in the right panel of Figure 3 depicting the evolution of the persistence parameter in the world-wide technology shock. The point estimate of the parameter jumped from pre-crisis value 0.2 to 0.6 in "crisis quarter" 2009.25. The persistence of whatever hits both economies from outside is therefore much more long-lasting. In the mechanics of the model, this increase was used to describe the economic crisis, which was understood by the model workings as negative world-wide technology shock. As the crisis subsided over time, it was not as deep as the model expected it to be and the persistence parameter went back down to 0.4.

The left panel therefore shows a case of changing preferences of one of the economic agents in the model and the right panel shows an important part of model's explaining current economic crisis.

6 Impulse response functions with different time-frames

Figure 4 displays impulse response functions as a result of an innovation with value -0.6 to the system. In this particular case, the unitary innovation is put to an AR1 process z_t describing evolution of growth rate of world-wide technology shock. The value 0.6 was estimated as the standard deviation of the shock and negative value was used to better mimic negative shock causing economic crisis.

Bottom-right axis in panels in Figure 4 is labeled "lag" and it is the lag of the impulse response function. Time 0 corresponds to the time of the innovation, time 1 is one period later and the lag goes up to 15. Each black line in the figure is an ordinary impulse response function that can be found in numerous scientific literature.

Bottom-left axis "time (end-of-period)" addresses different time frames on which the model was estimated. First observation for the estimation was always first quarter of 1999. Last observation varies from first quarter of 2007 to fourth quarter of 2010 and bottom-left axis denotes the time of last observation. More concretely, impulse response functions in the left-hand part of the figure are responses calculated on time frame from 1999.25 to 2007.25. It might be considered as the "oldest" impulse response function. Each other impulse response function more to the right corresponds to impulse response function calculated on time frame longer by one quarter. The second impulse response function from the right therefore corresponds to time frame from 1999.25 to 2007.5. The impulse response function most to the right uses most up-to-date data and is calculated on time frame from 1999.25 to 2011.

The z-axis denoted on the right of the figure is the value of the impulse response functions, which is common

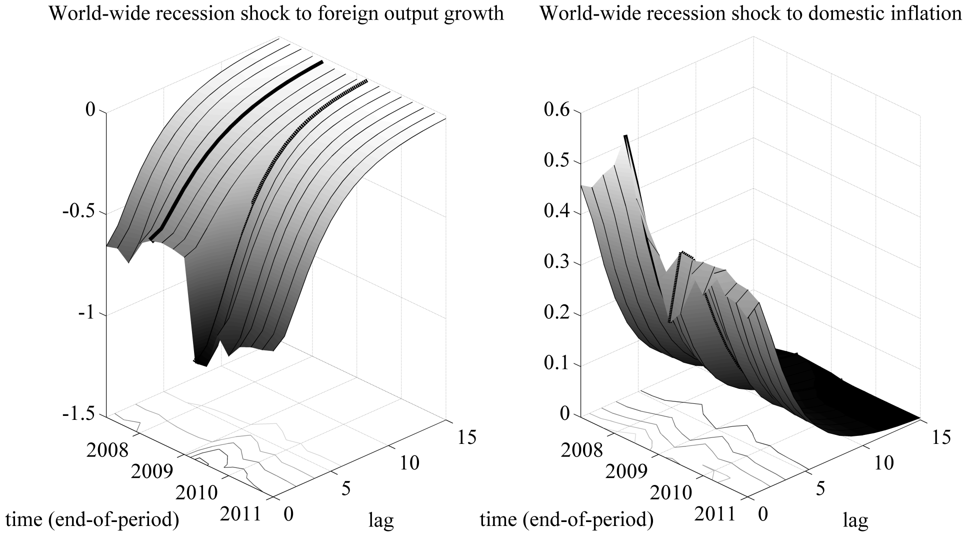


Figure 4: Impulse response functions with different time-frames

in depicting of impulse response functions.

Impulse responses function highlighted in bold are of special interest, since they exhibit biggest difference between each other. The bold impulse response function more to the left (the "older" one) corresponds to pre-crisis time-frame from 1999.25 to 2008.25. The second bold impulse response function is calculated on time frame ending in first quarter of 2009, that is, the period where is the oncoming recession most apparent in the data.

Changing structure of the economy described by model behavior exhibits approximately twice as low drop in foreign output in crisis than it was before crisis as a result of a negative world-wide technology shock. These two highlighted impulse response functions are also shown in left panel of Figure 5, together with 90% confidence bands. Dashed impulse response function corresponds to data set containing crisis period, the solid line depicts pre-crisis impulse response function.

Confidence bands are depicted by shading and dotted borders. Darker shading in the background corresponds to "crisis" dashed impulse response function, whereas lighter shading in the front belong to "pre-crisis" impulse response function. This depiction shows that point estimate of the IRF in crisis is outside pre-crisis confidence bands. On the other hand, upper part of the confidence bands is almost the same and major parts of the confidence intervals therefore overlap.

The right panels of Figures 4 and 5 show the same type of results that were discussed in previous paragraphs. The investigated case was again a world-wide technology shock, but this time the reaction under scrutiny was domestic inflation.

Right panel of Figure 4 displays that the reaction of inflation is much lower in crisis and moreover, the shape of the point estimate of the impulse response function is different. Again, in order to investigate the case in greater detail, right panel of Figure 5 can be used. It is now visible, that the reaction of inflation in crisis is almost nonexistent at the time when the shock occurs. In 3–4 quarters, the inflation rises by 0.3 percentage points. The reaction of inflation before crisis is quite different – inflation jumps by 0.6 percentage points and then it gradually diminishes. However, the confidence bands show a great deal of uncertainty about the point estimates of the impulse response functions. However, the confidence bands show that positive reaction of inflation is much more probable in period before crisis than in economic crisis, when the lower band of confidence interval drops in negative values.

7 Conclusions

The paper presents innovative approaches towards investigation of changes that may occur in the behavior of modeled economies during economic recession and also offers some economic examples.

Section 4 briefly introduced a methodology that could help unveiling different economic eras. The working

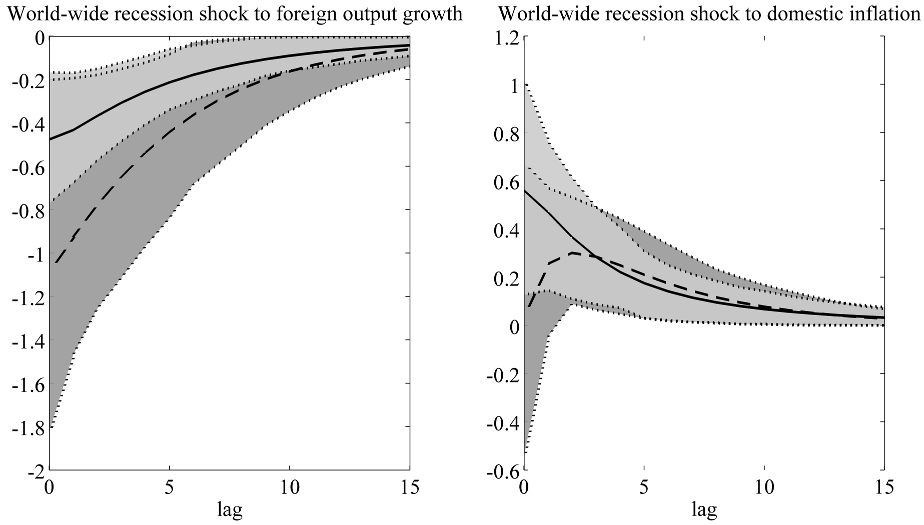


Figure 5: Impulse response functions with different time-frames with 90% confidence bands. Dashed line with background darker shaded confidence bands correspond to "crisis" time frame from 1999.25 to 2009.25. Solid line with lighter shaded confidence bands correspond to "pre-crisis" time frame from 1999.25 to 2008.25.

of the procedure is presented on a case with a resulting change to the data set, which was truncated by first three observations. Section 5 presents two exemplary results of changes in estimated parameter values together with the uncertainty of the point estimates. It turns out that e.g. foreign central bank lowers its preference about output growth while it engages in monetary policy. Section 6 uses parameter sets estimated on different time-frames a computes impulse response functions with different time-frames. 3-D graph of the impulse response functions unveil changes in the reactions of variables in question. Attention is also paid to uncertainty associated with the impulse response functions. The results show a significantly different reaction of foreign output growth to a world-wide technology shock. Another case shows a different shape of impulse response function of domestic inflation. Both of these cases indicate changes in the structure of the economies represented by the model.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

References

[1] Adolfson, M., Laséen, S., Lindé, J., and Villani, M. Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through. Working Paper Series 179, Sveriges Riksbank (Central Bank of Sweden), Mar. 2005.

[2] Galí, J. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press, 2008. ISBN: 978-1-4008-2934-7.

[3] Justiniano, A., and Preston, B. Small Open Economy DSGE Models: Specification, Estimation and Model Fit. Tech. rep., September 15, 2004.

[4] Liu, P. A Small New Keynesian Model of the New Zealand economy. Reserve Bank of New Zealand Discussion Paper Series DP2006/03, Reserve Bank of New Zealand, May 2006.

[5] Lubik, T., and Schorfheide, F. A Bayesian Look at the New Open Economy Macroeconomics. In *NBER Macroeconomics Annual 2005, Volume 20*, NBER Chapters. National Bureau of Economic Research, Inc, May 2006, pp. 313–382.

Economically Optimal Subnetwork for Passenger Transport

Anna Černá[1], Jan Černý[2]

Abstract. The paper deals with the following basic problem: Given the connected graph $GP = (V, EP, d, q)$ where $d(e)$ is the length of $e \in EP$ and $q(v)$ is the number of passengers having the vertex $v \in V$ as the origin or the destination of their trip. Given the connected graph $GT = (VT, ET, dT, cT)$ where VT is a subset of V, $dT(e)$ is the transit time through the edge $e \in ET$, $cT(e)$ is the cost of $e \in ET$. The functions d, dT and cT are extended to all paths on GP and GT respectively. Given positive numbers λA and λT. The problem is to find a connected subgraph $GTS = (VTS, ETS, dT, cT)$ of GT such that the average walking distance of passengers to the closest vertex from VTS does not exceed λA, the relative extension of mean travel time on GTS, with respect to the one on GT does not exceed λT and the total costs, i.e. the sum of all $cT(e)$ for $e \in ETS$ is minimal.
Exact methods and heuristics for the problem solution are presented. Finally, several particular cases and modifications of the problem are discussed.

Keywords: subnetwork, optimal, passenger transport, accessibility, cost.

JEL Classification: C65, L92, O18, R42
AMS Classification: 05C12, 90B06, 90B10

1 Introduction

The paper deals with problems belonging to the theory of network economy. This modern part of science can be divided into two main directions: The first one focuses networks of mutually connected and interacting enterprises, as described e.g. in the interesting books [7] and [4]. The second direction looks for economically optimal transport and telecommunication networks as described e.g. in the papers [9] and [11].
Our problem belongs to the family of network design problems, especially the ones seeking a subnetwork of the given network. We can mention several papers dealing with related topics, concerning a graph $G = (V, E, c)$ where $c(e) \geq 0$ is a general cost (e.g. toll or passing time or length) of the edge $e \in E$. E.g. the paper [5] aims to the maximum planar subgraph of G or [8], [3] look for a λ-edge connected spanning subgraph minimizing the costs of edges for the given integer $\lambda > 0$ and the same does [6] for $c(e) \equiv 1$ and [2] for $\lambda = 2$. Similar problems minimizing the total costs of edges are studies e.g. in [10], [1] or [13]. On the other hand, there are other papers look for maximum cost subgraphs, e.g. [12] and again [1].

The problems we describe in the sequel seek the minimum costs subgraph similarly as the abovementioned ones. The difference is in constraints. We are not oriented to λ-edge or k-vertex connectivity. Our basic problem requires that the accessibility of the subgraph vertices is in some limit. The modified problem requires that the distance between any pair of important vertices must not be more than q times longer on the subgraph as on the original graph.

1.1 The Basic Problem (BP)

Let $GP = (V,EP,d,q)$ be a connected non-oriented graph, let $d(e)$ be the generalized cost (e.g. the length) of $e \in EP$ and $q(v)$ be the demand (e.g. the number of passengers) of the vertex $v \in V$. Let $GT = (VT,ET,dT,cT)$ be a connected non-oriented graph, let $VT \subset V$, let $dT(e)$, $cT(e)$ be the generalized costs (e.g. the transit time and building costs) of $e \in EP$. Let the functions d, dT and cT be extended to all paths on GP and GT respectively. Let λA be positive integer and $\lambda T \geq 1$.

The **problem** is to find a subgraph $GTS = (VTS, ETS, dTS, cT)$ of GT such that

$$dTS(e) = dT(e) \text{ for all } e \in ETS \text{ and } dTS(v, w) \text{ is the distance of } v, w \in VTS \times VTS \text{ on } GTS \quad (1)$$

$$dTS(v, w) \leq \lambda T.\ dT(v, w) \text{ for all } v, w \in VTS \quad (2)$$

[1] University of Economics in Prague, Faculty of Management, Jarošovská Street 1117/II, 37701 Jindřichův Hradec, cerna@fm.vse.cz.
[2] University of Economics in Prague, Faculty of Management, Jarošovská Street 1117/II, 37701 Jindřichův Hradec, cerny@fm.vse.cz..

$$\mu(VTS) = \frac{1}{q}\sum_{v\in V} q(v)d(v,VTS) \le \lambda A, \tag{3}$$

$$cT(ETS) = \sum_{e\in ETS} cT(e) \to \min \tag{4}$$

Remark 1. The condition (2) requires that the relative extension of any distance in the graph *GTS* with respect to the one in *GT* does not exceed λT. Usually $\lambda T \in \langle 1; 2\rangle$.

Remark 2. One of practical urban applications is the following: *GP* is the network of footways, *GT* is the network of possible busways, *GTS* is the network which is chosen for bus transport, *VTS* is the set of all bus stops in the chosen network and the condition (3) assures that the average walking distance $\mu(VTS)$ does not exceed the value λA.

Remark 3. The condition (4) requires that the cheapest acceptable network is chosen.

1.2 Exact Solution of BP

Since the structure of BP is rather complex one can expect that only small instances are *exactly* solvable by PC in reasonable time. One possible way is **depth-first search technique**.

Variant 1. Start with $ETS = \emptyset$ and subsequently add the edges from *ET* to *ETS* until both (2) and (3) hold. Then compare the reached cost $cT(ETS)$ with the record and try the next branch of the solution tree.

Variant 2. Start with $ETS = ET$ and subsequently omit the edges from *ETS* until (2) or (3) is violated for the first time. Then turn to the predecessor, compare the reached cost $cT(ETS)$ with the record and try the next branch of the solution tree.

The authors believe that a linear programming solution is possible as well but the formulation of BP would be extremely complicated. Therefore the LP models are introduced only for particular problems.

2 Linear Subgraph Problem (LSP)

The first particular case of BP is reached if the subgraph *GTS* is required linear, i.e. the edges from *ETS* have to form a path in the graph *GT*. On the other hand the condition (2) is omitted ($\Leftrightarrow \lambda T =$ "∞") and therefore the function *dT* is inutile. It is obvious that the depth-first search technique is usable like before. The only change is that there is the additional condition of linearity.

2.1 Solution of LSP by Linear Programming (LP)

Before we formulate LSP by LP, we will precise denotations.

Suppose that $m \le n$ are positive integers, $V = \{1, 2, \dots, n\}$ is the set of all vertices with demand values $q_1, \dots, q_n$ and $VT = \{1, 2, \dots, m\} \subset V$. For each pair $i \in V$, $j \in V$, the given number d_{ij} represents the walking distances between the nodes *i* and *j*. Similarly, δ_{ij} represents the cost *cT* on the set $VT \times VT$. Of course, $d_{ii} = 0$ and $\delta_{jj} = 0$ for $i \in V$ and $j \in VT$. Moreover, $\lambda = \lambda A > 0$ represents the upper bound of the average walking distance. Finally, the set $F' \subset VT \times VT$ represents the arcs on the given digraph $DT = (VT, F', \delta)$ derived from *GT* in such a way that oriented arcs $(u, v) \in F'$ and $(v, u) \in F'$ if the edge $(u, v) \in ET$. The cost δ remains unchanged.

LP Problem – Formulation

Let $m \le n$ be positive integers. Let $\lambda > 0$. Let $d_{ij} \ge 0$, $d_{ii} = 0$ for $i, j \in V = \{1, 2, \dots, n\}$ and let $\delta_{ij} \ge 0$, $\delta_{ii} = 0$ for $i, j \in VT = \{1, 2, \dots, m\}$. Let $F' \subset VT \times VT$. Let

$$D = \sum_{i,j\in\{1,\dots,n\}} d_{ij} + \sum_{i,j\in\{1,\dots,m\}} \delta_{ij} \qquad Q = \sum_{i\in\{1,\dots,n\}} q_i \tag{5}$$

The problem is to find

$$u_j \in \{0, 1\}, j \in \{1, 2, \dots, m\},$$

$$v_{ij} \in \{0, 1\}, i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\},$$

$$x_{jk} \in \{0, 1\}, j \in \{1, 2, \ldots, m\}, k \in \{1, 2, \ldots, m\}, \ (j, k) \in F'$$

$$y_j \text{ non-negative integers,}$$

minimizing the objective function

$$z = \sum_{j,k\in\{1,\ldots,m\};(j,k)\in F'} \delta_{jk} x_{jk} \tag{6}$$

and fulfilling the following constraints:

$$\frac{1}{Q} \sum_{i\in\{1,\ldots,n\}} \sum_{j\in\{1,\ldots,m\}} q_i d_{ij} v_{ij} \le \lambda, \tag{7}$$

$$v_{ij} \le u_j \quad \text{for} \quad i \in \{1, 2, \ldots, n\}, j \in \{1, 2, \ldots, m\}, \tag{8}$$

$$\sum_{j\in\{1,\ldots,m\}} v_{ij} = 1 \quad \text{for} \quad i \in \{1, 2, \ldots, n\}, \tag{9}$$

$$\sum_{j\in\{1,\ldots,m\}} d_{ij} v_{ij} \le d_{ik} u_k + D(1 - u_k) \quad \text{for} \ i \in \{1, 2, \ldots, n\}, k \in \{1, 2, \ldots, m\}, \tag{10}$$

$$x_{jk} \le \frac{u_j + u_k}{2} \quad \text{for} \quad j \in \{1, 2, \ldots, n\}, k \in \{1, 2, \ldots, m\}, \ (j, k) \in F', \tag{11}$$

$$\sum_{k\in\{1,\ldots,m\};(k,j)\in F'} x_{jk} \ge \frac{1}{m}\left(\sum_{k\in\{1,\ldots,m\}} u_k - y_j \right) + D(u_j - 1) \quad \text{for} \ j \in \{1, 2, \ldots, m\}, \tag{12}$$

$$\sum_{j\in\{1,\ldots,m\};(k,j)\in F'} x_{jk} \ge \frac{1}{m}(y_k - 1) \quad \text{for} \ k \in \{1, 2, \ldots, m\}, \tag{13}$$

$$y_k \ge y_j + 1 + D(x_{jk} - 1) \quad \text{for} \quad j, k \in \{1, 2, \ldots, m\}, \ (j, k) \in F', \tag{14}$$

$$y_k \le y_j + 1 + D(1 - x_{jk}) \quad \text{for} \quad j, k \in \{1, 2, \ldots, m\}, \ (j, k) \in F', \tag{15}$$

$$y_k \le n\, u_k \quad \text{for} \quad k \in \{1, 2, \ldots, m\}, \tag{16}$$

LP Problem – Interpretation

In (5), the value D means a “big number”, i.e. a substitute of infinity. Q is the total number of passengers.

The variable $u_j = 1$ means that the j^{th} vertex is chosen to the route and $u_j = 0$ i the opposite, $v_{ij} = 1$ means that the j^{th} vertex is the closest vertex of the chosen path to the i^{th} vertex, $x_{jk} = 1 \Leftrightarrow k^{\text{th}}$ vertex is the (immediate) successor of the j^{th} vertex on the path, y_j is the order of the j^{th} vertex on the path, z in (6) represents the total cost of the resulting path.

The constraint (7) expresses the limit of accessibility, (8) finds the closest vertex among the ones of the path, (9) and (10) choose the nearest vertex of the path to the i^{th} vertex. Further, the constraint (11) assures that the k^{th} vertex succeeds the j^{th} vertex only if both vertices are chosen, (12) assures that the path does not continue after the ending vertex and (13) assures the similar fact before the first vertex, (14) and (15) assure that, on the path, the immediate successor of the vertex with order y_j is assigned the order $y_j + 1$. Finally, (16) assures that the vertices not belonging to the path have no orders.

2.2 Solution of LSP by a Heuristics

1st step (initial). Put $VTS = \{v_o\}$ where v_o is the median of GT.

2nd step (recursive). If VTS fulfils (3) then VTS is the solution. Otherwise, if $VT - VTS \neq \emptyset$ find such a $w \in VT - VTS$ that the cost $\mu(VTS \cup \{w\})$ is minimal and turn to the beginning of 2nd step again. If $VT - VTS = \emptyset$ then the heuristics fails in finding the solution.

3 Bus Subgraph Problem (BSP)

The second particular case of BP is reached if $GP = GT$, $d \equiv dT$ and c is a multiple of d. The condition (3) is omitted ($\Leftrightarrow \lambda A =$ "∞"). Obviously, the depth-first search technique is usable like before.

3.1 Solution of BSP by Linear Programming

Before the formulation of LP problem we declare a small modification. Since it may happen that several stops in V are not important for passengers, it would be inutile to ask for the property (2) for them. Therefore, we shall limit the condition (2) only for "important pairs of vertices" $W \subset V \times V$. Of course, if necessary, we can put $W = V \times V$ and return to non-modified BSP. In the sequel, we shall make some simplifications of denotations.

LP Problem – Formulation of BSP

Let $n \geq 2$ be a positive integer, let $V = \{1, 2, \ldots, n\}$ and let $G = (V, E, d)$ be a connected non-oriented finite graph with the length $d(e)$ for each $e \in E$. Further, let $d(v, w)$ denotes the distance between the vertices v and w on the graph G. Let $\emptyset \neq W \subset V \times V$ and finally let $\lambda \in \langle 1, \infty)$.

The problem is to find $x_{ij} \in \{0, 1\}$ for $(i, j) \in E$, $i < j$, $y_{vwij} \in \{0, 1\}$ for $(v, w) \in W$, $(i, j) \in E$, minimizing the objective function

$$\sum_{(i,j)\in E} d(i,j)x_{ij} \to \min \tag{17}$$

and fulfilling the following constraints

$$\sum_{(i,j)\in E} y_{vwij} \leq \lambda d(v,w) \quad \text{for } (v, w) \in W, \tag{18}$$

$$\sum_{(i,v)\in E} y_{vwiv} + 1 = \sum_{(v,k)\in E} y_{vwvk} \quad \text{for } (v, w) \in W, \tag{19}$$

$$\sum_{(i,w)\in E} y_{vwiw} = \sum_{(w,k)\in E} y_{vwwk} + 1 \quad \text{for } (v, w) \in W, \tag{20}$$

$$\sum_{(i,j)\in E} y_{vwij} = \sum_{(j,k)\in E} y_{vwjk} \quad \text{for } (v, w) \in W, j \neq v, j \neq w, \tag{21}$$

$$y_{vwij} \leq x_{ij} \text{ for } (v, w) \in W, (i, j) \in E, i < j \text{ and } y_{vwij} \leq x_{ji} \text{ for } (v, w) \in W, (i, j) \in E, j < i \tag{22}$$

LP Problem – Interpretation

The value $x_{ij} = 1 \Leftrightarrow$ the edge (i, j) remains in the reduced network, $x_{ij} = 0 \Leftrightarrow$ it is omitted. The value $y_{vwij} = 1 \Leftrightarrow$ the shortest path from v to w passes through (i, j) in the reduced network, $y_{vwij} = 0$ means the opposite.

The formula (17) expresses the total length of the subgraph, (18) assures that the relative extension of the distance between any $(v, w) \in W$ does not exceed the value λ, (19), (20) and (21) assure that the shortest path from v to w starts in v terminates in w and passes other vertices without branching or merging and finally, (22) ensures that the shortest path from v to w passes only through the subgraph.

3.2 Solution of BSP by Heuristics

The heuristics we are to use is based on the following lemma.

Lemma. Let $G = (V, E, d)$ be a connected non-oriented finite graph without loops with the length $d(e)$ for each $e \in E$. Let $d(p) = d(v_0, e_1, v_1, e_2, v_2, ..., v_{n-1}, e_n v_n) = d(e_1) + ... + d(e_n)$ be the length of the path $p = v_0, e_1, v_1,$

$e_2, v_2, ..., v_{n-1}, e_n\ v_n$ and let $d(v, w)$ denotes the distance between the vertices v and w on the graph G. Let $E' \subset E$, let $\lambda > 1$ and let $d'(e) = d(e)/\lambda$ for each $e \in E'$ and $d'(e) = d(e)$ for $e \in E - E'$. Let $p= v_0, e_1, v_1, e_2, v_2, ..., v_{n-1}, e_n, v_n$ be the shortest path connecting the vertices v, w in the graph $G_o = (V, E, d')$.

Then $d(p) \le \lambda.d(v, w)$.

Proof. It is obvious that

$$\lambda.d'(p) = \lambda(d'(e_1) + d'(e_2)+ d'(e_n)) \ge d(e_1) + d(e_2)+ d(e_n) = d(p)$$

Let r be the shortest path from v to w in the graph G. However, in G_o it need not be true and consequently

$$d'(p) \le d'(r) \Rightarrow\ d(p) \le \lambda.d'(p) \le \lambda.d'(r) \le \lambda.d(r) = \lambda.d(v, w)$$

since $d'(e) \le. d(e)$ for all $e \in E$. The proof is complete.

The heuristics

1st step (initial). Put $j = 0$, $E_j = \varnothing$, $G_j = G$.

2nd step (recursive). Find the shortest path $p(v, w)$ on G_j for each pair $(v, w) \in W \times W$. For each $e \in E - E_j$ put

$$m_j(e) = \text{card}\{(v, w) \in W \times W : e \in p(v,w)\}$$

Find such $e^* \in E - E_j$ that $m_j(e^*) = \max\{m_j(e): e \in E - E_j\}$. If $m_j(e^*) = 0$ then put $E' = E_j$ and **stop**. Otherwise put

$$E_{j+1} = E_j \cup \{e^*\},\ d_{j+1}(e^*) = d(e^*)/\lambda,\ d_{j+1}(e) = d_j(e) \text{ for each } e \in E - \{e^*\},$$

$$G_{j+1} = (V, E, d_{j+1}), \text{ add 1 to } j \text{ ("put } j = j + 1\text{")}$$

and go to the beginning of the 2nd step.

4 Application and Experience

Both problems LSP and BSP are taken from practice. Now we shall outline several applications.

4.1 Application of LSP

The problem LSP from the chapter 2 has an important practical application in **weak demand areas** where not yet exists any bus service. The resulting path can be used as a bus route.

However, some additional requirements, not yet counted, can arise. The most important are the following:

One compulsory terminal v' of the path may be requested. E.g. it is the railway station or a "border" point between the weak demand area and the strong demand one. This requirement can be introduced in the mode by an additional constraint. Similar situation occurs if **two compulsory terminals** v' and v'' are requested. It is obvious, that these constraints can be applied in all our methods in the form that at least one of the chosen edges has to be incident with v' (or one with v' and one with v'')

4.2 Application of BSP

The authors resolved the bus transport network in Slovak Town Piešťany by means of BSP. However, the application field seems to be quite wide, far beyond the bus transport. Let us imagine that some body, e.g. a regional administration, has too dense road network to maintain in winter. The optimal subgraph may be a good solution.

4.3 Computing experience

V. Přibyl, a senior lecturer of Informatics at the Prague University of Economics, Faculty of Management, has recently made wide computing experiments with depth-first search technique, LP model and heuristics both for LSP and BSP. These results have not yet been published but one can expect the publication within months. He shows that all methods are working well, the results of exact methods are correct and the heuristics give very good approximations. The exact methods seem to be usable on PC's until the number of vertices about 20. The heuristics seem to be able to solve any problem from real practice.

5 Conclusions

We have shown that the theory of economically optimal network design, as a part of the theory of network economy, can be developed in the direction of subnetwork optimization.

Two problems were identified and solved:

- The cheapest plausibly accessible path design.
- The cheapest subgraph where the relative trip elongation does not exceed the given limit.

For both problems, the depth-first search techniques, LP models and heuristics were proposed.

Finally, practical applications and computing experience were mentioned.

The authors hope that it is not the final word in this direction since there may occur many new practical constrains needeng new research work.

References

[1] Amini, O., Peleg, D., Pérennes, S., Sau, I., Saurabh, S.: Degree-Constrained Subgraph Problems: Hardness and Approximation Results. In: E. Bampis and M. Skutella (Eds.): *WAOA 2008, LNCS 5426*, Springer-Verlag Berlin Heidelberg 2009, pp. 29–42.

[2] Carr, R., Ravi, R.: A New Bound for the 2-Edge Connected Subgraph Problem. In: R. E. Bixby, E. A. Boyd, and R. Z. Rios-Mercado (Eds.): *IPCO VI, LNCS 1412*, Springer-Verlag Berlin Heidelberg 1998, pp. 112-125.

[3] Cheriyan, J., Thurimella, R.: Approximating Minimum-Size k-Connected Spanning Subgraphs via Matching, *SIAM J. Comput.*, 30, 2 (2000), pp. 528-560.

[4] De Man, A.P.: *The Network Economy: Strategy, Structure and Management*. Edward Elgar, Cheltenham, GB, 2004.

[5] Eades, P., Foulds, L. R., Giffin, J. W.: An efficient Heuristic for Identifying a Maximum Weight Planar Subgraph. *Combinatorial Mathematics IX, Lecture Notes in Mathematics No. 952*, Springer-Verlag, Berlin 1982.

[6] Fernandes, C.G.: A Better Approximation Ratio for the Minimum Size k-Edge-Connected Spanning Subgraph Problem, *J. Algorithms*, 28, 1 (1998), pp. 105-124.

[7] Fiala, P.: *Network Economics* (in Czech) Professional Publishing, Praha 2008

[8] Gabow, H.N., Goemans, M.X., Tardos, E., Williamson, D.P.: Approximating the Smallest k-Edge Connected Spanning Subgraph by LP-Rounding. In: *SODA '05 Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2005, pp. 562-571.

[9] Jun Xiao et al.: Comprehensive method on evaluation of distribution network planning. In: *Proc. DRPT 2008, Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, Nanjuing (China), April 2008, pp. 1249-1254.

[10] Safari, M.A., Salavatipour, M.R.: A Constant Factor Approximation for Minimum λ-Edge-Connected k-Subgraph with Metric Costs. In: A. Goel et al. (Eds.): *APPROX and RANDOM 2008, LNCS 5171*, Springer-Verlag Berlin Heidelberg 2008, pp. 233–246.

[11] Shimoda, Y.: Heat source network and heat transport. In: *District Heating & Cooling* Vol. 41 (2000), No. 3, pp.3-5.

[12] Suzuki, A. Tokuyama, T.: Dense subgraph problem revisited. Proc. Joint Workshop "New Horizons in Computing" and "Statistical Mechanical Approach to Probabilistic Information Processing" ,18-21 July, 2005, Sendai, Japan.

[13] Vassilevska, V., Williams, R.,Yuster, R.: Finding the Smallest H-Subgraph in Real Weighted Graphs and Related Problems. In: M. Bugliesi et al. (Eds.): *ICALP 2006, Part I, LNCS 4051*, Springer-Verlag Berlin Heidelberg 2006, pp. 262–273.

Comparison of Cobb-Douglas Production Functions of the Chosen Countries by Using Econometric Model

Josef Černý[1]

Abstract. Production functions of the state economy can be seen as a key tool for the description of the potential product of the economy. It is a basic and necessary mean for the description of how the economy of the state reaches or does not reach its potential by using labour force and capital as economic fundamentals of the production. This function form first tested by Cobb and Douglas can be used for both macro- and microeconomic perspectives such as a potential product of the economy of the state or production function of the business sector or the enterprise by itself.

This paper faces the goal to compare the production functions of the Czech Republic, Poland and Slovakia via econometric linear regression model using quarterly data of years 1998 and 2010.

This comparison supports the research initiatives in the supply chain management field in Moravian-Silesian (M-S) region in meaning to analyze the wide entrepreneurial macro-environment of the region with regard to the whole Czech Republic and neighbour states of M-S Region.

Keywords: Applied Econometrics, Cobb-Douglas Production Functions

JEL Classification: C50
AMS Classification: 91G70

1 Introduction

Production function belongs to the essential economic models, which have been developed to help forming both understanding production process in microeconomic perspective and economic policy of countries in macroeconomic perspectives. The easiest form of the production function is Cobb-Douglas production function (CDPF), developed by Cobb and Douglas, which can be found in most of textbooks of economics, for example in [1] or [2], or [3]. This function has also been examined by applied econometricians and therefore appears, for example in [4]. Production function of the economies of the Czech Republic, Slovak Republic and Poland are modelled in order to create a touch in macroeconomic business environment of the Moravian-Silesian (M-S) region, where the scope is taken not only for the Czech Republic, as the country, where M-S region lies, but for the Slovak Republic and Poland as neighbours of the region. With knowing the economies potential product and the structure of the influence of labour-force and capital, production functions can contribute to expectations of business environment and give the space of better strategic planning of partners for local, international and global supply chains. The purpose of this article is to develop CDPF, estimate the parameters by ordinary least squares (OLS) as a linear regression model (LRM), and other related procedures and testing for the model validation. This paper is spread out into several chapters. After the introduction, the CDPF is described in general with regard to some theoretical modifications and brief papers reviewed in section two. The third section brings a short inside into the methodology, which it was worked in this paper with. Practical applications, their summary and conclusions are coved in sections four and five. The last part of paper refers to the acknowledgment information and the literature sources.

2 Cobb-Douglas Production Function

CDPF leans on neoclassical approach and presents a causal relationship between output and alternative combinations of inputs [4]. There can be analyzed also relationship between input factors. Hušek [4] also adds that this relationship can be utilized both with the aim of explanation of the real output quantities dependent on used combination of inputs (labour and capital), and in perspective for explaining the potential product, resp. production gap.

[1] VŠB-Technical University of Ostrava/Faculty of Economics, Department of Mathematical Methods in Economics, Sokolská třída 33, Ostrava 1, 701 21, josef.cerny@st2.vsb.cz.

2.1 Mathematical Causality

As having said in the paragraph above, CDPF represents the causal relationship between output and inputs. In economic denotation, it is usually written as [2, p. 221]:

$$Q = AL^{\alpha} K^{\beta}. \quad (1)$$

In this formula (1), Q means output in physical units [2], (f. e. for firm utilization) or in currency units (f. e. for national economy production functions), L means quantity of labour and K is referred to be quantity of capital. A, α and β are positive parameters estimated in each case from the given data [4]. Generally parameter A refers to actual state of the technology, the higher value, the more developed technology. Parameter α represents the elasticity of labour-force and parameter β represents the elasticity of capital, with important application that if labour changes by 1 percent point product changes by α percent points and if capital changes by 1 percent point, production changes by β percent points. To parameters α and β also apply relations: $\alpha + \beta > 1$ means increasing returns to scale; $\alpha + \beta = 1$ provides constant return to scale; and $\alpha + \beta > 1$ represents decreasing returns to scale [2] or [3].

2.2 Linearization of the Cobb-Douglas Production function

From the equation (1) is obvious, that the function specification is non-linear in parameters but in this particular case, this form can be easily transform to linear form using logarithm transformation, which is proposed in, for example [2] or [4] and stochastic linear transformation is mentioned in [5]:

$$\ln Q = \ln A + \alpha \cdot \ln L + \beta \cdot \ln K. \quad (2)$$

For stochastic form containing residual compound (random error) and time component for the number of observation:

$$\ln Q_t = \ln A + \alpha \cdot \ln L_t + \beta \cdot \ln K_t + u_t. \quad (3)$$

For the estimation of parameters from the equation form (3), OLS procedure can be used [5].

2.3 Other Forms of Production Functions – Growth Models

CDPF model represented above is also used as a standard neoclassical growth model [6]. Traditionally, constant returns to scale are assumed in this model, i. e. $\alpha + \beta = 1$. There are mentioned other growth models, for example in [6], growth model (The Augmented Neoclassical Growth Model) includes factor of human capital with own parameter in addition to previous form. Another form is the growth model (The Broad Capital Endogenous Growth Model) also measures externalities of investments via including another multiplying parameter to capital. The Intentional Human Capital Endogenous Growth Model portrays technological progress as a result of research and education vie the accumulation of human capital. Schumpeterian or Innovation Endogenous Growth Model includes into a model a variable D with own parameter and this measures innovative progress increasing with the amount of labour allocated in R&D. Effect of technological progress is also pointed in [1] where the standard model includes time variable t in the production function, f. e. $Y = f(K, L, t)$.

From research papers dealing with CDPF applications, there can be included an application of CDPF in case of converging economies [7] where CDPF for the Czech Republic was compared with more general production function without any significant difference. From other papers, there can be depicted the case of analysis of total factor productivity contribution to economic growth via production function with input factors of labour and capital [8]. Special applications of CDPF can be also found in agricultural analysing agricultural technology progress with CDPF [9]. Upmann [10] analyses Laffer curve with specific utilization of CDPF.

3 Methods and Methodology

3.1 Description of variables

For the estimation of parameters of CDPF, there were obtained necessary variables for each country (i. e. the Czech Republic, Slovak Republic and Poland) from OECD statistics [11]. Gross domestic product (*GDP*) refers variable product, capital formation (*CAP*) is for capital and labour-force (*LF*) represents labours. Each of variables contains 52 observations, when time series range from 1998 to 2010, quarterly. Time series are seasonally adjusted. All variables have been transformed into logarithm form in order to be possible used equation (2), more preciously (3) and perform the OLS estimation of parameter in the LRM.

3.2 Classical Assumptions and used testing methods for LRM

In order to perform OLS estimation of parameter of linear regression model of Cobb-Douglas Production function and obtain valid model, there are several classical assumptions to be followed, tested and if not satisfied then correct the model. Assumptions are (f. e. in [5] or [12]):

- The average value of random error u is $E(u) = 0$ and the variance of random error $var(u) = \sigma^2$.
- The variance of random error is equal for u_i and u_j, where $i \neq j$ represent cases of observations (in time series models often denoted as t, such as in the paper application).
- The covariance between any pair of random error u_i and u_j is $cov(u_i, u_j) = 0$ for cases $i \neq j$.
- The random error is supposed to be normally distributed: $u \approx N(0;\sigma^2)$.
- The matrix X is not random.
- The explanatory variables X_1, ... X_2, ... X_n are not perfectly correlated.

3.3 Methods for Testing Model Validation and Assumption Violation

After the classical OLS Estimation, it is necessary to test parameters to be statistically significant and to test statistical significance of the whole model. For testing parameters, t-test will be used and for the whole model testing, F-test will be used. The chosen significance level of p-value is 0.05, i. e. 5%. This significance level will be used for all hypothesis testing [5].

For testing specification of the model, it will be used Reset Test, where the difference between original estimation an auxiliary regression is tested for significant via F-test. The auxiliary regression is used for examining the impact of adding new variables into original regression model to changes of R^2.

The test for autocorrelation, respectively for the first-order autocorrelation will be tested by using Durbin-Watson (DW) Statistic, f. e. described in [5] or [12] or [13]. Critical values of DW - d_L and d_U - can be found on web Standford.edu [13].

In CDPF, there are only two explanatory variables then Pearson correlation coefficient is enough to indentify pair collinearity between explanatory variables. Authors recommend different border for absolute value of Pearson correlation coefficient, for example Hušek [5] claims 0.8; Koop [15] writes 0.9. For heteroscedasticity testing, White-test will be used, with auxiliary regression of squared residuals in order to compute test statistic. [5]

The last assumption testing faces the test if random errors follow normal distribution with zero mean and specified variation. For this testing, Smirnov-Kolmogorov (KS) one sample test will be used, comparing theoretical and empirical distribution via differences. If the p-value of the test will be higher than significance level, we do not reject the null hypothesis about the normal distribution of the random errors.

For all testing, there can be used also an alternative way of hypothesis testing, via comparing computed p-value of the test statistic with significance level 0.05. When p-value exceeds the significance level 0.05, the null hypothesis is not rejected. For significant first-order (not only the first) autocorrelation of residuals, Cochrane-Orcutt iterative procedure will be used. This procedure lies in determination of the parameter ρ from the auxiliary regression [5]:

$$u_t = \rho \cdot u_{t-1} + \varpi_t \,. \tag{4}$$

And then to apply the parameter ρ into the following generally denoted regression [5]:

$$Y_t - \rho \cdot Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1,t} - \rho \cdot X_{1,t-1}) + \beta_2(X_{2,t} - \rho \cdot X_{2,t-1}) + u_t - \rho \cdot u_{t-1} \,. \tag{5}$$

Or possible to denote the equation (5) into form:

$$Y_t^* = \beta_0^* + \beta_1^* \cdot X_{1,t}^* + \beta_2^* X_{2,t}^* + u_t^* \,, \tag{6}$$

where $Y_t^* = Y_t - \rho \cdot Y_{t-1}$; $X_{1,t}^* = X_{1,t} - \rho \cdot X_{1,t-1}$; $X_{2,t}^* = X_{2,t} - \rho \cdot X_{2,t-1}$; $\beta_0^* = \beta_0(1-\rho)$ and $u_t^* = u_t - \rho \cdot u_{t-1}$.

4 Practical Cases of the Czech Republic, Slovak Republic and Poland

This section deals with the practical application of the econometric methodology for CDPF in real data of the Czech Republic, the Slovak Republic and Poland in time period from 1998 to 2010 as it was specified in the section 3. 1. The following table shows basic statistics of each of variables.

Variable	Units	No. Obs.	Min	Max	Mean	St.Dev.	Skew	Kurt
GDP_CZE_{SAS}	mil. USD	52	142934.53	275692.44	205126.99	47699.22	0.87	-1.52
CAP_CZE_{SAS}	mil. USD	52	28092.24	60720.31	44061.40	10298.18	0.18	-0.65
LF_CZE_{SAS}	Persons	52	4666537.00	5001074.00	4787916.00	103770.30	0.56	-1.02
GDP_SLV_{SAS}	mil. USD	52	54589.93	133043.57	87492.30	27174.06	0.52	-1.07
CAP_SLV_{SAS}	mil. USD	52	11205.52	25428.97	16877.68	4651.16	0.41	-1.38
LF_SLV	Persons	52	2089768.00	2460991.00	2330649.59	109431.60	0.64	-0.92
GDP_POL	mil. USD	52	356509.53	771248.73	527782.45	130988.83	0.45	-1.22
CAP_POL	mil. USD	52	61700.93	146981.42	95427.34	25746.06	0.49	-1.10
LF_POL	Persons	52	13557217.00	16013193.00	14746128.00	819090.46	0.15	-1.39

Table 1 Original variables description (before logarithm transformation)

4.1 Cobb-Douglas Production Function of the Czech Republic

Original CDPF for the Czech Republic follows the equation (1). After logarithm transformation OLS provides following 1st estimation of the CDPF for the Czech Republic:

$$ln\,GDP_{CZE,t} = -29.246 + 2.121\,ln\,LF_{CZE,t} + 0.827\,ln\,CAP_{CZE,t} + \hat{u}_{CZE,t} \tag{7}$$

All parameters are statistically significant since p-value is 0.000 for all and it is smaller than the significance level 0.05 and the null hypothesis about parameters equal to zero is rejected. The same result comes to the whole model significance. Coefficient of determination R_0^2 = 0.986, which stands for very good indicator of model fitting the data. Pearson correlation coefficient between explanatory variables is 0.672, which is lower than 0.8, then pair collinearity between capital and labour is rejected and this assumption is satisfied. Graphical plot of residuals showed problems and in addition with lower value of DW statistic, the problem with positive first-order autocorrelation is taken into account. Before full acceptance of the positive first-order autocorrelation, it was necessary to run Reset test. Since the R_1^2 computed from auxiliary regression reaches the same value as the original R^2 therefore F-statistic is equal to 0 and the null hypothesis of well-specified model is not rejected. Test statistic for White-test is equal to 0.468 falls into the null hypothesis acceptance region, since critical value $c^2(0.05; 3)$ is 7.814. KS one sample test proved the normality of random errors, when p-value 0.645 supports the null hypothesis about not rejecting normality. The only one of the classical assumptions was violated, i. e. autocorrelation. An attempt for better estimation will be provided by Cochrane-Orcutt procedure. After using transformations of variables in equation (5 resp. 6), the following 2nd estimation has been obtained:

$$ln\,GDP^*_{CZE,t} = -8.115 + 1.989\,ln\,LF^*_{CZE,t} + 0.778\,ln\,CAP^*_{CZE,t} + \hat{u}^*_{CZE,t} \tag{8}$$

The parameters after 1st Cochrane-Orcutt iteration are also statistically significant, coefficient of determination declined at value R^2 = 0.925 but Durbin-Watson statistic reached up to 1.847, which indicates no-autocorrelation of the first-order and partial autocorrelation function displayed no autocorrelation of other orders. Other tests showed the assumption satisfaction: Reset test proved the correct specification, White-Test supported homoskedasticity, explanatory variables are not perfectly correlated and the value of Pearson correlation is 0.716 lower than 0.8. KS test verified random errors to be normally distributed. The second Cochrane-Orcutt iteration led to small improvement of DW up to 1.899 and R^2 declines to 0.920 so the estimation (8) can be considered to be appropriate Cobb-Douglas production model of the Czech Republic.

4.2 Cobb-Douglas Production Function of the Slovak Republic

The function form for the Slovak Republic starts from the modified equation (1) transformed by using logarithm into the linear form of the equation. The first estimation obtained by the OLS:

$$ln\,GDP_{SLV,t} = -11.436 + 0.898\,ln\,LF_{SLV,t} + 0.994\,ln\,CAP_{SLV,t} + \hat{u}_{SLV,t} \tag{9}$$

The formula (18) describing Cobb-Douglas production function for the Slovak Republic discloses problems with nearly all areas of classical assumptions of the LRM. The only assumption of normality of random errors was satisfied. Else parameters and the model as a whole were statistically significant and R^2 with the value at 0.968 was relatively excellent, but DW statistic showed very low value 0.352 and plot of the residuals brought a suspicion of the very strong first-order autocorrelation or miss-specification of model. Reset test rejected the hypothesis about the correct specification. Another problem appeared in a strong correlation between explanatory variables, where Pearson correlation coefficient between labour-force and capital is 0.879. White-test and graphical scatter plot proved the rejecting the homoskedasticity null hypothesis, since $n \cdot R^2 = 52 \cdot 0.118 = 9.776$ and $c^2(0,05; 3) = 7.815$.

With regard to problems described above, an attempt for the Cochrane-Orcutt Iterative method was applied. After 2nd iteration, DW statistic reached the plausible value 1.831. R^2 is 0.824, which is plausible value with regard to increased value of DW. While most of tests fail to classical assumptions in the 1st estimated model, the production function after the second iteration of Cochrane-Orcutt procedure does not embody problem with collinearity of explanatory variable, since Pearson coefficient is 0.654. Autocorrelation of random errors has been shifted away in all orders ($DW = 1.831$) and heteroskedasticity has turned into homoskedasticity via White-test. Reset test did not reject the correct specification in the null hypothesis and the production function of the Slovak Republic is presented in following equation:

$$ln\,GDP^*_{SLV,t} = -4.374 + 2.225\,ln\,LF^*_{SLV,t} + 0.517\,ln\,CAP^*_{SLV,t} + \hat{u}^*_{SLV,t}\,. \tag{10}$$

Random errors are normally distributed with p-value 0.994 (via KS test).

4.3 Cobb-Douglas Production Function of Poland

The function form for Poland starts from the modified equation (1) transformed by using logarithm into the form with estimated parameters:

$$ln\,GDP_{POL,t} = 1.851 + 0.051\,ln\,LF_{POL,t} + 0.914\,ln\,CAP_{POL,t} + \hat{u}_{POL,t} \tag{11}$$

The coefficient of determination R^2 is highest of all CDPFs analyzed in this paper. The value at 0.996 is at excellent level. Despite of the determination, one serious problem appeared in the model. Model is significant, parameter of capital is also significant at 0.05 level but labour-force p-value is 0.293 which concludes in the insignificance of the parameter. The production function has been estimated without labour-force and cannot be compared with previous states. The final estimation after 1st Cochrane-Orcutt Iteration has arrived with following equation:

$$ln\,GDP_{POL,t} = 1.406 + 0.912\,ln\,CAP^*_{POL,t} + \hat{u}^*_{POL,t}\,. \tag{12}$$

This model estimation is statistically significant, R^2 lightly declined at the level 0.987 and Durbin-Watson statistic increased up to 1.940. Reset test proved correct specification with p-value 0.0512 which is higher than significance level 0.05. Random errors records the high improvement in graphical progress and the autocorrelation has been smoothed away, errors via White-test (and also graphical test) proved homoskedasticity and Kolmogorov-Smirnov test does not reject normal distribution of residuals with p-value 0.223 higher than significance level 0.05.

5 Summary and Conclusion

For the Czech Republic, the corrected by Cochrane-Orcutt Procedure estimated stochastic function is:

$GDP_{CZE,t} = 0.000299 \cdot LF^{1.989}_{CZE,t} \cdot CAP^{0.778}_{CZE,t} \cdot \hat{u}_{CZE,t}$. The elasticity of labour-force is more than 2.5 times higher than elasticity of capital. While taking into account $\alpha + \beta = 2.767$ and it means the increasing return to scale with stronger impact of labour-force.

The Slovak Republic's production function has following stochastic form:

$GDP_{SLV,t} = 0.0126 \cdot LF^{2.225}_{SLV,t} \cdot CAP^{0.517}_{SLV,t} \cdot \hat{u}_{SLV,t}$. The elasticity of labour-force is more than 4 times higher than elasticity of capital. While taking into account $\alpha + \beta = 2.742$ and it means the increasing return to scale with much stronger impact of labour-force than in the Czech Republic.

Production function of Poland can be estimated in following stochastic form:

$GDP_{POL,t} = 4.0796 \cdot LF^{0}_{POL,t} \cdot CAP^{0.912}_{POL,t} \cdot \hat{u}_{POL,t}$. Labour-force is in this econometric model insignificant, therefore parameter is set to be equal to zero. The returns to scale are decreasing but containing only capital in the model has to be considered because economically it is against economic assumptions that the labour-force would not be included in the model of the national economy product. The comparison of CDPF input factor parameters is showed in a table 2.

Country	A	α	β	α + β
CZE	0.000299	1.989	0.779	2.768
SLV	0.0126	2.225	0.517	2.742
POL	4.0796	0	0.912	0.912

Table 2 Comparison of Parameters given by Estimation of CDPF

The CDPFs have been estimated for the Czech Republic and Slovak Republic. Both Countries show the increasing returns to scale and the Slovak Republic involves more labour than Czech Republic. Supply Chains should consider business partners from the Slovak Republic in fields more using labour force and in the Czech Republic to find more capital involved fields of business. Poland's production model is hardly exploitable for comparison and analysis for labour or capital possibilities for partners, because the model does not include labour-force, which is against CDPF economic assumption. To develop other production function forms estimation can be suggested for better understanding and practical utilization by including other factors, such as human capital, research and development, or time component as a explanatory variable as it was suggested in paragraph 2. 3. With more precious models (closer to the real economies), productivities of included factors could be analyzed.

Acknowledgements

The computations in this paper were provided by software SPSS Statistics. This paper was prepared with the support of student research project grant competition SP2011/87. "Ways to reduce risks by the change of the organization of supplier-customer relationships in chosen selected supply chains."

References

[1] BARRO, Robert, J. *Macroeconomics*. 4th. Ed. New York: Wiley, 1993. 599 s. ISBN 0-471-59953-0.
[2] SALVATORE, Domicńick. *Microeconomics: Theory and Applications*. 4th. Ed. New York: Oxford University Press, 2008. 724 s. ISBN 0-19-513995-X.
[3] HOŘEJŠÍ, Bronislav, et al. *Mikroekonomie*. 4. vyd. Praha : Management Press, 2006. 573 s. ISBN 80-7261-150-X.
[4] HUŠEK, Roman. *Aplikovaná ekonometrie : Teorie a praxe*. 1. vyd. Praha: Oeconomica, 2009. 344 s. ISBN 978-80-245-1623-3.
[5] HUŠEK, Roman. *Ekonometrická analýza*. Praha : Oeconomica, 2007. 367 s. ISBN 978-80-245-1300-3.
[6] MARTIN, Ron; SUNLEY, Peter. Slow Convergence? The New Endogenous Growth Theory and Regional Development. *Economic Geography*. Jul., 1998, 74, s. 201-227. ISSN 0013-0095.
[7] HÁJKOVÁ, Dana; HURNÍK, Jaromír. Cobb-Douglass production function: The case of a converging economy. *Finance a úvěr*. 2007, 57, s. 465-476. ISSN 0015-1920.
[8] HÁJEK, Mojmír; MIHOLA, Jiří. Analysis of total factor productivity contribution to economic growth of the Czech Republic. *Politická ekonomie*. 2009, 57 (6), s. 740 - 753. ISSN 0032-3233.
[9] BATEN, Azizul; KAMIL, Anton Abdulbasah; HAQUE, Mohammad Anamul. Modeling technical inefficiencies effects in a stochastic frontier production function for panel data. *African Journal of Agricultural Research*. 2009, 12, s. 1374-1382. ISSN 1991-637X.
[10] OECD. *OECD.StatExtracts* [online]. 2011 [cit. 2011-04-15]. Complete Databases available via SourceOECD. Dostupné z WWW: <http://stats.oecd.org/index.aspx>.
[11] UPMANN, Thorsten. A positive analysis of labor-market institutions and tax reforms. *International Tax and Public Finance*. 2009, 16, s. 621 - 646. ISSN 1573-6970.
[12] HILL, R. Carter; GRIFFITHS, William E.; LIM, Guay C. *Principles of Econometrics*. 3rd. ed. Hoboken: Wiley, 2008. 579 s. ISBN 978-0-471-72360-8.
[13] GUJARATI, Damodar N. *Basic Econometrics*. 4th. Ed. New York: McGraw-Hill, Inc., 2004. 1002 s. ISBN 0-07-233542-4.
[14] *Stanford.edu* [online]. [cit. 2011-05-20]. Critical Values for the Durbin-Watson Test. Dostupné z WWW: <http://www.stanford.edu/~clint/bench/dwcrit.htm>.
[15] KOOP, Gary. *Introduction to Econometrics*. Chichester : Wiley, 2008. 371 s. ISBN 978-0-470-03270-1.

On estimation of the commodity hedge ratio in inventory management

Michal Černý[1]

Abstract. It is often the case that producers and users of commodities hedge the exposition in a commodity (e.g. an expected purchase, expected sale, commodity held on stock, commodity being processed) with a derivative position in another, highly correlated commodity. The basic risk-management issue is to estimate the hedge ratio properly. The hedge ratio is usually estimated from historical market data. We address the question of long-term stability of the hedge ratio and propose a method for estimation of the hedge ratio allowing to improve the hedging relationship using stability analysis.

Keywords: financial time series, hedge ratio, stability of estimators

JEL classification: C44
AMS classification: 62M10, 91B30, 91B84

1 Intoduction

It is often the case that commodity producers, consumers, processing companies and storekeepers hedge their future exposition in one market variable using derivatives with another, highly correlated underlying variable. Some examples include:

- an airline company hedges its expected purchases of kerosene with crude oil futures;
- a producer of high-quality wheat hedges its future sales with standardized wheat futures (such as CBOT futures as specified by the CBOT Rule 14104 *Wheat Futures — Grades/Grade Differentials*);
- a producer of wires hedges the aluminium held as inventory with aluminium futures, where the underlying aluminium is of different quality.

Another example is investigated in [7]. The main reason for this behaviour is that while the market-traded commodities are highly standardized, the companies need to purchase or sell commodities with different technological parameters for which the market is not available or is illiquid. Illiquid markets suffer from high bid-ask spreads; therefore, while hedging of e.g. kerosene exposition with kerosene futures could be theoretically possible, the spreads are so expensive that companies prefer highly liquid crude oil futures. Moreover, crude oil futures are available for more maturities than kerosene futures.

In general, the case is that an open position in a variable y originated from the core business of the industry is hedged via a derivative position in another variable x. Of course, in order the hedging relationship make sense, the variables x and y must be tightly interconnected.

From now on, we shall take the example $y =$ kerosene (say, expected future purchases of kerosene by an airline company) and $x =$ crude oil (say, price of crude oil futures to hedge the kerosene exposition).

When such a hedging relationship is established, a crucial question is to estimate the hedge ratio properly. For example, shall one ton of expected purchase of kerosene be hedged via purchased futures with the volume of 1.2, 1.3 or 1.4 tons of crude oil? The hedge ratio, i.e. the ratio between a unit of y and the number of units of x is the crucial factor determining the quality of the hedging relationship. If the hedge ratio is selected too low, then an unhedged position in kerosene remains open. On the other hand, if the hedge ratio is selected too high, then a new speculative position in crude oil originates. The risk-management aim is to select the hedge ratio in order the position be fully hedged. Henceforth, the hedge ratio will be denoted λ.

[1]University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague, Czech Republic, e-mail: cernym@vse.cz

2 Model for estimation of the hedge ratio

The hedge ratio λ is usually estimated from historical data. A long-term relationship between y and x is usually assumed to be of a form as

$$\begin{aligned}
\Delta y_t &= \lambda \Delta x_t + \varepsilon_t, \\
\Delta y_t &= \alpha + \lambda \Delta x_t + \varepsilon_t, \\
y_t &= \lambda x_t + \varepsilon_t, \\
y_t &= \alpha + \lambda x_t + \varepsilon_t, \\
\Delta \log y_t &= \lambda \Delta \log x_t + \varepsilon_t, \\
\Delta \log y_t &= \alpha + \lambda \Delta \log x_t + \varepsilon_t,
\end{aligned}$$

where t is the index of time, Δ denotes the difference operator and ε is the random error.

Let ν_t be homoskedastic with unit variance. The random errors ε_t are usually assumed in one of the forms

$$\begin{aligned}
\varepsilon_t &= \sigma \nu_t, \\
\varepsilon_t &= x_t \sigma \nu_t, \\
\varepsilon_t &= y_t \sigma \nu_t
\end{aligned}$$

where $\sigma > 0$ is a parameter. As an example, in this text we shall assume the model

$$y_t = \alpha + \lambda x_t + x_t \sigma \nu_t. \tag{1}$$

Figure 1, where prices of kerosene (y_t) and prices of crude oil (x_t) are plotted, shows that this assumption seems to be reasonable. Moreover, it is reasonable to assume that prices of kerosene, being an oil-based product, are driven by the prices of oil. The variance being proportional to the price level is a traditional feature of financial time series.

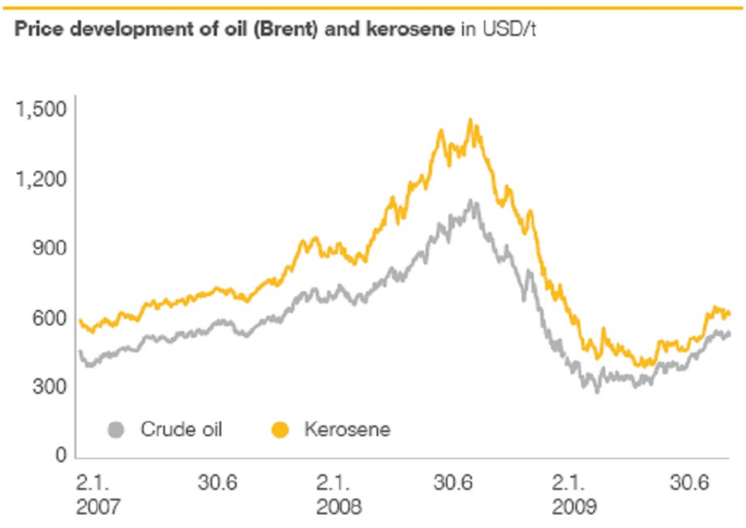


Figure 1: Long-term relation between prices of kerosene and Brent. Source: Lufthansa Annual Report.

Note that the long-term relationship (1) can be, with some simplification, also interpreted as a production function: on average, to produce y tones of kerosene, we need fixed costs α plus variable costs in the form of λ tons of crude oil per ton of kerosene.

The hedge ratio λ can be estimated as the absolute term in the homoskedastic model

$$\frac{y_t}{x_t} = \lambda + \alpha \cdot \frac{1}{x_t} + \sigma \nu_t \tag{2}$$

which is equivalent to (1).

We shall assume that ν_t's are such that the model (2) can be estimated with Ordinary Least Squares (OLS).

3 The main problem

Both x_t and y_t are high-frequency data. They are available on a daily basis (or even intraday). A long history of them is available; their prices have been quoted for several decades.

Our aim is to estimate λ as precisely as possible. It is well known that the variance of $\widehat{\lambda}$ ($\widehat{\lambda}$ is the OLS estimator of λ) decreases with the number of observations. Hence, if we assume a long-term relationship (1), to get the most exact estimate of λ, we should use all the historical data available.

However, on the other hand, it is doubtful whether e.g. kerosene/crude oil ratios from 1960's are relevant for estimation of the *contemporary* hedge ratio. It is not clear whether the assumption that the relationship (1) is stable over decades is valid. (For example, if the relationship (1) is interpreted as the production function, we may argue that technology changes time to time.) It seems more realistic to assume that the relationship (1) is stable in the short run (i.e., say months of a few years) while in the long run it may be subject to changes. The short-run stability assumption is essential; otherwise hedging of kerosene with oil futures would not make sense.

The main question now arises: how long history of data shall be taken into account when estimating the hedge ratio λ in (2) in order

- to minimize the variance of the estimator (i.e., to estimate λ as exactly as possible) and, simultaneously,
- to avoid estimation bias arising from possible instability of the value of λ in the long run.

We shall propose a method to deal with this problem. The problem has also been addressed, from a different perspective, in [3, 4, 5, 6, 8]. We shall introduce a method which is partially motivated by [2].

4 Step 1: A statistic for testing stability

Now assume that the set of historical data $x_1, \dots, x_n$ and $y_1, \dots, y_n$ is fixed. Let us test the hypothesis

$$H : \text{the relationship (2) is valid for all } t \in \{1, \dots, n\}$$

against the alternative

$$A : \text{there is a time } \tau \in \{3, \dots, n-3\} \text{ such that } \frac{y_t}{x_t} = \begin{cases} \lambda_0 + \alpha_0 \cdot \frac{1}{x_t} + \sigma\nu_t & \text{for } t \in \{1, \dots, \tau\}, \\ \lambda_1 + \alpha_1 \cdot \frac{1}{x_t} + \sigma\nu_t & \text{for } t \in \{\tau+1, \dots, n\}, \end{cases}$$

where α, λ, $(\alpha_0, \lambda_0) \neq (\alpha_1, \lambda_1)$, τ and $\sigma > 0$ are unknown parameters. Assuming that ν_t are $N(0,1)$ independent, we can construct the log-likelihood ratio

$$\begin{aligned}
\mathcal{L} &= \ln \frac{f_A}{f_H} \\
&= \ln \frac{\prod_{t=1}^{\tau} \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot \exp\left(-\frac{(y_t-\lambda_0 x_t-\alpha_0)^2}{2x_t^2\sigma^2}\right) \cdot \prod_{t=\tau+1}^{n} \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot \exp\left(-\frac{(y_t-\lambda_1 x_t-\alpha_1)^2}{2x_t^2\sigma^2}\right)}{\prod_{t=1}^{n} \frac{1}{\sigma\cdot\sqrt{2\pi}} \cdot \exp\left(-\frac{(y_t-\lambda x_t-\alpha)^2}{2x_t^2\sigma^2}\right)} \\
&= \frac{1}{2\sigma^2}\left[\sum_{t=1}^{n} \frac{(y_t - \lambda x_t - \alpha)^2}{x_t^2} - \sum_{t=1}^{\tau} \frac{(y_t - \lambda_0 x_t - \alpha_0)^2}{x_t^2} - \sum_{t=\tau+1}^{n} \frac{(y_t - \lambda_1 x_t - \alpha_1)^2}{x_t^2}\right],
\end{aligned}$$

where f_A and f_H denote the joint distribution of $\frac{y_t}{x_t}$ under A and H, respectively. If we assume that τ is fixed, we get the traditional log-likelihood test for the existence of change in the regression relationship in time τ of the form

$$V_\tau = \frac{RSS_{1:n} - RSS_{1:\tau} - RSS_{\tau+1:n}}{RSS_{1:n}},$$

where $RSS_{i:j}$ is the residual sum of squares from OLS-estimated regression $\frac{y_t}{x_t} = \lambda + \alpha \cdot \frac{1}{x_t} + \sigma\nu_t$ using the data $t \in \{i, i+1, \ldots, j\}$. Relaxing the assumption that τ is fixed, we obtain the statistic

$$V = \max_{t \in \{3,\ldots,n-3\}} V_t. \tag{3}$$

We shall need critical values for the statistic V under H. The statistic V, being the maximum of dependent $B_{1,\frac{n}{2}-2}$-distributed random variables, has a complicated distribution; in fact, an exact formula is not known.

Fortunately, the statistic V is essentially the same statistic as investigated by Worsley [9]. Worsley derived a Bonferroni-type approximation (see also [1]) of the distribution of V in the form

$$\Pr[V \leqslant z] \approx B_{1,\frac{n}{2}-2}(z) - \frac{2\beta_{\frac{3}{2},\frac{n}{2}-1}(z)}{\pi \cdot (n-2)} \cdot \left(\sum_{i=2}^{n-3} \xi_i - \frac{1}{6} \left(\frac{n-5}{6} \cdot \frac{z}{1-z} - 1 \right) \cdot \sum_{i=2}^{n-3} \xi_i^3 \right), \tag{4}$$

where β and B denote the density function and the cumulative distribution function of beta distribution, respectively, and

$$\xi_i = \sqrt{\boldsymbol{x}_{i+1}^{\mathrm{T}} (\boldsymbol{X}_{i+1:n}^{\mathrm{T}} \boldsymbol{X}_{i+1:n})^{-1} \boldsymbol{X}_{1:n}^{\mathrm{T}} \boldsymbol{X}_{1:n} (\boldsymbol{X}_{1:i+1}^{\mathrm{T}} \boldsymbol{X}_{1:i+1})^{-1} \boldsymbol{x}_{i+1}}, \quad \boldsymbol{x}_i = (1 \ \tfrac{1}{x_i}), \quad \boldsymbol{X}_{i:j} = \begin{pmatrix} \boldsymbol{x}_i \\ \boldsymbol{x}_{i+1} \\ \vdots \\ \boldsymbol{x}_j \end{pmatrix}.$$

Using (4) and binary search, it is computationally feasible to derive the z_0-quantile for V given z_0. This z_0-quantile will be referred to as the Worsley's z_0-critical value.

If H is rejected, then (3) also suggests a natural estimator of the unknown value τ:

$$\widehat{\tau} = \arg \max_{t \in \{3,\ldots,n-3\}} V_t. \tag{5}$$

So far we have treated the data set $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, to which the statistic V and the estimator $\widehat{\tau}$ is applied, as fixed. Instead of V and $\widehat{\tau}$, we could have also written $V(x_1, \ldots, x_n; y_1, \ldots, y_n)$ and $\widehat{\tau}(x_1, \ldots, x_n; y_1, \ldots, y_n)$ to emphasize the data set to which the statistic V and the estimator $\widehat{\tau}$ are applied. We shall also write

$$V(x_k, \ldots, x_\ell; y_k, \ldots, y_\ell) =: V_{k:\ell} \quad \text{and} \quad \widehat{\tau}(x_k, \ldots, x_\ell; y_k, \ldots, y_\ell) =: \widehat{\tau}_{k:\ell}$$

if the statistic V and the estimator $\widehat{\tau}$ are applied to the restricted data set $x_k, \ldots, x_\ell$, $y_k, \ldots, y_\ell$ (i.e., the observations $t \notin \{k, k+1, \ldots, \ell\}$ are omitted). Then, the Worsley's z_0-critical value will be denoted $W_{k:\ell}^{z_0}$.

5 Step 2: Estimating the period of stability

Assume that the data set $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ is available; the tuple (x_n, y_n) is the most recent observation (say, today's quotes) and (x_1, y_1) the endmost observation (say, quotes originated several decades ago).

We shall start from the most recent period and search for the last point of change in the hedge ratio. Let z_0 be fixed (e.g. $z_0 = 0.95$ corresponds to the 5% level of the test). We shall run the following procedure:

{1} **for** $k := n - 20$ **downto** 1 **do**
{2}   **if** $V_{k:n} \geqslant W_{k:n}^{z_0}$ **then** stop and output $\widehat{\tau} := k + \widehat{\tau}_{k:n} - 1$;
{3} **next** k
{4} stop and output "the entire data set is stable".

(The choice of $n - 20$ in {1} is arbitrary; any other value could be used.)

If the procedure {1} – {4} stops in the step {2}, then a point of change is detected and the procedure outputs its estimate $\widehat{\tau}$. Then, for estimation of the hedge ratio, we shall use the model (2) with the data set $t \in \{\widehat{\tau}+1, \widehat{\tau}+2, \dots, n\}$.

Moreover, it is suitable to take into account the fact that the estimator $\widehat{\tau}$ could have estimated the true point of change inexactly. Though the exact distribution of $\widehat{\tau}$ under A is not known, it seems to be reasonable to get over the loss of a reasonable number of observations, say m (determined heuristically), and estimate the hedge ratio using the model (2) with the data set $t \in \{\widehat{\tau}+m, \widehat{\tau}+m+1, \dots, n\}$. (A recommendation, how m should be chosen in practice, is subject to further research.)

6 Visualization of the method and an example

The method can be visualized in the following way. We plot the processes

$$\widetilde{V}_k := (n-k) \cdot V_{n-k:n}, \quad \widetilde{W}_k^{z_0} := (n-k) \cdot W_{n-k:n}^{z_0}, \quad \widetilde{\tau}_k := n-k+\widehat{\tau}_{n-k:n}$$

for $k = 20, 21, \dots, n$ with (say) $z_0 = 0.95$ and $z_0 = 0.99$. (The scaling factor $n-k$ in the definition of $\widetilde{V}_k$ and $\widetilde{W}_k^{z_0}$ has been added, without loss of generality, to make the pictures more transparent.) Such a plot also shows how stable the estimate of $\widehat{\tau}$ output in {2} is; ie whether the value $\widehat{\tau}$ detected with the procedure {1} – {4} remains stable even if we had not stopped in {2} and had iterated further.

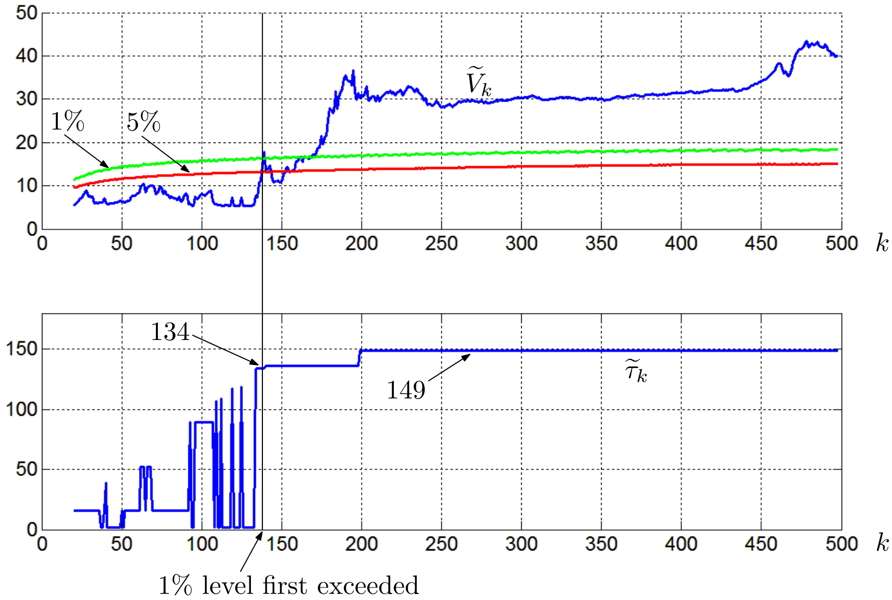


Figure 2: The process $\widetilde{V}_k$, Worsley's 5% and 1% critical bounds and the process $\widetilde{\tau}_k$.

In Figure 2, results of a simulated example are shown. We generated a trajectory of x_t for $t = 1, \dots, 500$ (which corresponds to four years if 1 year = 250 business days) as a lognormal random walk varying between \$20 and \$85, see Figure 3. The process y_t (the evolution of kerosene prices) was simulated using (1) with $\nu_t \sim N(0,1)$ independent and

$$\sigma = 0.1, \quad \alpha = 0, \quad \lambda = \begin{cases} 1.3 & \text{for } t = 1, \dots, 330, \\ 1.4 & \text{for } t = 331, \dots, 500. \end{cases} \tag{6}$$

Observe that the variance is quite high: if the price of crude oil is \$100, the standard error is \$10.

In Figure 2 it is apparent that the procedure {1} – {4} detects $\widehat{\tau} = 134$, which is an inexact estimate (by (6) the point of change appeared $500 - 330 = 170$ days ago). If we don't stop in the step {2} when the 1% level is first exceeded and iterate further, we arrive at the estimate $\widehat{\tau} = 149$ (which is a value closer to the true value 170). It can be seen that the estimate $\widehat{\tau} = 149$ is stable. Hence, the procedure

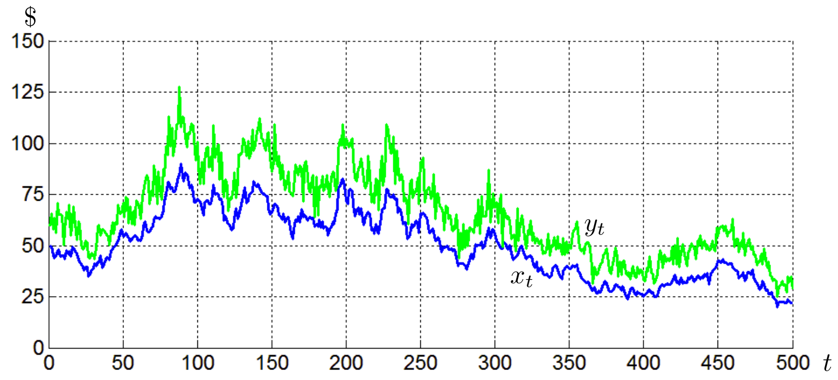


Figure 3: Simulated evolution of values of crude oil (x_t) and kerosene (y_t).

suggests to estimate the hedge ratio either using last 133 or last 148 observations, the remaining ("old" ones) being omitted.

We know that the true value of the hedge ratio is 1.4. The resulting OLS-estimates are

$$\widehat{\lambda}_{\text{from data } t \in \{500-133,\ldots,500\}} = 1.43 \quad \text{and} \quad \widehat{\lambda}_{\text{from data } t \in \{500-148,\ldots,500\}} = 1.46.$$

If we estimate the hedge ratio from the entire data set, we obtain a much worse value

$$\widehat{\lambda}_{\text{from data } t \in \{1,\ldots,500\}} = 1.23$$

which is clearly biased by the "old" history. If the latter value had been used, only 88% ($= 1.23 \div 1.4$) of the kerosene position would be hedged (on average), while with the method presented we get an "over-hedged" position of 102% ($= 1.43 \div 1.4$) or 104% ($= 1.46 \div 1.4$), respectively.

7 Conclusion

In this paper we introduced a method that helps in deciding how long data history shall be taken into account when the hedge ratio between two correlated commodities is estimated from historical data. The method is based on the theory of stability of regression models. We also suggested a natural visualization tool for the method. The method can be extended in different ways. It may be useful in on-line monitoring whether the hedge ratio remains stable or whether it should be adjusted during the existence of the hedging relationship. It can be also used for other market variables, such as hedging of an exposition in one currency by taking a derivative position in another currency. Further research shall focus on theoretical properties of the method, improvement of the approximation of critical values and testing in practice (including comparison with other methods known from literature in real-world case studies).

Acknowledgements

Supported by the Internal Grant Agency of the University of Economics, Prague.

References

[1] Černý, M.: Binary segmentation and Bonferroni-type bounds. *Kybernetika* **47** (2011), 38–49.

[2] Černý, M.: On estimation of volatility of financial time series for pricing derivatives. *Acta Oeconomica Pragensia* **4** (2008), 12–21.

[3] Černý, M., and Hladík, M.: The regression tolerance quotient in data analysis. In: *Proceedings of Mathematical Methods in Economics 2010* (J. Friebelová, and M. Houda, eds.), University of South Bohemia, České Budějovice, 2010, 98–104.

[4] Choudhry, T.: Short-run deviations and time-varying hedge ratios: Evidence from agricultural futures markets. *International Review of Financial Analysis* **18** (2009), 58–65.

[5] Hladík, M., and Černý, M.: Interval regression by tolerance analysis approach. *KAM-DIMATIA Series* 2010, paper 963.

[6] Hladík, M., and Černý, M.: New approach to interval linear regression. In: *24th Mini-EURO Conference on Continuous Optimization and Information-Based Technologies in the Financial Sector MEC EurOPT 2010. Selected Papers* (Kasimbeyli R. et al. eds.), Technika, Vilnius, 2010, 167–171.

[7] McMillan, D.: Time-varying hedge ratios for non-ferrous metals prices. *Resources Policy* **30** (2005), 186–193.

[8] Lien, D., and Shrestha, K.: Hedging effectiveness comparisons: A note. *International Review of Economics & Finance* **17** (2008), 391–396.

[9] Worsley, K.: Testing for a two-phase multiple regression. *Technometrics* **25** (1983), 35–42.

Optimal Assignment of Contest Objects to Expert Jurors

Jan Černý[1], Daniel Šťastný[2]

Abstract. The article deals with the optimization of a contest assessment process based on jury of experts, each of whom is able to consider only some (not all) contest participants. The point is to find an assignment of given objects of assessment to individual experts that would minimize the cost of assessment subject to the condition that each pair of objects is assigned to at least one expert.
We show the complexity of the problem and present a heuristics for its solution in a reasonable time. Moreover, two particular cases of the problem and its variations are analyzed and their solutions are described.

Keywords: contest design, covering of pairs.

JEL Classification: C65, J39
AMS Classification: 91B32, 05C70

1 Introduction

When designing contests and competitions (or, more generally, any selection process), one is ultimately interested in producing a ranking of contest participants while facing two crucial issues, the usual suspects: the cost and the quality of selection.

When objective selection criteria are missing, the assessment is based on juries – panels of experts. Sometimes the number of contest participants – objects of assessment – is so large that making each individual juror consider (rank) all of them would be excessively costly or practically impossible. Thus, in order to permit economic feasibility, individual jurors are assigned only a limited number of objects to consider.

However, in order not to compromise quality of assessment, one may require that any two objects must be considered (compared) against each other at least once. More formally, let us suppose that we are given n objects of assessment $a_1, \ldots, a_n$. We then may conceive of the selection process as a set of pairwise comparisons[3] and make sure that each possible pair a_i, a_j (for $j \neq k$, and $j, k = 1, \ldots, n$) belongs to at least one subset of objects assigned to individual jurors.

In this article we shall focus on the problem of designing the optimal assignment of contest objects to expert jurors with respect to the above requirements regarding the quality of the assessment process and its cost.

1.1 Experts and their Remuneration

Let us suppose that our n objects are to be evaluated by expert jurors belonging to a set $\boldsymbol{E}$. Each expert $E \in \boldsymbol{E}$ has a given capacity of $c(E) \geq 2$ (the maximal number of objects assignable to the expert E; the expert is then said to be $c(E)$-*valent*), and a given number $x(E)$ of objects that were actually assigned to the expert for evaluation, which cannot exceed the capacity ($2 \leq x(E) \leq c(E)$). The cost of assessment (evaluation), i.e. the expert's remuneration, is $p_i(x)$ taking positive values for all $i = 1, \ldots, m$ and all $x = 2, \ldots, c(E_i)$..

In practice, one can meet mainly two types of these cost functions:

a) Constant $p_i(x) = p > 0$ for all $i = 1, \ldots, m$ and all $x = 2, \ldots, c(E_i)$. This represents the case when each active expert receives a constant remuneration p not depending on the number of evaluated objects.

b) Linear $p_i(x) = px$ for all $i = 1, \ldots, m$ and all $x = 2, \ldots, c(E_i)$ where $p > 0$. This is the most frequent case in practice. Each expert receives the same remuneration p for each evaluated object.

[1] University of Economics in Prague, Faculty of Management, Jarošovská Street 1117/II, 37701, Jindřichův Hradec, cerny@fm.vse.cz.
[2] University of Economics in Prague, Faculty of Economics, W. Churchill Square 4, 13067, Prague 3, Daniel.Stastny@vse.cz.
[3] If we can say that two objects are equal or one of them is better, then there are methods of getting the desired ordering, as one can see e.g. in [1] or [6].
There may be some additional benefits of forcing experts to strictly rank objects (not allowing them to judge objects equal), consisting of the fact that it forces experts to make up their mind even in situations, where experts – to economize on effort or for other reasons – would choose equal rank.

The problem is to find such an assignment of objects to experts that each pair of objects is assigned at least to one expert while minimizing the cost of remuneration to experts.

1.2 Assignment of Objects to Experts and the General Problem

Let $\boldsymbol{E} = \{E_1, \dots, E_m\}$ be the set of available experts with capacities $c_1, \dots, c_m$ such that $c_1 \geq \dots \geq c_m > 1$ and prices $p_1(\bullet), \dots, p_m(\bullet)$ where $p_i(x_i)$ represents the remuneration of the expert E_i if assigned x_i objects. Let $I = \{1, 2, \dots, n\}$ be the set of indices of the objects to be evaluated, let I^k be the set of all k-element subsets of the set I and let $I^* = \varnothing \cup I^2 \cup I^3 \cup \dots \cup I^n$. Let α: $\boldsymbol{E} \to I^*$ be a mapping having the property that for each $(i, j) \in I^2$ there exists such $h \in \{1, ..., m\}$ that $(i, j) \in \alpha(E_h)$, i.e. each pair of object indices is assigned together to at least one expert. Then we call such mapping α the *proper assignment* (of objects to experts).

The General Problem (GP)

Let $\boldsymbol{E}$, c_i, $p_i(\bullet)$ and I^* be defined as above and let them be given. The **general problem** is to find such a proper assignment α that

$$\sum_{i=1}^{m} p_i(\operatorname{card}\alpha(E_i)) \to \min$$

where card$\alpha(E_i)$ means the number of elements in the set $\alpha(E_i)$.

2 Solution of the General Problem

In this chapter, we shall see that the problem can be solved by linear programming in the cases a) and b) and we shall say something about the complexity of the problem. Finally we shall present a heuristics.

2.1 Linear Programming Formulation of GP

Let $\boldsymbol{E}$, c_i and I be defined as above and let them be given. Let $p > 0$ and $p_i(x) = px$ for each $i = 1, \dots, m$ and each $x \in I^*$. Let x_{ij}, y_{ijk} be binary variables, $x_{ij} = 1$ means that $j \in \alpha(E_i)$, $x_{ij} = 0$ means the opposite, $y_{ijk} = 1$ means that $j,k \in \alpha(E_i)$, $y_{ijk} = 0$ means the opposite.

The problem is to find the values x_{ij}, y_{ijk}, $i = 1, \dots, m$ and $j,k = 1, \dots, n$, where $j \neq k$ fulfilling the following constraints:

$$x_{ij} + x_{ik} \geq 2y_{ijk}, \text{ for } j \neq k, j,k = 1, \dots, n \text{ and } i = 1, \dots, m. \tag{1}$$

$$\sum_{i=1}^{m} y_{ijk} \geq 1 \text{ for } j \neq k \text{ and } j,k = 1,\dots,n \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} \leq c_i \quad \text{for } i = 1,\dots,m \tag{3}$$

and minimizing the value of objective function

$$\sum_{i=1}^{m}\sum_{j=1}^{n} px_{ij} \to \min \tag{4}$$

The constraint (1) ensures that the pair j,k is assigned to the same expert, (2) ensures that each pair is assigned at least to one expert, (3) ensures that the capacity of the i-th expert is not exceeded and finally the condition (4) minimizes the total cost.

If we replace the linear cost function by the constant function, we need additional binary variables z_i, $i = 1, \dots, m$ fulfilling the constraints

$$\sum_{j=1}^{n} x_{ij} \leq (n+1)z_i \quad \text{for } i = 1,\dots,m \tag{3′}$$

and, moreover, we have to change the objective function

$$\sum_{i=1}^{m} pz_i \to \min \tag{4'}$$

We see that the GP in both cases a) and b) belongs to the family of constrained set covering problems. Therefore it is NP-hard and, moreover, it is NP-easy, since it is formulated as an integer linear programming problem.

2.2 Heuristic Method of Solution

The dimension of LP formulation of the GP is large regardless of the choice of the cost function. We have $mn(n – 1)$ variables y_{ijk}, mn variables x_{ij}, m variables z_i and at least $mn(n – 1) + n(n – 1) + m + 1$ conditions. We have to take into account the possible insufficiency of the available software. For that reason we can propose the following **heuristics:**

1st step (initial): Put $\alpha(E_i) = \emptyset$ for all $i = 1, \ldots, m$ and $J = \emptyset$. Find the greatest m' with the following property:

$$\sum_{i=1}^{m'} c(E_i) \le n$$

Then put $J = \{m'+ 1, \ldots, m\}$ and

$$\alpha(E_1) = \{1,\ldots,c(E_1)\},\quad \alpha(E_2) = \{ c(E_1) + 1,\ldots,c(E_1) + c(E_2)\},\ \ldots,$$

$$\ldots, \alpha(E_{m'}) = \{\sum_{i=1}^{m'-1} c(E_i) + 1, \ldots, \sum_{i=1}^{m'} c(E_i)\}$$

Finally, put

$$I_C = \left\{(j,k) \in I^2 : \exists_i (j \in \alpha(E_i) \text{ and } k \in \alpha(E_i))\right\}$$

2nd step (recursive): If $I_C = I^2$ then the mapping $\alpha(\bullet)$ represents the final result. If $I^2 - I_C \ne \emptyset$ and $J = \emptyset$ then the method cannot solve the problem. If $I^2 - I_C \ne \emptyset$ and $J \ne \emptyset$ then find such $i \in J$ and $I' \in I^*$ that $\text{card}I' = c(E_i)$ and $\text{card}\{(j, k) \in I^2 - I_C: j \in I', k \in I'\} \to \max$.[4]

Put $J = J - \{i\}$ and $I_C = I_C \cup \{(j, k) \in I^2 - I_C: j \in I', k \in I'\}$ and return to the beginning of the 2nd step.

3 Particular Cases

Now, we shall analyze two particular situations that are likely to arise when designing contests in real life.

3.1 2-valent Experts plus Limited Number of More-valent Experts

The first one, when $c(E_i) = 2$ for all $i = 1, \ldots, m$, is trivial. The solution exists if and only if $m \ge n(n – 1)/2$ and each pair of objects is assigned to a different expert. This case can be slightly modified. In practice, it can happen that the group of expert jurors can be divided into two subgroups:

- Internal jurors, i.g. employees of the submitter, having the capacity much greater than two (e.g. 10).
- External ones, having the capacity quite small, close to 2.

In the following paragraph, we shall deal with the special case where each external expert is 2-valent.

Lemma. *Let* $c_1 = c(E_1) > 2, \ldots, c_{m'} = c(E_{m'}) > 2$, $c_{m'+1} = c(E_{m'+1}) = 2$, $c_{m'+2} = c(E_{m'+2}) = 2, \ldots$ *Let the number of* 2-*valent experts is unlimited. Let* $p_i(x) = p$ *for all* $i = 1, 2, \ldots, m$ *and all* $x = 2, 3, \ldots, c_i$. *Finally, let* $c_1 + \ldots + c_{m'} \le n$. *Then the following assignment represents the optimal solution of GP:*

$$\alpha(E_1) = \{1, \ldots, c_1\},\ \alpha(E_2) = \{c_1 + 1, \ldots, c_1 + c_2\}, \ldots, \alpha(E_{m'}) = \{c_1 + \ldots + c_{m'-1} + 1, \ldots, c_1 + \ldots + c_{m'}\},$$

and not yet covered pairs from I^2 *are assigned to* 2-*valent experts arbitrarily*.

[4] The essence of the recursive step is that when $I_C \ne I^2$ a $J \ne \emptyset$, then we look for such an expert E_i and such a set of object indices I' that the number of newly covered pairs is maximized.

Proof. The sets $\alpha(E_1)$, $\alpha(E_2)$, …, $\alpha(E_{m'})$ are mutually disjoint. It is obvious that it represents the optimal solution any overlapping among the sets $\alpha(E_1)$, ..., $\alpha(E_{m'})$ causes a decrease in the total number of covered pairs and therefore an increase of necessary number of 2-valent experts.
Consequently it is not possible to cover all pairs by a smaller number of experts, which completes the proof.

Although the lemma describes a solution of the GP for the constant cost function, it remains true for the linear cost function as well, since no pair is assigned to more than one expert and one can easily see that the full use of more-valent experts is always cheaper than the substitution of them by 2-valent ones.[5]

3.2 Minimal Required Number of 3-valent Experts: Practical Application

When designing contests of this type, a collateral problem of finding the minimal number of jurors required to cover all the pairs, given their capacities and the number of objects.

In fact, this article is a product of the following problem the authors faced: in a paper contest, 30 student papers needed to be ranked in overall quality through pairwise comparisons by a large pool of other students, while each of these "judges" read and ranked 3 papers only. The problem was to determine a) the minimum number of experts required in order to make sure that each particular paper will be compared to all other papers at least once, and b) to design the assignment of paper triples to individual student-jurors that would meet the above condition.

It is obvious that it is a particular case of the GP: $n = 30$, $c(E_i) = 3$ for $i = 1, 2, …, m$ and the cost function is constant or linear. It represents a covering of pairs by triples as described in [2]. If it was $n = 31$ instead of 30, then the covering could be exact, i.e. the set of covering triples would represent the "classic" Steiner system triples, introduced more than 150 years ago [7], since, in [2], it is shown that such an exact covering exists if and only if $n = 6k + 1$ or $n = 6k + 3$ and $31 = 6*5 + 1$. On the other hand $n = 30$ does not fulfill this constraint and, therefore, some pairs ought to be covered twice.

For the concrete design of the covering (i.e. for the solution of this variant of the GP) the idea of [3] is used. This study included the problem of covering a n-vertices polygon on circle by polygons of much smaller number of vertices than n. In our case we are looking for such a covering of 30-gon by triangles in such a manner that each pair of the n vertices is covered by a triangle. The solution has the following structure: We started with triangles T_1= [1][2][10], T_2= [1][3][16], T_3= [1][4][15], T_4= [1][5][11] , T_5= [1][6][13] where e.g. [13] represents the vertex No. 13 of a polygon with 30 vertices numbered subsequently and located on a circle. Afterwards, we added 29 times shifted triangles to each one, e.g. [1 + 1][2 + 1][10 + 1] = [2][3][11], …, [1 + 29][2 + 29][10 + 29] = [30][1][2] after the reduction mod(30). In this way, having 5 initial triangles, each one plus 29 shifts we reached 5*30=150 triangles where the chords (pairs) [1][16]…[15][30] are covered twice. Thus, we found we needed 150 experts, each one assigned one of these triplets/triangles.

4 Conclusions

The paper introduced a new perspective on contest assessment design using pairwise comparisons by expert jurors. It was shown that the assignment of objects to experts has economic aspects as well, leading to optimization problems of combinatorial type. It was shown that these problems can be (theoretically) solved by linear programming or by the set covering procedure. However, in practice, they are usually too large for it, which is why a heuristics was proposed for the general case and for two particular cases.

The authors hope to have pointed out at useful directions to seek solutions for such problems. For example, the case of 4-valent and 5-valent experts could perhaps make use of [4] or [5]. This does not mean that the 3-valent expert problem is fully "exhausted". For instance, when number of experts (or contest budgets) allows the frequency of covering one particular pair of objects to be higher than one (e.g. when $|\boldsymbol{E}|$ is many times greater than $|I^2|$), we may want to introduce further condition assuring quality: a) requiring that the frequency of covering one particular pair is as uniform as possible, b) requiring the assignments of objects ($\alpha(E_i)$) not to be duplicated (or, if the number of jurors is so permitting, to be duplicated and further multiplied in a uniform way).

[5] In fact, for any integer $x > 2$ we have that the juror evaluating x objects covers $x(x – 1)/2$ pairs and receives the remuneration xp. Hence the comparison of one pair of objects costs $2p/(x – 1)$. On the other hand, if a pair is assigned to a 2-valent juror, then it costs $2p > 2p/(x – 1)$.
There is another independent case for making use of internal jurors: given their employee status, the cost of hiring them is – with respect to their participation in the assessment – essentially a sunk cost.

References

[1] Fiala, P., Jablonský, J., Maňas, M. : *Vícekriteriální rozhodování*. VŠE, Praha, 1997.

[2] Fort, M.K., Hedlund, G.A.: Minimal Coverings of Pairs by Triples. *Pacific J. Math.* Vol. 8. (1958) pp. 709-719.

[3] Guldan, F.: Maximization of distances of regular polygons on a circle. *Apl. Math.* 25 (1980), pp 182-195.

[4] Lamken, E.R., Mills, W.H., Mullen, R.C., Vanstone, S.A. Coverings of Pairs by Quintuples. *Journal of Combinatorial Theory*, Series A, Vol. 44, pp. 49-68.

[5] Mills, W.H. 1973. On the covering of pairs by quadruples. II. *Journal of Combinatorial Theory*, Series A, Vol. 15, pp. 138-166.

[6] Ramík, J.: *Analytický hierarchický proces (AHP) a jeho vyžití v malém a středním podnikání*. Slezská univerzita v Opavě, Obch.-podnik. fak. v Karviné, 2000.

[7] Steiner, J.: Combinatorische Aufgabe, *Journal für die Reine und Angewandte Mathematik* 45 (1853), 181–182 .

A Memetic Algorithm for Solving the Vehicle Routing Problem

Zuzana Čičková[1], Ivan Brezina, Juraj Pekár

Abstract: This article deals with evolutionary approach for solving the vehicle routing problem. The importance of that problem follows from many practical applications in the field of logistics, as well as from its computational complexity (NP-hardness); therefore the use of optimization techniques seems to be relatively complicated. Nowadays the popular way seems to be the use of alternative computational techniques that are inspired by evolutionary biology. Presented self organizing migrating algorithm (SOMA) belongs to the class of evolutionary techniques although one can classify it also as memetic algorithm.

Keywords: heuristics, vehicle routing problem with time windows, self organizing migrating algorithm, memetic algorithms

JEL Classification: C6, C61, C63
AMS Classification: 90B20, 90C90

Introduction

The vehicle routing problem (VRP) belongs to the class of combinatorial optimization. The interest in VRP is motivated by its practical relevance as well as by its computational complexity (NP–hardness). This problem consists in designing the optimal set of routes for a vehicle in order to serve a given set of customers ($j = 1, 2, ..., n$). Each customer has a certain demand q_j ($j = 1, 2, ..., n$). The vehicle has a certain capacity M (so that $q_j \leq$ M for all $j = 1, 2, ..., n$) and it is located in a certain depot. Further on there exists a matrix **C** (n + 1, n+1) that represents the minimum distance (length, cost, time) between all the pairs customers and also between the customers and the depot. The goal is to find optimal vehicle routes (usually minimum distance). The routes must be designed in such a way that each point is visited only once by exactly one vehicle, all routes start and end at the depot, and the total demands of all points on one particular route must not exceed the capacity of the vehicle.
Nowadays, there is no exact algorithm for solving so that problem in polynomial time. Even though that problems of small size could be solved by exact methods (branch and bound algorithms, cutting plane method etc.), the real-sized problems are still usually solved by heuristics.

1 Evolutionary algorithms

Evolutionary algorithms comprise a large number of nontraditional computing techniques whose common characteristic is that they are inspired by the observation of the nature processes (genetic algorithms, ant colony optimization, differential evolution, etc.), respectively from other disciplines (e.g. simulated annealing). Nowadays evolutionary algorithms are considered to be effective tools that can be used to search for solutions of optimization problems. The big advantage over traditional methods is that they are designed to find global extremes (with built-in stochastic component) and that their use does not require a priori knowledge of optimized function (convexity, differential etc.). To the date, the best known representatives are genetic algorithms.
Evolutionary algorithms differ from more traditional optimization techniques in that they involve a search from a "population" of individuals, not from a single one. Each individual represents one candidate solution for the given problem that is represented by parameters of individual. Associated with each individual is also the *fitness*, which represents the relevant value of objective function. A population can be viewed as np x d matrix (np - number of individuals in the population, d – number of parameter of individual). Every step involves a competitive selection that is carried out poor solutions.

Self organizing migrating algorithm (SOMA) can be classified as evolutionary algorithm, because it is based on the self organizing behavior of individuals in a social environment (e.g. a herd of animals looking for food).

[1] University of Economics in Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská 1, 852 35 Bratislava, cickova@euba.sk, brezina@euba.sk, pekar@euba.sk.

Secondly, SOMA can be also classified as memetic algorithm, where the basic characteristic is the use of competitive cooperative strategies with synergic attributes.

SOMA was introduced by Ivan Zelinka in 1999 and has been successfully tested on various types of test functions (e.g. Rosenbrock's saddle, De Jong functions, Schwefel's function etc.). Self organizing migration algorithm, as well as other evolutionary techniques, works well on solving non-constrained problems that contain continuous variables, but nowadays there were developed few approaches that involve the solving of constrained problems with integer or binary variables. This algorithm was used for solving many engineering problems e.g. active compensation in RF-driven plasmas, neural network learning, statistical optimization of chemical reactor etc. (see (Onwubolu & Babu, 2004), (Zelinka, 2002)). Varied approaches were used with more or less success for solving also NP-hard problems many real problems etc. traveling salesman problem, flow shop problem etc. (see (Brezina et al., 2009), (Davendra & Zelinka, 2008).

The steps of the algorithm are detailed described in (Zelinka, 2002), (Onwubolu, 2004), where it is also possible to found recommended values for different control parameters.

2 Solving the Vehicle Routing Problem by SOMA

Since SOMA is an algorithm, which works well in the case of non-constrained problems with continuous variables, in applying the algorithm for solving NP-hard problems, is necessary to consider the following facts:
- Selection of an appropriate representation of individual
- Formulation of objective function
- Transformation the parameters of individual to the real numbers
- Transformation of unfeasible solutions
- Setting of the control parameters of SOMA

Selection of an appropriate representation of individual. We chosen a natural representation for coding an individual. In response to the vehicle routing problem, each node (city) except center node (depot) is assigned with integer from 1 to n (n represents the number of nodes except depot), which represents corresponding node in individual. Each individual is then represented by n-dimensional vector of integers, representing the sequence of visiting of the nodes. Each individual in the population is also assigned with its fitness that represents total cost (total distance) of the routes.

Formulation of objective function. The computation of objective function value for an individual is carried out in two steps with respect the following facts:

1) The sum of all demands on the route must not exceed the capacity of vehicle
2) The supply nodes are included in the same route only in case of non negative value of savings (based on heuristic Clarke-Wright algorithm).

Transformation the parameters of individual to the real numbers. Because SOMA was originally designed to solve problems with continuous-time variables and the used natural representation consists of integer variables, it is desirable to transform integers to real numbers. The used method for transformation was presented in (Onwubolu & Babu, 2004) for solving traveling salesman problem. Let z_i, i = 1, 2, ...,n represents an integer number. The equivalent continuous variable for z_i is given as:

$$r_i = -1 + \frac{z_i * f * 5}{10^3 - 1} \tag{1}$$

where f is given, for example. f = 200.

Reverse transformation of real numbers to integers (used to evaluate the objective function):

$$z_i = \frac{(1 + r_i) * (10^3 - 1)}{5 * f} \tag{2}$$

$$\alpha_i = \mathrm{int}(z_i + 0,5) \tag{3}$$

$$\beta_i = \alpha_i - z_i \tag{4}$$

Transformation of unfeasible solutions. The use of SOMA for VRP does not require the formation of feasible solution in case of natural representation of individual; therefore it is necessary to choose an appropriate method of transformation of the unfeasible solutions. The problem of infeasibility occurs in two cases:

a) Parameter of individual after the transformation from real numbers to integers is less then 1 or greater then d, in this case the relevant parameter is replaced by new randomly generated parameters in range 1 to d.
b) Created individual does not comprise a permutation of integers d. In this case, the correction approach that was presented in (Brezina et al, 2009) was used.

1) Let **m** is the vector of parameters of the individual dimension of d with k different elements. If d - k = 0, go to step 4). Otherwise, go to step 2).
2) Create the vector **p** (dimension d - k) of random permutation of such d - k elements, which are not included in the vector **m**. If the number of non-zero components of the vector **p** = 0, go to step 4). Otherwise, find the first repeated element of vector **m**. Let this element be m_c and let the first nonzero element of vector **p** be p_k . Set $m_c = p_k$ and go to step 3).
3) Set $pk = 0$ and return to the step 2).
4) Return **m**

Setting of the control parameters of SOMA. Efficiency of SOMA has a slight dependence on the setting of control parameters. Recommended values for the parameters are usually derived empirically from experiments:
d – dimensionality. Number of parameters of individual (usually also number of arguments of objective function).
np – population size. Number of individuals in population. recommended setting is 5d to 30d, respectively 100d, in case the optimized function is multimodal (Zelinka, 2002) (Mařik, 2003).
mig – migrations. Represent the maximum number of iteration (*mig* is also stopping criterion).
mass – path length, $mass \in \langle 1,1;3 \rangle$. Represents how far an individual goes over the search space. Recommended value is 3.
step– $step \in \langle 0,11; mass \rangle$. Defines the granularity with what the search space is sampled. In case of simple optimizing function it is possible to use a large step to speed up search process, but if no apriori information are known, recommended value is 0,11.
prt – perturbation, $prt \in \langle 0,1 \rangle$. Determines whether an individual travel directly over the search space or not. If *prt* is equal to 1, SOMA use only deterministic principles in a search process, if it is equal to 0, SOMA is purely stochastic. *Prt* is one of the most sensitive control parameters.

To discover the effective set of control parameters of SOMA, the free available test data *Eil17* from OR library were used [2] and the 1260 simulations were carried out to determine the effective setting of control parameters *prt.* The best results were evidently obtained with the use of *prt* = 0,7.

3 Experiments

To discover the effectiveness of the presented techniques, the free available test data from OR library *Eil23, Eil30* were used. These instances were investigated by many researchers who applied a variety of techniques to solve it (and also there is a known optimal value of objective function). Based on before mentioned experiments we used a set of control parameters: *np* = 10**d* (where *d* represents number of nodes), *mig* = 5000 and *step* = 0,9, *prt* = 0,7. For each problem ten simulations were done. Results are presented in Table 1.

From the Table 1 it is evidently seen that SOMA was able to find optimal solution for that problems. Based on these results it can be stated that SOMA presents a powerful approach for solving vehicle routing problem.

[2] http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/ (15.9.2010)

Problem	Eil23	Eil 30
Simulation 1	569	548
Simulation 2	570	548
Simulation 3	569	548
Simulation 4	569	539
Simulation 5	570	534
Simulation 6	600	535
Simulation 7	570	589
Simulation 8	570	537
Simulation 9	572	562
Simulation 10	576	589
Optimum	569	534
Minimum	569	534
Mean	573,5	552,9
Perc. deviation	0,79%	3,54%

Table 1 Results for Eil Instances

Conclusion

Vehicle routing problem is one of NP-hard problems. A typical conflict in dealing with this kind of problems arises between the time available for the calculation and the quality of the solution. While the exact methods (eg, branch and bound algorithm) are often able to identify the optimum in unacceptably long time, the quality of solutions obtained by heuristic may be disputable. Nowadays, we follow the increased interest in methods, which are inspired by different biological evolutionary processes in nature. This technology is covered by the common name of "evolutionary algorithms". But their application to constrained problems requires some additional modifications of theirs basic versions. The paper was focused on application of self organizing migrating algorithm to vehicle routing problem. The special factors that involve the use of that algorithm were presented and the efficiency of calculations has been validated on the basis of publicly available instances. Based on presented results it can be concluded that the proposed approach is quite powerful in dealing with vehicle routing problem.

References

[1] Brezina, I.– Čičková, Z.– Gežík, P.– Pekár, J. 2009. *Modelovanie Reverznej Logistiky – Optimalizácia Procesov Recyklácie A Likvidácie Odpadu.* Bratislava: Ekonóm, 2009. 222 S. Isbn 978-80-225-2825-2.

[2] Brezina, I. – Čičková, Z.– Reiff, M. 2009. Kvantitatívne Metódy Na Podporu Logistických Procesov. Bratislava, Ekonóm, 2009. 191 S. ISBN 978-80-225-2648-7.

[3] Čičková, Z – Brezina, I. – P. Pekár. 2008. Alternative Method For Solving Traveling Salesman Problem By Evolutionary Algorithm. In *Management Information Systems : The International Scientific Journal*. - Subotica ; Bratislava. University Of Novi Sad, The Faculty Of Economics : University Of Economics Bratislava, Faculty Of Economic Informatics: 2008. ISSN 1452-774x. - No. 1, Pp. 17-22.

[4] Davendra, D. – Zelinka, I. 2008. Flow Shop Scheduling Using Self Organizing Migration Algorithm. In *22nd European Conference On Modelling And Simulation*. Nicosia: 2008. Isbn 0-9553018-5-8, Pp 195-200.

[5] Onwubolu, G.C. – Babu B.V. 2004. *New Optimization Techniques In Engineering, Studies In Fuzziness And Soft Computing*. Springer, 2004. 712 P. ISBN 3-540-20167-X.

[6] Palúch, S. – Peško, Š. 2006. *Kvantitatívne metódy v logistike*. Žilina: Edis Vydavateľstvo ŽU, 2006. 185 S. ISBN 80-8070-636-0.

[7] Sekaj, I. 2005. *Evolučné výpočty a ich využitie v praxi*. Bratislava: IRIS, 2005. 157 S. ISBN 80-89018-87-4.

[8] Števo, S. 2008. Optimalizácia chodu robota pomocou evolučných metód. In: *Technical Computing Prague* 2008 : 16th Annual Conference Proceedings. Prague, Czech Republic, 11.11.2008. - Praha : Humusoft s.r.o, 2008. - ISBN 978-80-7080-692-0.

[9] Zelinka, I. 2002. *Umělá inteligence v problémech globální optimalizace*. Praha: Ben-Technická Literatúra, 2002, 232 S. ISBN 80-7300-069-5.

The Net Present Value of Investment on Oligopoly Markets

Martin Dlouhý[1], Zuzana Fíglová[2]

Abstract. The investment is usually valued by the expected cash flow that is measured by the net present value (NPV). We show that the NPV does not value investment correctly because it does not include the comprehensive impact of investment on the market. Assuming that the investment will change the position of firm on the market we can make a decomposition of the NPV into various effects. The direct (static) effect of the investment is the change in profit compared to the base-case scenario (no investment is made, constant outputs of both investor and competitor). The strategic effect results from changes in output of firms due to transition of the Cournot oligopoly (base-case scenario) to a new equilibrium. The structural effect of the investment measures an impact of the change in the structure of the market.

Keywords: Net Present Value, Investment, Oligopoly, Game Theory.

JEL Classification: C70; L13
AMS Classification: 91A18

1 Introduction

Game theory is a discipline of mathematical economics that studies situations in which decision-makers (players) interact (Osborne, 2004, Maňas, Dlouhý, 2009). The main idea of game theory is that in making my decision I have to take into account your decision, and vice versa. Theory of oligopoly that studies decisions of firms on the imperfect markets is a part of the game theory that is especially interesting for economists. We will show that strategic interactions between firms on the oligopoly market may significantly change the expected value of the investment project.

A value of the investment can be estimated by various methods: the internal rate of return, the payback period, the net present value, real options and other methods (Scholleová, 2009, Valach, 1997). Probably most frequently, the investment is valued by the expected cash flow that is measured by the net present value (*NPV*). Net present value is the value of discounted cash flow (*DCF*) minus the value of discounted investment expenditures (*DIN*):

$$NPV = DCF - DIN = \sum_t CF_t/(1+r)^t - \sum_t IN_t/(1+r)^t$$

where t is a time period 0, 1,..., T, r is a discount rate, CF_t is a nominal cash flow in period t and IN_t is an investment expenditure in period t. If the *NPV* is positive, the investment project is profitable and should be realized.

In this study we present a broader perspective according to which a strategic investment is valued by the comprehensive impact it has on the oligopoly market as a whole. The key strategic project cannot be considered as separate investment decision, but as a strategic move that changes a position of firm on the oligopoly market. We believe that such framework is potentially very helpful to top managers when strategic investment has to be made.

Most markets are oligopolies with significant entry barriers. In such imperfect markets, the right investment decision at right time can offer large returns and can even lead to a change on the whole market. Thus the *NPV* valuation should be combined with the game theory. From this combined perspective, an investment decision includes not only direct returns from the investment, but also a gain of strategic competitive advantage over rival firms.

2 Model

Economic theory defines an oligopoly as a market structure in which an industry is dominated by a small number of firms (Nicholson, 2002, Samuelson, Nordhaus, 1995). To simplify, we assume that there are only two sellers on the market. This situation is known as a duopoly. Let us have a duopoly market with two firms denoted as 1

[1] University of Economics in Prague, Faculty of Informatics and Statistics, 4 Winston Churchill Sq., 130 67 Prague 3, Czech Republic, e-mail: dlouhy@vse.cz.
[2] University of Economics in Prague, Faculty of Informatics and Statistics, 4 Winston Churchill Sq., 130 67 Prague 3, Czech Republic, e-mail: figlova@vse.cz.

and 2. The outputs of firms are denoted as x_1 and x_2. The cost functions are $C_1(x_1) = c_1x_1$ and $C_2(x_2) = c_2x_2$, and profit functions are $M_1(x_1, x_2)$ and $M_2(x_1, x_2)$. We assume the decreasing price function $p = (D - x_1 - x_2)$. We assume that the market can be described as the Cournot oligopoly. In this model, each firm recognizes that its decision affects the market price, but not the output of the competing firm (e.g. Dlouhý, Fiala, 2009, Fendeková, 2006, Soukup, 2003). The profit functions of firms are defined as follows:

$$M_1(x_1, x_2) = f(x_1, x_2)\, x_1 - C_1(x_1) = (D - x_1 - x_2)\, x_1 - c_1x_1,$$
$$M_2(x_1, x_2) = f(x_1, x_2)\, x_2 - C_2(x_2) = (D - x_1 - x_2)\, x_2 - c_2x_2.$$

The first derivatives of profit functions give the system of equations:

$$\partial M_1(x_1, x_2)/\partial x_1 = 0,$$
$$\partial M_2(x_1, x_2)/\partial x_2 = 0.$$

The equilibrium outputs are then:

$$x_1 = (D - 2c_1 + c_2)/3,$$
$$x_2 = (D + c_1 - 2c_2)/3.$$

If the market structure is best described as the Stackelberg oligopoly, we assume that one firm is a leader and the other is follower. The equilibrium outputs of a leader (firm 1) and a follower (firm 2) are given:

$$x_1 = (D - 2c_1 - c_2)/2,$$
$$x_2 = (D - x_1 - 2c_2)/2.$$

Assuming that the investment will change a position of the firm on the market we can make a decomposition of the net present value into various effects:

1. The *direct (static) effect* of the investment is a change in profit compared to the base-case scenario (no investment is made, constant outputs of both investor and competitor).
2. The *strategic effect* results from changes in output of firms due to transition of the current Cournot oligopoly equilibrium (base-case scenario) to a new equilibrium.
3. The *structural effect* of the investment measures an impact of the change in the structure of the market. In this model, the *structural effect* may express the effect of transformation of the Cournot oligopoly to Stackelberg oligopoly, with firm 1 as a leader and firm 2 as a follower.

3 Illustrative Example

Let us assume the Cournot duopoly. Let us have two firms denoted as firm 1 and firm 2. The outputs are denoted as x_1 and x_2, the cost functions are $C_1(x_1) = 5x_1$ and $C_2(x_2) = 5x_2$. The price function is $p = 20 - x_1 - x_2$. The profit functions are:

$$M_1(x_1, x_2) = (20 - x_1 - x_2)\, x_1 - 5x_1,$$
$$M_2(x_1, x_2) = (20 - x_1 - x_2)\, x_2 - 5x_2.$$

The reaction functions are:

$$x_1 = 7.5 - 0.5x_2,$$
$$x_2 = 7.5 - 0.5x_1.$$

We obtain a following equilibrium solution: the output of the first firm x_1 is 5 and its profit is 25, the output of the second firm x_2 is 5 and its profit is 25, the total output of the industry is 10 and equilibrium price is 10.

Now let us suppose that firm 1 can make an investment that costs 20 in the first year, but in the next four years it will decrease variable cost by 20%. The new cost function of firm 1 will be $C_1(x_1) = 4x_1$. The profit functions will be:

$$M_1(x_1, x_2) = (20 - x_1 - x_2)\, x_1 - 4x_1,$$
$$M_2(x_1, x_2) = (20 - x_1 - x_2)\, x_2 - 5x_2.$$

We obtain a following equilibrium: the output of the first firm x_1 is 17/3 and its profit is 32.1, the output of the second firm x_2 is 14/3 and its profit is 21.8, the total output of the industry is 31/3 and the equilibrium price is 29/3.

Finally, let us suppose that the investment that costs 20 in the first year will lead to a new market situation. The firm 1 will become a leader in the Stackelberg oligopoly. In this case, the output of the first firm x_1 is 8.5 and its profit is 36.1 the output of the second firm x_2 is 3.25 and its profit is 10.6, the total output of the industry is 11.75 and the equilibrium price is 8.25.

The net present value of the investment under these three models is summarised in Table 1. We assume the discount rate 5%. The net present value of the investment estimated by the traditional net present value method is negative, which means that the top management should not realize this investment. However, if we take into account the new equilibrium on the Cournot oligopoly market, the net present value of the investment is positive. In case of the Stackelberg oligopoly, the net present value is even higher.

Period	1. Investment with no change in output	2. New equilibrium in the Cournot oligopoly	3. The Leader in the Stackelberg oligopoly
0	–20.00	–20.00	–20.00
1	4.76	6.77	10.60
2	4.54	6.45	10.09
3	4.32	6.14	9.61
4	4.11	5.85	9.15
Net Present Value	**–2.27**	**5.22**	**19.45**

Table 1 Net Present Value of Investment under Alternatives Assumptions

The decomposition of the net present value in case of the Stackelberg oligopoly is described in Table 2. The direct (static) effect of the investment is -2.27. The strategic effect is a difference between the first and second column in Table 1 (5.22–2.27). The structural effect of the investment is a difference between the second and third column (19.45–5.22).

Type of effect	Value
Direct (static) effect	–2.27
Strategic effect	7.49
Structural effect	14.23
Total net present value	19.45

Table 2 The Decomposition of the Net Present Value

The numerical example shows that the net present value of the investment project has to take into account a change on the oligopoly market initiated by the investment. With different assumptions we have obtained different decisions whether to invest or not. One can imagine many other scenarios that can be investigated, for example:

- The investment will be successful only with certain probability, for example with probability 80%.
- If the investment is successful, the competing firm will make the same investment next year.
- If the investment is successful, the firm becomes a monopoly.

- The cash flows and/or investment expenditures are stochastic.
- The growth or decline of the market is expected.

4 Two-Stage Game

Smit and Trigeorgis (2004) describe an example of investment decision on the duopoly market as a two-stage game. A pioneering firm 1 can make or not the first-stage strategic R&D investment. At the first stage of the game, the firm 1 faces uncertainty about future market demand (high or low demand). A committing strategy at the first stage is risky, however, it represents a strategic advantage from which firm 1 can benefit at next stage of the game, At the second stage, the market opens and both firms can invest in production capacity. If one firm invests, the firm will act as a Stackelberg leader or even as a monopolist. When both firms invest or both firms do not invest at the second stage, the game ends in a Cournot duopoly. In their model of two-stage game, the game theory and theory of real options are combined.

To measure a comprehensive impact the investment, Smit and Trigeorgis (2004) define the expanded version of the net present value that is decomposed into four effects. First, Smit and Trigeorgis distinguish between *commitment effect* from making investment earlier and *flexibility value effect* from postponing irreversible investment. Second, the total commitment effect can be broken down into a *direct effect* reflecting reduction in operating cost, a *strategic reaction value* reflecting the impact of competitor's reaction, and a *strategic pre-emption value* from deterring competitive entry and leading to a change in the market structure. Hence, their model is even more complex than the model described in section 2.

5 Conclusion

As we showed in the illustrative example that net present value in its simple form may not measure the value of the investment properly. The expanded version of the net present value can be divided into three different effects. The direct (static) effect of the investment is the change in profit compared to the base-case (no investment), with constant outputs (both investor and competitor). The strategic effect results from changes in output of firms due to transition of the Cournot oligopoly to a new equilibrium. The structural effect of the investment measures the impact of the change in the structure of the market. In the numerical example above, the structural effect will measure the impact of transformation of the Cournot oligopoly to Stackelberg oligopoly with firm 1 being a leader and firm 2 being a follower. In the two-stage game, Smit and Trigeorgis (2004) defined the expanded version of the net present value that is decomposed even into four effects.

In the paper, we showed that the net present value as a criterion for strategic investment decisions is not able to value correctly investment projects on the oligopoly markets. We argue that the top management has to take into account the impact of investment decision on the whole oligopoly market. The calculations presented here can be applied also in case of other methods investment valuation with minimal changes.

Acknowledgements

The research was funded by the Internal Grant Agency of the University of Economics in Prague, project no. F4/16/2011 Models of Operational and Financial Management.

References

[1] Dlouhý, M., Fiala, P.: *Úvod do teorie her*. Second Edition. Prague: Nakladatelství Oeconomica, 2009.

[2] Fendeková, E.: *Oligopoly a regulované monopoly*. Bratislava: IURA EDITION, 2006.

[3] Fiala, P., Dlouhý, M.: *Základy kvanitativní ekonomie a ekonomické analýzy*. Prague: Nakladatelství Oeconomica, 2006.

[4] Holman, R.: *Ekonomie*. 3rd Edition. Prague: C. H. Beck, 2002.

[5] Maňas, M., Dlouhý, M.: *Games and Economic Decisions*. 4th Edition. Prague: Nakladatelství Oeconomica, 2009.

[6] Nicholson, W.: *Microeconomic Theory: Basic Principles and Extensions.* Eighth Edition. South Western, 2002.

[7] Osborne, M.J.: *An Introduction to Game Theory*. New York: Oxford University Press, 2004.

[8] Samuelson, P. A., Nordhaus, W.D.: *Economics*. Fifteenth Edition. New York: McGraw-Hill, 1995.

[9] Scholleová, H.: *Investiční controlling: Jak hodnotit investiční záměry a řídit podnikové investice*. Prague: Grada, 2009.

[10] Smit, H.T.J, Trigeorgis, T.: *Strategic Investment: Real Options and Games*. Princeton, Princeton University Press, 2004.

[11] Soukup, J.: *Mikroekonomická analýza*. Slaný: Melandrium, 2003.

[12] Valach, J. a kol.: *Finanční řízení podniku*. Prague: EKOPRESS, 1997.

Cost Analysis of Finite Multi-server Markov Queueing System Subject to Breakdowns

Michal Dorda[1], Dušan Teichmann[2]

Abstract. In the paper we introduce a mathematical model of a finite multi-server Markov queueing system M/M/*n*/*m* with servers which are subject to breakdowns. We assume that broken server is being repaired by a single repairman, where the number of repairmen is less than the number of the system servers. The system is modelled as a two-dimensional Markov process presented by a state transition diagram and a finite system of linear equations that describes the behaviour of the system in steady state. The steady state probabilities are obtained by numerical solving by means of Matlab.

On the basis of the steady state probabilities we can compute several performance measures. The mathematical model is supported by a coloured Petri net model in order to validate the analytical outcomes. Further we present a cost function which can serve as an optimization criterion for the optimization of the studied system parameters.

Keywords: Unreliable M/M/*n*/*m*, cost analysis, coloured petri net.

JEL Classification: C44
AMS Classification: 60K25

1 Introduction

The server that is working without failures is usually assumed in the queueing theory. But in many practical cases this assumption is not correct; servers are often technical devices and every technical device can be broken. Many authors studied the behaviour of single server queueing systems subject to breakdowns under diverse assumptions. Modelling of multi-server queueing systems is not so often done due to the mathematical complexity of their analysis.

Mitrani and Avi-Itzak [1] investigated a multi-server Markov queueing model with an infinite queue capacity, server breakdowns and an ample repair capacity. Neuts and Lucantoni [2] and Wartenhorst [3] considered a multi-server Markov queueing model with an infinite queue capacity and server breakdowns as well, but with a limited repair capacity. Some experimental outcomes related to the behaviour of an unreliable multi-server Markov queueing system were presented by Mitrani and King in paper [4]. Queueing systems with two unreliable servers were studied by Madan, Abu-Dayyeh and Gharaibeh [5] and by Yue D., Yue W., Yu and Tian [6]. In paper [5] the authors assumed homogeneous servers, in paper [6] heterogeneous servers were considered. Wang and Chang [7] studied a finite Markov queueing system with balking, reneging and server breakdowns and introduced a cost analysis of it.

2 Assumptions of the model

Let us consider a finite Markov multi-server queueing system consisting of n homogenous parallel placed servers subject to breakdowns. Incoming customers can wait for the service in a basic queue with a finite capacity equal to m-n, customers are served one by one according to FIFO (First In – First Out) discipline. Thus there are in total m places in the queueing system.

Customers come to the system according to the Poisson process with the parameter λ, the customers service times follow the exponential distribution with parameter μ.

Each of the servers is successively failure-free and broken; let us assume that failures of the servers are mutually independent. Furthermore let us consider that a server breakdown can occur at any time. That means the server can break if it is busy or idle. Time of failure-free state is an exponential random variable with the parameter η. The time to repair is an exponential random variable as well, but with the parameter ζ. Let us assume

[1] VŠB – Technical University of Ostrava /Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, tel. number: +420 597 325 754, e-mail: michal.dorda@vsb.cz.
[2] VŠB – Technical University of Ostrava /Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava – Poruba, tel. number: +420 597 324 575, e-mail: dusan.teichmann@vsb.cz.

there are r repairmen, where $r<n$; broken server is being repaired by one of the free repairmen. If there is no free repairman, broken server must wait for its repair in the queue of broken servers.

And finally let us assume the so called preemptive-repeat discipline that means the service of the customer is interrupted after the occurrence of the server breakdown (if the server is busy at this moment) and its service starts from the beginning. Generally one of the two different events can occur:

- If there is an idle server in the system, the customer will go immediately over to the idle server.
- If there is not an idle server in the system, the customer goes back to the queue. We assume the customer which service has been interrupted always finds a free place in the queue. If there is a free place in the basic queue, the costumer goes back to the basic queue. If there is no free place in the basic queue, he goes to an extra queue designated for waiting such type of customers. Since we have 4 servers, we must consider up to 4 extra places in the queue; therefore the maximum capacity of both queues together is equal to m. Further we assume that the parameter λ is zero when there are at least 4 customers in the queue that means the basic queue is full.

3 Mathematical model

Let us consider a random variable X(t) being the number of broken servers and a random variable Y(t) being the number of the customers finding in the system at the time t. On the basis of the assumptions established in Section 2 it is clear that {X(t), Y(t)} is a two-dimensional Markov process with the state space

$$\Omega = \{(i,j), i = 0,1,\ldots,n;\ j = 0,1,\ldots,m\}.$$

The system is found in the state (i,j) at the time t if X(t)=i and Y(t)=j, let us denote the corresponding probability $P_{(i,j)}(t)$.

Let us illustrate the queueing model graphically as a state transition diagram (see figure 1). The vertices represent the particular states of the system and oriented edges indicate the possible transitions with the corresponding rate. Please notice that in figure 1 there are depicted only selected states that are necessary for set-up of the equation system.

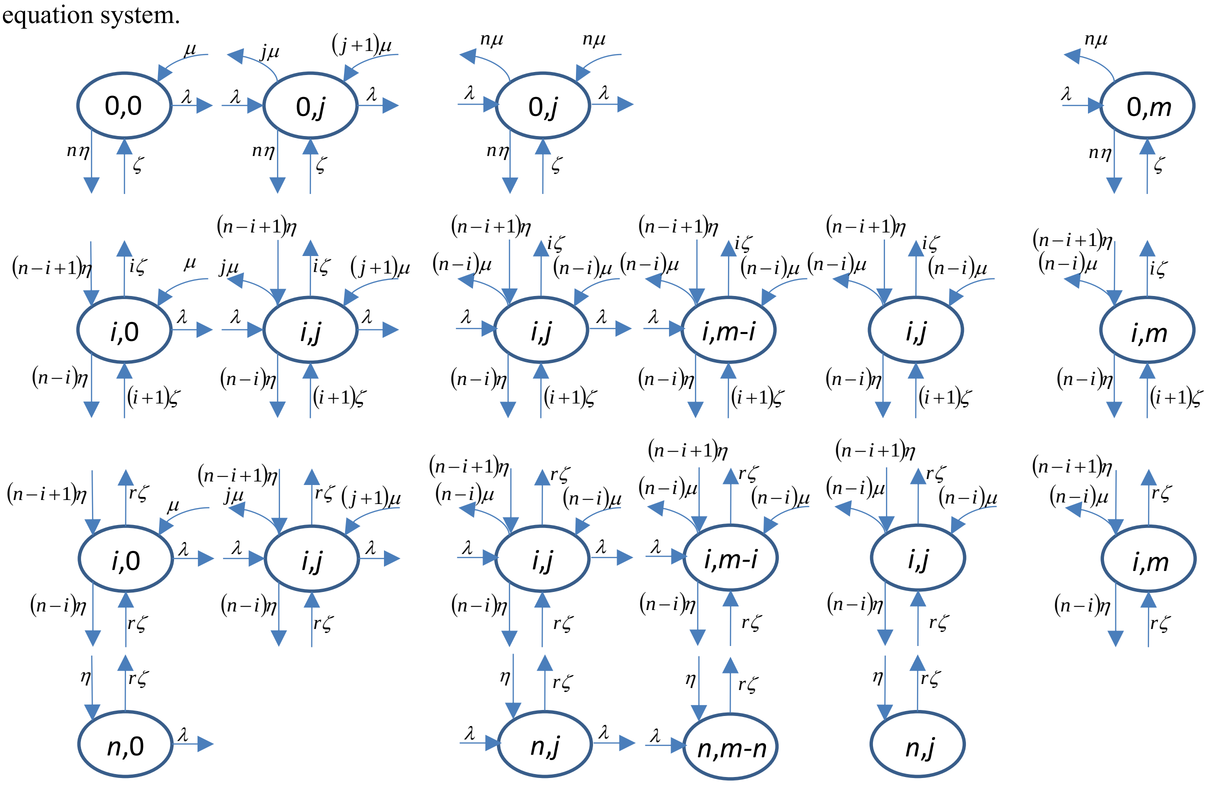


Figure 1 The state transition diagram

On the basis of the state transition diagram we can obtain a finite system of the differential equations for the probabilities $P_{(i,j)}(t)$ depending on the time t. For $t\to\infty$ we get the system of the linear equations for steady state probabilities $P_{(i,j)}$ that are not dependent on the time t. The steady state balance equations are:

$$(\lambda + n\eta)P_{(0,0)} = \mu P_{(0,1)} + \zeta P_{(1,0)}, \tag{1}$$

$$(\lambda + j\mu + n\eta)P_{(0,j)} = \lambda P_{(0,j-1)} + (j+1)\mu P_{(0,j+1)} + \zeta P_{(1,j)} \text{ for } j = 1,\dots,n-1, \tag{2}$$

$$(\lambda + n\mu + n\eta)P_{(0,j)} = \lambda P_{(0,j-1)} + n\mu P_{(0,j+1)} + \zeta P_{(1,j)} \text{ for } j = n,\dots,m-1, \tag{3}$$

$$(n\mu + n\eta)P_{(0,m)} = \lambda P_{(0,m-1)} + \zeta P_{(1,m)}, \tag{4}$$

$$[\lambda + (n-i)\eta + i\zeta]P_{(i,0)} = \mu P_{(i,1)} + (n-i+1)\eta P_{(i-1,0)} + (i+1)\zeta P_{(i+1,0)} \text{ for } i = 1,\dots,r-1, \tag{5}$$

$$[\lambda + j\mu + (n-i)\eta + i\zeta]P_{(i,j)} = \lambda P_{(i,j-1)} + (j+1)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + (i+1)\zeta P_{(i+1,j)} \\ \text{for } i = 1,\dots,r-1 \text{ and } j = 1,\dots,n-i-1, \tag{6}$$

$$[\lambda + (n-i)\mu + (n-i)\eta + i\zeta]P_{(i,j)} = \lambda P_{(i,j-1)} + (n-i)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + (i+1)\zeta P_{(i+1,j)} \\ \text{for } i = 1,\dots,r-1 \text{ and } j = n-i,..m-i-1, \tag{7}$$

$$[(n-i)\mu + (n-i)\eta + i\zeta]P_{(i,m-i)} = \lambda P_{(i,m-i-1)} + (n-i)\mu P_{(i,m-i+1)} + (n-i+1)\eta P_{(i-1,m-i)} + (i+1)\zeta P_{(i+1,m-i)} \\ \text{for } i = 1,\dots,r-1, \tag{8}$$

$$[(n-i)\mu + (n-i)\eta + i\zeta]P_{(i,j)} = (n-i)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + (i+1)\zeta P_{(i+1,j)} \\ \text{for } i = 1,\dots,r-1 \text{ and } j = m-i+1,..m-1, \tag{9}$$

$$[(n-i)\mu + (n-i)\eta + i\zeta]P_{(i,m)} = (n-i+1)\eta P_{(i-1,m)} + (i+1)\zeta P_{(i+1,m)} \text{ for } i = 1,\dots,r-1, \tag{10}$$

$$[\lambda + (n-i)\eta + r\zeta]P_{(i,0)} = \mu P_{(i,1)} + (n-i+1)\eta P_{(i-1,0)} + r\zeta P_{(i+1,0)} \text{ for } i = r,\dots,n-1, \tag{11}$$

$$[\lambda + j\mu + (n-i)\eta + r\zeta]P_{(i,j)} = \lambda P_{(i,j-1)} + (j+1)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + r\zeta P_{(i+1,j)} \\ \text{for } i = r,\dots,n-1 \text{ and } j = 1,\dots,n-i-1, \tag{12}$$

$$[\lambda + (n-i)\mu + (n-i)\eta + r\zeta]P_{(i,j)} = \lambda P_{(i,j-1)} + (n-i)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + r\zeta P_{(i+1,j)} \\ \text{for } i = r,\dots,n-1 \text{ and } j = n-i,..m-i-1, \tag{13}$$

$$[(n-i)\mu + (n-i)\eta + r\zeta]P_{(i,m-i)} = \lambda P_{(i,m-i-1)} + (n-i)\mu P_{(i,m-i+1)} + (n-i+1)\eta P_{(i-1,m-i)} + r\zeta P_{(i+1,m-i)} \\ \text{for } i = r,\dots,n-1, \tag{14}$$

$$[(n-i)\mu + (n-i)\eta + r\zeta]P_{(i,j)} = (n-i)\mu P_{(i,j+1)} + (n-i+1)\eta P_{(i-1,j)} + r\zeta P_{(i+1,j)} \\ \text{for } i = r,\dots,n-1 \text{ and } j = m-i+1,..m-1, \tag{15}$$

$$[(n-i)\mu + (n-i)\eta + r\zeta]P_{(i,m)} = (n-i+1)\eta P_{(i-1,m)} + r\zeta P_{(i+1,m)} \\ \text{for } i = r,\dots,n-1, \tag{16}$$

$$(\lambda + r\zeta)P_{(n,0)} = \eta P_{(n-1,0)}, \tag{17}$$

$$(\lambda + r\zeta)P_{(n,j)} = \lambda P_{(n,j-1)} + \eta P_{(n-1,j)} \text{ for } j = 1,\dots,m-n-1, \tag{18}$$

$$r\zeta P_{(n,m-n)} = \lambda P_{(n,m-n-1)} + \eta P_{(n-1,m-n)}, \tag{19}$$

$$r\zeta P_{(n,j)} = \eta P_{(n-1,j)} \text{ for } j = m-n+1,\dots,m. \tag{20}$$

Clearly, the probabilities $P_{(i,j)}$ must satisfy the normalization equation:

$$\sum_{i=0}^{n}\sum_{j=0}^{m} P_{(i,j)} = 1 . \tag{21}$$

By solving the linear equations system obtained from equations (1) – (21) we get steady state probabilities of the particular states of the system that are needed for performance measures computing. On the basis of steady state probabilities the following performance measures can be computed. Please notice that we present only selected performance measures that we will use for the cost analysis of the studied queue. The mean number of the costumers in the service, denoted as *ES*, can be computed according to formula:

$$ES = \sum_{i=0}^{n-1}\sum_{j=1}^{n-i} jP_{(i,j)} + \sum_{i=0}^{n-1}(n-i) \sum_{j=n-i+1}^{m} P_{(i,j)} . \tag{22}$$

The mean number of the waiting customers *EL* can be expressed by formula:

$$EL = \sum_{i=0}^{n} \sum_{j=n-i+1}^{m} (j-n+i) P_{(i,j)} . \tag{23}$$

For the mean number of the broken servers *EP* it can be written:

$$EP = \sum_{i=1}^{n} i \sum_{j=0}^{m} P_{(i,j)} . \tag{24}$$

The probability that the incoming customer will be rejected is equal to:

$$P_{ODM} = \sum_{i=0}^{n} \sum_{j=m-i}^{m} P_{(i,j)} . \tag{25}$$

It is clear that the mean number of the idle servers *EI* can be computed:

$$EI = n - ES - EP . \tag{26}$$

Now we develop a steady state cost function which can serve as a criterion for the optimization of the system parameters. Let us establish:

- *P* as profit per one served customer,
- C_1 as cost per unit time when one customer is waiting in the queue,
- C_2 as cost per unit time when one server is idle,
- C_3 as cost per unit time when one server is busy,
- C_4 as cost per unit time when one server is broken,
- C_5 as cost per unit time and one repairman (time wage).

Then the cost function C(*x*) based on the difference between profit and costs of the system can be expressed by the formula:

$$C(x) = P \cdot (1 - P_{ODM}) - C_1 \cdot EL - C_2 \cdot EI - C_3 \cdot ES - C_4 \cdot EP - C_5 \cdot r \rightarrow \max , \tag{27}$$

where *x* is the parameter which we want to optimize.

4 Simulation model

In order to validate the outcomes, which were reached by solution of the above-mentioned mathematical model, Petri net model of the studied queueing system was created by using CPN Tools – Version 3.0.4. The software CPN Tools is designed for editing, simulating and analyzing coloured Petri nets. The created simulation model in initial marking is shown in figure 2. The model is compound of 16 places and 12 transitions. The Petri net presented in figure 2 models the unreliable M/M/4/8 (thus n=4, m=8) queueing system fulfilling the conditions mentioned in Section 2. In figure 2 the applied values of individual parameters are λ=10 h^{-1}, μ=3 h^{-1}, η=0,01 h^{-1}, ζ=0,1 h^{-1} and r=3.

The concrete values of the random variables are generated during the simulation through the defined function fun ET(EX) = round(exponential(1.0/EX)). We apply a minute as the unit of time and the parameter of the defined function is equal to the mean value of the exponential distribution. So the mean value of each random variable must be expressed in [min].

The created Petri net works with following tokens:

- tokens C represent incoming customers and customers waiting for the service in the basic queue,
- tokens CI represent customers waiting for the service in the extra queue,
- tokens (C,i), where $i \in \{1,2,3,4\}$, represent customers as well, but in the phase of the service process (each customer in the service is labeled by a number of the server where is serviced); these tokens are used for modelling available and working servers as well,
- tokens (F,i), where $i \in \{1,2,3,4\}$, model failures of the servers,
- auxiliary tokens P serve for modelling for example free queue places and unavailable servers.

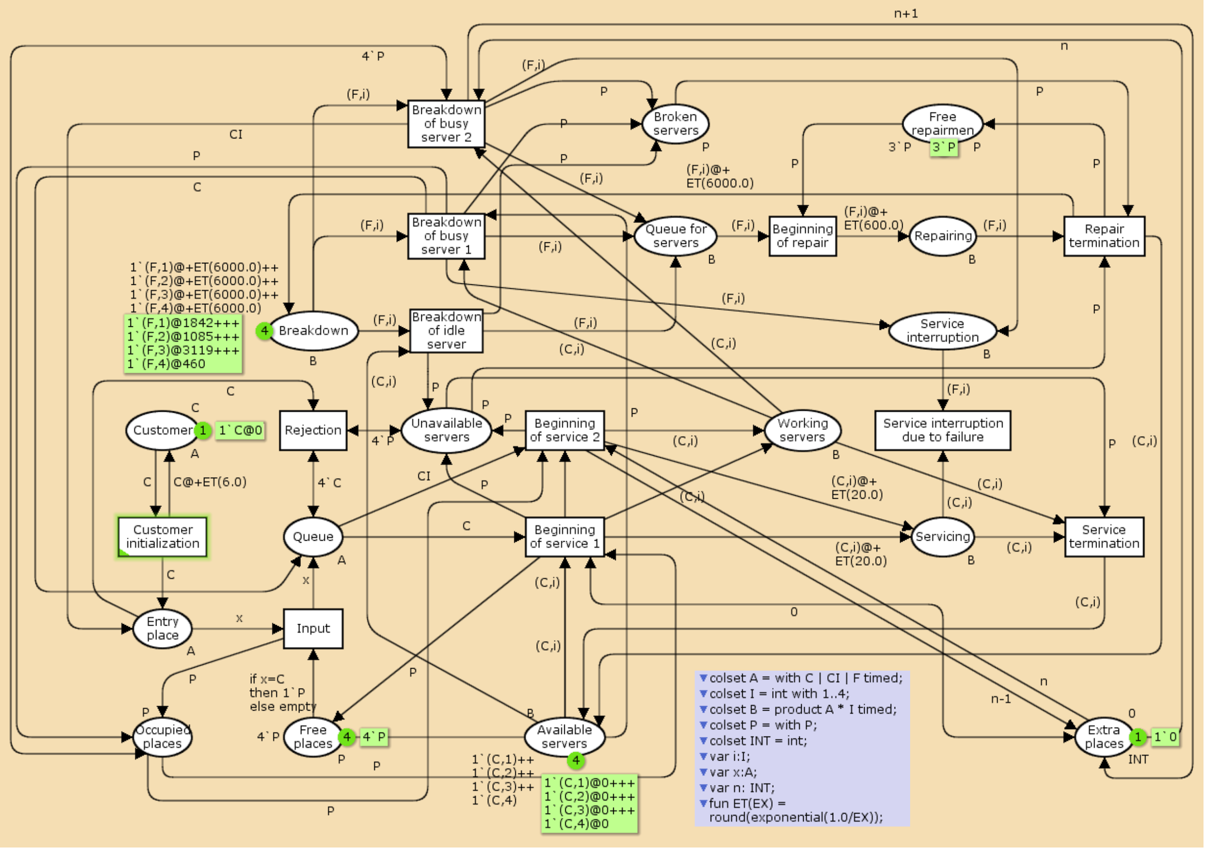


Figure 2 The simulation model (CPN Tools Version 3.0.4)

5 Evaluation of executed experiments

Let us consider the studied queuing system with 4 parallel servers (n=4) and in total 8 places in the system (m=8), therefore the basic queue capacity is equal to 4 (with up to 4 extra places for waiting of the customers which service has been interrupted). Let us consider the applied values of the random variables parameters being λ=10 h^{-1}, μ=3 h^{-1}, η=0,01 h^{-1} and ζ=0,1 h^{-1}, the number of repairmen r can be equal to 1, 2 or 3. In the table 1 there are shown the applied values of the coefficients for the cost analysis.

P [Kč/cust.]	C_1 [Kč/h·cust.]	C_2 [Kč/h·serv.]	C_3 [Kč/h·serv.]	C_4 [Kč/h·serv.]	C_5 [Kč/h·rman.]
100	50	50	100	50	100

Table 1 The applied coefficients for the cost analysis

The steady state probabilities were obtained by the solution of the linear equations system introduced in Section 3, the equations system was solved numerically by using the software Matlab. On the basis of steady state probabilities knowledge we can compute the performance measures according to the formulas (22) up to (25).

The experimental estimation of the performance measures we got by simulation of the coloured Petri net presented in Section 4. Ten independent experiments were executed for each model configuration; each experiment was terminated after a million steps (a step corresponds to a transition firing). The reached outcomes are shown in table 2. As we can see, there are no essential differences between analytic and simulation outcomes; please notice that more accurate simulation outcomes we would probably obtain by using lower unit of time, for example a second. The cost function takes the maximum value for r=1.

r	**Performance measure**	**Analytic result**	**Simulation**			**Difference [%]**	**C(*r*)**
			Average	**Standard deviation**	**95% confidence interval**		
3	*ES*	2,93108	2,93407	0,01043	(2,92661; 2,94153)	-0,10219	172,6013
	EL	1,20335	1,21330	0,01722	(1,20098; 1,22561)	-0,82672	
	EP	0,36372	0,36043	0,01978	(0,34629; 0,37458)	0,90338	
	P_{ODM}	0,12068	0,12087	0,00237	(0,11836; 0,12175)	0,51952	
2	*ES*	2,92811	2,93245	0,00666	(2,92768; 2,93721)	-0,14825	271,677
	EL	1,20699	1,21846	0,01202	(1,20986; 1,22706)	-0,95047	
	EP	0,36773	0,36633	0,01083	(0,35859; 0,37408)	0,37892	
	P_{ODM}	0,12157	0,12087	0,00207	(0,11938; 0,12235)	0,57746	
1	*ES*	2,86367	2,86449	0,01686	(2,85243; 2,87655)	-0,02849	350,8573
	EL	1,30123	1,31338	0,02308	(1,29687; 1,32989)	-0,93408	
	EP	0,46663	0,46837	0,02540	(0,45020; 0,48654)	-0,37220	
	P_{ODM}	0,14090	0,14119	0,00394	(0,13838; 0,14401)	-0,20857	

Table 2 The reached outcomes and their comparison

6 Conclusions

There were presented two models of the M/M/n/m queueing system with the servers subject to breakdowns and a repair capacity $r<n$ in this paper – the mathematical model and the simulation model created as coloured Petri net by using CPN Tools. The major part of the paper was focused on the mathematical model of the studied system. We presented the cost function based on the difference between profit and costs of the queueing system. At the end of the paper reached outcomes were evaluated for the exemplary unreliable M/M/4/8 queueing system with the concrete parameters of the random variables. For our example system we will make the biggest profit with only 1 repairmen; this is caused by the fact that there are no substantial differences in the performance measures.

References

[1] Mitrani, I. L., Avi-Itzhak, B.: A many-server queue with service interruptions. *Operations Research 16* (1968), 628–638.

[2] Neuts, M. F., Lucantoni, D. M.: A Markovian queue with *N* servers subject to breakdowns and repairs. *Management Science 25* (1979), 849–861.

[3] Wartenhorst, P.: *N* parallel queueing systems with server breakdown and repair. *European Journal of Operational Research 82* (1995), 302–322.

[4] Mitrani, I., King, J. B.: Multiserver systems subject to breakdowns: an empirical study. *IEEE Transactions on Computers C-32:1* (1983), 96–98.

[5] Madan, K. C., Abu-Deyyeh, W., Gharaibeh, M.: On two parallel servers with random breakdowns. *Soochow Journal of Mathematics 29:4* (2003), 413–423.

[6] Yue, D., Yue, W., Yu, J., Tian, R.: A heterogeneous two-server queueing system with balking and server breakdowns. In: *Proceedings of the Eighth International Symposium on Operations Research and Its Applications*, Zhangjiajie, China, 2009, 230–244.

[7] Wang, K., Chang, Y.: Cost analysis of a finite M/M/R queueing system with balking, reneging, and server breakdowns. *Mathematical Methods of Operations Research 56* (2002), 169–180.

Column generation method for the vehicle routing problem – the case study

Jan Fábry[1], Václav Kořenář[2], Maria Kobzareva[3]

Abstract. Vehicle routing problem is frequently solved problem in logistic management and the optimal solution to the problem can substantially decrease the transportation costs. Because of the NP hardness of the problem there is no polynomial method available for its solution. It is observed in practice that the optimal solution cannot be obtained in the reasonable time with use of the branch-and-bound method. One way to solve the problem is the column generation method. In the paper we use this method for solving the case study vehicle routing problem with a set of identical vehicles with the given capacity. There is a set of customers with given demand and the transportation cost matrix. The goal is to find a set of routes to cover all demands with minimal costs. The method stepwise generates columns (routes) of the set covering problem. Two problems and models are solved by turns - the cover problem with a reduced set of routes and the problem which generates a new route to decrease the transportation costs of the cover problem. The computation experiences are shown in the paper.

Keywords: column generation method, case study, vehicle routing problem.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction – vehicle routing problem and description of case study

Transportation costs are a great deal of total production and distribution costs. Therefore, each tool designed to reduce transportation costs increases enterprise efficiency and optimization of logistics chains. Nowadays, the optimization of pick-up and delivery is wide-spread problem in practice. Mathematical models for logistic problems enable enterprises to decrease transportation costs without high investments into research.

One of the most recent problems is traveling salesman problem (TSP) and its modification known as vehicle routing problem (VRP). Its basic form is the following. There is a set of n nodes, denoted as $\{1,2,...,n\}$, where node 1 corresponds to the depot, in which the set of vehicles, each with capacity V, is situated. Each node has a demand of goods to be delivered from depot. Let the demand of node i be denoted as q_i. The aim is to find a set of routes, each of which starts and ends at the depot, which will satisfy the demand of all nodes. There are several mathematical models that use integer linear programming designed to solve this problem. Because VRP belongs to the class of NP hard problems we face computational difficulties.

Vehicle routing problem and traveling salesman problem have a lot of modifications, since there are a lot of further operational conditions for route generation. If each node has a time interval, inside which the vehicle has to arrive to the node and unload the shipment, the problem is called vehicle routing problem (or traveling salesman problem) with time windows. These constraints limit the time that vehicle travels on routes (Pelikán, 2006). In some problems there are several depots and/or more vehicles with different capacities and transportation costs.

[1] Jan Fábry, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, fabry@vse.cz

[2] Václav Kořenář, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, korenar@vse.cz

[3] Maria Kobzareva, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, student, mariakobzareva@gmail.com

2 Column generation method

Column generation method solves linear programming problem with a large number of variables, columns of simplex table. The main idea of this method is to solve the reduced problem instead of primary problem, which contains only particular part of variables instead of all variables in the primary problem. Thus, in simplex table the great number of columns is missing. The optimal solution to the reduced problem obviously does not need to match the optimal solution to the primary problem. If we checked the fact that after adding the rest of columns to the reduced problem, the optimal solution to the reduced problem would remain the optimal solution to the primary problem, then the optimal solution to the reduced problem would be the optimal solution to the primary problem. Thus, we should prove that all *z*-coefficients (reduced costs) for the missing columns satisfy the optimality test. This method is used in many problems, not only in logistic ones (Pelikán, 2005 and 2011).

The main question of the method is to find *z*-coefficients for columns, which are not included in the reduced problem. If we have the optimal solution to the reduced problem including the dual solution π, than for *z*-coefficients (both included in and excluded from the reduced problem) the following relation is valid:

$$z_j = \pi . a_j - c_j \, , \tag{1}$$

where π is dual solution, a_j denotes coefficients of *j*-th column in the mathematical model and c_j is the cost of *j*-th variable in the model. If we find, according to (1), such *z*-coefficient that violates the optimalization test, then the optimal solution to the reduced problem is not the optimal solution to the primary problem and it is necessary to add the current column to the reduced problem and to resolve it.

3 Set covering algorithm for vehicle routing problem

On the contrary to the standard model of this problem the vehicle routing problem is formulated as a problem of covering the demand with the subset of all routes. Let *R* denote the set of all routes, which satisfy the vehicle capacity constraint, i.e. the total demand of all nodes visited on a route does not exceed the vehicle capacity. Let R' denote the subset of *R*.

***CG* algorithm** (Bramel and Simchi-Levi, 2002)**:**

Step 1: generate an initial set of columns R' ,

Step 2: solve reduced problem P' (see below); as the result we get optimal values of primal variables *p* and optimal values of dual variables π,

Step 3: solve problem *CG* (see below); if the resulting route *r* satisfies the condition $z_r < 0$ add the column *r* to R' (i.e. $R' := R' \cup \{r\}$) and go to Step 2, otherwise stop, *p* is the optimal solution to P' . If *p* is binary, it is also optimal solution to VRP, otherwise, the value *dc* is the lower bound on the optimal solution to VRP.

Note.

If optimal solution *p* to problem P' is binary then this vector from set of variables R' selects the optimal set of routes, i.e. the optimal solution to problem VRP with the optimal objective value *dc*. If the optimal solution *p* is not binary then *dc* is only the lower bound on the objective value in VRP and in case we added binary condition for *p* in problem P' then we would get suboptimal solution.

Problem P' :

$$dc = \min \sum_{r \in R'} d_r p_r \, , \tag{2}$$

$$\sum_{r \in R'} a_{ir} p_r \geq 1, \; i = 1,2,\ldots,n \, , \tag{3}$$

$$p_r \geq 0, \; r \in R', \tag{4}$$

where

- p_r is a binary variable, if $p_r = 1$ then route *r* is selected, 0 otherwise,

- d_r is the length of route r, counted as $\sum_{i,j} c_{ij} x_{ij}$, where x_{ij} is the optimal solution to problem *CG* and c_{ij} is the value of arc (i, j),
- a_{ir} is a binary constant, which is equal to 1 if route r contains node i, 0 otherwise,
- dc is an objective function which is equal to the sum of lengths of all routes.

Problem *CG*:

$$z_r = \min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} - \sum_{i=1}^{n} \pi_i y_i , \tag{5}$$

$$\sum_{j=1}^{n} x_{ij} = y_i ,\ i = 1,2,\ldots,n , \tag{6}$$

$$\sum_{j=1}^{n} x_{ji} = y_i ,\ i = 1,2,\ldots,n , \tag{7}$$

$$\sum_{i=1}^{n} q_i y_i \le V , \tag{8}$$

$$x_{ij}, y_i \in \{0,1\}, \tag{9}$$

where

- z_r is the optimal objective value,
- c_{ij} is the value of arc (i, j),
- x_{ij} is a binary variable, which is equal to 1 if arc (i, j) is included in a route, 0 otherwise,
- y_i is a binary variable, which is equal to 1 if node i is included in a route, 0 otherwise,
- π_i is a optimal value of dual variable to the optimal solution to P',
- q_i is the demand of node i,
- V is the vehicle's capacity.

4 Solving the case study using set covering algorithm

First we generate a set of initial routes so that the problem has a feasible solution. For this purpose, we used the heuristic algorithm, concretely Fletcher Clark savings algorithm. The outcome are routes in Table 1, where first eight routes form the initial set of routes, other routes had been generated by problem *CG* based on the dual solutions to problem P'. Each route is associated with length d_r and value z_r, which, in case it is positive, means that the route must be included in the set of routes in problem P'. Then, problem P' is resolved with a new set of routes; if the value of z_r is non-positive, algorithm stops.

Table 1 contains a list of routes, which were generated with *CG* algorithm, their length and the objective values of the optimal solution of problem P'.

Table 2 contains dual values obtained by solution of problem P', then the objective value, i.e. the upper bound on the optimal objective value of VRP; in case the optimal solution to P' is binary, then this solution is optimal solution to VRP; otherwise, the optimal solution can be found using different approaches, such as cutting plane method, branch and bound algorithm or branch and price algorithm. Cutting plane method does not guarantee the optimal solution to VRP, while branch and bound and branch and price algorithms do.

Table 3 contains a list of routes and their length, which were generated by branch and bound algorithm which was used in the case study to find the optimal solution to VRP.

Route	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	d_r	z_r
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	10	0,5
2	0	0	1	0	0	1	0	0	0	0	1	0	0	0	184	0
3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	66	1
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	88	1
5	0	0	0	0	0	0	1	1	0	0	0	0	0	0	230	0,5
6	0	0	0	0	0	0	0	0	1	0	0	0	1	0	236	0
7	0	0	0	0	0	0	0	0	0	1	0	1	0	0	105	0,5
8	0	0	0	0	0	0	0	0	0	0	0	0	0	1	54	1
9	0	0	1	0	0	0	0	0	1	0	0	0	0	0	225	0
10	0	0	0	0	0	1	0	0	0	1	0	0	1	0	229	0
11	0	0	0	0	0	0	0	0	0	0	0	1	1	0	54	0
12	0	0	0	0	0	0	0	0	0	0	1	0	1	0	142	0
13	0	0	0	0	0	1	0	0	1	0	0	0	0	0	270	0,5
14	0	0	1	0	0	0	0	0	0	1	0	0	1	0	139	0
15	0	0	0	0	0	0	0	0	0	1	0	0	1	0	104	0
16	0	0	1	0	0	0	1	0	0	0	0	0	0	0	228	0,5
17	0	0	1	0	0	1	0	1	1	0	0	0	0	0	275	0
18	0	0	0	0	0	1	0	0	0	0	1	0	0	0	160	1
19	0	0	0	0	0	0	0	0	1	1	0	0	0	0	277	0,5
20	0	1	0	0	0	0	0	0	0	0	0	0	1	0	27	0,5
21	0	0	1	0	0	0	0	1	1	0	0	0	0	0	225	0,5

Table 1 List of routes for CG algorithm

Number of routes	Obejctive value	Dual values													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
9	973	0	-10	0	-66	-88	-184	-230	0	0	-105	0	0	-236	-54
10	973	0	-10	0	-66	-88	0	-230	0	-7	0	-184	-105	-229	-54
11	973	0	-10	0	-66	-88	0	-230	0	-182	-105	-184	0	-54	-54
12	973	0	-10	0	-66	-88	-156	-230	0	-225	-62	-28	-43	-11	-54
13	973	0	-10	-37	-66	-88	-82	-230	0	-188	-99	-65	-6	-48	-54
14	973	0	-10	-8	-66	-88	-74	-230	0	-196	-91	-102	-14	-40	-54
15	973	0	-10	-15	-66	-88	-60	-230	0	-209	-77	-108	-27	-26	-54
16	973	0	-10	-15	-66	-88	-60	0	-230	-209	-77	-108	-27	-26	-54
17	968	0	-10	-1,5	-66	-88	-67	-226,5	-3,5	-203	-77,5	-115,5	-27,5	-26,5	-54
18	958,5	0	-10	-9,5	-66	-88	-44,5	-218,5	-11,5	-209,5	-77,5	-115,5	-27,5	-26,5	-54
19	953,5	0	-10	-14,5	-66	-88	-44,5	-213,5	-16,5	-199,5	-77,5	-115,5	-27,5	-26,5	-54
20	953,5	0	-10	-14,5	-66	-88	-35	-213,5	-16,5	-209	-68	-125	-37	-17	-54
21	946	0	-10	-7	-66	-88	-50	-221	-9	-209	-68	-110	-37	-17	-54
21 int.	953	-	-	-	-	-	-	-	-	-	-	-	-	-	-
opt.	949	-	-	-	-	-	-	-	-	-	-	-	-	-	-
21	950,75	0	0	-11,75	-66	-88	-50	-216,25	-13,75	-199,5	-77,5	-110	-27,5	-26,5	-54
22	947,5	0	0	-8,5	-66	-88	-50	-219,5	-10,5	-206	-71	-110	-34	-20	-54
23	949	0	0	0	-66	-88	-54	-228	-2	-216	0	-106	0	-20	-54
24	949	0	0	0	-66	-88	-50	-228	-2	-220	0	-110	0	-16	-54

Table 2 Dual solution to problem P' with optimal objective values dc

Route	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	d_r
22	0	0	0	0	0	0	0	0	0	0	0	0	1	0	20
23	0	0	1	0	0	0	0	0	1	0	0	0	1	0	236
24	0	0	0	0	0	1	0	0	1	0	0	0	1	0	286

Table 3 List of routes for branch and bound algorithm

CG algorithm has not found a binary solution for the problem P' , therefore, for finding the optimal solution to VRP we used branch and bound method. Three more routes were added to the model and optimal solution was found. The vector of the objective values of the optimal solution is the following zr^{opt}=(1;0;1;1;1;0;1;1;0;0;0;0;0;0;0;0;0;1;0;0;0;0;1;0), the objective function value is 949.

5 Computation experiments

To solve problem *CG* we designed a VBA-application that uses the Lingo solver. Its principle is the following. After running the application, a user is asked to define a few ranges, which are used in the model. The description appears on the screen and helps the user to get the correct size and values for each range. Defining all ranges, the user has to start the VBA-module with the appropriate button to solve problem *CG*. The application runs the Lingo solver automatically and exports results into defined ranges in Excel's list, where user can analyze them. In case the optimal solution to problem *CG* is binary, then this solution is optimal solution to VRP. Otherwise, the user has to start the second VBA module using branch and bound algorithm to find the optimal solution to VRP by generating new routes to optimal solution to problem *CG*. When branch and bound algorithm is finished, the optimal solution to VRP will appear on the screen.

In the case study the optimal solution to VRP could be found using the standard optimization mathematical model. Thus, our application did not show great time savings. Therefore, we made several computation experiments and solved other generated problems with a large number of variables which were impossible to solve using the standard optimization model of VRP. These cases were solved with use of our application and the optimal solutions were found.

6 Conclusions

The paper illustrates the use of set-covering algorithm to solve VRP on a case study, where the optimal solution could be found. Beside detailed description, we also demonstrated weak points of this method of solving VRP. Great time savings are expected when solving VRP problems with large number of variables with *CG* technique. This is another way how to find the optimal or suboptimal solution to this NP hard problem.

Acknowledgements

This publication was supported by the project 402/09/0041 founded by the Grant Agency of the Czech Republic and the project F4/18/2011 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Bramel, J., Simchi-Levi, D. Set-Covering-Based Algorithms for Capacitated VRP. In: Toth, P., Vigo, D. (ed.). *The Vehicle Routing Problem*, Philadelphia, USA, 2002, pp. 85–100.

[2] Pelikán, J. Hybrid flow-shop with adjustment. *Kybernetika*, vol. 47, 1/2011, pp. 50–59.

[3] Pelikán, J. A Time Limited Vehicle Routing Problem. *Communications*, vol. 7, 3/2006, pp. 9–11.

[4] Pelikán, J. Scheduling Serial-Parallel Processors – A Case Study. Hradec Králové 14.09.2005 – 16.09.2005. In: Skalská, H. (ed.). *Mathematical Methods in Economics 2005*. Hradec Králové : Gaudeamus, 2005, pp. 283–298.

Quantitative Methods, Normative Economy and Regional Development

Ivana Faltová Leitmanová[1], Jana Friebelová[2], Jana Klicnarová[2], Renata Klufová[2], Marek Šulista[2]

Abstract. The aim of the paper is to demonstrate the possibilities of using the quantitative methods for the evaluation of regional stabilisation on the level of administrative districts of municipalities with extended powers (MEPs). The stabilisation is characterized here by means of relative net migration, as its share in the total population increase of regions in last decades considerably exceeds the share of natural increase. MEPs were chosen for this analysis because they seem to be the most natural and best operating regions.

The rate of effectiveness and the order of MEPs were assessed on the base of criteria available from the Czech Statistical Office (unemployment rate, businesses with head offices in MEPs, total dwellings complete, coefficient of ecological stability, and distance to district towns) with help of methods DEA, and WSA. The results achieved by these methods were than compared each other and their validity was further analysed in connection to the net migration flows intensity as a basis for adequate regional strategies creation in the frame of the normative economy.

Keywords: quantitative methods, normative economy, regional development, DEA, WSA, net migration.

JEL Classification: C61, R11

AMS Classification: 90B50

1 Introduction

The development at the level of particular regions within the administrative division of the Czech Republic was directed with an effort to eliminate differences between the regions. However, certain differences at the regional level prevail. It does not mean that the convergence of socio-economic characteristics does not make sense, or it is not possible. It cannot be approached without prior knowledge of causes of the present differences. The identification and quantification of these differences as well as the role of their partial causes can be conducted sophistically using suitable quantitative methods.

1.1 Normative Economy and Regional Development

“Positive”, “descriptive” and “empirical” theories are frequently promoted as being more realistic, factual and relevant than normative approaches. We argue that labels such as “positive” and “empirical” emanate from a Realist theory of knowledge; a wholly inadequate epistemological basis for a social science ([16]).

One of the characteristic features of the change of spatial differentiation is the increasing polarization between important regional centres on one side and relatively rural areas on the other ([11]). One of the significant impacts of the economic transformation is a noticeable change in the spatial distribution of labour opportunities. Regional disparities can be divided into two groups - those which are fixed and cannot be changed or their change is hardly possible within the reasonable time or at reasonable cost and those which may be influenced by local government or inhabitants. Aforementioned standards of living along with work opportunities, traffic serviceability, opportunities for community, cultural and sport activities, health services, schools etc. Are circumstances of considerable impact on comfortable life of region’s inhabitants and thus its attractiveness and possibilities of further development ([18]).

A standardly used tool, influencing the labour market flexibility regarding the labour market mobility, is dwelling policy ([12], [14], [3]).

[1] University of South Bohemia, Faculty of Economics, Department of Economics, Studentská 13, 370 05 České Budějovice, Czech Republic, leitman@ef.jcu.cz.

[2] University of South Bohemia, Faculty of Economics, Department of Mathematics and Informatics, Studentská 13, 370 05 České Budějovice, Czech Republic, friebelova@ef.jcu.cz, janaklic@ef.jcu.cz, klufova@ef.jcu.cz, sulista@ef.jcu.cz.

Already at the beginning of the 1980s, scientists were interested how the dwelling situation and the situation in the labour market in the context of the labour market migration are influenced from the regional point of view. The analyses of the labour market in the metropolitan area of London, focusing on the mutual dependencies between partial part of the agglomeration in the context of dwelling, employment, migration, and commuting were conducted by Gordon and Lamont ([9]). They identified three different migration flows depending on the distance, the employment growth, unemployment, construction and prices of houses. Important bounds between the accessibility of rented dwelling, labour migration and also unemployment, and between migration to long distances and the rate of private construction in the area of the original destination as well as between the movement inside of an urban centre and the following unemployment decentralization were identified. The distribution of new dwelling and labour opportunities is perceived as a key factor of changes more than residential preferences.

In 1990s, Jackman and Savouri ([13]) investigated if differences in prices of realties and migration of labour differ in accordance with the fact if regions are neighbouring or non-neighbouring. The authors found out that the flexibility of prices of realties increases with the length of the mutual regional border. It is likely that individuals can change their dwelling without changing their jobs. Concurrently, migration in order to improve relative labour opportunities in neighbouring regions occurs less frequently than the same migration in non-neighbouring regions.

Empirical findings by Eliason, Lindgren, and Westerlund ([6]) from Sweden in years 1994–95 show that probability of inter-regional mobilities of labour decreases unexpectedly with accessibility of labour opportunities in neighbouring regions. Accessibility of job vacancies in surrounding regions increases significantly probability of commuting as a means of mobility. In addition to this, individual experience with unemployment increases probability of a mobility and migration.

Based on the data from the last decade of the last century and the first five years of the 21st century, Zabel ([19]) investigated how home-ownership rate, price elasticity of dwelling offer and growth rate of dwelling prices influence the flow of labour. The influence can be seen in two aspects: in relative mobility of home-owners versus renters and also in relative dwelling costs crosswise markets (towns). Rabe and Taylor ([15]) made a conclusion, based on data from years 1992–2007, that mainly regional differences in expected salaries and labour opportunities influence unemployment of individuals. Risk of employment loss of a spouse discourages households from migration more than salary differences.

2 Material and Methods

The assessment of the effectiveness of administrative districts of municipalities with extended power (MEP) was based on accessible statistical data from 2008 [2], when the net migration in 2008 was not available at the level of MEP. For this reason, data from 2009 were used. The investigation limited by one year time horizon does not allow to determine time changes and to filter out single random economic and political swings which could possible influence the observed data. This fact does not have to be detrimental after all, as the explaining variables can influence the explained variable with a time delay. The regional capitals were excluded from the set, because their posts and roles are, from the administrative area division point of view, incomparable with the other MEP. The analysis was conducted using these variables: unemployment rate, distance from district towns (modelled in GIS environment), coefficient of ecological stability, the number of registered business subjects (standardized to 1.000 inhabitants) and the number of finished flats and houses (standardized to 1.000 inhabitants).

The primary step necessary for using other methods was to conduct regression analysis investigation the relationship between the above mentioned variables and net migration as explained variable

To assessed municipalities with extended power, we used first all five above mentioned criteria. From methods of multi-criteria decision making, we used Data Envelopment Analysis (DEA) – CCR model with non-discretionary inputs (distance from district towns) and Weighted sum average (WSA).

DEA method divides assessed units to effective and ineffective according to the size of consumed inputs and produced outputs (unemployment rate and the distance to district towns were considered as inputs, while the coefficient of ecological stability, the number of business subjects, and the number of finished flats and houses were taken as outputs (for more details, see [5]).

Method WSA is based on the utility function. Each of the criteria value is assigned utility (the utility function is considered to be linear), utilities of all criteria values for each decision making are added up and the assessed units are ordered from the highest to the lowest value of this aggregated utility (for more details, see [7], [17]). Calculation was made using our own SW created in Maple (DEA), MS Excel (WSA).

Spatial distribution of resulting DEA scores and WSA utility function values was assessed by the Moran´s autocorrelation coefficient ([1], [8]).

3 Results and Discussion

As the conducted regression analysis did not prove a significant influence of the distance of MEPs to district towns and the ratio of the registered business subjects on the level of net migration, these two variables were omitted in the following analyses. Using regression analysis, when net migration was modelled using explaining variables of unemployment, coefficient of ecological stability and the number of finished flats and houses, determination index R^2=0.55 was reached, which represents (considering socio-economic character of the data) relatively high rate of dependence (see Table 1).

	B	p-value
Abs. term	-0.729	0.592
Rate of unemployment	-0.377	0.012
Coefficient of ecological stability	-0.729	0.002
Built flats	1.956	0.000

Table 1 Results of regression analysis

The values of net migration in 2009 (Fig. 1) were sorted into classes by the methods of quantiles (The boundary between migration loss and migration gain was set artificially.). Core areas can be divided into two groups. Also in 2009, it is evident that the first group consists of MEPs in hinterland of the biggest towns and cities, which are lucrative regarding labour opportunities in quickly developing progressive fields, tertiary and quartery sector. The second group is MEPs with a high concentration of investment activities, very often with a participation of foreign investors (e.g. Tatra Kopřivnice, Hyundai Nošovice, Třinec industrial zone Baliny, CT Park in Bor u Tachova, Šlapanice u Brna, Ledčice u Prahy and others).

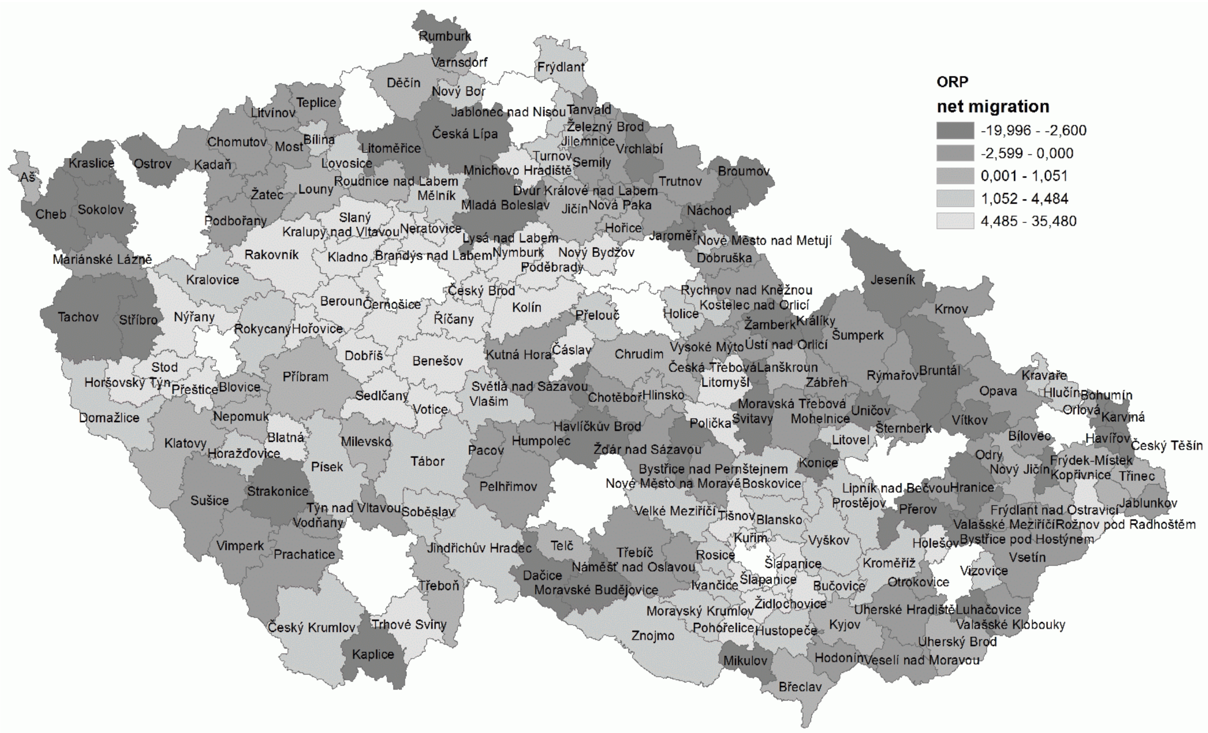


Figure 1 Net migration 2009

Although the correlation rate between net migration and resulting efficiencies obtained using DEA method could indicate that there is a low dependence (R=0.26), the use of this method can offer, to a certain extent, information about inter-regional disparities. The presented example represents the impact of housing construction, ecological stability and the situation in the labour market to the migration behaviour of inhabitants. MEPs showing the highest values of DEA efficiencies can be, according to the analysed variables, divided into two different groups represented in Figure 2 (The cartograms of DEA scores (and WSA utility function values in Figure 3) were made in ArcGIS. The values of variables mentioned above were sorted into classes by the method of standard deviation.):

- In „background“ of bigger towns and cities, mainly Prague, Pilsen, Brno, and partially Ostrava, Pardubice, and Hradec Králové the resulting above-average efficiency of micro-regions is caused mainly by housing construction, i.e. the number of finished flats and houses. In these areas, it is necessary to focus on the future regional policy towards the respect to environmental characteristics. It means that the ratio of built-up areas should not be extended and other areas should be re-cultivated (e .g. brown-fields).
- The second group represents remaining MEPs, for which the ecological stability influences crucially the resulting efficiency.

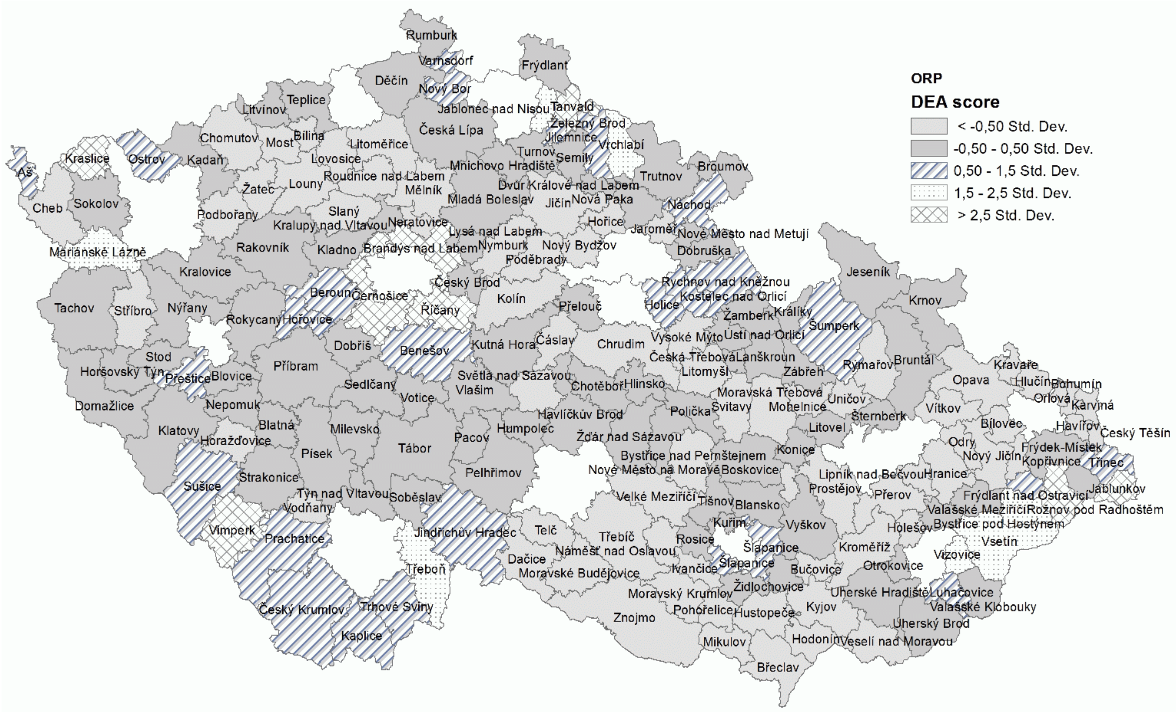


Figure 2 DEA scores

The use of WSA method brought better results regarding the correlation between net migration and resulting utilities of MEPs (R=0,496). The spatial distribution of the result values is very similar to the results obtained using DEA method.

Almost one half of the observed MEPs do not differ statistically significantly from the average value of utility of the whole set (see Figure 3). Comparison of spatial distribution of DEA scores and utilities of WSA using Moran coefficient of spatial autocorrelation came with the conclusion that the utility values obtained using WSA method show a bigger rate of spatial autocorrelation (Z=0.33, resp. Z=0.43), which means that micro-regions with similar values of the resulting efficiencies tend to make clusters and make more compact units. This fact is important for the following formulation of measurements and tools of regional policy, because it is possible to use returns to scale.

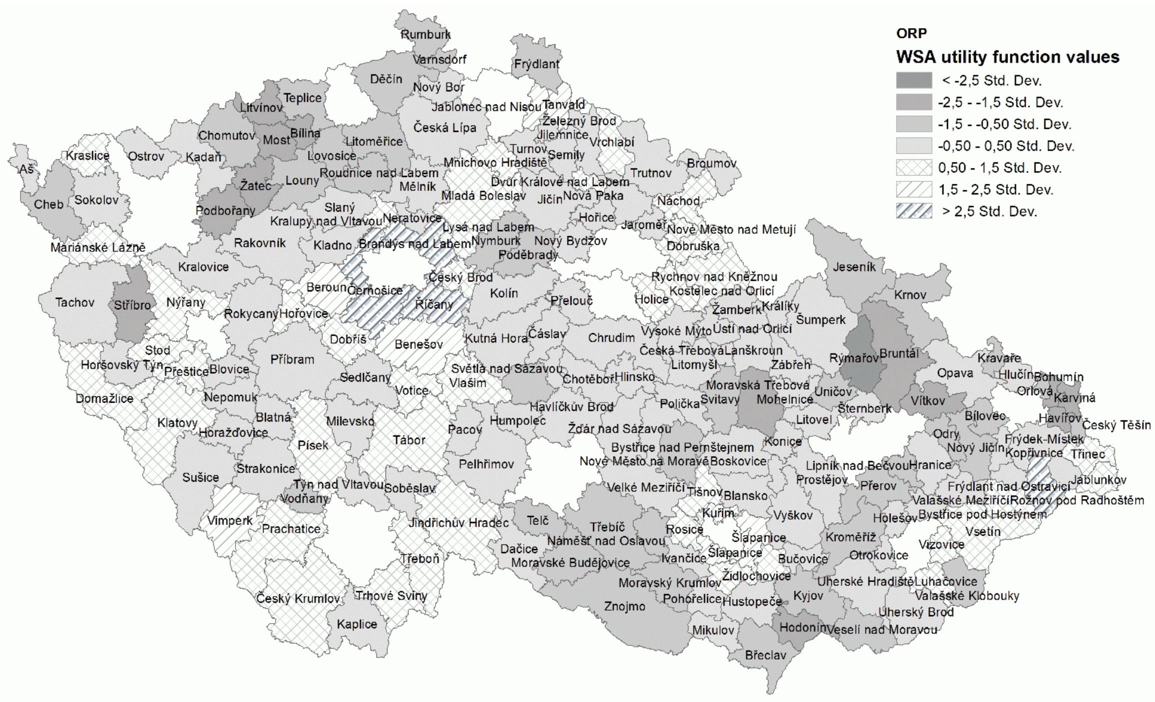


Figure 3 WSA scores

Although, there are partial differences in the results obtained with the used methods, we can say that these are not random findings, as the application of both used methods leads, in some cases, to the same conclusions, resp. one method proves the results of the other. The results obtained using both methods prove that above-average net migration in the following MEPs: Černošice, Říčany, Brandýs nad Labem, Lysá nad Labem, and Frýdlant nad Ostravicí is significantly influenced by the location regarding the core areas and connected housing construction in case of MEP in backgrounds of metropolitan areas of the Czech Republic ([10], [11], [4]) or ecological stability, resp. by impacts which had not been included in the analysis. The findings prove that there is an inverse relationship between migration and unemployment. Traditionally developed regions as well as newly supported regions show low values of unemployment rate and, at the same time, high values of net migration and the number of finished flats and houses. On the other hand, economically weak regions face higher and high unemployment and are attractive only for their ecologically stability.

Complete results are available on http://home.ef.jcu.cz/research/DEA/regions/.

4 Conclusion

Using the chosen quantitative methods enabled us to identify almost equally the situation in MEPs in the Czech Republic regarding net migration in 2009 and selected factors which could possibly influence it.

It is possible to say that in case of MEPs, we can observe, considering explaining variables, a dual segmentation. First, there is a segment where the level of labour opportunities has the most significant influence on migration of inhabitants and connected dwelling conditions. Similarly, it is possible to identify the segment where migration is subject to micro-economic factors represented in the paper by the coefficient of ecological stability.

Even though, the findings are based on data only form one year and therefore they could be connected with the risk of a random influence, the findings imply possible solutions for the development and stabilization of MEPs. A wider practical use of the above mentioned conclusions can be expected when conducting a more detailed analysis considering other factors or variables on one hand and a time extension of the investigated time period on the other.

Acknowledgements

This paper was supported by the project of Ministry of Agriculture of the Czech Republic NAZV QI92A023 and the project of Ministry of Education of the Czech Republic OPVK 0178.

References

[1] Anselin, L.: Exploring Spatial Data with GeoDa: A Workbook, Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana-Champaign, 2005, 244 pp. – available on-line: http://geodacenter.asu.edu/system/files/geodaworkbook.pdf

[2] Up-to-date data for all SO MEP, available on-line: http://www.czso.cz/csu/redakce.nsf/i/csu_a_uzemne_analyticke_podklady_za_obce_ceske_republiky

[3] Blannchflower, D.G., Shadforth, C.: Fear, unemployment and migration. The Economic Journal, 119 (February), 2009, F139-F182.

[4] Blažek, J.: Meziregionální rozdíly v České republice v transformačním období, Geografie – Sborník ČGS 101, č. 4, 1996, str. 265-277.

[5] Cooper, W. W., Seiford, L. M., Tone, K: *Data Envelopment Analysis. A Comprehensive Text with Models, Applications, Refernces and DEA-Solver Software*. Springer, 2007, p. 477.

[6] Eliasson, K., Lindrgren, U. and Westerlund, O.: Geographical labour mobility: migration or commuting?, Reg. Studies 37 , 2003, 827-837

[7] Figueira, José; Salvatore Greco, Matthias Ehrgott. *Multiple Criteria Decision Analysis: State of the Art Surveys*. New York,: Springer Science + Business Media Inc., 2005.

[8] Getis, A.: Reflections on spatial autocorrelation. Geographical analysis 37, 2007, pp. 491–496.

[9] Gordon I R, Lamont D.: "A model of labour-market interdependencies in the London region". *Environment and Planning, A* 14(2), 1982, 237–264

[10] Hampl, M.: Teorie geografické organizace společnosti. In: Hampl, M. a kol.: Geografická organizace společnosti a transformační procesy v České republice. Univerzita Karlova v Praze, Přírodovědecká fakulta, katedra sociální geografie a regionálního rozvoje, Praha. 1996, str. 13-34.

[11] Hampl, M.: Geografická organizace společnosti v České republice: transformační procesy a jejich obecný kontext. Univerzita Karlova v Praze, Přírodovědecká fakulta, katedra sociální geografie a regionálního rozvoje, 2005, Praha.

[12] Hughes, G.A. and McCormick, B., 1987. Housing markets, unemployment and labour market flexibility in the UK. *European Economic Review* 31 (1987), pp. 615–645

[13] Jackman, R., Savouri, S.: Regional Migration versus Regional Commuting: The Identification of Housing and Employment Flows. *Scottish Journal of Political Economy*, vol. 39, issue 3, 1992, pages 272-87

[14] Leitmanová, I.: Problémy trhu práce v Evropské unii a České republice. Politická ekonomie, Vol.1998, issue 4, 1998

[15] Rabe, B., Taylor, M. P.: Differences in opportunities? Wage, unemployment and house-price effects on migration ISER Working Paper Series 2010-05

[16] Tinker, A. M., Merino, B. D., Neimark, M. D.: Accounting, Organizations and Society Volume 7, Issue 2, 1982, Pages 167-200, The normative origins of positive theories: Ideology and accounting thought

[17] Triantaphyllou, E.: *Multi-Criteria Decision Making: A Comparative Study*. Dordrecht, The Netherlands: Kluwer Academic Publishers (now Springer), 2000, pp. 320

[18] Vojvodíková, B., Vojvodík, M.: Regional disparities as an aspect of regional strategic planning, XXI AESOP Conference, University of Napoli, 2007

[19] Zabel, J.: The Role of the Housing Market in the Migration Response to Employment Shocks. New England Public Policy Center Working Paper 09-2, 2009

Modeling of network revenue management problems with customer choice behavior

Petr Fiala[1]

Abstract. Revenue management is the process of understanding, anticipating and influencing customer behavior in order to maximize revenue. A significant limitation of the applicability of classical models is the assumption of independent demand. Network revenue management models attempt to maximize revenue when customers buy bundles of multiple resources. The dependence among the resources in such cases is created by customer demand. The common modeling approaches assume that customers are passive and they do not engage in any decision-making processes. This simplification is often unrealistic for many practical problems. In response to this, interest has arisen in recent years to incorporate customer choice into these models, further increasing their complexity. Today's customers actively evaluate alternatives and make choices. Revenue management pays increasing attention to modeling the behavior of individual customers. Strategic customer behavior is analyzed. A modeling approach for strategic customer behavior is proposed.

Keywords: revenue management, network problems, customer choice behavior.

JEL Classification: C44
AMS Classification: 90B50

1 Introduction

The general problem is about how companies should design their selling mechanisms in order to maximize expected revenue. Revenue management (RM) is the process of understanding, anticipating and influencing customer behavior in order to maximize revenue. RM is to sell the right product, to the right customer at the right time, for the right price through the right channel by maximizing revenue. RM is the art and science of predicting real-time customer demand and optimizing the price and availability of products according to the demand.

The RM area encompasses all work related to operational pricing and demand management. This includes traditional problems in the field, such as capacity allocation, overbooking and dynamic pricing, as well as newer areas, such as oligopoly models, negotiated pricing and auctions.

Recent years have seen great successes of revenue management, notably in the airline, hotel, and car rental business. Currently, an increasing number of industries is exploring to adopt similar concepts. What is new about RM is not the demand-management decisions themselves but rather how these decisions are made. The true innovation of RM lies in the method of decision making.

Network revenue management models attempt to maximize revenue when customers buy bundles of multiple resources. The dependence among the resources in such cases is created by customer demand. The basic model of the network revenue management problem is formulated as a stochastic dynamic programming problem whose exact solution is computationally intractable. Most approximation methods are based on one of two basic approaches: to use a simplified network model or to decompose the network problem into a collection of single-resource problems. The Deterministic Linear Programming (DLP) method is a popular in practice. The DLP method is based on an assumption that demand is deterministic and static.

Network problems are computationally intensive even without consideration of customer choice behavior, thus good heuristics need to be found. The common modeling approaches assume that customers are passive and they do not engage in any decision-making processes. This simplification is often unrealistic for many practical problems. In response to this, interest has arisen in recent years to incorporate customer choice into these models, further increasing their complexity. Today's customers actively evaluate alternatives and make choices. Revenue management pays increasing attention to modeling the behavior of individual customers. Strategic customer behavior is analyzed. A modeling approach for strategic customer behavior is proposed.

Among the efficient techniques that have been proposed is the so-called choice-based linear program (CDLP) of Gallego et al. [3]. Based on this work, van Ryzin and Liu [10] present an extension of the standard deterministic linear program approach to include choice behavior.

[1] University of Economics, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, pfiala@vse.cz.

2 Network revenue management problems

The quantity-based revenue management of multiple resources is referred as network revenue management. This class of problems arises for example in airline, hotel, and railway management. In the airline case, the problem is managing capacities of a set of connecting flights across a network, so called a hub-and-spoke network. In the hotel case, the problem is managing room capacity on consecutive days when customers stay multiple nights.

Network revenue management models attempt to maximize some reward function when customers buy bundles of multiple resources. The interdependence of resources, commonly referred to as network effects, creates difficulty in solving the problem. The classical technique of approaching this problem has been to use a deterministic LP solution to derive policies for the network capacity problem. Initial success with this method has triggered considerable research in possible reformulations and extensions, and this method has become widely used in many industrial applications. A significant limitation of the applicability of these classical models is the assumption of independent demand. In response to this, interest has arisen in recent years to incorporate customer choice into these models, further increasing their complexity. This development drives current efforts to design powerful and practical heuristics that still can manage problems of practical scope.

The basic model of the network revenue management problem can be formulated as follows (see [2], [9]): The network has m resources which can be used to provide n products. We define the incidence matrix $A = [a_{ij}]$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, where

$a_{ij} = 1$, if resource i is used by product j, and

$a_{ij} = 0$, otherwise.

The j-th column of A, denoted a_j, is the incidence vector for product j. The notation $i \in a_j$ indicates that resource i is used by product j.

The state of the network is described by a vector $x = (x_1, x_2, \ldots, x_m)$ of resource capacities. If product j is sold, the state of the network changes to $x - a_j$.

Time is discrete, there are T periods and the index t represents the current time, $t = 1, 2, \ldots, T$. Assuming within each time period t at most one request for a product can arrive.

Demand in time period t is modeled as the realization of a single random vector $r(t) = (r_1(t), r_2(t), \ldots, r_n(t))$. If $r_j(t) = r_j > 0$, this indicates a request for product j occurred and that its associated revenue is r_j. If $r_j(t) = 0$, this indicates no request for product j occurred. A realization $r(t) = 0$ (all components equal to zero) indicates that no request from any product occurred at time t. The assumption that at most one arrival occurs in each time period means that at most one component of $r(t)$ can be positive. The sequence $r(t)$, $t = 1, 2, \ldots, T$, is assumed to be independent with known joint distributions in each time period t. When revenues associated with product j are fixed, we will denote these by r_j and the revenue vector $r = (r_1, r_2, \ldots, r_n)$.

Given the current time t, the current remaining capacity x and the current request $r(t)$, the decision is to accept or not to accept the current request. We define the decision vector $u(t) = (u_1(t), u_2(t), \ldots, u_n(t))$ where

$u_j(t) = 1$, if a request for product j in time period t is accepted, and

$u_j(t) = 0$, otherwise.

The components of the decision vector $u(t)$ are functions of the remaining capacity components of vector x and the components of the revenue vector r, $u(t) = u(t, x, r)$. The decision vector $u(t)$ is restricted to the set

$$U(x) = \{u \in \{0, 1\}^n,\ Au \le x\}.$$

The maximum expected revenue, given remaining capacity x in time period t, is denoted by $V_t(x)$. Then $V_t(x)$ must satisfy the Bellman equation

$$V_t(x) = E\left[\max_{u \in U(x)} \{r(t)^T u(t, x, r) + V_{t+1}(x - Au)\}\right] \tag{1}$$

with the boundary condition

$$V_{T+1}(x) = 0, \forall x.$$

A decision u^* is optimal if and only if it satisfies:

$u_j(t, x, r_j) = 1$, if $r_j \ge V_{t+1}(x) - V_{t+1}(x - a_j)$, $a_j \le x$,

$u_j(t, x, r_j) = 0$, otherwise.

This reflects the intuitive notion that revenue r_j for product j is accepted only when it exceeds the opportunity cost of the reduction in resource capacities required to satisfy the request.
The equation (1) cannot be solved exactly for most networks of realistic size. Solutions are based on approximations of various types. There are two important criteria when judging network approximation methods: accuracy and speed. Among the most useful information provided by an approximation method are estimates of bid prices (see [9]).

Most approximation methods are based on one of two basic approaches:

- to use a simplified network model or
- to decompose the network problem into a collection of single-resource problems.

The first approach is to use a simplified network model, for example posing the problem as a static mathematical program. We introduced Deterministic Linear Programming (DLP) method (see).

The DLP method uses the approximation

$$V_t^{LP}(x) = \max r^T y$$

subject to (2)

$$Ay \le x$$
$$0 \le y \le E[D]$$

where $D = (D_1, D_2, \ldots, D_n)$ is the vector of demand over the periods $t, t+1, \ldots, T$, for product j, $j = 1, 2, \ldots, n$, and $r = (r_1, r_2, \ldots, r_n)$ is the vector of revenues associated with the n products. The decision vector $y = (y_1, y_2, \ldots, y_n)$ represent partitioned allocation of capacity for each of the n products. The approximation effectively treats demand as if it were deterministic and equal to its mean $E[D]$.

The optimal dual variables, π^{LP}, associated with the constraints $Ay \le x$, are used as bid prices. The DLP was among the first models analyzed for network RM (see [9]). The main advantage of the DLP model is that it is computationally very efficient to solve. Due to its simplicity and speed, it is a popular in practice. The weakness of the DLP approximation is that it considers only the mean demand and ignores all other distributional information. The performance of the DLP method depends on the type of network, the order in which fare products arrive and the frequency of re-optimization.

3 Strategic customer

Customer behavior modeling has been gaining increasing attention in the revenue management (see [7]). Because customers will exhibit systematic responses to the selling mechanisms, firms are responsible for anticipating these responses when making their pricing decisions.

The problems can be divided into two parts. The first part examines the effect of inter-temporal substitution by customers. Customers may choose when to buy a particular product in response to firms' dynamic pricing practices. When they anticipate price reductions, consumers may choose to wait for the sale. Other relevant issues include capacity rationing, valuation uncertainty, and consumer learning effects. The dynamics of consumer demand depend directly on the seller's dynamic pricing strategies. The second part studies customer choice in multi-product revenue management settings. The focus is on how customers choose which product to buy. A common approach is to use discrete choice models to capture multi-product consumer demand. Substitution and complementary effects across multiple products are studied also.

Inter-temporal Substitution and Strategic Customer Behavior

Inter-temporal substitution refers to the practice of delaying purchases to a future point in time. Until recently, the revenue management literature has almost completely neglected this issue. The standard modeling paradigm is to assume that demand arriving at each instance in time is either realized or lost forever. Recognizing that this assumption is unrealistic, recent work has begun to pay increasing attention to this issue.

Modeling approaches are concentrated on following issues:

- Customer Response to Dynamic Pricing
- Dynamic Pricing under Strategic Customer Behavior
- Capacity Rationing
- Valuation Uncertainty

The first step in understanding strategic customer behavior in revenue management is to develop models of how customers respond to firms' pricing strategies. The first approach is to assume that the firm sets prices dynamically according to some established policy and then investigate the optimal behavior of a single strategic customer in this setting.

With some understanding of how customers respond to well-known pricing policies, the natural next step is to study how firms should set their prices in the presence of strategic customers. One of the earliest papers to investigate optimal dynamic pricing under strategic customer behavior was by Aviv and Pazgal [1]. They analyze a model with a single price reduction at a fixed point in time *T*. There is a Poisson stream of strategic customers, with valuations drawn from some distribution. Customers arriving before time *T* may strategically wait to purchase at time *T* if doing so is beneficial, but customers arriving after *T* have no incentive to wait. The seller's problem is to choose the discount price optimally. The authors demonstrate that the losses resulting from neglecting strategic customer behavior can be substantial.

Another approach is focusing on how firms can use pricing and rationing to extract maximum revenue. Because capacity is scarce, rationing is almost inevitable, so it is quite natural for firms to use this as a strategic tool in the face of strategic customers. The common approach is to start with a two-period model; prices are higher in the first period but there may be limited availability in the second period. When capacity is observable, customers can perfectly anticipate the probability of availability in the second period. However, when capacity is unobservable, customers infer the fill rate through a learning process that converges after multiple repetitions of the underlying two-period model. The models investigate how rationing affects strategic demand by making customers more inclined to purchase earlier at higher prices. Liu and van Ryzin [4] study the effects of capacity rationing in a two-period model in which all customers are present at the start of the horizon. Customers who buy in the first period pay a premium and customers who attempt to buy in the second period may be rationed. Demand is deterministic and prices are fixed. Customers observe capacity levels, so fill rates in the second period can be perfectly anticipated. The authors find that the effectiveness of rationing depends on consumer risk preferences.

There may be other reasons for delaying purchase. In particular, when customers are uncertain about their valuation for the product, it may be wise to wait until more information is available. For example, in the context of airlines, travelers who are not certain about their plans may choose to wait. This situation lends itself naturally to advance purchase discounts, in which customers who buy early are compensated for bearing risk. In the revenue management literature, a few papers examine customer behavior in the presence of valuation uncertainty (see Xie and Shugan see [11]).

Customer behavior and multiple products

Firms often sell multiple products that exhibit demand dependencies. The question is how the firm should make revenue management decisions. The answer depends on the nature of these demand dependencies. There are two broad settings:

- demand dependencies may arise due to substitution or complementarity effects across products,
- demand dependencies across products may be driven by customer choice.

Substitution and complementarity effects across different products can be captured using multi-dimensional demand functions. There are also other related scenarios that fall under this category. A customer who has purchased a particular product may also be willing to purchase a related product, especially if a discount is offered. Such practices are called cross-selling or up-selling and are quite common in practice. The questions here would be how to choose the accompanying product and what price to charge for the bundle. As another example, consider the following situation. When the product requested by a customer is sold out, the customer may still be willing to accept a substitute. In such situations, should these substitution offers be made, and how should they be priced?

Maglaras and Meissner [5] consider a multiproduct revenue management problem with multidimensional demand functions that map prices (for each product) into demand rates (for each product). There is a common resource, and different products deplete the resource at different rates. With this model, the authors formulate a dynamic pricing problem and a capacity allocation model and demonstrate that both formulations can be reduced to a common framework in which the firm controls the aggregate rate at which the resource is depleted. The authors characterize the optimal controls in closed-form and suggest several static and dynamic heuristics. These heuristics are also shown to be asymptotically optimal.

Netessine, Savin, and Xiao [6] introduce an investigation of cross-selling practices into the revenue management literature. In the model, the firm manages a set of products, faces stochastic customer arrivals, and makes dynamic cross-selling decisions based on current inventory levels of each product. In this setting, firms must select the complementary product to offer as well as the optimal price for the packaged offer; this leads to a combinatorial optimization problem.

Customer choice from a set of products can be modeled using a discrete choice framework. This may be a general choice model or may also be specialized to more commonly used models such as the multinomial logit model. The customer's choice depends critically on the set of available products. Therefore, there are two related questions: which set of options to make available and how to price each of these options. The customer choice behavior will be investigated in section 4.

4 Customer choice behavior

Potential customers usually do not come with a predetermined idea of which product to purchase. Rather, they only know some particular features that the product should possess and compare several alternatives that have these features in common before coming to a purchase or non-purchase decision. For example, a customer might be interested in a flight from A to B, but considers several flights with close-by departure times, or several class options. This issue of customer choice was first investigated by Talluri and van Ryzin [8], who study a revenue management problem under a discrete choice model of customer behavior. There are n fare products, each associated with exogenous revenue r_j, j = 1,2,…n.. At each point in time, the firm chooses to offer a subset of these fare products. Given the subset of offered products, customers choose an option (which may also be a no purchase option) according to some discrete choice model. Gallego et al. [3], van Ryzin and Liu [10] extend this analysis to the network setting. Each product consists of a fare class and an itinerary, which may use up resources on multiple legs of the network. The dynamic program of finding the optimal offer sets becomes computationally intractable. The authors adopt a deterministic approximation by reinterpreting the purchase probability as the deterministic sale of a fixed quantity (smaller than one unit) of the product. Under this interpretation, the revenue management problem can be formulated as a linear program, and there is possible demonstrate that the solution is asymptotically optimal as demand and capacity are scaled up. There is possible to design implementation heuristics to convert the static LP solution into dynamic control policies.

Choice-Based Deterministic LP (CDLP)

The probability that the customer chooses product j given the set of offered fares S (conditioned to arrival of a customer) is denoted by $P_j(S)$. Time is discrete and partitioned into T time periods that are small enough such that there is at most one customer arrival with probability λ and no arrival with probability $1-\lambda$. The network has m resources which can be used to provide n products. The incidence matrix A = $[a_{ij}]$, i = 1, 2, …, m, j = 1, 2, …, n, introduced in network revenue management problems, is used. Demand is treated as known and being equal to its expected value. The problem reduces then to an allocation problem where we need to decide for how many time periods a certain set of products S shall be offered, denoted by $t(S)$. Denote the expected total revenue from offering S by

$$R(S) = \sum_{j \in S} P_j(S) r_j$$

and the expected total consumption of resource i from offering S by

$$Q_i(S) = \sum_{j \in S} P_j(S) a_{ij}, \quad \forall i,$$

Then the choice-based deterministic linear program (3) is given by

$$\begin{gathered} V^{CDLP} = \max \sum_{S \subseteq N} \lambda R(S) t(S) \\ \sum_{S \subseteq N} \lambda A P(S) t(S) \le x, \\ \sum_{S \subseteq N} t(S) = T \\ t(S) \ge 0, \quad \forall S \subseteq N \end{gathered} \tag{3}$$

The objective is to maximize total revenue under constraints that consumption is less than capacity and total time sets offered are less than horizon length. Decision variables are total time subset S is offered $t(S)$.

There are two basic possibilities how to we use the CDLP solution. First one is to directly apply time variables $t^*(S)$ (Gallego et al. [3]). For certain discrete-choice models it is possible to efficiently use column generation to solve the CDLP model to optimality. It returns a vector with as many components as there are possible offer sets, and each component represents the number of time periods out the finite time horizon that the corresponding offer set should be available. The notion of efficient sets introduced by Talluri and van Ryzin [8] for the single leg case is translated into the network context and the authors show that CDLP only uses efficient sets in its optimal solution. Second one is to use dual information in a decomposition heuristic (van Ryzin & Liu [10]). The dual variables of the capacity constraints can be used to construct bid prices.

5 Conclusions

Network revenue management models attempt to maximize revenue when customers buy bundles of multiple resources. The basic model of the network revenue management problem is formulated as a stochastic dynamic programming problem whose exact solution is computationally intractable. The Deterministic Linear Programming (DLP) method is a popular in practice. The DLP method is based on an assumption that demand is deterministic and static. The common modeling approaches assume that customers are passive and they do not engage in any decision-making processes. This simplification is often unrealistic for many practical problems. In response to this, interest has arisen in recent years to incorporate customer choice into these models, further increasing their complexity. In the paper strategic customer behavior was analyzed. A modeling approach for strategic customer behavior based on deterministic linear programming (CDLP) was investigated. This area is promising for next scientific research.

Acknowledgements

The research project was supported by Grant No. 402/10/0197 „Revenue management – models and analyses“ from the Grant Agency of the Czech Republic and by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.
.

References

[1] Aviv, Y., Pazgal A.: Optimal pricing of seasonal products in the presence of forward-looking consumers. *Manufacturing & Service Operations Management* 10 (2008), 339-359.

[2] Fiala, P.: Approximations for solving network revenue management problems. In: *Mathematical Methods in Economics 2010*. České Budějovice : University of South Bohemia, 2010, s. 132–137.

[3] Gallego, G., Iyengar, G., Phillips, R. and Dubey, A.: Managing flexible products on a network. Tech. Rep. TR-2004-01, Department of Industrial Engineering and Operations Research, Columbia University, 2004.

[4] Liu, Q., and van Ryzin, G. J.: Strategic capacity rationing to induce early purchases. Working paper, Columbia University, 2007.

[5] Maglaras, C., and Meissner, J.: Dynamic pricing strategies for multi-product revenue management problems. *Manufacturing & Service Operations Management* 8 (2006), 136–148.

[6] Netessine, S., Savin, S., and Xiao, W. Q.: Revenue management through dynamic cross-selling in e-commerce retailing. *Operations Research* 54 (2006), 893–913.

[7] Shen, Z.-J.and Su, X.: Customer Behavior Modeling in Revenue Management and Auctions. *Production and Operations Management* 16 (2007), 713–728,

[8] Talluri, K. and van Ryzin G. J.: Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50, (2004), 15–33.

[9] Talluri, K. T., and van Ryzin, G. J.: *The Theory and Practice of Revenue Management*. Kluwer Academic Publishers, Boston, MA, 2004.

[10] van Ryzin, G. J. and Liu, Q: On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management* 10 (2008), 288–311.

[11] Xie, J., and Shugan, S. M.: Electronic tickets, smart cards, and online prepayments: When and how to advance sell. *Marketing Science* 20 (2001), 219–243.

The Evaluation of the Entropy of Decision Makers' Preferences in Ordinal Consensus Ranking Problem

Jan Fiedor[1], Jiří Mazurek[2]

Abstract. The aim of the article is to introduce Shannon entropy and the relative entropy into ordinal consensus ranking problem (OCRP) context as a measure of uncertainty of preferences within a group of decision makers (DMs). An existence of a consensus in the OCRP crucially depends on initial decision makers' preferences: rankings of *n* objects from the 1st to the *n*th place. When those rankings are markedly different, there is high entropy (uncertainty) associated with DMs' preferences and a unique consensus within a group is unlikely to be achieved. However, when DMs' rankings are similar, the entropy is small and a consensus is much more likely to be obtained. In the article a relationship between the relative entropy of DMs preferences and an agreement among four classic methods' for OCRP solutions is discussed on several examples. Moreover, experimental evaluation of the relationship is performed and our results indicate that there is a negative relation between the relative entropy and the agreement among methods' solutions, e.g. the higher is the relative entropy, the lower is the methods' agreement. The question remains whether a theoretical relation connecting the entropy and a solution existence (or agreement among methods' solutions) to OCRP can be derived in general.

Keywords: entropy, group decision analysis, ordinal consensus ranking problem, preferences, uncertainty.

JEL Classification: D71, C44, C63
AMS Classification: 90B50, 91B08, 91B10, 94A17

1 Introduction

In ordinal consensus ranking problem (OCRP), a set of *k* decision makers (DMs) rank *n* objects from the 1st to the nth place according to some overall criterion or a set of criteria. The goal is to find a solution – a consensus ranking. To solve OCRP, many methods were proposed, which can be divided into *ad hoc methods* (such as Borda-Kendall's method of marks) and *distance based methods* (such as MAH, CRM or DCM) in general. These four methods are described in Section 2.

The existence of a consensus in OCRP crucially depends on initial structure of decision makers' preferences. When decisions of DMs are similar (information of objects' rankings is more specific), then there is a good chance of a meaningful consensus to be achieved. However, when DMs' decisions differ markedly (information of objects' rankings is uncertain within a group), a consensus might not exist at all, or it will depend on a method used (see Section 5). The problem of experts' decision agreement or the entropy has been largely ignored (not only in OCPR, but in general as well) so far, in spite of the fact experts' decisions (which constitute an input of a problem) are apparently fundamental to a problem's solution (a problem's output). Generally, a decision making process can be regarded as a process of entropy's reduction: from many alternatives or objects only one (or few) is chosen finally. Moreover, recent research revealed the methods of OCRP differ in their solutions, but it is an open question to what degree ([2], [10]) and under which conditions. It can be hypothesized, that low entropy problems would lead to the same consensus under different methods, while high entropy problems would yield different results under different methods (or there would be no consensus at all).

The aim of the article is to introduce Shannon entropy and the relative entropy into OCRP context as a measure of uncertainty of preferences within a group of decision makers, and to demonstrate a relationship between the entropy of initial decision makers' preferences and OCRP solutions by four selected methods described in Section 2.

This paper is organized as follows. In Section 2, four classic methods for OCRP solution are described. In Section 3, Shannon entropy and the relative entropy in OCRP framework is introduced. In Section 4, Kendall's coefficient of concordance (*W*) is described and an empirical relationship between the relative entropy of DMs'

[1] Brno University of Technology, Faculty of Information Technology, Department of Intelligent Systems, DITS FIT BUT, Božetěchova 2, 612 66 Brno, e-mail: ifiedor@fit.vutbr.cz.
[2] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics, Univerzitní náměstí 1934, 733 40 Karviná, e-mail: mazurek@opf.slu.cz.

preferences and the OCRP solutions' agreement is discussed in Section 5. Experimental evaluation of this relationship is performed in Section 6. Conclusions close the article.

2 Classic methods for OCRP solution

For a comparison of solutions to the OCRP problem, following methods were selected:

- In *Borda-Kendall's method of marks* (*BAK*) [4] each alternative is given a number of points corresponding to its position assigned by each DM. The best alternative is the alternative with the lowest total count (mark) or with the lowest average (which is equivalent).
- *Maximize agreement heuristic* (*MAH*) [1] evaluates the number P_i which expresses how many times an alternative i is preferred to all other alternatives for all DMs. If $P_i = 0$, then the alternative i is ranked the last (the worst) in the final consensus ranking. If some alternative j is not preferred to any other alternative ($N_j = 0$), then it is ranked the first (the best). If P_i and N_i are non-zero for all alternatives, then an alternative with the lowest value of ($P_i - N_i$) is placed the last and eliminated from the list. The algorithm repeats until all alternatives are ranked.
- *Consensus ranking model* (*CRM*) [3] starts with a preference intensity matrix A (a_{ij}) for each DM. In the matrix element a_{ij} gives the number of ranks (positions) an alternative i is preferred to an alternative j. Next, a total agreement frequency matrix for each possible rank permutation is generated and distances between the total agreement frequency matrix and the preference intensity matrices are evaluated. Consensus ranking is the ranking with the minimum distance from preference intensity matrices.
- In *distance-based ideal-seeking consensus ranking model* (*DCM*) [10] a single preference matrix (for all DMs) is constructed. Then a matrix with identical preferences of all DMs (ideal consensus matrix) is generated for all possible rankings (permutations). All ideal consensus matrices are compared with the initial preference matrix. The matrix representing the final consensus ranking is the matrix with the lowest distance from the initial preference matrix.

In MAH, CRM and DCM methods Kemeny-Snell's distance is used to measure the distance between matrices A and B [2]:

$$d_{KS}(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| a_{ij} - b_{ij} \right| \tag{1}$$

A more detailed review of these and other methods can be found in [2], [8] or [10].

3 The entropy of decision makers' preferences

To evaluate the entropy of DMs' preferences in OCRP, each DM's ranking (which is a permutation of n objects) is turned into a square matrix π of the n-th order in the first step. For instance, the permutation (1, 3, 2) is represented by a matrix π with rows corresponding to alternatives and columns to positions:

$$\pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Every matrix π is bistochastic (the sum over all rows and columns is 1). In the second step, all DMs' rankings (permutation matrices π) are converted into one matrix Π. The matrix Π can be considered a fuzzy (or group) ranking, as it represents an average ranking of a group of DMs. The matrix Π is defined as follows:

Definition 1. *Let π_i be a permutation matrix of order n given by i-th decision maker, $i \in \{1, 2, ..., K\}$. Then, a fuzzy permutation matrix $\Pi(\pi_1, ..., \pi_K)$ is a bistochastic matrix of order n given as:*

$$\Pi(\pi_1, ..., \pi_K) = \frac{\sum_{i=1}^{K} \pi_i}{K} \tag{2}$$

Definition 2. *The Shannon entropy of a probability distribution p(x) on a finite set X ($x \in X$) is given as [9]:*

$$H\left(p(x)/x \in X\right) = -\sum_{x \in X} p(x)\log_2 p(x), \tag{3}$$

where by definition $p(x)\cdot\log_2 p(x) = 0$ *for p(x)* = 0.

It follows from (2) that *i*-th row of the fuzzy permutation matrix Π with elements p_{ij} can be considered a probability distribution of *i*-th alternative over all positions (a result of averaging DMs' preferences). To evaluate Shannon entropy of decision makers' preferences, the relation (3) is used for each *i*-th row of the matrix Π, and the total entropy H_T is given as a sum over all matrixes rows (over all alternatives):

$$H_T = \sum_{i=1}^{n}\sum_{j=1}^{n} H\left(p_{ij}\right), \; p_{ij} \in \Pi \tag{4}$$

If a probability distribution is uniform: $p_{ij} = \frac{1}{N}, \; j \in \{1,2,...N\}$ for *i*-th alternative, then the entropy is maximal and equal to the Hartley's information *I* (*N*) (Hartley's measure of *nonspecifity*) [6]:

$$I(N) = \log_2 N \tag{5}$$

Hence, the total maximum entropy H_{MAX} for all *N* alternatives is given as:

$$H_{MAX} = N\log_2 N \tag{6}$$

And the relative entropy H_r is given as:

$$H_r = \frac{H_T}{H_{MAX}} \tag{7}$$

4 A concordance among methods

To measure an agreement among methods, Kendall's *W* (Kendall's coefficient of concordance) is used [5]:

$$W = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{\frac{1}{12}k^2\left(n^3 - n\right)} \tag{8}$$

In the relation (8), *k* is the number of rankings, X_i is the sum of placings for *i*-th object by all DMs, and *n* the number of objects or alternatives compared, and $W \in \langle 0,1 \rangle$. The higher is value of *W*, the higher is agreement among decision makers. *W* = 0 in a case of absolute disagreement among decision makers, while *W* = 1 expresses an absolute agreement among DMs.

However, Kendall's *W* is not an appropriate measure of concordance when there is a large number of tied rankings. In such case, the relation (8) has to be corrected (for details see [7]).

5 Numerical examples

In this section numerical examples are provided. The entropy of DMs' preferences is evaluated, the final consensus ranking is achieved by all four methods and subsequently the agreement among methods' solutions is estimated via Kendall's coefficient of concordance.

Example 1. *Four decision makers (DM1 to DM4) give rankings of four alternatives A, B, C and D from the best to the worst in two cases a) and b). Find the consensus ranking.*
a) DM1 = (A, B, C, D), DM2 = (B, C, D, A), DM3 = (D, A, B, C) and DM4 = (C, D, A, B).
b) DM1 = (A, B, C, D), DM2 = (B, A, D, C), DM3 = (A, D, B, C) and DM4 = (A, D, C, B).

a) All alternatives are ranked equally, so the consensus ranking cannot be found.

b) Alternative A is the best ranked alternative, as it is ranked the first three times and the second once. The methods give the following final consensus:

- BAK: (A, B, D, C),
- MAH: (A, B, D, C),

- CRM: (A, D, B, C),
- DCM: (A, B, D, C) and (A, D, B, C).

The difference between a) and b) rests in the fact that the information on objects' preferences in the case b) is more specific than in the case a), which, in fact, demonstrates an absolute uncertainty within a group.

Example 2. *Evaluate Shannon's entropy of cases a) and b) from Example 1.*

a) The fuzzy ranking (obtained from the definition 1) is shown in Table 1. From (4) and $N = 4$ we get $H = 2$ for each row (column) of the table, hence the total entropy $H_T = 8$ and the relative entropy $H_r = 1$.

Alternative/position	1.	2.	3.	4.
A	0.25	0.25	0.25	0.25
B	0.25	0.25	0.25	0.25
C	0.25	0.25	0.25	0.25
D	0.25	0.25	0.25	0.25

Table 1. The fuzzy rankings of alternatives A, B, C and D in the case a).

b) The fuzzy ranking is shown in Table 2. From (4) and $N = 4$ we get $H_1 = 0.811$, $H_2 = 2$, $H_3 = 1$ and $H_4 = 1.5$, the total entropy $H_T = 5.311$ and the relative entropy $H_r = 0.664$.

Alternative/position	1.	2.	3.	4.
A	0.75	0.25	0	0
B	0.25	0.25	0.25	0.25
C	0	0	0.5	0.5
D	0	0.5	0.25	0.25

Table 2. The fuzzy rankings of alternatives A, B, C and D in the case b).

Example 3. *In Tavana et al. [10], five projects A, B, C, D, and E were ranked by seven DMs (see Table 3). Evaluate the entropy of decision makers' preferences and estimate Kendall's W.*

Projects' fuzzy rankings are shown in Table 4. The relative entropy of decision makers' preferences estimated from (7): $H_r = 0.837$. As for the problem's solution, we get the following consensus rankings:

- BAK: (C,D,A,E,B),
- MAH: (D,C,A,E,B),
- CRM: (C,E,B,A,D),
- DCM: (D,C,E,A,B).

The agreement among methods evaluated by (8): $W = 0.538$. Thus, the high relative entropy of initial preferences resulted in only moderate agreement among methods' solutions.

DM/ranking	1.	2.	3.	4.	5.
DM1	C	A	B	D	E
DM2	C	B	E	A	D
DM3	D	C	A	E	B
DM4	D	E	B	A	C
DM5	C	B	E	D	A
DM6	A	D	C	E	B
DM7	D	E	A	C	B

Table 3. DMs' preferences of of projects A, B, C, D, E.

Project/ranking	1.	2.	3.	4.	5.
A	1/7	1/7	2/7	2/7	1/7
B	0	2/7	2/7	0	3/7
C	3/7	1/7	1/7	1/7	1/7
D	3/7	1/7	0	2/7	1/7
E	0	2/7	2/7	2/7	1/7

Table 4. Projects fuzzy rankings.

6 Experimental evaluation

To evaluate the relationship between the relative entropy (H_r) and the agreement among methods (W) more precisely, we implemented a tool for solving the OCRP problem. This tool generates an OCRP problem and then solves it using the four methods described in Section 2, namely the BAK, MAH, CRM and DCM methods. The generation is done using the Monte Carlo algorithm, which randomly generates preferences of each decision maker. Then, the relative entropy of the input data (DMs' preferences) is evaluated and each of the four methods computes the final consensus. At last step, the correlation between the relative entropy (H_r) and the agreement among the four methods (W) is evaluated. The results are summarized in Table 5 below, where the correlation between the H_r and W can be found.

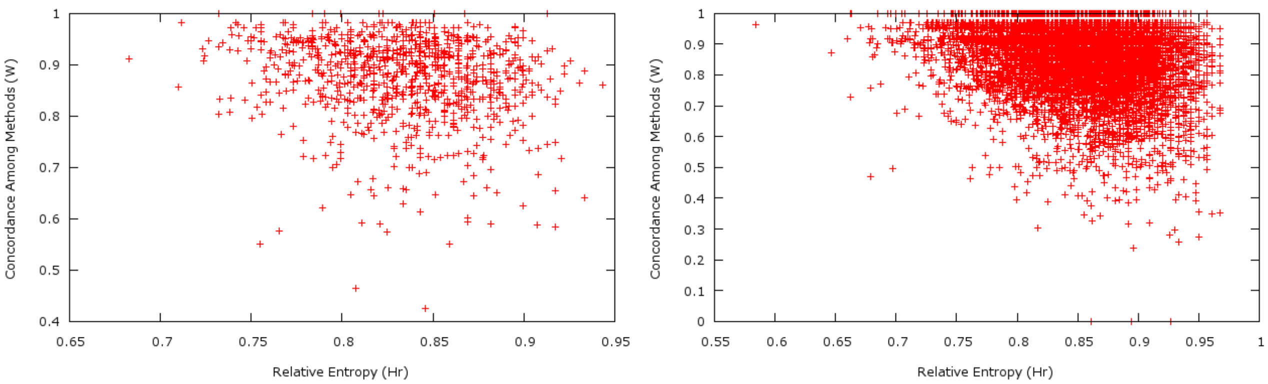
Each row of the table contains a correlation between the relative entropy and the agreement among methods (W(ALL)) as well as the agreement in within each method (W(BAK), W(MAH), W(CRM), W(DCM)). The correlation was evaluated on a huge set of randomly generated problems (at least 1000 problems) for various number of alternatives and decision makers. The last row then shows the average correlation between H_r and W.

Alternatives	**Decision Makers**	**Generated Problems**	**Correlation between H_r and W**				
			***W*(BAK)**	***W*(MAH)**	***W*(CRM)**	***W*(DCM)**	***W*(ALL)**
5	4	1000	-0.1128	-0.1203	-0.1214	-0.2257	-0.2444
5	10	1000	-0.1186	-0.1489	-0.0948	-0.1378	-0.1683
5	100	1000	-0.0753	-0.0871	-0.0711	-0.1234	-0.2126
6	4	1000	-0.1725	-0.1137	-0.0094	-0.0827	-0.1466
6	10	1000	-0.1161	-0.1089	-0.0688	-0.1240	-0.1736
6	100	1000	-0.0555	-0.0611	-0.0169	-0.0994	-0.1453
7	4	1000	-0.0885	-0.0948	-0.0865	-0.1134	-0.1213
7	10	1000	-0.1415	-0.0653	-0.0499	-0.1645	-0.1912
7	50	1000	-0.0563	-0.0197	0.0082	-0.0679	-0.1529
4	4	10000	-	-	-	-	-0.2564
4	10	10000	-	-	-	-	-0.1969
5	4	10000	-	-	-	-	-0.2017
5	10	10000	-	-	-	-	-0.1905
6	4	10000	-	-	-	-	-0.1872
6	10	10000	-	-	-	-	-0.1550
		Average	-0.1041	-0.0910	-0.0567	-0.1265	-0.1829

Table 5. The correlation between the relative entropy and the concordance among methods.

As can be seen from Table 5, the average concordance among methods (W(ALL)) was inversely proportional to the relative entropy with the Pearson's $r = -0.18$, which was, for N = 69 000 generated problems, statistically significant at $\alpha = 0.01$ level. The larger was the number of decision makers, the better was the concordance among methods. But this relation might be a little distorted because with the increasing number of decision makers the relative entropy falls into a very narrow interval. So, in practice, the relationship between the relative entropy and the concordance among methods might be more significant, i.e., more negative.

In case of the concordance in within each of the four methods, it is evident that some methods are dependent on the entropy more than the others, i.e., the higher the entropy, the more and more distinct solutions to the



a) 6 alternatives, 10 decision makers, 1000 problems **b)** 5 alternatives, 10 decision makers, 10000 problems

Figure 1. The relationship between the relative entropy and the concordance among methods.

OCRP problem is achieved. From the four studied methods, the DCM method seems to be the most dependant one. This method often gives many solutions where e.g. the CRM method produces only one. On the other hand, although the CRM method computes mostly a single solution to the OCRP problem, its computation complexity is significantly higher than with the other methods.

More detailed analysis of the relationship between the relative entropy (H_r) and the concordance among methods (W) reveals that the entropy influences the probability of obtaining similar solutions from all of the methods, i.e., the higher the entropy, the higher is the probability to obtain distinct solutions from each of the methods. Figure 1 shows the relationship between the relative entropy (H_r) and the concordance among methods (W) for two of the tests performed. From Figure 1(a) it may not be obvious how the entropy influence the probability of distinct solutions, but in Figure 1(b) can be clearly seen the increasing number of low values of concordance between the methods when the value of entropy gets higher. So, higher entropy does not mean that one cannot obtain the same single solution (consensus) from all of the four methods, but there is a lower probability of getting such a solution if the entropy is high.

7 Conclusions

The aim of the article was to introduce Shannon's entropy and the relative entropy into ordinal consensus ranking problem framework, and to empirically explore the relationship between the relative entropy of DMs' preferences (H_r) and the concordance (W) among four selected methods for OCRP solution by Monte Carlo simulations. Our results indicate there is a statistically significant negative relationship between H_r and W, that is when the entropy of initial preferences is lower, the consensus is much more likely to be achieved than in cases with the higher entropy. Also, it was found that the larger is the number of decision makers, the better is the concordance among methods. As for the selected methods, the DCM method produced the largest number of solutions globally (it didn't produced a single consensus in many cases), while the CRM method was the most computationally complex, but produced the smallest number of solutions.

References

[1] Beck, M. P., Lin, B. W. Some heuristics for the consensus ranking problem. *Computers and Operations Research* 1983; Vol. 10, Nr. 1, p. 1–7.

[2] Cook, W.D. Distance-Based and Ad Hoc Consensus Models in Ordinal Preference Ranking. *European Journal of Operational Research*, Nr. 172 (2006), p. 369-385.

[3] Cook, W. D., Kress, M. Ordinal ranking with intensity of preference. *Management Science*, 1985, vol. 31, No. 1, p. 26-32.

[4] Kendall, M. *Rank Correlation Methods*, Hafner, New York, NY, USA, 3rd edition, 1962.

[5] Kendall, M., Babington-Smith, B. The Problem of *m* Rankings. 1939, *The Annals of Mathematical Statistics*, 1939, Vol. 10, Nr. 3, p. 275–287.

[6] Klir, G., Folger, T. A. *Fuzzy Sets, Uncertainty, and Information*. State University of New York: 1988.

[7] Legendre, P. Species Associations: the Kendall Coefficient of Concordance Revisited. *Journal of Agricultural, Biological and Environmental Statistics*, 2005, vol. 10, Nr. 2, p. 226–245.

[8] Mazurek, J. Group Decision Making in Ordinal Consensus Ranking Problem. In: *Proceedings of the 28th International Conference on Mathematical methods in Economy*, České Budějovice, 2010.

[9] Shannon, C .E. The mathematical theory of communication. *The Bell System Technical Journal*, 1948, Nr. 27, p. 379-423, 623-656.

[10] Tavana, M., LoPinto, F., Smither, J. W. A Hybrid Distance-Based Ideal-Seeking Consensus Ranking Model. *Journal of Applied Mathematics and Decision Sciences*, vol. 2007, 18 pages.

Efficiency evaluation of selected agricultural enterprises considering also negative data

Ludvík Friebel[1], Jana Friebelová[1]

Abstract: Data Envelopment Analysis (DEA) is a non-parametric method for measuring the efficiency of decision making unit (DMU) based on chosen inputs and outputs. The originally proposed model requires positive values of the inputs and the outputs of evaluated DMU. Later, different alternative approaches applicable to DMU characterized by negative inputs or outputs have been described. The following models are involved: Generic directional distance model, Range directional measures and Semi-oriented radial measure.

The aim of this work is to evaluate chosen agricultural enterprises in the Czech Republic. Some of these enterprises make negative profits. We would like to compare the possibilities either using the above mentioned alternative methods or using the basic DEA method considering the division of negative profit into two parts: the yield and the cost (both positive values). The obtained results and recommendations are further discussed in the paper.

Keywords: data envelopment analysis (DEA), Generic directional distance model (GDDM), Range directional measures (RDM), Semi-oriented radial measure (SORM)

JEL Classification: C02, C44, C60
AMS Classification: 90C05, 90B50

1. Introduction

DEA models serve for assessment of technical efficiency of production units based on consumed inputs and produced outputs. One requirement for basic models CCR [5] and BCC [2] is that the values of inputs and outputs cannot be negative, more for example [9]. Later, different ways, which allow to assess also units with negative inputs/outputs, were proposed, for example Modified slacks-based model (Sharp et al [12]), Range directional model - RDM (Portela et al. [11]), Semi-oriented radial measure - SORM (Emrouznejad et al. [6]).

The paper deals with assessment of efficiency of farms in potato growing areas, regarding hectare yield from the most common grown plants such as wheat, barley, and rape. The chosen potato growing region was chosen, because that area constitutes the largest part of the arable land in the Czech Republic. The area is characterized by typical altitudes of 400-650 m above the sea level, its total cultivated land is 60-80% and the average slope is 5-12%. Another reason for this choice was the homogeneity regarding the grain production structure.

Many authors have devoted their works to the application of DEA method the agricultural context. See, for instance, Gaspar [7], Ahmad and Bravo-Ureta [1], Wilson, Hadley and Asby [13], Iráizoz, Rapún and Zabaleta [8], Lansink and Reinhard [10], among others.

In our work, the assessment was conducted on the basis of three inputs – shares on the area of the arable land for particular plants, and three outputs – the hectare yield of these plants. Some farms record a negative profit and therefore it is necessary to use the above mentioned models, which can work with negative data. As we had records about costs and revenues, we divided profit, in the next part of our paper, into two parts which led to the fact that there were no negative values anymore and we could apply the basic DEA model. All assessments were made using output-orientated models considering Variable Returns to Scale - VRS.

The aim of the paper is to assess efficiency of chosen farms using various DEA models, and to find recommendations for inefficient farms and to compare the obtained results between each other.

2. Methodology

Let us consider a set of n DMUs $(\mathrm{DMU}_j\,(j=1,\dots,n))$ using m inputs $(X_{ij}, i=1,\dots,m)$ to secure s outputs $(Y_{rj}, r=1,\dots,s)$.

[1] University of South Bohemia, Faculty of Economics, Department of Applied mathematics and informatics, Studentská 13, 370 05 České Budějovice, Czech Republic, friebel@ef.jcu.cz, friebelova@ef.jcu.cz

BCC model

Output oriented BCC model was developed by Banker et al. [2] and is defined as follows:

$$
\begin{aligned}
&\text{Max } h \\
&\text{s.t. } \sum_j \lambda_j X_{ij} \le X_{ij_o}; \quad \forall i \\
&\sum_j \lambda_j Y_{rj} \ge h Y_{rj_o}; \quad \forall r \\
&\sum_j \lambda_j = 1 \\
&\lambda_j \ge 0,
\end{aligned} \tag{1}
$$

where j_o symbolizes evaluated unit. This model enables us to assess only units with positive values of inputs and outputs. Efficiency is determined as $1/h^*$, where h^* is the optimal value of objective function.

In case that a negative output can be divided into a positive input and a positive output, we can consider to use the standard BCC model (1). After solving the model, it is necessary recalculate the resulting recommendation to the original output.

SORM model

SORM model was developed by Emrouznejad et al. [6]. This model is derived from the classical BCC model and its improvement is based on the special treatment of the inputs or outputs with negative values (see bellow). Output oriented SORM model in case of VRS is defined as follows:

$$
\begin{aligned}
&\text{Max } h \\
&\text{s.t.} \sum_j \lambda_j X_{ij} \le X_{ij_o}; \forall i \\
&\sum_j \lambda_j Y_{rj} \ge h Y_{rj_o}; \forall r \ne k \\
&\sum_j \lambda_j Y_{kj}^1 \ge h Y_{kj_o}^1 \\
&\sum_j \lambda_j Y_{kj}^2 \le h Y_{kj_o}^2 \\
&\sum_j \lambda_j = 1 \\
&\lambda_j \ge 0; \forall j, h \text{ free.}
\end{aligned} \tag{2}
$$

Y_k and Y_r represent outputs where Y_k is negative for some units and positive for others and Y_r is positive for all DMU. This model can be also used for assessment of units, some of whose outputs can be for some units positive and for the other negative. Regarding these outputs, we differentiate whether outputs for the given unit are positive or negative and two variables Y_k^1 and Y_k^2 are introduced:

$$
\begin{aligned}
&Y_k^1 = Y_k \text{ and } Y_k^2 = 0 \text{ if } Y_k \ge 0, \\
&Y_k^2 = -Y_k \text{ and } Y_k^1 = 0 \text{ if } Y_k < 0.
\end{aligned} \tag{3}
$$

The resulting efficiency is for model (2) calculated as $1/h^*$, where h^* is the optimal value of objective function.

RDM$^+$ Model

Model RDM$^+$ was introduced by Portela [11] in year 2004. This model comes from GDDM (Chambers., [3], [4]) and is defined in the following way:

$$
\begin{aligned}
&\text{Max } \beta_0 \\
&\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{r0} + \beta_0 R_{r0};\ r = 1,\ldots,s \\
&\sum_{j=1}^{n} \lambda_j X_{ij} \leq X_{i0} - \beta_0 R_{i0};\ i = 1,\ldots,m \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j, \beta_0 \geq 0,
\end{aligned} \tag{4}
$$

where (R_{i0}, R_{r0}) are the directions of DMU*j* to the ideal point:

$$
\begin{aligned}
R_{i0} &= X_{i0} = \min\{X_{ij};\ j = 1,\ldots,n\}, i = 1,\ldots,m, \\
R_{r0} &= \max\{Y_{rj};\ j = 1,\ldots,n\} - Y_{r0}, r = 1,\ldots,s.
\end{aligned} \tag{5}
$$

Efficiency is calculated as 1-β^*, where β^* is the optimal value of objective function of model (4).

In order to compare RDM$^+$ model with the above mentioned output oriented models, we set $R_{i0} = 0$, then we have β only for outputs. The final output oriented RDM$^+$ model is defined as follows:

$$
\begin{aligned}
&\text{Max } \beta_0 \\
&\sum_{j=1}^{n} \lambda_j Y_{rj} \geq Y_{r0} + \beta_0 R_{r0};\ r = 1,\ldots,s \\
&\sum_{j=1}^{n} \lambda_j X_{ij} \leq X_{i0};\ i = 1,\ldots,m \\
&\sum_{j=1}^{n} \lambda_j = 1 \\
&\lambda_j;\ \forall j, \beta_0 \geq 0.
\end{aligned} \tag{6}
$$

3. Results and discussion

We considered a group of 31 farms of similar characteristics (in the potato growing region). As inputs, we took the ratio of land area designated for growing wheat, barley and rape, or also possibly costs of particular plants per hectare in CZK. As outputs, we took yield of the afore-mentioned crops in CZK per hectare, or also possibly revenues from their sale calculated per hectare (see Table 1). The source data are available at http://home.ef.jcu.cz/research/DEA/negative/.

The tasks for of above mentioned models were made in Maple 14 environment. Solutions were obtained using function LPSolve and checked using Simplex library.

Model with negative inputs/outputs		Model with non-negative inputs/outputs	
Symbol	Description	Symbol	Description
x_1	ratio of soil for growing wheat	x_1	ratio of soil for growing wheat
x_2	ratio of soil for barley	x_2	ratio of soil for barley
x_3	ratio of soil for rape	x_3	ratio of soil for rape
y_1	profit per hectare of growing wheat [CZK]	x_4	cost per hectare of growing wheat [CZK]
		y_1	revenue per hectare of growing wheat [CZK]
y_2	profit per hectare barley [CZK]	x_5	cost per hectare barley [CZK]
		y_2	revenue per hectare barley [CZK]
y_3	profit per hectare rape [CZK]	x_6	cost per hectare rape [CZK]
		y_3	revenue per hectare rape [CZK]

Table 1: Description of inputs and outputs

i	SORM	RDM	BCC	*i*	SORM	RDM	BCC
1	1.0000	0.9977	1.0000	17	0.6654	0.6802	1.0000
2	1.0000	0.8621	1.0000	18	0.0481	0.4452	1.0000
3	0.3307	0.4272	0.9043	19	0.1448	0.3573	0.8808
4	0.0661	0.3539	0.7764	20	1.0000	1.0000	1.0000
5	1.0000	1.0000	1.0000	21	1.0000	1.0000	1.0000
6	0.3717	0.6576	1.0000	22	1.0000	1.0000	1.0000
7	0.8006	0.8202	1.0000	23	1.0000	1.0000	1.0000
8	1.0000	1.0000	1.0000	24	1.0000	1.0000	1.0000
9	0.1142	0.3873	0.8630	25	0.0001	0.6264	0.8951
10	0.2237	0.3432	0.8918	26	0.5213	0.5006	1.0000
11	0.1293	0.4480	0.8284	27	1.0000	0.6992	1.0000
12	0.1050	0.4640	0.9596	28	0.3599	0.6047	0.8876
13	0.5692	0.6962	0.9448	29	0.1578	0.5492	0.8541
14	1.0000	1.0000	1.0000	30	1.0000	1.0000	1.0000
15	1.0000	1.0000	1.0000	31	1.0000	1.0000	1.0000
16	1.0000	1.0000	1.0000				

Table 2: Efficiency comparison

In table 2, there are presented efficiency units where at least one of their outputs was positive. Units, whose outputs with only negative values (profit per hectare of particular plant was negative) had to be excluded from the analysis. The reason was SORM model, which was in this case unbounded.

SORM marked 15 units as fully effective. As problematic, we considered, using SORM model, extremely low resulting efficiency of units 4 and 25. The efficiency level of units which are not at the efficiency frontier is significantly lower in total than in the case of other two models.

Model RDM$^+$ marked 12 units as effective. Three units, which were SORM effective, are not present here. The efficiency level of ineffective units is higher than in model SORM.

Model BCC marked 20 units as effective, which can be explained by doubling the number of inputs which is caused by the division of every outputs into one input and one output in order to remove negative values, which has been mentioned above. The total efficiency level is naturally higher.

	Change of profit for wheat			Change of profit for barley			Change of profit for rope		
i	SORM	RDM	BCC	SORM	RDM	BCC	SORM	RDM	BCC
1	0	8	0	0	13	0	0	18	0
2	0	1992	0	0	1200	0	0	853	0
3	3182	4686	3468	5411	6539	1032	10507	6019	2478
4	5067	6527	5296	5400	7976	2485	13670	9518	5717

6	2162	2144	0	2022	4115	0	5819	5263	0
7	663	877	0	2870	1637	0	7607	7407	0
9	2528	4599	2919	7396	7816	2128	12954	8596	3612
10	3994	5906	3898	10081	10566	2951	12187	8005	2690
11	4301	3809	3564	3286	8186	6121	9258	11653	4073
12	3634	4672	1137	1792	5326	707	11886	7667	1172
13	5246	5246	5311	9167	9167	8319	3403	3403	1491
17	1230	1629	0	1883	1883	0	3502	2793	0
18	4953	4044	0	992	6609	0	8679	9620	0
19	6451	7349	4646	5680	7206	1755	12752	8703	3151
25	4562	2816	1790	438	3601	1567	1995	5874	2639
26	3609	1802	0	5429	5774	0	4558	7219	0
27	0	1830	0	0	3515	0	0	3438	0
28	3378	3421	2696	3295	3076	1823	6509	6583	5982
29	3122	3726	3206	1483	5050	1677	9685	6259	3725

Table 3: Necessary changes for reaching efficiency

Table 3 presents proposed changes of outputs for particular plants and models. Model BCC can be, logically, considered as the least fierce, as its corrections are caused by the higher number of inputs. The resulting values of profit for this model were obtained by subtracting of the resulting value of revenue and the original costs. On the contrary, model SORM proposes very often an increase of profit by more than 10 thousand CZK, e. g. for unit 4 (5067, 5400, 13670), whose efficiency was assessed to mere 6%; similarly unit 19, whose efficiency was 14%. Regarding the efficiency, relatively low changes are proposed by this model for unit 25 (4562, 438, 1995), whose efficiency was determined to 0.1 ‰.

4. Conclusion

Regarding functionality, we can say that all three analyzed models can cope with negative values. A certain exception is SORM which does not offer solutions for units with all the values of inputs, resp. outputs negative. These units were excluded from the conducted comparison. BCC model can be successfully used in those cases, where it is possible to divide negative outputs into a positive input and a positive output, as it is in our case. Generally speaking, it would be possible, similarly, to divide also a negative input.

Regarding the assessment, it is interesting that model SORM marks as effective more units than model RDM+. Model SORM gives very low efficiency values in case of low effective units. This phenomenon can be explained by a different calculation. The meaning of efficiency, regarding this model, is rather controversial. Model BCC assesses as effective much more units, which caused by doubling the number of inputs thanks to the division of negative outputs into a positive input and a positive output.

Particular models give different results regarding the recommendations for the ineffective units. Models RDM and BCC show a certain relation with the reached efficiency. It is not the case of model SORM.

When speaking about the mentioned facts, we would propose in cases of negative outputs to use model RDM+, which seems to be more robust and it can process also units with outputs, whose outputs are all negative. Also efficiency values are more easily interpreted.

References

[1] Ahmad, M., Bravo-Ureta, B.: Technical efficiency measures for dairy farms using panel data: a comparison of alternative model specifications. *The Journal of Productivity Analysis* 7 (1996), 399–415.

[2] Banker, R. D., Charles, A., Cooper, W. W.: Some models for estimating technical and scale inefficiencies in DEA. *Management Science*, vol. 30 (1984), 1078–1092.

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

[4] Chambers, R. G., Chung, Y., Färe, R.: Profit, directional distance functions and Nerlovian efficiency. *Journal of Optimization Theory and Applications* 98(2) (1998), 351–364.

[5] Charnes, A., Cooper, W. W., Rhodes, E.: Measuring the efficiency of decision making units, *European Journal of Operational Research*, vol. 2, no. 6 (1978), 429–444.

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

[7] Gaspar, P., Mesias, F.J., Escribano, M., Pulino, F.: Assesing the technical of extensive livestock farming systems in Extremadura, Spain. *Livestock Science*, vol. 121(1) (2009), 7–14.
[8] Iráizoz, B., Rapún, M., Zabaleta, I.: Assessing the technical efficiency of horticultural production in Navarra, Spain. *Agricultural Systems* 78 (2003), 387–403.
[9] Jablonský,J.: Models for Efficiency Evaluation of Decision Making Units. *Politická ekonomie* 52(2) (2004), 206–220.
[10] Lansink, A. O., Reinhard, S.: Investigating technical efficiency and potential technological change in Dutch pig farming. *Agricultural Systems* 79 (2004), 353–367.
[11] Portela, M.C.A.S., Thanassoulis, E. and Simpson, G.P.M.: Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society* 55(10) (2004), 1111–1121.
[12] Sharp. J. A., Liu, W. B., Meng, W.: A modified slacks-based measure model for data envelopment analysis with 'natural' negative outputs and inputs. *Journal of the Operational Research Society* 57(11) (2006), 1–6.
[13] Wilson, P., Hadley, D., Asby, C.: The influence of management characteristics on the technical efficiency of wheat farmers in Eastern England. *Agricultural Economics* 24 (2001), 329–338.

Financial frictions during the period of the global economic recession: A DSGE model of the Czech economy

Pavel Fryblík[1], Osvald Vašíček[2]

Abstract. This paper deals with analysis of financial market frictions of Czech economy using DSGE model framework. The methodology is borrowed from Alpanda, Koetze, Woglom (2010) who based their model on Justiniano, Preston (2009) and specifics of the financial market during global economic crisis are examined. This model features staggered wages and prices, incomplete pass-through of exchange rate and incomplete international asset markets. Computations and results are provided by Matlab's toolbox Dynare. The research is divided into two parts. In the first one, Czech economy from year 1996 till the end of 2010 is analyzed by DSGE model with financial accelerator using shock decompositions. This gives us a prove about impact of risk premium and financial accelerator on the output. Furthermore, it determines also effect of shocks on observed variables. In the second part, method of recursive estimations is used. Thus, we are able to study quantity of financial accelerator and other parameters which are now changing across the time. Those parameters change the structure of model and influence development of the economy.

Keywords: DSGE model, Small Open Economy, Financial Friction, Financial Accelerator, Bayesian Estimation

JEL classification: C44
AMS classification: 90C15

1 Introduction

Main goal of this paper is to investigate the Czech economy during global financial crisis. First, we look on the shocks, which have influenced Czech output and have caused the recent crisis in the Czech economy. Then, we observe changes of structural parameters values before and during the worldwide financial crisis and their impact on significant macroeconomic variables. Moreover, we expect to find possible influences and changes of the parameter, which has the role of financial accelerator.

To analyze the economic development we use small open economy DSGE (Dynamic Stochastic General Equilibrium) model. Following model was developed by Alpanda *et al.*(2010). This model uses also less common feature of both incomplete international asset markets and the risk premium on domestic assets relative to foreign assets. Following Bernake *et al.* (1994), the main idea of financial accelerator is that fluctuations in borrowers' net worth lead to fluctuations in real activity. This mechanism is transferred by debt-elasticity premium in budget constraints of households.

2 Model

Following Steinbach *et al.* (2009), the model is small open economy New Keynesian DSGE model. Monopolistically competitive firms are price-setters in the goods market, and households are wage-setters in the labor market. Nominal rigidities are introduced through staggered price and wage-setting and through indexation of domestic goods prices and wages to past inflation.

[1]Masaryk University, Faculty of economics and administration, Department of economics, Lipova 41a, 60200, Brno, Czech Republic, e-mail: 364089@econ.muni.cz

[2]Masaryk University, Faculty of economics and administration, Department of economics, Lipova 41a, 60200, Brno, Czech Republic, e-mail: osvald.vasicek@econ.muni.cz

Domestic sector consists of households and three types of firms: intermediate goods producers, final goods producers and domestic importing retailers.

Representative household seeks to maximize its lifetime utility function

$$E_t \sum_{t=0}^{\infty} \beta^t \tilde{\mu}_t^d \left[\frac{(C_t - \upsilon C_{t-1})^{1-\sigma}}{1-\sigma} - \frac{N_t(i)^{1+\varphi}}{1+\varphi} \right], \tag{1}$$

where parameter β is discount factor, N_t is per capita employment, υC_{t-1} is an external habit taken as exogenous by the household. Parameter σ stands for risk aversion and parameter φ is inverse of labor supply elasticity. Domestic demand shock is represented by $\tilde{\mu}_t^d$, whose natural logarithm is an AR(1) process. Households face sequence of budget constraints

$$P_{h,t}C_{h,t} + P_{f,t}C_{f,t} + D_t + \tilde{e}_t B_t = D_{t-1}(1+i_{t-1}) + \tilde{e}_t B_{t-1}(1+i_{t-1}^*)\phi_t(NFA_t) + W_t N_t, \tag{2}$$

for all $t > 0$, where D_t represents one period domestic bond which yields domestic nominal interest rate i_t. B_t is one period foreign bonds. Yield of foreign bonds is given by foreign nominal interest rate i_t^* multiplied by the nominal exchange rate $\tilde{e}_t$ and is influenced by debt-elastic interest rate premium, which is described below. Following Justiniano *et al.*(2010), the function $\Phi_t(.)$ can be interpreted as the debt elastic interest rate premium and follows the function

$$\Phi_t = exp\left[-\chi NFA_t + \tilde{\phi}_t\right], \tag{3}$$

where NFA_t is net foreign assets and is given as

$$NFA_t = \frac{\tilde{e}_{t-1}B_{t-1}}{\bar{Y}P_{t-1}}. \tag{4}$$

Net foreign assets is the real quantity of outstanding foreign debt expressed in terms of domestic currency as a fraction of steady state output and $\tilde{\phi}_t$, a risk premium shock, whose natural logarithm is AR(1) process. χ means a debt-elasticity of risk premium. According to Bernanke *et al.* (1994), debt-elasticity provides connection between macroeconomic variables and domestic households‘ debt abroad. Thus, debt-elasticity χ can be regarded as the financial accelerator. In our model we have two different bonds, namely domestic and foreign bonds. We have to place restriction on the relative movements of domestic and foreign interest rate, and changes in the nominal exchange rate. We assume incomplete financial markets and the uncovered interest rate parity condition can be written as

$$E_t\left[(1+r_t) - (1+r_t^*)\frac{\tilde{e}_{t+1}}{\tilde{e}_t}\Phi_{t+1}\right] = 0, \tag{5}$$

where r_t and r_t^* are domestic and foreign real interest rate, respectively. The real exchange rate is defined as $\tilde{q}_t = \tilde{e}_t P_t^*/P_t$, where P_t^* is foreign price index and P_t is domestic price index. Then $log(\tilde{q}_t) = log(\tilde{e}_t) + log(P_t^*) - log(P_t)$ and we write $q_t = e_t + p_t^* - p_t$.

The production is assumed to take place in two stages. First, each one of a continuum of identical intermediate goods producing firms (indexed by j, where $j \in [0,1]$) produces a differentiated goods with the same technology level. These firms are monopolistically competitive and set their prices in a staggered manner á la Calvo(1983). Second, final goods producers operate in a perfectly competitive environment and combine the differentiated intermediate goods into final product, which is then sold to households.

Each intermediate goods producing firm produces a differentiated goods $Y_t(j)$ using linear technology production function

$$Y_t(j) = A_t N_t(j), \tag{6}$$

where A_t is technology used in period t. $a_t = log(A_t)$ follows an AR(1) process. The labor input of each firm $N_t(j)$ is described by the constant elasticity of substitution (CES) composite of individual households‘ labor supply.

Firms are facing a problem in form

$$\max_{\bar{P}_{h,t}} E_t \sum_{k=0}^{\infty} \theta_h^k Q_{t,t+k}[Y_{t+k}(j)\bar{P}_{h,t}(j)\Pi_{h,t+k-1}^{\delta} - TC_{t+k}(Y_{t+k}(j))], \tag{7}$$

subject to the demand for intermediate goods function by final goods producers

$$Y_t(j) = \left[\frac{\bar{P}_{h,t}(j)\Pi^{\delta}_{h,t-1}}{P_{h,t}}\right]^{-\xi_h} Y_t, \tag{8}$$

where δ is the degree of partial indexation to the last period domestic inflation $\Pi_{h,t-1}$, $\bar{P}_{h,t}(j)$ is the price set by firm j and $TC_{t+k}(Y_{t+k}(j))$ is function of total costs.

All final goods producing firms choose the same price since they face an identical problem and are perfectly competitive. Final goods is produced by using intermediate goods and for final goods producers it is their only input. They employ the following CES technology production function

$$Y_t = \left[\int_0^1 Y_t(j)^{\frac{\xi_h - 1}{\xi_h}} dj\right]^{\frac{\xi_h}{\xi_h - 1}}, \tag{9}$$

where $\xi_h > 1$ is the elasticity of substitution between types of differentiated domestic goods. Final goods producers demand for intermediate goods is given by (8).

The last type of firms in our model are the retailers who import foreign goods. Furthermore, when importing foreign goods they pay the world-market price in terms of domestic currency. Accordingly, these are their only costs. Similarly to the domestic firms, the domestic importing retailers exhibit Calvo-style price setting behaviour, and hence are able to reset the domestic retail price of the imported good in period t with probability $1 - \theta_f$. Thus, the importing retailer needs to set price $\bar{P}_{f.t}(j)$ for imported good j such as to maximize

$$\max_{\bar{P}_{f,t}} E_t \sum_{k=0}^{\infty} \theta_f^k Q_{t,t+k} C_{f,t+k}[\bar{P}_{f.t}(j) - \tilde{e}_{t+k} P^*_{t+k}(j)], \tag{10}$$

subject to demand function for imported goods

$$C_{f,t}(j) = \left[\frac{\bar{P}_{f.t}(j)}{P_{f,t}}\right]^{-\xi_f} C_{f,t}, \tag{11}$$

where $\xi_f > 1$ is the elasticity of substitution between types of differentiated foreign goods, $P^*_{t+k}(j)$ is the price paid for good j in the world market, and hence $\tilde{e}_{t+k}P^*_{t+k}(j)$ is the domestic currency price paid by local importer.

Goods market clearing in the domestic economy requires

$$Y_{h,t} = C_{h,t} + C^*_{h,t}. \tag{12}$$

Foreign demand for the domestically produced goods $C^*_{h,t}$ is specified as

$$C^*_{h,t} = \left(\frac{P_{h,t}}{P^*_t}\right)^{-\eta} Y^*_t,$$

where $\eta > 0$ is the elasticity of substitution between domestic and foreign goods.

We can close our model by specifying a monetary policy which follows Taylor rule in form

$$i_t = \rho_t i_{t-1} + (1 - \rho_t)[\phi_\pi \pi_t + \phi_y(\Delta y_t)] + \epsilon^i_t, \tag{13}$$

where ρ_t is degree of interest rate smoothing, ϕ_π and ϕ_y are the relative weights of inflation and output growth in the response function. $\epsilon^i_t \sim i.i.d.\ N(0, \sigma^2_i)$ captures innovations. An analogous Taylor rule holds for the foreign sector.

Following Steinbach *et al.* (2009), the foreign economy in the context of this model is modeled as closed economy version of the domestic economy, hence $\gamma = 0$. The domestic economy cannot influence the foreign economy, thus $y^*_t = c^*_t$. Since the foreign economy is closed, there are no imported goods, and hence no imported inflation component, so $\pi^*_t = \pi^*$.

The model features ten exogenous stochastic shocks, which causes dynamics in the model.

1. Demand shock in domestic μ_t^d and foreign $\mu_t^{d^*}$ economy are AR(1) processes. These shocks influence utility function and Euler function, respectively, and have direct effect on consumption.
2. Technology shock in domestic (a_t) and foreign (a_t^*) economy are AR(1) processes and influence technological level of each country. Positive shock of this type decreases marginal costs and is responsible for inflation decrement.
3. Domestic cost-push shock η_t^p is linked to marginal costs and is AR(1) process. As well as technology shock, cost-push shock influences costs and inflation in the country.
4. Wage cost-push shock in domestic (η_t^w) and foreign ($\eta_t^{w^*}$) economy are AR(1) processes and are included in wage markup equation. Wage cost-push shocks are connected to wage inflation.
5. Risk premium shock ϕ_t is AR(1) process and is included simultaneously with debt and debt-elasticity in uncovered interest parity equation. Therefore, the shock influences the real exchange rate.
6. Monetary shocks in domestic (ϵ_i) and foreign economy (ϵ_{i^*}) are innovations and are linked directly to Taylor rule of the country.

3 Data and Estimation

The domestic sector is represented by the Czech economy and the foreign sector by EA15. The model was executed on quarterly data from 1996Q1 to 2010Q4 (60 observations). We measure domestic and foreign GDP, domestic and foreign consumer price inflation, domestic and foreign nominal interest rate (three months PRIBOR and three months EURIBOR) and nominal exchange rate. All data were seasonally adjusted (except the nominal interest rates and the exchange rate). Domestic and foreign GDP were detrended from linear trend and other time series were demeaned. Used data are freely downloadable materials of Czech National Bank.

Computations and results are provided by Matlab's toolbox Dynare [1]. Estimation uses method of Random Chain Walk Metropolis Hastings. Estimation was done in two Metropolis-Hastings runs generating 1,250,000 draws each. Next, we use method of recursive estimation. In each estimation, two runs of Metropolis-Hastings algorithm were conducted, each generating 100,000 draws. Using this method, we can see changes of parameters during time. Furthermore, changes of impulse response functions due to changes of parameters can be observed.

4 Results

4.1 Shock decomposition

In Figure 1 there is depicted historical decomposition of Czech output. One can summarize the decomposition as follows. In years 2009 - 2010, the fall was driven by lower foreign demand shock and the cost-push shock, that acted at the beginning of the year 2009. Higher risk premium of the Czech economy caused depreciation of real exchange rate and higher competitiveness of Czech exporters on foreign markets.

4.2 Recursive estimation

In Figure 2 one can notice the changing value of debt-elasticity χ (our 'financial accelerator') that declined till 2008. The accelerator rose around 7% in the year 2008 till third quarter of 2009. The debt-elasticity parameter slowly fell from 2010 and reached its pre-crisis value in last quarter of 2010. We can see that Czech households debt was considered as riskier at the end of the year 2008 and the year 2009. However, year 2010 showed the opposite. Since the end of 2009, the impact of household debt on risk premium of the country declined. Czech economy net foreign assets from equation (4) took values around 42 at the beginning of the year 2008, around -5 in first quarter of the year 2009 and around -10 in third quarter of the year. Thus, households debt/investment abroad (net foreign assets) pushed depreciation rate up by 0.4242 at the beginning of the year 2008, by -0.0535 in first quarter of the year 2009 and -0.107 in third quarter

[1] Dynare version 4.1.3. and Matlab version R2010b are used.

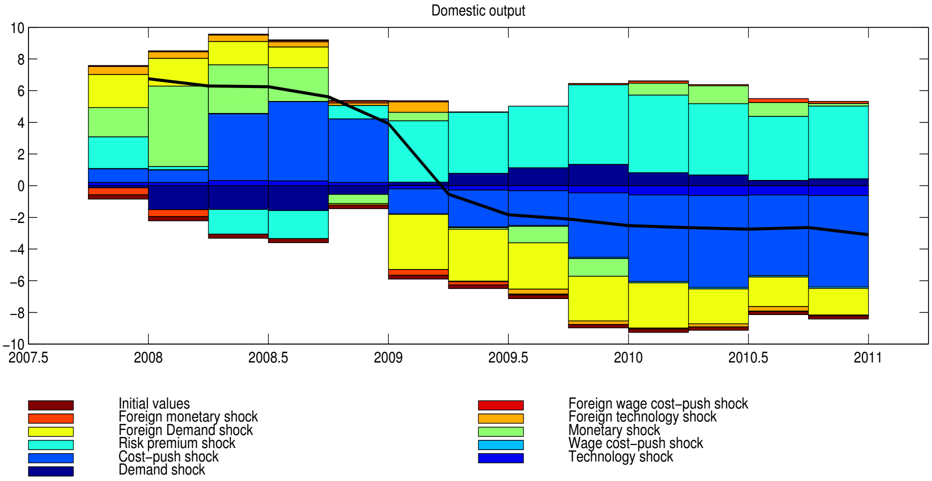


Figure 1: Historical shock decomposition of domestic output

of the year 2009. Depreciation rate in Czech economy took average values around 0.6510 from 2008Q1 till 2010Q4. Figure 3 depicts the influence of net foreign assets and financial accelerator on depreciation rate. Although, financial accelerator rose from 2008Q1 till 2009Q3, the impact on depreciation rate is negligible.

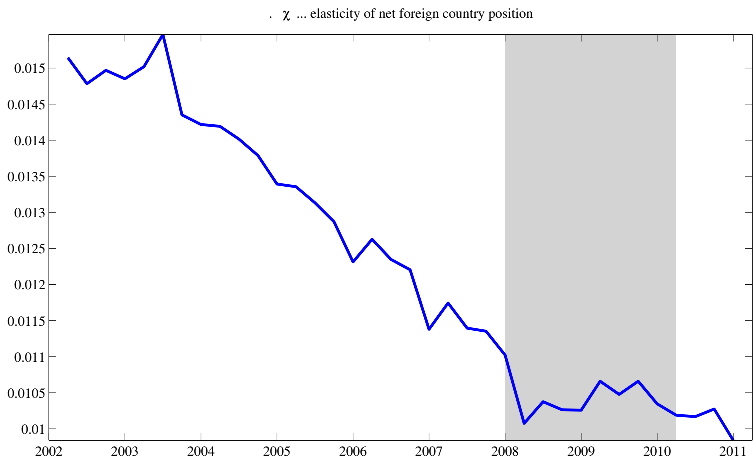
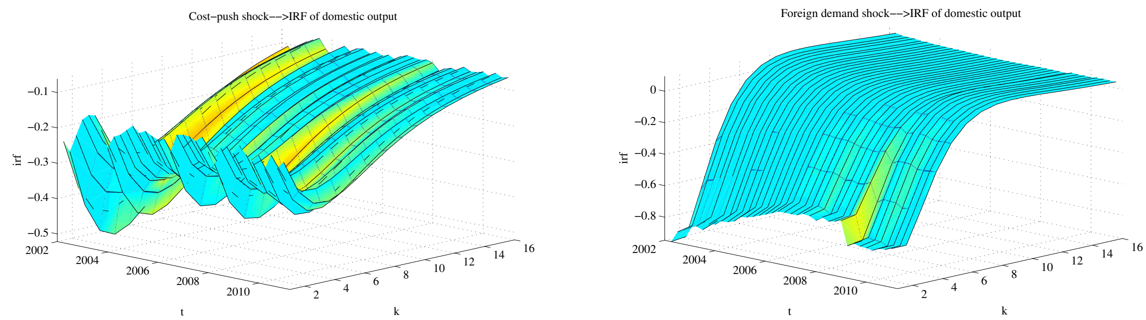


Figure 2: Recursively estimated development of financial accelerator

Figure 4 depicts recursive estimation of impulse response function development of domestic output to positive cost-push and negative foreign demand. We can see that changes in structural parameters influence the output and it is one of the reasons that led to deeper fall of Czech output during global financial crisis. Figure 4a depicts interesting changes of impulse response functions to cost-push shock at the beginning of the year 2008. The economy became more sensitive to this type of shock. The highest impact on domestic output came after 3-4 periods because of a delay of the response. In Figure 4b is shown development of impulse response functions of Czech output to negative foreign demand shock which was leading shock at the beginning of the year 2009 and an increase of economic sensitivity is obvious. Clearly, the foreign demand shock and the change in structure of the economy caused significant fall in output from year 2009. It implies that changes in structural parameters caused deeper fall of domestic output when these shocks acted.

Figure 3: The influence of net foreign assets and financial accelerator on depreciation rate



(a) Recursive estimation of impulse response function development of domestic output to positive cost-push shock

(b) Recursive estimation of impulse response function development of domestic output to negative foreign demand shock

Figure 4: Recursive estimation of impulse response function development of domestic output

5 Conclusion

In this paper, we use a small open economy New Keynesian DSGE model for Czech republic and incorporate a country risk premium shock that allows deviations from uncovered interest rate parity. First, using shock decomposition global financial crisis has different features. The main shocks, which influenced Czech output, were foreign demand shock and cost-push shock. This suggest that recent crisis spread to Czech economy partly from abroad. Thus, Czech National Bank could accommodate shocks in European Union. Next, we explored changes of structural parameters across time using method of recursive estimation. Modification of the economic structure had slight impact on the output and can deepen recent crisis. Moreover, we focused also on debt-elasticity premium of risk premium that plays the role of financial accelerator in the model. The parameter rose at the beginning of the crisis and then slowly went down. This means that domestic debt abroad was considered more riskier at the beginning of 2009 and then little by little improved its risk. However, the impact of households debt/investment on depreciation rate was not significant.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

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*, Economic Society of South Africa, vol. 78(2), pages 170-191, 06.

[2] Bernanke, B. - Gertler, M. - Gilchrist, S. (1994): "The Financial Accelerator and the Flight to Quality," *Working Papers 94-24*, C.V. Starr Center for Applied Economics, New York University.

[3] Calvo, G. (1983): "Staggered Prices in a Utility Maximizing Framework,: *Journal of Monetary Economics*, vol. 12: 383398.

[4] 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.

[5] Steinbach, R. - Mathuloe, P. - Smit, B. (2009): "An open economy New Keynesian DSGE model of the South African economy," *South African Journal of Economics*, vol. 77: 207-227.

Queues in Deterministic Linear Manufacturing Systems

Martin Gavalec[1], Zuzana Němcová[2]

Abstract. Development of the queues in a manufacturing system is important for the costs optimization of the manufacturing process. Waiting costs represent a significant part of total manufacturing costs, therefore any reduction of queue lengths leads to considerable savings. Linear manufacturing systems are frequently used in assembly line production and their optimization can strongly influence the total production costs. Queues in special types of linear systems were investigated earlier using max-plus algebra. Closed and open tandem networks were considered with special starting conditions. In the paper the control of linear queueing systems is studied, under assumption that the flow of service requests is stable for some time period. The objective of the control is to minimize the variable, fixed and waiting costs in the considered linear manufacturing system. Interesting results in this problem were found, such as an algorithm for computing the variable queue lengths in front of each server during the considered time period. The information helps to adjust the service times at servers in the system and minimize the total costs.

Keywords: queueing systems, lengths of the queues, production costs.
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Open linear system considered in this paper occurs in manufacturing systems, where the sequence of several successive operations has to be observed. Naturally there are forming the queues. The knowledge of progression of the lengths of the queues for the managers of the system is strategically important because of possibility to significantly reduce the total manufacturing costs by changing the service times at several servers.

Similar system were described by Heidergott [4] with support of the max-plus algebra. Evolution of the system in discrete time is computed using a state vector and a transition matrix in max-plus algebra. The state vector in time k represents the moments of the completion of the kth service at the nodes of the system, and the sequence of the vectors describes the time evolution of the system. However, the lengths of the queues are not considered in Heidergott's work.

General queuing systems are considered in [1], [3], [5], [6]. Closed queuing system was investigated using max-plus algebra in [2].

2 Considered queuing system

Throughout the paper we use a fixed natural number n and notation $N = \{1, 2, \ldots, n\}$. The time of one period is denoted by $t = \{1, 2, \ldots, T\}$. We denote the set of natural numbers (including 0) by notation $N_0 = \{0, 1, 2, \ldots\}$.

We consider an open linear system of n servers (see Figure 1). The incoming requests have to pass through the series of servers (denoted by numbers from 1 to n), and then they leave the system. Each server can serve only one customer at a time. At the start of the system, there is a queue in front of

[1]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, e-mail: martin.gavalec@uhk.cz
[2]University of Hradec Králové/Faculty of Informatics and Management, Rokitanského 62, Hradec Králové 3, 500 03, e-mail: zuzana.nemcova@uhk.cz

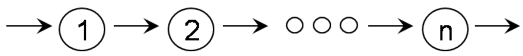


Figure 1: Considered system

each server, where the lengths of the starting queues are arbitrary. For simplicity we assume that the arrival intervals and the service times at each server are constant, i.e. the system is deterministic. The development of the queue lengths will be studied, in order to indicate when some of the queues are longer than prescribed bounds, or some of the servers with empty queue are idle. Both cases are economically undesirable, hence the service times should be changed. Similarly, change of service times is needed when the intensity of arrivals has increased, or decreased. The analysis performed in this paper can be useful in optimization of linear manufacturing systems.

The first server in the linear system controls the arrivals of demands into the system, i.e. the service intensity at first server represents the intensity of arrivals. We assume that the service time at each server remains stable for some considered period of time - the stage. When changes of service times on some servers are caused by an external reason (on the first server), or by decision of the system manager (to avoid creating too long queues on other servers), then next stage begins, which will be analyzed by the same way.

3 Types of the queues

Generally, due to changing service times of servers, the queues behave differently in each stage. Some of them remain stable, some of them increase or decrease. In so called transient time, it can happen that the queue is temporarily increasing or decreasing. This depends on the order of the servers (sequence of the service times). After the transient time, in which the queues can possibly change the sign of their growth, the stable state is considered. Then the queues do not change their sign of growth i.e. each queue either remains stable, or constantly increases or decreases, except the situation where the queue becomes empty.

In the stable state four types of queues are distinguished

- constant queue
- decreasing queue
- increasing queue
- empty queue

The empty queue is a special case of constant queue, or emerges as a result of decreasing queue.

Let us denote the vector of service times as $\mathbf{s} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, i.e. the service time at server i is σ_i, then the service intensity at server i is $1/\sigma_i$. The length of the queue in front of server i in time t is denoted by $f_i(t)$, and is contained in the vector $\mathbf{f}(t) = (f_1(t), f_2(t), \ldots, f_n(t))$ of waiting requests.

Following theorem describes various situations that may occur after changing the service times of ith server. Some of them can be seen in Figure 2. The letter f indicates the queue, the arrows its tendency. The upward continuous arrow means that the queue will increase in the long term, the downward continuous arrow indicates that the queue will decrease in the long term; similarly the dashed arrows represents the short term tendencies of the queues - temporary growth.

Theorem 1. *The queue in front of ith server in the system remains constant in stable state, if* $\sigma_i = \max(\sigma_1, ..., \sigma_{i-1})$.

The queue in front of the ith server in the system will decrease in stable state, if $\sigma_i < \max(\sigma_1, ..., \sigma_{i-1})$.

The queue in front of the ith server in the system will increase in stable state, if $\sigma_i > \max(\sigma_1, ..., \sigma_{i-1})$.

The queue in front of the ith server in the system will temporarily increase, if $\sigma_{i-2} > \sigma_{i-1} < \sigma_i$.

Proof. If the $\sigma_i = \max(\sigma_1, \sigma_2, ..., \sigma_{i-1})$, then clearly the intensity of arrivals of the requests at queue in front of the server i (queue i) is the same as the service intensity of server i. In other words the server

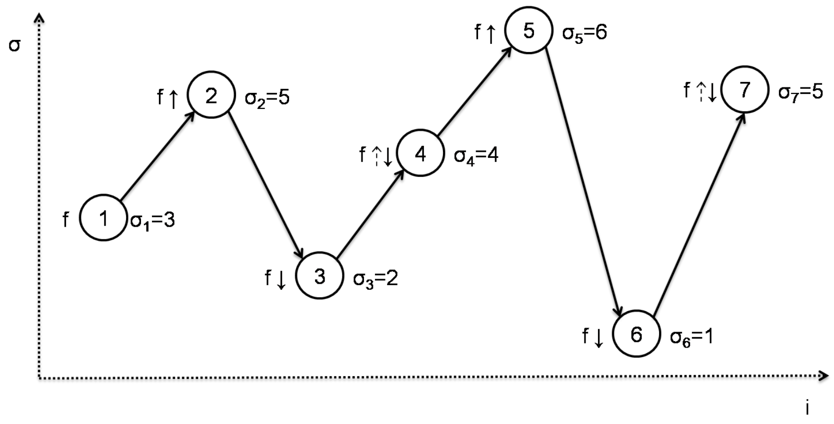


Figure 2: Changes of lengths of the queues

serves the requests by the same speed as the requests come at the queue in front of the server - the length of the queue in stable state does not change.

If the $\sigma_i < \max(\sigma_1, ..., \sigma_{i-1})$, it holds, that the service intensity of server i is higher than the intensity of arrivals of the requests at queue i - the queue will decrease in the stable state.

If $\sigma_i > \max(\sigma_1, ..., \sigma_{i-1})$, the service intensity of server i is lower than the intensity of arrivals of the requests at queue i. Thus it holds that the queues will continuously grow just in front of the servers, which precede faster servers (permanently-growing queues are created in front of the servers, which are slower than all servers before them).

If $\sigma_{i-2} > \sigma_{i-1} < \sigma_i$, the ith queue will temporarily increase, because the queue in front of the $(i-1)$th (faster) server overflows to the ith queue. But in the long term the queue will decrease because the intensity of arrivals at queue i will be thanks to server σ_{i-2} lower than service intensity of server i. □

4 Lengths of the queues

The rate of change of the length of the queue at each server (except the first server) can be calculated via following theorem.

Theorem 2. *Let R be the rate of change of the length of the queue at server i. The rate of change in the stable state is computed by formula*

$$R = \frac{\sigma_i - \sigma_{i-1}}{\sigma_{i-1}\sigma_i} \tag{1}$$

Proof. Let us denote σ_i and σ_{i-1} as service times at servers (i) and $(i-1)$. Denote $1/\sigma_i$ and $1/\sigma_{i-1}$ as service intensities of the servers. Then it is easily seen, that the queue in front of server (i) enlarges by $1/\sigma_{i-1}$ and lessens by $1/\sigma_i$. Then it can be written

$$\frac{1}{\sigma_{i-1}} - \frac{1}{\sigma_i} = \frac{\sigma_i - \sigma_{i-1}}{\sigma_{i-1}\sigma_i}$$

□

The length of the queue at server i changes on an average in $\sigma_{i-1} \cdot \sigma_i$ time units about $\sigma_i - \sigma_{i-1}$ requests. If $\sigma_i - \sigma_{i-1} > 0$, the type of the queue is increasing, in reverse, if $\sigma_i - \sigma_{i-1} < 0$, the type of the queue is decreasing.

The time period, during which the service times are constant, has been denoted above as a stage. Transition from one stage to another can be caused by two reasons - internal or external. External reason means that the intensity of arrivals, σ_1, is changed. Internal reason means that the service intensity, σ_i for $i > 1$, is changed by managerial decision as a response to changes in queue lengths.

The aim of the optimization is to adapt the service times to the arrival intensities and keeping the length of the queues as well as the number of empty queues to minimize the total costs in the system.

Every stage is characterized by

- the vector of service times $\mathbf{s}$,
- the vector of number of waiting (yet not served) requests in time t denoted as $\mathbf{f}(t)$.

The elements of the vector $\mathbf{f}(t)$ need not be integer numbers. The integer part stands for the number of requests physically waiting in a queue in front of the server. The decimal part of the element stands for not yet served part of the request just being served. The rate of change of the decimal part is equal to the service intensity of the server. The vector $\mathbf{f}(t)$ is computed by following theorem.

Theorem 3. *The number of waiting requests at server i in time t, $f_i(t)$, is computed by following rules.*

For the first server
$f_1(t) = f_1(t-1) - 1/\sigma_1$

For other servers $(i > 1)$
If $f_{i-1}(t) \in N_0$
If $f_{i-1}(t) > 0$ *or* $(f_{i-1}(t) = 0$ *and* $f_{i-1}(t-1) > 0)$ *then*
$f_i(t) := f_i(t-1) - 1/\sigma_i + 1$
endif
else
$f_i(t) := f_i(t-1) - 1/\sigma_i$
endif
If $f_i(t) < 0$, *then* $f_i(t) := 0$

Proof. At the beginning of each stage in front of each server there can be some queue. We suppose, that the queue in front of the first server is potentially infinite. The value $f_1(t)$ is computed by subtracting the service intensity of the first server from the number of waiting requests at server 1 in time $t-1$ (length of the queue in previous time unit).

For other servers there can arise two cases. The first case is if the number of waiting requests at server $i-1$ in time t, $f_{i-1}(t)$, is a integer number or if value $f_{i-1}(t)$ is zero, but in previous time value $f_{i-1}(t-1)$ was not zero. In this case the zero as a number of waiting requests of previous server means, that previous server finished service of last request (if $f_{i-1}(t-1) = 0$, it holds, that the server did not serve any request). This means that the previous server had launched just served request. Then the length of the queue $f_i(t)$ is computed by addition 1 (newly arrival request) and by subtracting the service intensity $1/\sigma_i$ (just served part of the request at server i) from the number of waiting requests at server i in time $t-1$. The second case is that if $f_{i-1}(t)$ is a decimal - the request being served at previous server is not finished. The length of the queue $f_i(t)$ is computed by subtracting the service intensity of i from the value $f_i(t-1)$. □

Service time at each server can be changed for two reasons: because of not fully used server or because the queue in front of the server is too long. The capacity of the server should be decreased in the first case and increased in the second case.

Service time at the server i can be changed after the current request is completed. It means that the request is handled, is sent to the next station, and another one is still waiting for service at this point. In other words, the service time at server i can be modified at time t when the element of the vector $\mathbf{f}(t)$ is an integer (when served request is launched and the service of the next has not started yet). If the value $f_i(t)$ is not an integer, the modification of the service time at server i will be possible at earliest in $x\sigma_i$ time units, where x is a decimal part of the value $f_i(t)$.

Example 1. Consider a deterministic linear manufacturing system of four servers with the vector of service times $\mathbf{s} = (2, 4, 1, 2)$ and the vector of number of waiting requests $\mathbf{f}(t) = (20, 2, 5, 4)$. With use of described algorithm the evolution of the lengths of the queues in time t, $f_i(t)$, is computed in the table below.

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$f_1(t)$	20	19,5	19	18,5	18	17,5	17	16,5	16	15,5	15	14,5	14	13,5	13
$f_2(t)$	2	1,75	2,5	2,25	3	2,75	3,5	3,25	4	3,75	4,5	4,25	5	4,75	5,5
$f_3(t)$	5	4	3	2	2	1	0	0	1	0	0	0	1	0	0
$f_4(t)$	4	4,5	5	5,5	6	6,5	7	6,5	6	6,5	6	5,5	5	5,5	5

The rate of change of the length of the queue at server 2 is $\frac{1}{4}$, at server 3 is $-\frac{3}{4}$ and at server 4 is $\frac{1}{2}$. In the stable state the queues in front of the first and the third server are decreasing, the queue in front of the second server is increasing and the queue in front of fourth server is temporarily increasing (for $t \leq 7$), but in the stable state the queue is decreasing.

5 Production costs

The system can be affected by the manager by changing the service times of the servers. The main reason is to reduce total costs and optimize the manufacturing process. The decision is divided into two parts. First, it must be decided in what moment will be done the change of service times in order to minimize the average production costs. Second, it has to be decided which of the service times will be changed. If the queue at server i is too long and is still increasing, than the service time σ_i should be diminished. This can be done e.g. by adding a parallel service machine at position i. On the other side if server i is idle for large part of its working time, then σ_i can be increased.

The decision lays in determining the optimal time T, when the stage should be finished and some changes in service times will be performed. The criterion for the decision is minimal value of the average production costs. The optimization runs in each stage separately. The total costs in the system consist of three parts: queueing costs, idle costs and change costs. The length of the stage is denoted as T.

The function of average production costs is a sum the average queuing, idle and change costs.

$$C = Q + I + G \tag{2}$$

Queueing costs, Q, are costs resulted from excessively long queues. In each stage the mean total costs will be minimized. The situation, when the length of the queue is larger than some maximal tolerable limit, M, represents undesirable extra costs. The part of the queue, that overpasses a given limit, creates queuing costs.

Theorem 4. *Let us denote M as the maximal tolerable limit of the length of the queue and q_i as a queuing costs of server i. Then the average queuing costs in the stage of length T are computed by following formula.*

$$Q = \frac{1}{T} \sum_{i=1}^{n} q_i \sum_{t=1}^{T} \max(f_i(t) - M, 0) \tag{3}$$

Proof. The length of the queue $f_i(t)$ decreased about M reflects the rate of overshooting of the maximal tolerable limit. If the limit is not overshot, the difference is negative number and thats why the $\max(f_i(t) - M,\ 0)$ is considered. The scalar product of the normalized vector $\mathbf{f}(t)$, $\max(f_i(t) - M,\ 0)$, with vector of queuing costs, $\mathbf{q}$, represents the total queuing costs expended on overshooting the maximal tolerable limit in the stage of length T. The average queuing costs are computed as the total queuing costs divided by the stage length T. $\square$

Idle costs, I, are considered when server's capacity is not fully used (it is the amount of money which is paid if the server is not working).

Theorem 5. *Let us denote p_i as an idle costs of the ith server. Then the average idle costs in the stage of length T are computed by following formula.*

$$I = \frac{1}{T} \sum_{i=1}^{n} p_i \sum_{t=1}^{T} |f_i(t) = 0| \tag{4}$$

Where $|f_i(t) = 0| = 1$ if $f_i(t) = 0$ and $|f_i(t) = 0| = 0$ otherwise.

Proof. The sum of $|f_i(t) = 0|$ reflects occurrence of the empty queue - if $f_i(t) = 0$, the value of $|f_i(t) = 0|$ is 1. In other cases the value of $|f_i(t) = 0|$ is 0. The scalar product of the vectors above represents the total idle costs for the stage of length T. The average idle costs are computed as the total idle costs divided by the stage length T. □

Change costs, G, poses for each stage one-shot costs, which emerges at the end of the stage. These costs are connected with discharging of the change of the service time.

Theorem 6. *Let us denote* $\mathbf{h}(T)$ *as a vector indicating if the service time at server i was changed at the end of the stage. If the service time at the ith server was changed, then* $h_i(T) = 1$. *If the service time at server i was not changed, than* $h_i(T) = 0$. *Denote the* r_i *as a change costs at server* i. *The average change costs in the stage of length* T *are computed as*

$$G = \frac{1}{T} \sum_{i=1}^{n} r_i h_i(T) \tag{5}$$

Proof. The vector $\mathbf{h}(T)$ represents the information about number and position of the servers, where the service times were changed at the end of the stage - in time T. The scalar product of the vector $\mathbf{h}(T)$ with vector of change costs, $\mathbf{r}$, represents the total change costs in the stage. The average change costs are computed as the total change costs divided by the stage length T. □

6 Conclusions

The formulas for optimal control of service times in a linear system were suggested. The formulas show average production costs in dependence of the length of the period without changes in service times. Formulas at the same time enable optimal flexible reaction to changes in external environment.

Acknowledgements

The support of the project INKOV # CZ.1.07/2.3.00/20.0001, financed from EU and Czech Republic funds, and the Czech Science Foundation project # 402/09/0405 is kindly acknowledged.

References

[1] Baccelli, F. L., and Cohen, G., and Olsder, G. J., and Quadrat, J. P.: *Synchronization and Linearity. An Algebra for Discrete Event Systems.* Wiley, Chichester, 1992.

[2] Gavalec, M., and Němcová, Z.: Eigenproblem for queueing systems in max-plus algebra, *Proc. of the 28th Conf. Mathem. Methods in Economics,* University of South Bohemia, České Budějovice (2010), 156–161.

[3] Gross, D. and Harris, C.M.: *Fundamentals of Queuing Theory.* John Wiley and Sons, New York, 1974.

[4] Heidergott, B.: A characterisation of (max,+)-linear queueing systems. In: *Annual report of the TMR programme ALAPEDES, contract ERB-FMRX-CT96-0074*, European Commission, 1999.

[5] Heidergott, B.: *Max-Plus linear Stochastic Systems and Perturbation Analysis*, The International Series of Discrete Event Systems **15**, Springer–Verlag, Berlin 2007.

[6] Robertazzy, T.G.: *Planning Telecomunication Networks.* IEEE: Piscataway, NJ. 1999.

Eigenspace of interval matrices in max-min algebra

M. Gavalec[1], J. Plavka[2], H. Tomášková[3]

Abstract. The input data in real problems are usually not exact values and can be characterized by intervals. Then the investigation of stable states leads to computing interval eigenvectors. The eigenspace structure of a given interval matrix A in max-min algebra is studied in the paper. Monotone interval eigenvectors x fulfilling the equation $Ax = x$ are described. By max-min algebra we understand a linear structure on a linearly ordered set with two binary operations maximum and minimum, used similarly as addition and multiplication in the classical linear algebra. The operations max and min are extended to matrices and vectors in a natural way.

The characterization of interval eigenvectors presented earlier in [7] for strictly increasing and for constant eigenvectors is extended now to general interval eigenvectors. Six types of general interval eigenvectors are studied according to classification of monotone interval eigenvectors and characterization of all six types is described.

Keywords: Max-min algebra, eigenspace, eigenvector, eigenproblem

JEL classification: C44
AMS classification: 90C15

1 Introduction

By max-min algebra we understand a triple $(\mathcal{B}, \oplus, \otimes)$, where $\mathcal{B}$ is a linearly ordered set, and $\oplus = \max$, $\otimes = \min$ are binary operations on $\mathcal{B}$. The notation $\mathcal{B}(m,n)$ ($\mathcal{B}(n)$) denotes the set of all matrices (vectors) of given dimension over $\mathcal{B}$. Operations $\oplus$, $\otimes$ are extended to matrices and vectors in a formal way. The linear ordering on $\mathcal{B}$ induces partial ordering on $\mathcal{B}(m,n)$ and $\mathcal{B}(n)$ and the notation $\wedge$ ($\vee$) is used for the binary operation of meet (join) in these sets. Similarly, $\bigwedge$ ($\bigvee$) denotes the subset version of meet (join) operation.

The eigenproblem for a given matrix $A \in \mathcal{B}(n,n)$ in max-min algebra consists of finding a value $\lambda \in \mathcal{B}$ (eigenvalue) and a vector $x \in \mathcal{B}(n)$ (eigenvector) such that the equation $A \otimes x = \lambda \otimes x$ holds true. It is well-known that the above problem in max-min algebra can be reduced to solving the equation $A \otimes x = x$. The eigenproblem in max-min algebra has been studied by many authors. Interesting results were found in describing the structure of the eigenspace (the set of all eigenvectors), and algorithms for computing the largest eigenvector of a given matrix were suggested, see e.g. [1], [4].

A classification consisting of six different types of interval eigenvectors is presented in this paper, and detailed characterization of all described types is given for strictly increasing interval eigenvectors using methods from [6].

2 Interval eigenvectors in applications

Eigenvectors of matrices in max-min algebra are useful in applications such as automata theory, design of switching circuits, logic of binary relations, medical diagnosis, Markov chains, social choice, models of organizations, information systems, political systems and clustering. In practice, the values of vector

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[2]Technical University in Košice, Faculty of Electrical Engineering and Informatics, B. Němcovej 32, 04200 Košice, Slovak Republic, e-mail: jan.plavka@tuke.sk
[3]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: hana.tomaskova@uhk.cz

(matrix) inputs are not exact numbers and often they are rather contained in some interval. Considering matrices and vectors with interval coefficients is therefore of great practical importance, see [2], [3], [5], [8], [9]. This paper investigates monotone interval eigenvectors of interval matrices in max-min algebra.

As an example of application we can consider the political system in a country, as it is influenced by other countries. The state of the political system consists of several components, such as culture, economics, technology, science, and the level of each component i is described by some value x_i, which is influenced by the situation in the neighboring countries. The influence depends on the proximity of the country j expressed by a constant factor a_{ij}. Hence the state vector x at time $t+1$ can be expressed by equalities $x_i(t+1) = \max_j(\min(a_{ij}, x_j(t))) = \bigoplus_j (a_{ij} \otimes x_j(t))$ for every i, or shortly $x(t+1) = A \otimes x(t)$, in the matrix notation. The steady state of the political system in the given country is then described by the equation $x(t+1) = x(t)$, i.e. $A \otimes x(t) = x(t)$. In other words, $x(t)$ is eigenvector of matrix A, which means that the steady states exactly correspond to eigenvectors. In real application, interval eigenvectors and interval matrices are considered.

Further examples are related to business applications describing mutual influence of projects adopted by a company, or planning different form of advertisement. Similar applications can be found in social science, biology, medicine, computer nets, and many other areas.

3 Interval eigenvectors classification

Let n be a given natural number. We shall use the notation $N = \{1, 2, \ldots, n\}$. Similarly to [3], [5], [8], we define interval matrix with bounds $\underline{A}, \overline{A} \in \mathcal{B}(n,n)$ and interval vector with bounds $\underline{x}, \overline{x} \in \mathcal{B}(n)$ as follows

$$[\underline{A}, \overline{A}] = \left\{ A \in \mathcal{B}(n,n);\ \underline{A} \le A \le \overline{A} \right\} \ , \quad [\underline{x}, \overline{x}] = \left\{ x \in \mathcal{B}(n);\ \underline{x} \le x \le \overline{x} \right\} \ .$$

We assume in this section that an interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ and an interval vector $\mathbf{X} = [\underline{x}, \overline{x}]$ are fixed. The interval eigenproblem for $\mathbf{A}$ and $\mathbf{X}$ consists in recognizing whether $A \otimes x = x$ holds true for $A \in \mathbf{A}$, $x \in \mathbf{X}$. In dependence on the applied quantifiers, we get six types of interval eigenvectors.

Definition 1. If interval matrix $\mathbf{A}$ is given, then interval vector $\mathbf{X}$ is called

- *strong eigenvector* of $\mathbf{A}$ if $(\forall A \in \mathbf{A})(\forall x \in \mathbf{X})[\, A \otimes x = x \,]$,
- *strong universal eigenvector* of $\mathbf{A}$ if $(\exists x \in \mathbf{X})(\forall A \in \mathbf{A})\, [A \otimes x = x \,]$,
- *universal eigenvector* of $\mathbf{A}$ if $(\forall A \in \mathbf{A})(\exists x \in \mathbf{X})[\, A \otimes x = x \,]$,
- *strong tolerance eigenvector* of $\mathbf{A}$ if $(\exists A \in \mathbf{A})(\forall x \in \mathbf{X})[\, A \otimes x = x \,]$,
- *tolerance eigenvector* of $\mathbf{A}$ if $(\forall x \in \mathbf{X})(\exists A \in \mathbf{A})[\, A \otimes x = x \,]$,
- *weak eigenvector* of $\mathbf{A}$ if $(\exists A \in \mathbf{A})(\exists x \in \mathbf{X})[\, A \otimes x = x \,]$.

Analogously as in [6], we denote the set of all strictly increasing vectors of dimension n as

$$\mathcal{B}^{<}(n) = \{\, x \in \mathcal{B}(n);\ (\forall i, j \in N)[i < j \Rightarrow x_i < x_j] \,\} \ ,$$

and the set of all increasing vectors as

$$\mathcal{B}^{\le}(n) = \{\, x \in \mathcal{B}(n);\ (\forall i, j \in N)[i \le j \Rightarrow x_i \le x_j] \,\} \ .$$

Further we denote the eigenspace of a matrix $A \in \mathcal{B}(n,n)$ as

$$\mathcal{F}(A) = \{\, x \in \mathcal{B}(n);\ A \otimes x = x \,\} \ ,$$

and the eigenspaces of all strictly increasing eigenvectors (increasing eigenvectors) as

$$\mathcal{F}^{<}(A) = \mathcal{F}(A) \cap \mathcal{B}^{<}(n) \ , \quad \mathcal{F}^{\le}(A) = \mathcal{F}(A) \cap \mathcal{B}^{\le}(n) \ .$$

It is clear that any vector $x \in \mathcal{B}(n)$ can be permuted to an increasing vector. Therefore, in view of the next theorem, the structure of the eigenspace $\mathcal{F}(A)$ of a given $n \times n$ max-min matrix A can be described by investigating the structure of monotone eigenspaces $\mathcal{F}^{<}(A_{\varphi\varphi})$ and $\mathcal{F}^{\le}(A_{\varphi\varphi})$, for all permutations φ on N.

Theorem 1. *[6] Let $A \in \mathcal{B}(n,n)$, $x \in \mathcal{B}(n)$ and let φ be a permutation on N. Then $x \in \mathcal{F}(A)$ if and only if $x_\varphi \in \mathcal{F}(A_{\varphi\varphi})$.*

For $A \in \mathcal{B}(n,n)$, the structure of $\mathcal{F}^<(A)$ has been described in [6] as an interval of strictly increasing vectors in the following theorem. The vector bounds $m^\star(A)$, $M^\star(A) \in \mathcal{B}(n)$ are defined for any $i \in N$ as follows

$$m_i^\star(A) := \max_{j\le i} \max_{k>j} a_{jk} \ , \qquad M_i^\star(A) := \min_{j\ge i} \max_{k\ge j} a_{jk} \ .$$

Theorem 2. *[6] Let $A \in \mathcal{B}(n,n)$ and let $x \in \mathcal{B}(n)$ be a strictly increasing vector. Then $x \in \mathcal{F}(A)$ if and only if $m^\star(A) \le x \le M^\star(A)$. In formal notation,*

$$\mathcal{F}^<(A) = \langle m^\star(A), M^\star(A)\rangle \cap \mathcal{B}^<(n) \ .$$

The structure of the increasing eigenspace $\mathcal{F}^{\le}(A)$ is described analogously, considering every increasing eigenvector as strictly increasing on some interval partition of the index set N. By interval partition, D, we understand a set of disjoint integer intervals in N, whose union is $\bigcup D = N$. The set of all interval partitions on N will be denoted as $\mathcal{D}_n$. For every interval partition $D \in \mathcal{D}_n$ and for every $i \in N$, we denote by $D[i]$ the partition class containing i, in other words, $D[i]$ is an integer interval $I \in D$ with $i \in I$. For $i, j \in N$ we write $D[i] < D[j]$, if $i' < j'$ holds for any $i' \in D[i]$, $j' \in D[j]$. If $D[i] < D[j]$ or $D[i] = D[j]$, then we write $D[i] \le D[j]$,

For a given partition $D \in \mathcal{D}_n$ and for a vector $x \in \mathcal{B}(n)$ we say that x is strictly D-increasing if for any indices $i, j \in N$

$$\begin{aligned} x_i = x_j &\quad \text{if and only if} \quad D[i] = D[j] \ , \\ x_i < x_j &\quad \text{if and only if} \quad D[i] < D[j] \ . \end{aligned}$$

The set of all strictly D-increasing vectors in $\mathcal{B}(n)$ will be denoted as $\mathcal{B}^<(D,n))$. If $A \in \mathcal{B}(n,n)$, then the set of all strictly D-increasing eigenvectors of A is denoted by $\mathcal{F}^<(D,A)$.

It is easy to see that for every increasing vector $x \in \mathcal{B}^{\le}(n)$, there is exactly one interval partition $D \in \mathcal{D}_n$ such that x is strictly D-increasing. If vector x is strictly increasing on N, then the corresponding partition is equal to the finest partition consisting of all singletons $D = \big\{\{i\};\, i \in N\big\}$. On the other hand, if x is constant, then $D = N$.

The interval bounds described in Theorem 2 for strictly increasing eigenvectors, are generalized in the next theorem for D-increasing eigenvectors with arbitrary interval partition $D \in \mathcal{D}_n$. This allows to describe the structure of the increasing eigenspace $\mathcal{F}^{\le}(A)$ of a given matrix A as the union of intervals in $\mathcal{B}^{\le}(n)$. Using Theorem 1, this characterization can be further extended to the whole eigenspace $\mathcal{F}(A)$.

For a given interval partition $D \in \mathcal{D}_n$, the vector bounds $m^\star(D,A)$, $M^\star(D,A) \in \mathcal{B}(n)$ are defined for any $i \in N$ as follows

$$m_i^\star(D,A) := \max_{j\in N, D[j]\le D[i]} \ \max_{k \in N, D[k]>D[j]} a_{jk} \ , \quad M_i^\star(D,A) := \min_{j\in N, D[j]\ge D[i]} \ \max_{k\in N, D[k]\ge D[j]} a_{jk} \ .$$

Theorem 3. *[6] Let $A \in \mathcal{B}(n,n)$, $D \in \mathcal{D}_n$ and let $x \in \mathcal{B}(n)$ be a strictly D-increasing vector. Then $x \in \mathcal{F}(A)$ if and only if $m^\star(D,A) \le x \le M^\star(D,A)$. In formal notation,*

$$\mathcal{F}^<(D,A) = \langle m^\star(D,A), M^\star(D,A)\rangle \cap \mathcal{B}^<(D,n) \ .$$

Vector bounds $m^\star(D,A), M^\star(D,A)$ will be used in the next section for characterization of all six types of interval eigenvectors.

4 Interval eigenvectors characterization

Definition 2. Let $\mathbf{X} = [\underline{x}, \overline{x}]$ be an interval vector with bounds $\underline{x}$ and $\overline{x}$ and let $D \in \mathcal{D}_n$ be an interval partition. Then we denote

$$\underline{x}_D = \bigotimes \{x \in \mathcal{B}^<(D,n);\, \underline{x} \le x\} \ , \quad \overline{x}_D = \bigoplus \{x \in \mathcal{B}^<(D,n);\, x \le \overline{x}\} \ .$$

Proposition 4. *For given $\mathbf{X} = [\underline{x}, \overline{x}]$ and $D \in \mathcal{D}_n$, the following statements hold true*

(i) $\underline{x}_D \in \mathcal{B}^{\leq}(D,n)$, $\overline{x}_D \in \mathcal{B}^{\leq}(D,n)$,

(ii) $\underline{x} \leq \underline{x}_D$, $\overline{x}_D \leq \overline{x}$,

(iii) *if* $x \in \mathcal{B}^{<}(D,n)$, *then* $x \in \mathbf{X}$ *if and only if* $\underline{x}_D \leq x \leq \overline{x}_D$.

Proof. (i) Let $i, k \in N, D[i] = D[k]$. Then $x_i = x_k$ for every $x \in Bless(D,n)$. Hence

$$(\underline{x}_D)_i = \bigotimes\{x_i;\, x \in Bless(D,n),\, \underline{x} \leq x\} = \bigotimes\{x_k;\, x \in Bless(D,n),\, \underline{x} \leq x\} = (\underline{x}_D)_k\ ,$$

analogously for $\overline{x}_D$. The inequalities (ii) and (iii) follow directly from the definition. □

Theorem 5. *(T1) Let interval matrix $\mathbf{A} = [\underline{A}, \overline{A}]$ and interval vector $\mathbf{X} = [\underline{x}, \overline{x}]$ be given. Then $\mathbf{X}$ is strong $D-$increasing eigenvector of A if and only if*

$$[\underline{x}_D, \overline{x}_D] \cap \mathcal{B}^{<}(D,n) = \emptyset \tag{1}$$

or

$$m^{\star}(\overline{A}, D) \leq \underline{x}_D, \quad \overline{x}_D \leq M^{\star}(\underline{A}, D). \tag{2}$$

Proof. Let conditions (2) be satisfied, let $x \in \mathbf{X} \cap \mathcal{B}^{<}(D,n), A \in \mathbf{A}$. Then we have

$$m^{\star}(D,A) \leq m^{\star}(\overline{A}, D) \leq \underline{x}_D \leq x$$

and

$$x \leq \overline{x}_D \leq M^{\star}(\underline{A}, D) \leq M^{\star}(D,A).$$

Therefore, $m^{\star}(D,A) \leq x \leq M^{\star}(D,A)$, which implies $x \in \mathcal{F}(A)$.

For the converse implication, let us assume, that the inequalities (2) hold true.
Then for $(\forall A)\forall x \in \mathcal{B}^{<}(D,n)$ we get the rules

$$m^{\star}(D,A) \leq x.$$

Then $\overline{A} \in \mathbf{A} \Rightarrow m^{\star}(\overline{A}, D) \leq \min x_{x \in \mathcal{B}^{<}(D,n) \cap \mathbf{X}} = \underline{x}_D$.
Second rule is $(\forall A \in \mathbf{A})(\forall x)$

$$x \leq M^{\star}(D,A)$$

$\overline{x}_D = \max x_{x \in \mathcal{B}^{<}(D,n) \cap \mathbf{X}} \leq M^{\star}(\underline{A}, D)$. □

Theorem 6. *(T2) Let $\mathbf{X} = [\underline{x}, \overline{x}]$, $\mathbf{A} = [\underline{A}, \overline{A}]$. Then $\mathbf{X}$ is strong universal eigenvector of $\mathbf{A}$ if and only if*

$$\langle m^{\star}(\overline{A}, D) \oplus \underline{x}_D, M^{\star}(\underline{A}, D) \otimes \overline{x}_D \rangle \cap \mathcal{B}^{<}(D,n) \neq \emptyset \tag{3}$$

and

$$m^{\star}(\overline{A}, D) \leq \overline{x}_D, \quad \underline{x}_D \leq M^{\star}(\underline{A}, D). \tag{4}$$

Proof. Let $\exists x \in \mathcal{B}^{<}(D,n), \quad \underline{x}_D \leq x \leq \overline{x}_D, \quad x \in \langle m^{\star}(\overline{A}, D), M^{\star}(\underline{A}, D) \rangle$.
Let $A \in [\underline{A}, \overline{A}]$

$$m^{\star}(\overline{A}, D) \leq x' \leq M^{\star}(\underline{A}, D) x = (x' \wedge \overline{x}_D) \vee \underline{x}_D = ((x' \vee \underline{x}_D) \wedge \overline{x}_D \underline{x}_D \leq x \leq \overline{x}_D$$

$$\left.\begin{array}{ll} m^{\star}(\overline{A}, D) & \leq x' \leq (x' \vee \underline{x}_D) \\ m^{\star}(\overline{A}, D) & \overline{x}_D \end{array}\right\} \Rightarrow m^{\star}(\overline{A}, D) \leq x$$

analogously $M^{\star}(\underline{A}, D) \geq x \Rightarrow$

$$m^{\star}(D,A) \leq m^{\star}(\overline{A}, D) \leq x \leq M^{\star}(\underline{A}, D) \leq M^{\star}(D,A)$$

□

The Theorems 7–10 below are proved analogously as corresponding theorems in [7], with similar modifications as shown above.

Theorem 7. *(T3) Let interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and interval vector* $\mathbf{X} = [\underline{x}, \overline{x}]$ *be given. Then* $\mathbf{X}$ *is a universal eigenvector of* $\mathbf{A}$ *if and only if*

$$m^\star(\overline{A}, D) \leq \overline{x}_D \ , \quad \underline{x}_D \leq M^\star(\underline{A}, D) \ , \tag{5}$$

and

$$(\forall A \in \mathbf{A}) \left[\langle m^\star(D, A) \oplus \underline{x}_D, M^\star(A, D) \otimes \overline{x}_D \rangle \cap \mathcal{B}^<(D, n) \neq \emptyset \right] \ . \tag{6}$$

Theorem 8. *(T4) Let interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and interval vector* $\mathbf{X} = [\underline{x}, \overline{x}]$ *be given. Further let us denote*

$$\tilde{\mathbf{A}} = \{ A \in \mathbf{A};\ m^\star(D, A) \leq \underline{x}_D \} \ , \quad \tilde{A} = \bigvee \tilde{\mathbf{A}} \ .$$

Then $\mathbf{X}$ *is a strong tolerance eigenvector of* $\mathbf{A}$ *if and only if*

$$m^\star(\underline{A}, D) \leq \underline{x}_D \ , \quad \overline{x}_D \leq M^\star(\tilde{A}) \ . \tag{7}$$

Theorem 9. *(T5) Let interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and interval vector* $\mathbf{X} = [\underline{x}, \overline{x}]$ *be given. Then* $\mathbf{X}$ *is a tolerance eigenvector of* $\mathbf{A}$ *if and only if*

$$m^\star(\underline{A}, D) \leq \underline{x}_D \ , \quad \overline{x}_D \leq M^\star(\overline{A}, D) \ . \tag{8}$$

Theorem 10. *(T6) Let interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and interval vector* $\mathbf{X} = [\underline{x}, \overline{x}]$ *be given. Then* $\mathbf{X}$ *is a weak eigenvector of* $\mathbf{A}$ *if and only if*

$$\underline{x}_D \leq M^\star(\overline{A}, D) \ , \quad m^\star(\underline{A}, D) \leq \overline{x}_D \ , \tag{9}$$

and

$$\langle m^\star(\underline{A}, D) \oplus \underline{x}_D, M^\star(\overline{A}, D) \otimes \overline{x}_D \rangle \cap \mathcal{B}^<(D, n) \neq \emptyset \tag{10}$$

5 Relations between various types of $D-$increasing eigenvectors

Theorem 11. *Let interval matrix* $\mathbf{A} = [\underline{A}, \overline{A}]$ *and interval vector* $\mathbf{X} = [\underline{x}, \overline{x}]$ *with* $D-$*increasing bounds* $\underline{x}_D$, $\overline{x}_D$ *be given. Then the following implications hold true.*

- $(T1 \Rightarrow T2)$ *If* $\mathbf{X}$ *is a* $D-$*increasing strong eigenvector of* $\mathbf{A}$*, then* $\mathbf{X}$ *is a* $D-$*increasing strong universal eigenvector of* $\mathbf{A}$*,*
- $(T1 \Rightarrow T4)$ *If* $\mathbf{X}$ *is a* $D-$*increasing strong eigenvector of* $\mathbf{A}$*, then* $\mathbf{X}$ *is a* $D-$*increasing strong tolerance eigenvector of* $\mathbf{A}$*,*
- $(T3 \Rightarrow T6)$ *If* $\mathbf{X}$ *is a* $D-$*increasing universal eigenvector of* $\mathbf{A}$*, then* $\mathbf{X}$ *is a* $D-$*increasing weak eigenvector of* $\mathbf{A}$*,*
- $(T5 \Rightarrow T6)$ *If* $\mathbf{X}$ *is a* $D-$*increasing tolerance eigenvector of* $\mathbf{A}$*, then* $\mathbf{X}$ *is a* $D-$*increasing weak eigenvector of* $\mathbf{A}$*.*

Proof. The implications follow directly from Definition 1. □

Remark 1. The converse implications to those in Theorem 11 do not hold true.

Acknowledgement

The support of the project INKOV # CZ.1.07/2.3.00/20.0001, financed from EU and Czech Republic funds, the Czech Science Foundation project # 402/09/0405 and APVV #0073-07 is kindly acknowledged.

References

[1] K. Cechlárová: Eigenvectors in Bottleneck Algebra. Lin. Algebra Appl. **175** (1992), 63–73.

[2] K. Cechlárová: Solutions of Interval Linear Systems in $(\max, +)$-algebra. In: Proceedings of the 6th International Symposium on Operational Research Preddvor, Slovenia 2001, pp. 321–326.

[3] K. Cechlárová, R.A. Cuninghame-Green: Interval Systems of Max-separable Linear Equations. Lin. Algebra Appl. **340** (2002), 215–224.

[4] R.A. Cuninghame-Green: Minimax Algebra. Lecture Notes in Economics and Mathematical Systems **166**, Springer–Verlag, Berlin 1979.

[5] M. Fiedler, J. Nedoma, J. Ramík, J. Rohn, K. Zimmermann: Linear Optimization Problems with Inexact Data. Springer–Verlag, Berlin 2006.

[6] M. Gavalec: Monotone eigenspace structure in max-min algebra. Lin. Algebra Appl. **345** (2002), 149–167.

[7] M. Gavalec, J. Plavka: Monotone interval eigenproblem in max-min algebra. Kybernetika **46** (2010), 387–396.

[8] M. Gavalec, K. Zimmermann: Classification of solutions to systems of two-sided equations with interval coefficients. Inter. J. of Pure and Applied Math. **45** (2008), 533–542.

[9] J. Rohn: Systems of Linear Interval Equations. Lin. Algebra Appl. **126** (1989), 39–78.

Complementarity in dual max-separable optimization problems

Martin Gavalec [1], Karel Zimmermann[2]

Abstract. We consider a general duality theory for the class of max-separable optimization problems. In such problems so called max-separable functions occur both as objective functions and as equality or inequality constraint functions. A max-separable function is defined as a function of n variables equal to the maximum of n functions each depending on a different variable. All functions involved are assumed to be continuous and strictly increasing. A pair of optimization problems is introduced, for which the strong duality property can be proved. Complementarity conditions for such dual problems are derived. and demonstrated on numerical examples. The results have applications in post-optimal sensitivity analysis and in other problems of operations research.

Keywords: (max, min)-linear equations, two-sided system
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Extending duality theory to different classes of nonlinear problems and using the results to provide a deeper insight into properties of optimal solutions is until now the subject of many papers (see e.g. [1], [2], [9]). Max-separable problems considered in this contribution generalize problems on certain special algebraic structures considered e.g. in [3], [7], [8]. The results presented in this paper are based on the results of [4], [5], [6].

We consider a general duality theory for a class of so called "max-separable" optimization problems. In such problems so called max-separable functions $h : R^k \to R$ of the form

$$h(x_1, \dots,\ x_k) = \max_j (h_j(x_j)),$$

occur both as objective functions and as constraint functions (h_j are given continuous functions of one variable).

We consider the following two optimization problems:

$$\max_{j\in J} w_j \to \max$$

subject to

$$\max_{j\in J}(r_{ij}(w_j)) \leq 1, \quad \forall i \in I, \quad w \geq 0$$

and

$$\max_{i\in I} u_i \to \min$$

subject to

$$\max_{i\in I}(r_{ij}(u_i)) \geq 1, \quad \forall j \in J, \quad u \geq 0$$

Here $I,\ J$ are finite index sets and functions $r_{ij} : R^1 \to R^1$ are assumed to be strictly increasing real functions, are positive for all $i \in I,\ j \in J$. We will show that the problems represent the pair of symmetric dual problems, which have the strong duality property.

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[2]Charles University Prague, Faculty of Mathematics and Physics, Malostranské nám. 25, 11800 Praha 1, Czech Republic, e-mail: karel.zimmermann@mff.cuni.cz

Let R_+ be the set of non-negative real numbers, $R_+^k \equiv R_+ \times \cdots \times R_+$ (k-times) for any natural k. We will assume that $x \in R_+^k$ is a column vector so that $x = (x_1, \ldots, x_k)^T$, where superscript T denotes the transposition. Let $I \equiv \{1, \ldots, m\},\ J \equiv \{1, \ldots, n\}$ be finite index sets. Let for all $i \in I,\ j \in J$ functions $q_{ij} : R_+ \longrightarrow R_+$ be continuous and strictly increasing satisfying the conditions $q_{ij}(0) = 0,\ q_{ij}(R_+) = R_+$, where $q_{ij}(R_+)$ is the range of q_{ij}.

2 Notation, duality results

Let us consider the following optimization problem:

$$f(x) \equiv \max_{j \in J}(f_j(x_j)) \to \min$$

subject to

$$\max_{j \in J}(q_{ij}(x_j)) \leq b_i, \forall i \in I,$$

where functions$f_j,\ j \in J$ are strictly increasing, $f_j(0) = 0$, and b_i are positive for all $i \in I$.

By multiplying the inequalities with b_i^{-1} and setting $w_j = f_j(x_j)\ \forall j \in J$, and $r_{ij}(w_j) = q_{ij}(f_j^{-1}(w_j))/b_i$ for all $i \in I,\ j \in J$ we transform the problem to the form:

PROBLEM I

$$\max_{j \in J} w_j \to \max$$

subject to

$$\max_{j \in J}(r_{ij}(w_j)) \leq 1, \quad \forall i \in I, \quad w \geq 0$$

Let us further consider optimization problem

PROBLEM II

$$\max_{i \in I} u_i \to \min$$

subject to

$$\max_{i \in I}(r_{ij}(u_i)) \geq 1, \quad \forall j \in J, \quad u \geq 0.$$

Let us note that for all $i \in I,\ j \in J$ functions r_{ij} have the same properties as functions q_{ij}.

We proved in [5], [6] that the two problems have under the given assumptions always optimal solutions w^{opt}, u^{opt} respectively and the optimal values of the objective functions are equal. In other words, it was proved that the problems behave similarly like the symmetric dual problems of the linear programming. We will prove in the sequel that similarly to linear programming there exist optimal solutions of PROBLEM I and PROBLEM II satisfying also conditions, which are similar to complementarity conditions of optimal solutions of dual problems in linear programming. For this purpose we introduce some notation and recall some results of [5], [6].

Let us define for each $j \in J$ index $k(j) \in I$ as follows:

$$\min_{k \in I}(r_{kj}^{-1}(1)) = r_{k(j)j}^{-1}. \tag{1}$$

Let us set

$$J_i \equiv \{j \in J\ ;\ k(j) = i\} \quad \forall i \in I. \tag{2}$$

It was proved ([5], [6]) that the optimal solution of PROBLEM II is determined as follows:

$$u_i^{\text{opt}} = \max_{j \in J_i}(\min_{k \in I}(r_{kj}^{-1}(1))) \quad \text{ if } \quad J_i \neq \emptyset, \tag{3}$$

$$u_i^{\text{opt}} = 0 \quad \text{ otherwise }. \tag{4}$$

The optimal solution w^{opt} of PROBLEM I is:

$$w_j^{\text{opt}} = \min_{k\in I}(r_{kj}^{-1}(1)) \quad \forall j \in J. \tag{5}$$

It was proved (see [5], [6]) that the optimal values of the objective functions of PROBLEM I and PROBLEM II are equal, i.e.

$$\max_{j\in J} w_j^{\text{opt}} \;=\; \max_{i\in I} u_i^{\text{opt}}. \tag{6}$$

3 Complementarity conditions

In what follows, we will show that there exist optimal solutions of PROBLEM I, PROBLEM II satisfying complementarity conditions similar to those of linear programming.

Lemma 1. Let u^{opt}, w^{opt} be defined by (3), (4), (5). Let the inequality

$$\max_{i\in I}(r_{ij_0}(u_i^{\text{opt}})) \;>\; 1 \tag{7}$$

hold for some $j_0 \in J$.

Then there exist an index $j_1 \neq j_0$ such that $w_{j_0}^{\text{opt}} \;<\; w_{j_1}^{\text{opt}}$.

Proof: Let $i_0 \in I$ be any index such that $r_{i_0j_0}(u_{i_0}^{\text{opt}}) > 1$ and let $j_1 \in J_{i_0}$ be such index that

$$u_{i_0}^{\text{opt}} = \max_{j\in J_{i_0}}(\min_{k\in I} r_{kj}^{-1}(1)) = \max_{j\in J_{i_0}} r_{k(j_1)j_1}^{-1}(1) = \max_{j\in J_{i_0}} r_{i_0j_1}^{-1}(1).$$

It follows from the construction of u^{opt} that under the given assumptions we have $j_0 \in J_{i_0}$ and

$$\max_{j\in J_{i_0}} r_{i_0j_1}^{-1}(1) > r_{i_0j_0}^{-1}(1).$$

It follows further from the definition of set J_{i_0} and (5) that

$$w_{j_1}^{\text{opt}} = \min_{k\in I} r_{kj_1}^{-1}(1) = r_{k(j_1)j_1}^{-1}(1) = r_{i_0j_1}^{-1}(1),$$

and similarly

$$w_{j_0}^{\text{opt}} = \min_{k\in I} r_{kj_0}^{-1}(1) = r_{k(j_0)j_0}^{-1}(1) = r_{i_0j_0}^{-1}(1).$$

Therefore we have $w_{j_0}^{\text{opt}} < w_{j_1}^{\text{opt}}$. □

As a consequence of Lemma 1 we obtain that

$$w_{j_0}^{\text{opt}} < w_{j_1}^{\text{opt}} \leq \max_{j\in J} w_j^{\text{opt}},$$

i.e. $w_{j_0}^{\text{opt}}$ does not influence the optimal value of the objective function of PROBLEM I.

Let us define $\tilde{w}^{\text{opt}}$ as follows:

$$\tilde{w}_j^{\text{opt}} = 0 \quad \text{if} \quad \max_{i\in I} r_{ij}(u^{\text{opt}}) > 1,$$

$$\tilde{w}_j^{\text{opt}} = w_j^{\text{opt}} \quad \text{otherwise} .$$

It follows immediately from Lemma 1 that $\tilde{w}^{\text{opt}}$ is another optimal solution of PROBLEM I.

Theorem 1. *Optimal solutions $\tilde{w}^{\text{opt}}$, u^{opt} of PROBLEM I, PROBLEM II satisfy the following conditions:*

$$\tilde{w}_j^{\text{opt}} = 0 \quad \textit{if} \quad \max_{i\in I}(r_{ij}(u^{\text{opt}})) > 1, \tag{8}$$

$$u_i^{\text{opt}} = 0 \quad \textit{if} \quad \max_{j\in J}(r_{ij}(w_j^{\text{opt}})) < 1. \tag{9}$$

Proof:

Conditions (8) follow immediately from the construction of $\tilde{w}^{\mathrm{opt}}$. It remains to prove (9).

Let us assume that

$$\max_{j\in J}(r_{ij}(w_j^{\mathrm{opt}})) \ < \ 1$$

for some fixed index $i \in I$. We have therefore $r_{ij}(w_j^{\mathrm{opt}}) \ < \ 1 \ \ \forall j \in J$.

Let us recall that for any $j \in J$ either $\tilde{w}_j^{\mathrm{opt}} = 0$ or

$$\tilde{w}_j^{\mathrm{opt}} = \min_{k\in I}(r_{kj}^{-1}(1)) = r_{k(j)j}^{-1},$$

and we have therefore $j \in J_{k(j)}$ and

$$r_{k(j)j}(\tilde{w}_j^{\mathrm{opt}}) = 1$$

Since we assumed that $r_{ij}(\tilde{w}_j^{\mathrm{opt}}) < 1 \ \ \forall j \in J$, we have $k(j) \neq i \ \ \forall j \in J$ and thus $J_i = \emptyset$. Therefore we obtain from (3), (4) that $u_i^{\mathrm{opt}} = 0$ and the proof is completed. $\square$

Remark 1.
Let us remark that similar results can be obtained for alternative optimal solution $\hat{w}^{opt}$ defined as follows (with arbitrarily chosen values a_j, $j \in J$ in the indicated interval):

$$\hat{w}_j^{opt} = \begin{cases} a_j \in [0, w_j^{opt}] & \text{if} \quad \max_{i\in I}(r_{ij}(u^{opt})) > 1, \\ w_j^{opt} & \text{otherwise} \end{cases}$$

Remark 2.
The above results imply that if $\hat{w}_j^{opt} > 0$, then $\max_{i\in I}(r_{ij}(u_i^{opt})) = 1$, and further if $u_i^{opt} > 0$, then $\max_{j\in J}(r_i j(\hat{w}_j^{opt})) = 1$.

The result can be used for the post-optimal sensitivity analysis of the problems. For instance, if we want to change the optimal value $\max_{j\in J}(\hat{w}_j^{opt})$, then we must change all right-hand sides of the constraints $\max_{i\in I}(r_{ip}(u_i)) \geq 1$, for which the equality $\hat{w}_p^{opt} = \max_{j\in J}(\hat{w}_j^{opt})$ holds true.

4 Illustrating numerical example

We will illustrate the theoretical results by a small numerical example:

Example 1. Let us consider the problem

$$f(w) \equiv \max(w_1,\ w_2,\ w_3) \quad \longrightarrow \quad \max$$

subject to

$$\max(7w_1,\ 4w_2,\ 1w_3) \leq 1$$

$$\max(5w_1,\ 6w_2,\ 8w_3) \leq 1$$

$$\max(1w_1,\ 2w_2,\ 1w_3) \leq 1$$

$$w_j \geq 0,\ j = 1,\ 2,\ 3$$

The dual problem will have the following form:

$$g(u) \equiv \max(u_1,\ u_2,\ u_3) \quad \longrightarrow \quad \min$$

subject to

$$\max(7u_1,\ 5u_2,\ 1u_3) \geq 1$$

$$\max(4u_1,\ 6u_2,\ 2u_3) \geq 1$$

$$\max(1u_1,\ 8u_2,\ 1u_3) \geq 1$$

$$u_i \geq 0 \ i = 1,\ 2,\ 3$$

Using the same notation as above we have:

$w^{\text{opt}} = (1/7,\ 1/6,\ 1/8),\ \tilde{w}^{\text{opt}} = (1/7,\ 1/6,\ 0), u^{\text{opt}} = (1/7,\ 1/6,\ 0)$

$f(w^{\text{opt}}) = f(\tilde{w}^{\text{opt}}) = g(u^{\text{opt}}) = 1/6$

and the complementarity conditions are fulfilled:

$$\max(1\tilde{w}_1^{\text{opt}},\ 2\tilde{w}_2^{\text{opt}},\ 1\tilde{w}_3^{\text{opt}}) = 2/6 < 1 \qquad u_3^{\text{opt}} = 0$$

$$\max(1u_1^{\text{opt}},\ 8u_2^{\text{opt}},\ 1u_3^{\text{opt}}) = 8/6 > 1 \qquad \tilde{w}_3^{\text{opt}} = 0$$

Acknowledgements

The support of Czech Science Foundation # 402/09/0405 and MSM0021620838 is kindly acknowledged.

References

[1] Baccelli, F.L., Cohen, G., Olsder, G.J., Quadrat, J.P.: Synchronization and Linearity, An Algebra for Discrete Event Systems, Wiley, Chichester, 1992.

[2] Chadha, S.S., Chadha, V.: Linear Fractional Programming and Duality, CEJOR(2007), 15, pp. 119 - 125.

[3] Cuninghame-Green, R.A.: Minimax Algebra, Lecture Notes in Economics and Mathematical Systems, 166, Springer Verlag, Berlin, 1979.

[4] Gavalec, M., Zimmermann, K.: Dual Max-prod Optimization Problems, Proceedings Mathematical Methods in Economics 2010, Part I, September 2010, pp. 168 - 176.

[5] Gavalec, M., Zimmermann, K.: Duality for Max-separable Problems, CEJOR (2011), doi 10.1007/s10100-011-0203-x

[6] Gavalec, M., Zimmermann, K.: Duality of Optimization Problems with Generalized Fuzzy Relation Equation and Inequality Constraints, Information Sciences (2011), doi 10.1016/j.ins.2011.04.019.

[7] Litvinov, G.L., Maslov, V.P., Sergeev, S., N.(eds.): Idempotent and Tropical Mathematics and Problems of Mathematical Physics, vol. I, Independent University Moscow, 2007

[8] Maslov, V.P., Samborskij, S.N.: Idempotent Analysis, Advances in Soviet Mathematics, 13, AMS, Providence 1992.

[9] Zhang, Q.: Uniform LP Duality for Semidefinite and Semi-infinite Programming, CEJOR(2008), 16, pp. 205-213.

Efficient Algorithm for Systems of Two-Sided Max-Min Equations

Martin Gavalec [1], Karel Zimmermann[2]

Abstract. A simple iteration method for solving systems of (max, min)-linear equations is presented. The systems have variables on both sides of the equations. The algorithm is a simplified version of an earlier result of the authors. The simplification procedure can be extended to wider classes of problems with a similar structure.

Keywords: (max, min)-linear equations, two-sided system
AMS classification: 08A72, 90B35, 90C47

1 Introduction

Practical problems, such as maximization of the minimal profile of routes, minimization of the maximal time required for prescribed activities, and many others, often lead to mathematical models described by algebraic structures, in which pairs of operations $(\max, +)$ or $(\max, \min)$ replace addition and multiplication of the classical linear algebra. Such algebraic structures have been investigated in the literature approximately since the sixties of the last century (see e.g. [1], [5], [11]). In these publications among other problems, systems of the so called $(\max, +)$- or $(\max, \min)$-linear equations with variables on only one side of the equations were considered. Since the operation max replacing addition is not a group operation, but only a semigroup one, the difference between solving systems with variables on one side and systems with variables occurring on both sides of the equations is substantial. Special two-sided systems were studied in [4], [5], [10] in connection with the so called $(\max, +)$- or $(\max, \min)$-eigenvalue problem. General two-sided $(\max, +)$-linear systems were studied in [2], [3] and, in a more general form, in [6], [9]. In [7] a polynomial method for finding the maximum solution of the $(\max, \min)$-linear system was proposed.

In the paper, the algorithm described in [7] is modified and a more efficient version is presented. The new algorithm uses linear orderings of the input values, which substantially simplifies computation of maximal values needed in the algorithm. The computational complexity of the simplified version is $O(mn \log mn)$. The new method can also be used for solving optimization problems with objective function and the set of feasible solutions described by a system of two-sided linear equations over max-min algebras, or even distributive lattices, see [8]. Hence, the method simplifies computations in various types of the above problems.

2 Notions and notation

A *max-min algebra* $\mathcal{B}$ is defined as a linearly ordered set $(B, \leq)$ with the binary operations of maximum and minimum, denoted by $\oplus$ and $\otimes$. In this notation we do not strictly distinguish between the max-min algebra $\mathcal{B}$ and its carrier set denoted by the same symbol $\mathcal{B}$, e.g. we shall write $x \in \mathcal{B}$. For given natural numbers $m, n > 0$, we denote $M = \{1, \ldots, m\}$, $N = \{1, \ldots, n\}$. The notation $\mathcal{B}(n)$ is used for the set of all n-dimensional column vectors over $\mathcal{B}$, and $\mathcal{B}(m, n)$ denotes the set of all matrices of type $m \times n$ over $\mathcal{B}$. The matrix operations $\oplus$ and $\otimes$ for matrices and vectors over algebra $\mathcal{B}$ are defined in the usual way as the matrix operations over any field, with respect to max-min operations.

$\mathcal{B}(n)$ can also be considered as a partially ordered set with the product ordering, i.e. with the ordering

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[2]Charles University Prague, Faculty of Mathematics and Physics, Malostranské nám. 25, 11800 Praha 1, Czech Republic, e-mail: karel.zimmermann@mff.cuni.cz

induced by the ordering defined in $\mathcal{B}$. Namely, for any vectors $b, c \in \mathcal{B}(n)$, $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$, we write $x \leq y$ in $\mathcal{B}(n)$, if $x_i \leq y_i$ holds in $\mathcal{B}$ for every $i \in N$.

3 Two-sided systems of equations

We shortly describe the polynomial method for finding the maximum solution of a (max,min)-linear two-sided system over a given bounded max-min algebra, which has been proposed in [7]. We remark that in [7], the algebra is equal to the set $\overline{\mathcal{R}}$ of reals with added infinite elements ∞ and $-\infty$, but all definitions and proofs from [7] can also be formulated for arbitrary bounded max-min algebra $\mathcal{B}$ with bound elements O, I.

Let matrices $A, B \in \mathcal{B}(m,n)$ be given. The following two-sided system of (max, min)-linear equations over $\mathcal{B}$ with variable vector $x \in \mathcal{B}(n)$ will be considered

$$\bigoplus_{j \in N} (a_{ij} \otimes x_j) = \bigoplus_{j \in N} (b_{ij} \otimes x_j) \quad \text{for } i \in M \ . \tag{1}$$

The set of all solutions of the system (1) will be denoted by $S(A,B)$. Further we define the sets $S(A,B,\overline{x})$ $S(A,B,\underline{x},\overline{x})$ for any $\underline{x}, \overline{x} \in \mathcal{B}(n)$ as follows

$$S(A,B,\overline{x}) := \{x;\, x \in S(A,B),\, x \leq \overline{x}\} \ . \tag{2}$$

$$S(A,B,\underline{x},\overline{x}) := \{x;\, x \in S(A,B),\, \underline{x} \leq x \leq \overline{x}\} \ . \tag{3}$$

If no confusion arises, we often use shorter notation $S = S(A,B)$, $S(\overline{x}) = S(A,B,\overline{x})$ and $S(\underline{x},\overline{x}) = S(A,B,\underline{x},\overline{x})$.

Remark 1. The solution sets S, $S(\overline{x})$ are always nonempty, since $x(\alpha) \equiv (\alpha, \ldots, \alpha)$ belongs to $S(\overline{x})$, if $\alpha \leq \min\{a_{ij} \otimes b_{ij} \otimes \overline{x}_j;\, (i,j) \in M \times N\}$. Further, for $\overline{x} \equiv (I, \ldots, I)$ we have $S(\overline{x}) = S$.

Every element $\tilde{x} \in S$ ($\in S(\overline{x})$) such that for every $x \in S$ ($\in S(\overline{x})$) the inequality $x \leq \tilde{x}$ holds true is called the *maximum solution* in S (in $S(\overline{x})$). A polynomial algorithm for computing the maximum solution in $S(\overline{x})$ is described in [7]. The formulation below is modified according to the notation introduced in this paper.

Theorem 1. *[7] There exists an $O(mn(m+n))$ algorithm $\mathcal{A}_{\max}$ which finds the maximum solution $x^{\max}$ in $S(\overline{x})$ for any given $\overline{x} \in \mathcal{B}(n)$.*

Remark 2. It follows from the proof of Theorem 5 presented in [7] that all vector $x_j^{\max}$, $j \in N$ belong to the set of all matrix inputs $\{\, a_{ij},\, b_{ij};\, (i,j) \in M \times N\,\}$.

Remark 3. The solution set $S(\underline{x},\overline{x})$ is nonempty if and only if $\underline{x} \leq x^{\max}$.

4 Solving two-sided systems

An efficient method for computing the maximum solution of a given two-sided system is described in this section. We start with an easy observation.

Theorem 2. *Let $A, B \in \mathcal{B}(m,n)$, $\overline{x} \in \mathcal{B}(n)$ then*

$$S(A,B) = S(A, A \oplus B) \cap S(B, A \oplus B) \ , \tag{4}$$

$$S(A,B,\overline{x}) = S(A, A \oplus B, \overline{x}) \cap S(B, A \oplus B, \overline{x}) \ . \tag{5}$$

Proof. Let $x \in S(A,B,\overline{x})$, then $A \otimes x = B \otimes x$ and $x \leq \overline{x}$. By distributivity law we get $(A \oplus B) \otimes x = (A \otimes x) \oplus (B \otimes x) = A \otimes x$. Therefore, $x \in S(A, A \oplus B, \overline{x})$. Analogously we prove $x \in S(B, A \oplus B, \overline{x})$. Hence, $S(A,B,\overline{x}) \subseteq S(A, A \oplus B, \overline{x}) \cap S(B, A \oplus B, \overline{x})$.

For the converse inclusion, let us suppose that $x \in S(A, A \oplus B, \overline{x}) \cap S(B, A \oplus B, \overline{x})$. Then $x \leq \overline{x}$ and $A \otimes x = (A \oplus B) \otimes x = B \otimes x$. Therefore, $x \in S(A,B,\overline{x})$. We have proved equality (5). The previous equality (4) follows in view of Remark 1. □

Theorem 3. *Let $A, B \in \mathcal{B}(m,n)$, $\overline{x} \in \mathcal{B}(n)$, let matrices $\tilde{A}, \tilde{B} \in \mathcal{B}(m,n)$ be defined by formulas*

$$\tilde{a}_{ij} = a_{ij} \otimes \overline{x}_j \ , \quad \tilde{b}_{ij} = b_{ij} \otimes \overline{x}_j \quad \text{for every } i \in M,\ j \in N \ . \tag{6}$$

then

$$S(A, B, \overline{x}) = S(\tilde{A}, \tilde{B}, \overline{x}) \ , \tag{7}$$

Proof. For every $x \leq \overline{x}$ we have $x = x \otimes \overline{x}$. In view of (6) we get for $i \in M, j \in N$

$$a_{ij} \otimes x_j = a_{ij} \otimes (x_j \otimes \overline{x}_j) = (a_{ij} \otimes \overline{x}_j) \otimes x_j = \tilde{a}_{ij} \otimes x_j \ ,$$

therefore $A \otimes x = \tilde{A} \otimes x$. Similarly we get $B \otimes x = \tilde{B} \otimes x$. Hence $S(A, B, \overline{x}) = S(\tilde{A}, \tilde{B}, \overline{x})$. □

Theorem 4. *Let $A, B \in \mathcal{B}(m,n)$, $\overline{x} \in \mathcal{B}(n)$, let*

$$\bigoplus_{i \in M} a_{ij} \oplus \bigoplus_{i \in M} b_{ij} \leq \overline{x}_j \quad \text{for every } j \in N \ . \tag{8}$$

Then $\overline{x}$ is the maximum element in $S(A, B, \overline{x})$ if and only if

$$\bigoplus_{j \in N} a_{ij} = \bigoplus_{j \in N} b_{ij} \quad \text{for every } i \in M \ . \tag{9}$$

Proof. Condition (8) is equivalent with the set of inequalities $a_{ij} \leq \overline{x}_j$, $b_{ij} \leq \overline{x}_j$ for every $i \in M, j \in N$, which imply $a_{ij} \otimes \overline{x}_j = a_{ij}$ and $b_{ij} \otimes \overline{x}_j = b_{ij}$. Hence, the condition $\overline{x} \in S(A, B, \overline{x})$ is equivalent the equality

$$\bigoplus_{j \in N} (a_{ij} \otimes \overline{x}_j) = \bigoplus_{j \in N} (b_{ij} \otimes \overline{x}_j) \quad \text{for every } i \in M \ ,$$

which is equivalent with (9). □

Remark 4. The idea of the new algorithm for computing the maximum element in $S(A, B, \overline{x})$ uses Theorem 2 and Theorem 4. The method is based on modifying given matrices A, B with $A \leq B$, and given vector $\overline{x}$ in successive steps, so that the modified matrices $\tilde{A}, \tilde{B}$ and vector $\tilde{x}$ would satisfy the condition (8), and in each step the set of the indices $i \in M$ satisfying the equality in (9) would increase, until it is equal to M. Moreover, the modified system should have the same maximum solution as the original one, i.e. $\max S(A, B, \overline{x}) = \max S(\tilde{A}, \tilde{B}, \tilde{x})$. The details are described in what follows.

Let matrices $A, B \in \mathcal{B}(m,n)$ with $A \leq B$ and vector $\overline{x} \in \mathcal{B}(n)$ satisfying condition (8) be fixed. We denote by m^A, m^B the vectors containing the row maxima of matrices A, B, i.e.

$$m_i^A = \bigoplus_{j \in N} a_{ij} \ , \quad m_i^B = \bigoplus_{j \in N} b_{ij} \quad \text{for every } i \in M \ . \tag{10}$$

If condition (9) is not satisfied, then there is at least one index $i \in M$ such that $m_i^A < m_i^B$ holds. Then we define value $\alpha = \alpha(A, B)$ by putting

$$\alpha = \min \left\{ m_i^A;\, m_i^A < m_i^B,\, i \in M \right\} \ . \tag{11}$$

Further we define index sets

$$M^{<}(\alpha) = \left\{ i \in M;\, m_i^A < \alpha \right\} \ , \tag{12}$$

$$M^{==}(\alpha) = \left\{ i \in M;\, \alpha = m_i^A = m_i^B \right\} \ , \tag{13}$$

$$M^{=<}(\alpha) = \left\{ i \in M;\, \alpha = m_i^A < m_i^B \right\} \ . \tag{14}$$

By the above definitions, the equality in condition (9) is fulfilled for all $i \in M^{<}(\alpha) \cup M^{==}(\alpha)$. Moreover, $M^{=<}(\alpha) \neq \emptyset$.

For every index $i \in M^{=<}(\alpha)$ we have $\alpha = m_i^A < m_i^B = \bigoplus_{j \in N} b_{ij}$, hence there is some index $j \in N$ with $\alpha < b_{ij}$. We write $i \in M_j(\alpha)$, using notation

$$M_j(\alpha) = \left\{ i \in M^{=<}(\alpha);\, \alpha < b_{ij} \right\} \ . \tag{15}$$

Vector $\tilde{x} \in \mathcal{B}(n)$ is defined by putting

$$\tilde{x}_j = \begin{cases} \alpha & \text{if} \quad M_j(\alpha) \neq \emptyset \ , \\ \overline{x}_j & \text{otherwise} \ . \end{cases} \tag{16}$$

Finally, matrices $\tilde{A}, \tilde{B} \in \mathcal{B}(m, n)$ are defined similarly as in (6), with $\overline{x}$ replaced by $\tilde{x}$,

$$\tilde{a}_{ij} = a_{ij} \otimes \tilde{x}_j \ , \quad \tilde{b}_{ij} = b_{ij} \otimes \tilde{x}_j \quad \text{for every } i \in M,\ j \in N \ . \tag{17}$$

Lemma 1. Let $A, B \in \mathcal{B}(m, n)$ with $A \leq B$ and $\overline{x} \in \mathcal{B}(n)$ with (8) are given, let $\tilde{A}, \tilde{B} \in \mathcal{B}(m, n)$, $\tilde{x} \in \mathcal{B}(n)$ are defined by formulas (16) and (17). Then $\tilde{A} \leq \tilde{B}$, $\tilde{A} \leq A$, $\tilde{B} \leq B$, $\tilde{x} \leq \overline{x}$ and the following conditions hold

$$S(A, B, \overline{x}) = S(\tilde{A}, \tilde{B}, \tilde{x}) \ , \tag{18}$$

$$\bigoplus_{i \in M} \tilde{a}_{ij} \oplus \bigoplus_{i \in M} \tilde{b}_{ij} \leq \tilde{x}_j \qquad \text{for every } \ j \in N \ , \tag{19}$$

$$\bigoplus_{j \in N} \tilde{a}_{ij} = \bigoplus_{j \in N} \tilde{b}_{ij} \qquad \text{for every } \ i \in M^{<}(\alpha) \cup M^{==}(\alpha) \cup M^{=<}(\alpha) \ , \tag{20}$$

$$\alpha(A, B) \leq \alpha(\tilde{A}, \tilde{B}) \quad \text{if the equality in (20) does not hold for every } i \in M \ . \tag{21}$$

Proof. The inequalities $\tilde{A} \leq \tilde{B}$, $\tilde{A} \leq A$, $\tilde{B} \leq B$, $\tilde{x} \leq \overline{x}$ and the condition (19) follow directly from the assumptions set on A, B, $\overline{x}$ and from definitions (16) and (17).

Replacing $\overline{x}$ by $\tilde{x}$ in Theorem 3 we get $S(A, B, \tilde{x}) = S(\tilde{A}, \tilde{B}, \tilde{x})$. Therefore, the equality (18) is equivalent with $S(A, B, \overline{x}) = S(A, B, \tilde{x})$. By an easy computation we can verify that $S(A, B, \tilde{x}) \subseteq S(A, B, \overline{x})$.

For the proof of the converse inclusion, let us suppose that $x \in S(A, B, \overline{x})$ for some $x \in \mathcal{B}(n)$. Then $A \otimes x = B \otimes x$ and $x \leq \overline{x}$ hold true. We only have to verify that $x \leq \tilde{x}$. Let $j \in M$ be arbitrary, but fixed. If there exist $i \in M_j(\alpha)$, then in view of (14), (15) and (16) we have $x_j = \alpha = m_i^A < m_i^B$. By assumption $x \in S(A, B, \overline{x})$ we then have $A_i \otimes x = B_i \otimes x$. From $m_i^A = \alpha$ we get $A_i \otimes x \leq \alpha$, which implies $B_i \otimes x \leq \alpha$ and $b_{ij} \otimes x_j \leq \alpha$. Then, in view of $\alpha < b_{ij}$, the inequality $x_j \leq \alpha = \tilde{x}_j$ must hold true. On the other hand, if $M_j(\alpha) = \emptyset$, then $\tilde{x}_j = \overline{x}_j$, hence $x_j \leq \tilde{x}_j$. We have proved $S(A, B, \overline{x}) \subseteq S(A, B, \tilde{x})$.

The condition (20) easily follows from definitions (13)-(17), and the condition (21) is a consequence of (11) and (20). $\square$

5 Efficient algorithm

Using the above results we suggest the following algorithm $\mathcal{A}^{\star}_{\max}$ for computation of the maximum solution of a two-sided system of max-min equations.

ALGORITHM $\mathcal{A}^{\star}_{\max}$

$\boxed{1}$ Input $A, B \in \mathcal{B}(m, n)$, $\overline{x} \in \mathcal{B}(n)$;
$\boxed{2}$ Put $c := 1$, $N^{\star} := N$, $C := B$, $B := A \oplus B$ and $\tilde{x} := \overline{x}$;
$\boxed{3}$ For all $i \in M$, $j \in N$ put $a_{ij} := a_{ij} \otimes \tilde{x}_j$ and $b_{ij} := b_{ij} \otimes \tilde{x}_j$;
$\boxed{4}$ For all $i \in M$ put $m_i^A := \bigoplus_{j \in N} a_{ij}$ and $m_i^B := \bigoplus_{j \in N} b_{ij}$;
$\boxed{5}$ Put $\alpha := \min \left\{ m_i^A ;\ m_i^A < m_i^B,\ i \in M \right\}$;
$\boxed{6}$ For all $j \in N^{\star}$, $i \in M$ with $m_i^A = \alpha$ do
if $\alpha < m_i^B$ then put $\tilde{x}_j := \alpha$ and $N^{\star} := N^{\star} \setminus \{j\}$;
$\boxed{7}$ If $N^{\star} \neq \emptyset$ then goto $\boxed{3}$
$\boxed{8}$ Put $c := c + 1$, $N^{\star} := N$ and $A := C$;
$\boxed{9}$ If $c = 2$ then goto $\boxed{3}$
$\boxed{10}$ STOP

Theorem 5. *Algorithm $\mathcal{A}^{\star}_{\max}$ works correctly and finds the maximum solution $x^{\max} \in S(A, B, \overline{x})$ of a two-sided system of max-min equations for any given inputs $A, B \in \mathcal{B}(m, n)$, $\overline{x} \in \mathcal{B}(n)$ in $O(mn \log mn)$ time.*

Proof. Assertion of the theorem follows from considerations in Section 4. Computation and updating of the row maxima m^A and m^B is performed by arranging initially the row values in the increasing order, so that the maximal value can be easily found as the last one in the ordering. Updating of the ordering must be done in every cycle, but taking in account that every variable is changed at most once, we see that every value in a row changes its position in the ordering at most once. Hence the necessary updating can be done in $O(mn)$ time, while the initial ordering requires $O(mn \log n)$ time. Similarly, ordering of rows in the increasing order of values m_i^A is used. Then every $i \in M$ is considered at most once in step $\boxed{6}$ which can be therefore performed in total $O(mn)$ time, with the initial ordering of rows taking $O(m \log m)$ time. Hence, the computational complexity of $\mathcal{A}^{\star}_{\max}$ is $O(mn \log mn)$. $\square$

In the example below we describe a transport problem in which the method described in this paper is applied (the problem is a slight modification of the example presented in [7]).

Example 1. Let us consider means of transport of different sizes, which are transporting goods from places $i \in M$ to a common terminal T. The goods are unloaded in T and the transport means (possibly with other goods uploaded in T) will return to i.

We assume that the connection between every place i and T is only possible via one of the places (e.g. cities) $j \in N$, the roads between i and j are one-way roads, and the capacity of the road between $i \in M$ and $j \in N$ is equal to a_{ij}. We have to join every place j with T by a two-way road with a capacity x_j in both directions. The maximal capacity of the connection between i and T is therefore equal to $\max\{a_{ij} \otimes x_j;\, j \in N\}$. The transport from T to i is carried out via other one-way roads between places $j \in N$ and $i \in M$ with capacities between j and i equal to b_{ij}. Since the roads between T and j are two-way roads, the maximal capacity of the connection between T and i is equal to $\max\{b_{ij} \otimes x_j;\, j \in N\}$ for all $i \in M$.

We assume that the means of transport can only pass through some roads with the capacity which is not smaller than the capacity of the transport mean and our task is to choose appropriate capacities x_j, $j \in N$. In order that each of the transportation means could return to i, it is natural to require for each i that the maximal attainable capacity of connections between i and T via j is equal to maximal attainable capacity of connections between T and i on the way back. In other words, we have to choose x_j, $j \in N$ in such a way that

$$\max\{a_{ij} \otimes x_j;\, j \in N\} = \max\{b_{ij} \otimes x_j;\, j \in N\} \quad \text{for all} \quad i \in M \ .$$

We see that the problem transforms to solving system (1) with some finite upper bounds on x, since in reality the chosen capacities cannot be unbounded.

Remark 5. The above model is flexible enough to include different real situations. E.g., if the road between i and j does not exist, we set simply $a_{ij} = 0$. If the road between i and j is a two way road with equal capacity in both directions, we set $b_{ij} = a_{ij}$, or, if we do not want to connect j with T, we set the lower and upper bound on x_j equal to 0. Further, the connections of T with places $j \in N$ need not have the same capacity x_j in both directions: in such case we insert different variables y_j on the right-hand side of the system. Then the transformation to the form (1) is only a technical problem: we shall use a vector of variables $(x, y) \in R^{2n}$ and we shall introduce additional coefficients $a_{ij} = -\infty$ for $j = n + 1, \ \ldots, \ 2n$ on the left hand side and $b_{ij} = -\infty$ for $j = 1, \ \ldots, \ n$ on the right hand side of every equation.

Acknowledgements

The support of Czech Science Foundation # 402/09/0405 and MSM0021620838 is kindly acknowledged.

References

[1] Baccelli, F. L., Cohen, G., Olsder, G. J., and Quadrat, J. P.: *Synchronization and Linearity. An Algebra for Discrete Event Systems.* Wiley, Chichester, 1992.

[2] Butkovič, P., and Hegedüs, G.: An Elimination Method for Finding All Solutions of the System of Linear Equations over an Extremal Algebra, *Ekonomicko–matematický obzor* **20** (1984), 203–215.

[3] Butkovič, P., and Zimmermann, K.: A Strongly Polynomial Algorithm for Solving Two-sided Linear Systems in Max-algebra, *Discrete Applied Mathematics* **154** (2006), 437–446.

[4] Cechlárová, K.: Efficient Computation of the Greatest Eigenvector in Fuzzy Algebra, *Tatra Mt. Math. Publications* **12** (1997), 73–79.

[5] Cuninghame-Green, R. A.: *Minimax Algebra.* Lecture Notes in Economics and Mathematical Systems **166**, Springer–Verlag, Berlin 1979.

[6] Cuninghame-Green, R. A., and Zimmermann, K.: Equation with Residual Functions, *Comment. Math. Univ. Carolinae* **42** (2001), 729–740.

[7] Gavalec, M., and Zimmermann, K.: Solving Systems of Two-Sided $(\max, \min)$-Linear Equations, *Kybernetika* **46** (2010), 405–414.

[8] Gavalec, M., and Zimmermann, K.: Optimization problems with two-sided systems of linear equations over distributive lattices, *Proc. Conf. Mathematical Methods in Economics* (2010), 172–176.

[9] Litvinov, G. L., Maslov, V. P., and Sergeev, S. N. (eds.): *Idempotent and Tropical Mathematics and Problems of Mathematical Physics, vol.1.* Independent University Moscow, 2007.

[10] Sanchez, E.: Resolution of Eigen Fuzzy Sets Equations, *Fuzzy Sets and Systems* **1** (1978), 69–74.

[11] Vorobjov, N. N.: Extremal Algebra of positive Matrices, *Datenverarbeitung und Kybernetik* **3** (1967), 39–71 (in Russian).

Options for return of products to manufacturer

Pavel Gežík[1], Anna Hollá[2]

Abstract. Main topic of reverse logistics is return of end-of-life product from the point of final consumption back to the manufacturer. The reverse logistics deals primarily with the optimization of return processes or more precisely with the processes of collection and return of end-of-life products back to the manufacturer for the purpose of their reuse and remanufacturing.

This return can be facilitated by wide range of subjects and in different ways. It can be provided by the manufacturer itself or by other subjects of supply chain. There are many ways and corresponding processes which can facilitate this return. Particular technique, by which the return is facilitated, is dependent on the nature of products. Product can be damaged, destroyed beyond use or simply outdated. Condition of the product determines possibilities of return related to the purpose of collection of these products. Purpose can range from repair, maintenance or upgrade to another form of modernization. Product can be also returned for remanufacturing, reuse of certain parts, recycling or environmental friendly liquidation. Last factor influencing the return is timescale and repetition, which can be one time, repeated, long or short term.

Keywords: return, reverse logistics, remanufacturing, recycling

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Reverse Logistics is a new concept that deals with the management of the products in the reverse way, i.e. it is the process of managing all the flow of returned products and information from the point of consumption to the origin. The first by Stock (1992) recognized the field of reverse logistics as being relevant for business and society in general. One year later Kopicki et al. (1993) paid attention to the discipline and practice of reverse logistics, pointing out opportunities on reuse and recycling. In the late nineties, several other studies on reverse logistics appeared. The actual development of Reverse Logistics strategies in many companies across the world makes it a very interesting topic to work with.

Rogers and Tibben-Lembke (1999) presented a broad collection of reverse logistics business practices, giving special attention to the US experience, where the authors carried out a comprehensive questionnaire. It is expected that in few years it will become a crucial element in determining the way in which products will be designed, produced and distributed. The actual legislation of the European Union gives high importance to the recycled products and, in some cases, it has established the responsibility for the end of life products to the manufacturers. Therefore, in the future most of the European companies will have to think about incorporating Reverse Logistics activities in their business operations.

A number of factors, including the life-cycle stage of a product and the rate of technological change, influence the quality and quantity of the returns. This characteristic has a marked impact on demand management, and inventory control. The high level of uncertainty arises given the different characteristics in terms of quantity and quality of the returned products, makes the production planning task more complicated and increases the complexity of the inventory control process. To understand this complexity, it is a good idea to analyze why people return the products.

This paper is giving complex view of reverse logistics in a field of inventory management. It describes the return of End-of-Live products, the reasons of this return and ways of return. There is list of recovery options for returned products from resale through remanufacturing to recycling. Next part of paper deals with material requirements planning (MRP) and rate of return EOL product for remanufacturing or recycling. Last part of paper focuses on inventory in process of remanufacturing considering about two kinds of stock are supplemented by opportunity stock which is typical for remanufacturing.

[1] University of Economics Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, pavelgezik@gmail.com

[2] University of Economics Bratislava, Faculty of Economic Informatics, Department of Operations Research and Econometrics, Dolnozemská cesta 1/b, 852 35 Bratislava, Slovakia, annaholla@gmail.com

2 Return

Typical reverse logistics activities would be the processes a company uses to collect used, damaged, unwanted (stock balancing returns) or outdated products, as well as packaging and shipping materials from the end-user or the reseller.

Once a product has been returned to a company, the firm has many disposal options from which to choose. If the product can be returned to the supplier for a full refund, the firm may choose this option first. If the product has not been used, it may be resold to a different customer, or it may be sold through an outlet store. If it is not of sufficient quality to be sold through either of these options, it may be sold to a salvage company that will export the product to a foreign market.

If the product cannot be sold "as is," or if the firm can significantly increase the selling price by reconditioning, refurbishing or remanufacturing the product, the firm may perform these activities before selling the product. If the firm does not perform these activities in-house, a third party firm may be contracted, or the product can be sold outright to a reconditioning/remanufacturing/refurbishing firm. After performing these activities, the product may be sold as a reconditioned or remanufactured product, but not as new.

If the product cannot be reconditioned in any way, because of its poor condition, legal implications, or environmental restrictions, the firm will try to dispose of the product for the least cost. Any valuable materials that can be reclaimed will be salvaged.

Based on these facts Dekker (2004) created recovery option inverted pyramid. Recovery is actually only one of the activities involved in the whole process of reverse logistics. First there is collection, next there is the combined inspection/selection/sorting process, thirdly there is recovery (which may be direct or may involve a form of reprocessing), and finally there is redistribution. Collection refers to bringing the products from the customer to the point of recovery. At this point the products are inspected, i.e. their quality is assessed and a decision is made on the type of recovery. Products can then be sorted and routed according to recovery that follows.

If the quality is (close to) as-good-as-new, products can be fed into market almost immediately thorough reuse, resale, and redistribution. If not, another type of recovery may be involved that now demands more action, i.e. a form of reprocessing. Reprocessing can accrue at different levels: product level (repair), module level (refurbishing), component level (remanufacturing), selective part level (retrieval), material level (recycling), and energy level (incineration).

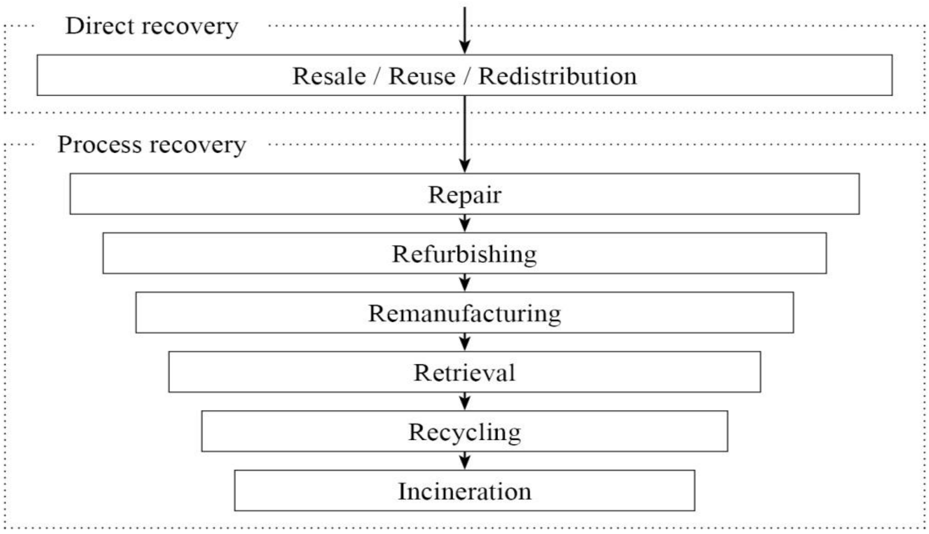


Figure 1 Recovery option inverted pyramid – (Dekker et al. 2004)

2.1 Reason of return

Tibben-Lembke (2002) describes the principal causes for what the people return the products. These ways of return of the products to the manufacturer may be divided into two main categories. In the first one the products do not change their character. We can include here return for the repair: *Factory Repair* (Return to vendor for repair), *Service, Maintenance, Agent Order Error* (Sales agent ordering error) *Customer Order Error* (Ordered wrong material), *Entry Error* (System processing error), *Shipping Error* (Shipped wrong material), *Incomplete Shipment* (Ordered items missing), *Wrong Quantity*, *Duplicate Shipment, Duplicate Customer Order, Not Ordered, Missing Part*.

If the product cannot be repaired the reason for the return may be the product is *Damaged* (Cosmetic), *Dead on Arrival* (Did not work), *Defective* (Not working correctly). Next reasons for the products return may be the contractual agreements: *Stock Excess* (Too much stock on hand), *Stock Adjustment* (Rotation of stock) or that the product is *Obsolete* (Outdated). Other reasons are *Freight Claim* (Damaged during shipment) or *Miscellaneous*.

Second category is comprised of returns when the products change their character. It can mean *Disposal of the product, Scrap, Destroy, Secure Disposal, Donate to Charity, Third Party Disposal, Salvage, Third Party Sale* (Secondary Markets) or modification of the product: *Rework, Remanufacture, Refurbish, Modify* (Configurable or Upgradable Products), *Repair, Return to Vendor*. At least the reasons of the return can be associated with *Use as Is, Resale, Exchange and Miscellaneous*.

2.2 Ways of return

In a truly integrated supply chain, everyone in the supply chain can track product as it moves forward through the channel. While there are very few supply chains that really function this well, there are virtually none that work in reverse. Most firms cannot track returns within their own organization, much less somewhere outside of their firm.

Depending on the reason of return ways the products are returned change as well. These ways depend on the subjects which are processing the return. Based on the relationship between manufacturer and retailer we can assume four possible ways of product return back to manufacturer. First is the closed-loop supply chain where the manufacturer collecting used products is retailer itself, the second is when the products are returned directly to the manufacturer, which is not the retailer. In the third way the products are returned through retailer which is collecting used products and the fourth, when the return is facilitated by a third party. These ways of products return are displayed in the figure bellow.

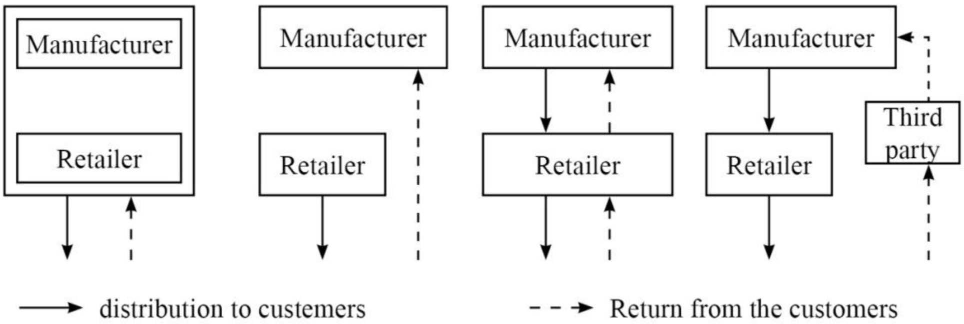


Figure 2 Ways of return (Savaskan et al. 2004)

Retailer

In a returns processing system that may reside at a centralized return center, several transactions can occur. A good system might include the following steps. The first transaction will likely be financial, where an inventory category will be updated. A chargeback to reconcile with the vendor, or something similar, will occur. A retailer may want to reorder first quality product from its supplier immediately. Then, routing for processing or a storage location within the processing center will be determined. A reverse warehouse management system may be required for this step.

Manufacturer

The manufacturer will generate a return authorization (RA). This is often a manual process. RAs could be generated electronically, including an automatic check to see if the return should be authorized. Next, the likely financial impact of the return could be generated. These capabilities would be very helpful in better managing returns. The next step is to automate pickup of product and an advanced shipping notification (ASN) could be cut. After it is shipped, it is received. Currently, most manufacturers receive returns manually. Once the material is received, a database is created for reconciliation. Because most manufacturers receive material manually, this database is created slowly-if it is created at all. This sluggishness results in slowing of the reconciliation and the disposition of the returns.

It is important to remark that not only the final customers can return the products, but retailers and distributors as well. The process of reception of the returned products implies some different activities of revision and control that determine the actual state of the product. Only after that it is possible to determine the best strategy how to dispose of product, in the most cost effective manner.

3 Material Requirements Planning

In reverse logistics, as a consequence of the return flow, the inventory level between new component replenishments is no longer necessarily decreasing but may increase also. This loss of monotonicity significantly complicates the underlying mathematical models. A possible starting point for a closer analysis of this aspect is the cash balancing models comprising in and outbound flows.

There are two alternatives for fulfilling the demand that impose an additional set of decisions to be taken. External orders and recovery have to be coordinated. Daniel Guide, V. et al (2000) determine seven characteristics of the recoverable manufacturing systems that complicate the management, planning, and control of supply chain functions. They are:
- The uncertain timing and quantity of returns
- The need to balance demands with returns
- The need to disassemble the returned products
- The uncertainty in materials recovered from returned items
- The requirement for a reverse logistics network
- The complication of material matching restrictions

The problems of stochastic routings for materials for repair and remanufacturing operations are highly variable processing times. As we have mentioned before, it is possible that we have to use different parts from different returned products to produce a specific product during the remanufacturing process and also to mix them with new parts. This also complicates the production process.

Guide et al (2000) shows a comparison between manufacturing and remanufacturing environment and the impact it has over the functional areas within an organization. Traditional MRP-systems are not feasible for recovery situations for several reasons. One of the main problem is the mismatch of supply and demand, due to the simultaneous release of 'wanted' and 'unwanted' components in the disassembly of returned products. A second major problem is the tradeoff between reusing return components and outside procurement. In Remanufacturing, new products are manufactured using three kinds of components:
- Components that are always retrieved from return products (the quantity is unknown)
- Components that are always purchased new
- Components that can either be purchased new or retrieved from return products, depending on availability and costs.

3.1 Rate of return

Fleischmann et al. (1997) states the following arguments, the repair operations needed to convert a returned product back to an 'as new' state depend on the actual condition of the product. This may vary from instance to instance and generally can be decided only after a number of testing and disassembly operations.

Therefore, in contrast with the traditional manufacturing, no well-determined sequence of production steps exists in remanufacturing. This exposes planning in a remanufacturing environment to a much higher uncertainty. A high level of coordination required in remanufacturing is therefore a result of interdependence between different parts and subassemblies. Disassembly of a returned product is not a procurement source for one part but releases various.

The relationship between manufacturing and remanufacturing and their uncertainty describe figure, where is:
- d - demand
- α – rate at which products are returned to manufacturer ,
- β – rate at which returned products are used,
- r_{tr} – quantity of returned products, $r_{tr} = \alpha d$
- r_r – quantity of used returned products $r_r = \alpha\beta d$
- ω – disposal (it means the quantity of products which are not returned or reused), $\omega = d - (r_{tr} - r_r)$

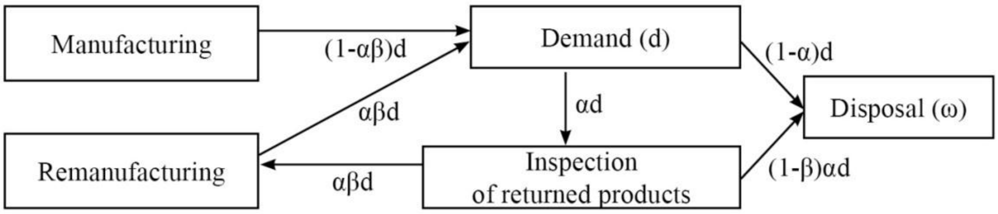


Figure 3 Manufacturing vs. Remanufacturing

3.2 Inventory in process of remanufacturing

In the process of manufacturing there is classic cycle stock of material needed for production and safety stock related to the fluency of production. When the returned products are used in the manufacturing these two kinds of stock are supplemented by opportunity stock which is typical for remanufacturing.

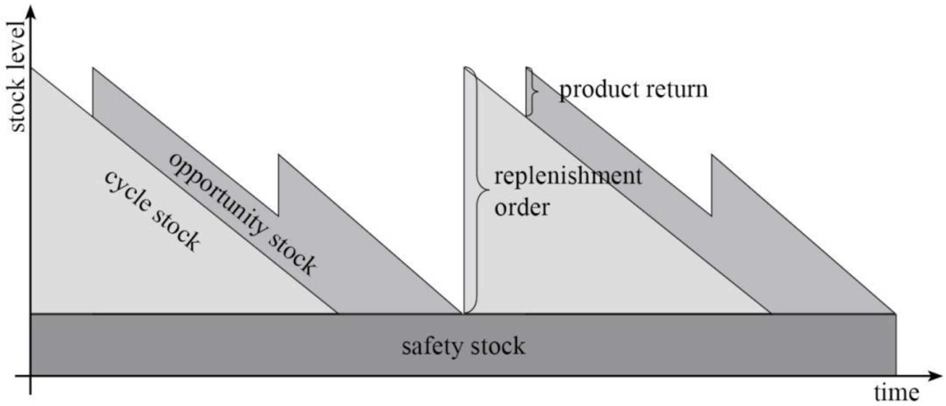


Figure 4 Inventory (Dyckhoff et al. 2003)

These opportunity stocks can be classified based on the quantity of returned product. Figure below illustrates state where the quantity of returned product is constant in every period (a)). However the quantities of return may be increasing in every consecutive period, (as seen in b)) or quantity is constant in every period but is returned more frequently (c)).

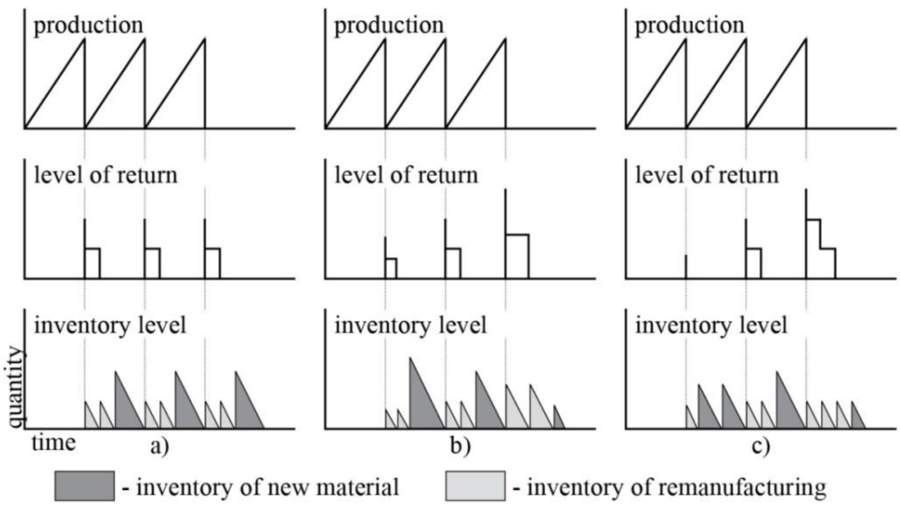


Figure 5 Inventory level in planning in a remanufacturing (Schulz & Ferretti 2008)

4 Conclusion

As we can see the return of products back to the manufacturer is effective not only in economic sense but environmentally as well. Based on these two facts the companies are focusing more on the planning of processes related to the return of the end-of-life products for their remanufacturing. Options for the returned products are numerous and include processes ranging from resale, reuse or redistribution, to repair, modification, remanufacturing, refurbishment, recycling, salvage and environmental friendly liquidation.

However the quantity and quality of returned products is uncertain which makes material requirements planning difficult. This uncertainty led to the development of models dealing with return, quantity and rate of return and also inventory management connected with inventory return.

References

[1] DEKKER, R. et al. 2004. *Reverse Logistics*. Berlin : Springer-Verlag, 2004. 436 s. ISBN 3-540-40696-4.

[2] DYCKHOFF, H. – LACKES, R. – REESE, J. 2003. *Supply Chain Management and Reverse Logistics*. Berlin : Springer, 2003. 426 s. ISBN 3-540-40491-0.

[3] FLEISCHMANN, M. et al. 1997. *Quantitative models for reverse logistics: A review*, In European Journal of Operational Research, Vol. 103, 1997. 1-17 s. ISSN: 0377-2217.

[4] GUIDE, V. D. R. Jr. et al. 2000. *Production planning and control for remanufacturing: industry practice and research needs*. In Journal of Operations Management, Vol. 18, Issue 4, 2000. ISSN 0272-6963..

[5] SAVASKAN, R. C. – BHATTACHARYA, S. – WASSENHOVE, L. N. V. 2004. *Closed-Loop Supply Chain Models with Product Remanufacturing*, Management Science, Vol. 50,No. 2, 2004. ISSN (printed): 0025-1909. ISSN (electronic): 1526-5501.

[6] SCHULZ, T. - FERRETTI, I.: *On the alignment of lot sizing decisions in a remanufacturing system in the presence of random yield*. Faculty of Economics and Management Magdenburg Working Paper No. 34, 2008

[7] TIBBEN-LEMBKE, R. S. 2002. *Life after death: reverse logistics and the product life cycle*. In International Journal of Physical Distribution & Logistics Management, Vol. 32. 2002. ISSN 0960-0035.

The stability investigation of the three large Czech banks within Z - metrics methodology

Petr Gurný[1]

Abstract. The paper is devoted to the investigation of the Czech banks health, which we can regard as one of the most important tasks in the time of the financial crisis. The main goal of the paper is an estimation of the future probability of default (PD) for three key Czech banks. At first the model (built on the basis of the Z-metrics methodology) for prediction of bank failure will be presented. Afterwards the relevant financial indicators needed for estimation of the future PD will be simulated via Lévy processes and their dependencies will be captured via gaussian copula function.

Keywords: Z-metrics, logit model, VG process, Gaussian copula function.

JEL Classification: C01, C51, G17
AMS Classification: 91G40, 91G60

1 Introduction

The internationalization of financial markets has significantly expanded investment opportunities and risk. Importance of coherent estimation of PD is manifold. Credit risk assessment and especially the estimation of default probabilities have relevance not only to asset prices in credit and debt markets, but also in equity and in many types of derivate markets. Designing of efficient techniques for estimation of PD is in the spotlight of many research units especially since 2008, when most of the world has been going through a period of financial and economic turmoil.

The goal of the paper is to estimate the probability distribution of PD for several Czech banks assuming that the significant financial indicators follow multidimensional subordinated VG process. To do that, we will first design the model for estimation of PD on the basis of the Z-metrics methodology (logit approach). Subsequently we will apply this model on three Czech commercial banks with a view to estimate their probability distribution of PD with quarterly prediction. It's generally expected that Czech financial market is better regulated than American one. At the conclusion the stress test on the impact of the extreme GDP downs on the analyzed banks stability will be performed.

For estimation the prediction of PD it is necessary to simulate particular financial indicators. Presently it is tendency to analyze and to simulate the time series of the particular variables by means of Lévy processes as it could be helpful to better predictor ability of the models, see Cont and Tankov [4] or Schoutens [10] In this paper we will simulate particular financial indicators by means of Variance Gamma (VG) process. Crucial factor within modeling of the particular indicators is also dependence. There will be used the Gaussian copula function approach to capture the dependencies among indicators in the paper.

We proceed as follows. In the theoretical part of the paper we expand in more detail the Z-metrics methodology (logit approach), Lévy processes and copula function. Further we will estimate and verify the mode and we will apply it on the portfolio of three Czech commercial banks to determine their distributions of the PDs and their stability.

2 Z-metrics methodology

Z-metrics methodology, presented by Altman [1], is approach based on combination of the RiskMetrics and more recent analysis utilized a multivariate logistic regression structure to construct the model. The crucial point of the approach is the inclusion of the macroeconomic data into the sample of the independent variable, with a view to catch the influence of the main macro indicators on the firms' PD.

Multivariate logistic regression is the multivariate techniques which allows for estimating the probability that an event occurs or not, by predicting a binary dependent outcome from a set of independent variables.

The response, y_i, is given

[1] VSB – TUO, Department of Finance, Sokolska 33, Ostrava, 701 21, Czech Republic, petr.gurny@vsb.cz.

$$y_i = \begin{cases} 1 & \text{if default occurs, with probability } P_i \\ 0 & \text{if default does not occur, with probability } 1\text{-}P_i \end{cases}$$

Thus the goal is to model probability P_i that default occurs by specifying the model

$$P_i = f(\alpha + \beta x_i),$$

where x_i are particular financial indicators and α, β are estimated parameters.

Now we will specify P_i as a logit transformation

$$p_i = E(Y|x)$$
$$\log\left(\frac{p_i}{1-p_i}\right) = w_0 + w_1x_1 + w_2x_2 + \cdots + w_px_p = \mathbf{w} \cdot \mathbf{x}^{\mathbf{T}}, \tag{1}$$

where $\left(\frac{p_i}{1-p_i}\right) \in (0;\infty)$ and $\log\left(\frac{p_i}{1-p_i}\right) \in (-\infty;\infty)$. So

$$p_i = \frac{e^{\mathbf{w}\cdot\mathbf{x}}}{1+e^{\mathbf{w}\cdot\mathbf{x}}} = \frac{1}{\frac{1+e^{\mathbf{w}\cdot\mathbf{x}}}{e^{\mathbf{w}\cdot\mathbf{x}}}} = \frac{1}{1+\frac{1}{e^{\mathbf{w}\cdot\mathbf{x}}}} = \frac{1}{1+e^{-(\mathbf{w}\cdot\mathbf{x})}}. \tag{2}$$

Due to nonlinear features of these models it is not possible to use ordinary least squares method for estimation of the parameters and instead of this we have to use the log likelihood function. Given P_i, we can form the logharithm of the likelihood function as

$$\ln L = \sum_{i=1}^{n} y_i \ln P_i + \sum_{i=1}^{n}(1-y_i)\ln(1-P_i). \tag{3}$$

For selection of the variables we can use the stepwise method and for testing of the models significance the log-likelihood ratio test and Wald test are usually used (see any econometric textbook).

3 Lévy processes

We can define a Lévy process $\{X(t)\}_{t\geq 0}$ as cádlág[2] real value stochastic process with $X(0) = 0$ which is stochastically continuous and has stationary independent increments. An important feature of Lévy processes is their intimate link to infinite divisible distributions (e.g., Sato [9]): If $\{X(t)\}_{t\geq 0}$ is a Lévy process, then every process value $X(t)$ is infinitely divisible. Conversely, to each infinitely divisible distribution there exists a unique in law Lévy process $\{X(t)\}_{t\geq 0}$ such that $X(t)$ has that distribution.

From the above it follows that a Lévy process $\{X(t)\}_{t\geq 0}$ has a unique so called characteristic exponent in form of a continuous function $\varphi: R \to R$ such that that the characteristic function of $X(t)$ is given by

$$E\{e^{iuX(t)}\} = e^{t\varphi(u)} \text{ for } u \in R \text{ and } t \geq 0,$$

where $\varphi(u)$ is the cumulant characteristic function $\varphi(u) = log\emptyset(u)$, which satisfies so called Lévy-Khintchine formula (see Schoutens [10]).

For a given infinitely divisible distribution, we can define the triplet of Lévy characteristics,

$$\{\gamma, \sigma^2, \nu(\mathrm{d}x)\}.$$

The former two define the drift of the process (deterministic part) and its diffusion. The latter is a Lévy measure. If it can be formulated as $\nu(\mathrm{d}x) = u(x)\mathrm{d}x$, it is a Lévy density. It is similar to the probability density, with the exception that it need not be integrable and zero at origin.

Let X be a Brownian motion. If we replace standard time t in Brownian motion X,

$$X(t;\mu,\sigma) = \mu dt + \sigma\mathcal{Z}(t), \tag{4}$$

by its suitable function $\ell(t)$ as follows:

$$X(\ell(t);\theta,\vartheta) = \theta\ell(t) + \vartheta\mathcal{Z}(\ell(t)) = \theta\ell(t) + \vartheta\varepsilon\sqrt{\ell(t)}, \tag{5}$$

[2] Right continuous with left limits.

we get a subordinated Lévy model. Due to the simplicity (tempered stable subordinators with known density function in the closed form), one of the most suitable candidates for the function $\ell(t)$ seem to be either the variance gamma model – the overall process is driven by a gamma process from the gamma distribution with shape a and scale b depending solely on variance κ, $G[a, b]$.

The final step to get a model for a marginal distribution depends on the issue we are going to solve. For example, if the task is to model the prices of a financial asset, ie. strictly positive value, we should evaluate the Lévy model (5) in the exponential part:

$$S(t) = S_0 e^{\mu t + X\ell(t) + \omega t}, \tag{6}$$

where μ states a long-term drift of the price (average return) and ω is the mean correcting parameter. By contrast, if we model a variable, which can be both positive and negative (eg. price returns), we can proceed as follows:

$$x(t) = \mu t + \square X(\ell(t)) - \theta t, \tag{7}$$

so that the long-term drift is fit again.

In order to estimate the parameters of marginal distributions, generalized method of moments will be used in this paper (see Schoutens [10]).

4 Copula function

A useful tool of dependency modeling is the copula function.[3] First we will define a n-variate copula C as the joint distribution function of n Uniform (0,1) random variables. If we label the n random variables as $U_1, U_2, \ldots, U_n$ then we can write down the copula C as

$$C(u_1, \ldots u_n) = Pr(U_1 \leq u_1, \ldots, U_n \leq u_n)$$

Now consider any continuous random variables $X_1, \ldots, X_n$ with distributions functions $F_1, \ldots, F_n$ respectively and joint distribution function F. Then (due to Sklar's theorem) for any joint distribution function F, there is a unique copula C that satisfies

$$F(x_1, \ldots x_n) = C\big(F_1(x_1), \ldots, F_n(x_n)\big), \tag{8}$$

Sklar's theorem proves that in examining multivariate distributions, we can separate the dependence structure from the marginal distributions. Conversely, we can construct a multivariate joint distribution from a set of the marginal distributions, and a selected copula. The dependence structure is captured in the copula function and is independent of the form of the marginal distributions.

Let R be a symmetric, positive definite matrix with $diag(R) = 1$ and let Φ_R be a standardized multivariate normal distribution with correlated matrix R. Then the multivariate Gaussian copula is defined as

$$C(u_1, \ldots u; R) = \Phi_R\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\big) \tag{9}$$

where $\Phi^{-1}(u)$ denotes the inverse of the normal cumulative distribution function.

5 The stability investigation of the portfolio of three Czech banks

In this part of the paper we will apply the theoretical assumptions presents above on the portfolio of the three Czech banks (ČSOB a. s., KB a. s., GE Money bank a. s.) in order to estimate their probability distribution of the PD. First we will derive the logit model from empirical data. Afterwards we will simulate the developed financial indicators by means of Variance gamma process. Dependences among the financial indicators will be captured by means of the multivariate Gaussian copula.

[3] In this paper, we restricted ourselves to ordinary copula functions. Basic reference for the theory of copula functions is Nelsen [7], while Rank [8] and Cherubini et al. [3] target mainly on the application issues in finance.

5.1 Data description

In this paper we will work with sample of 298 financial institutions. The defaulted banks are thought the financial institutions which have gone into liquidation or undergone financial restructuring processes (e.g. take-over by another company or by government). The samples of the financial institutions were chosen randomly pursuant to the publicly available information.[4] As a second step we identified the financial indicators from financial statements, see Karminsky and Peresetsky [6]. Table 1 shows the chosen financial indicators for both groups of companies including their mean values. According to the theoretical assumptions presented above we add to independent variable also macroeconomics variable, particularly the year-on-year growth rate of the GDP (we state as GDP in rest of the text), long-term interest rate and unemployment rate. We can observe the evolution of these macro indicators in Appendix A.

Non-default banks				Default banks			
fin. indicator	*mean value*	*fin. indicator*	*mean value*	*fin. indicator*	*mean value*	*fin. indicator*	*mean value*
x_1: LTA	15,8	***x_9: PE OI***	23,21%	***x_1: LTA***	11,91	***x_9: PE OI***	26,62%
x_2: YAEA	5,81%	***x_{10}: PL GL***	3,71%	***x_2: YAEA***	6,65%	***x_{10}: PL GL***	15,15%
x_3: CIBL	3,30%	***x_{11}: LLR GL***	1,96%	***x_3: CIBL***	3,69%	***x_{11}: LLR GL***	3,24%
x_4: NIM	3,56%	***x_{12}: PL EQ LLR***	27,42%	***x_4: NIM***	3,15%	***x_{12}: PL EQ LLR***	39,09%
x_5: ROAA	1,14%	***x_{13}: T1***	10,70%	***x_5: ROAA***	-4,31%	***x_{13}: T1***	7,84%
x_6: ROAE	7,62%	***x_{14}: EQ TA***	10,92%	***x_6: ROAE***	-47,53%	***x_{14}: EQ TA***	5,58%
x_7: IE II	37,87%	***x_{15}: CAR***	12,60%	***x_7: IE II***	55,43%	***x_{15}: CAR***	8,35%
x_8: CIR	86,81%	***x_{16}: D EQ***	7,86	***x_8: CIR***	115,54%	***x_{16}: D EQ***	15,32

Table 1 The mean values of financial indicators for good and bad banks

5.2 The logit model estimation

By means of the methodology from the theoretical part we have estimated following model:

```
Logistic regression                               Number of obs   =        298
                                                  LR chi2(4)      =     340.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -35.190074                       Pseudo R2       =     0.8288

------------------------------------------------------------------------------
      Status |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     x17_GDP |  -154.0935   39.94292    -3.86   0.000    -232.3802   -75.80681
     x5_ROAA |   -66.6987   18.24757    -3.66   0.000    -102.4633   -30.93411
   x10_PL_GL |   42.02622   9.718514     4.32   0.000     22.97828    61.07416
     x2_YAEA |   46.32679   22.38053     2.07   0.038     2.461753    90.19183
       _cons |  -4.485473    1.69872    -2.64   0.008    -7.814904   -1.156043
------------------------------------------------------------------------------
Note: 0 failures and 18 successes completely determined.
```

We can see that only four significant indicators remained in the model and that the model is significant (see log-likelihood ratio test). Also we have following model

$$\widehat{y_i} = PD_i = \frac{1}{1 + e^{-(-4{,}5 - 154GDP - 67ROAA + 42PLGL + 46YAEA)}}$$

where GDP denotes year-on-year growth rate of the Gross domestic product, ROAA states Return on average assets, PL GL is Problem loans on gross loans and YAEA is Yield on average interest earning assets.

5.3 Application of the model and determination of distributions of PDs

We now use the model defined in previous section to determine the distributions of the PDs of three Czech banks, ČSOB a. s., KB a. s. a GE Money Bank a. s., whereas these distributions we will predicate on one quarter. First we have to choose the random process and then to model the particular financial indicators. How it was said in the introduction we chose the VG model because they enable to model also the higher moments of the probability distribution, namely skewness and kurtosis. For estimation of the parameters we have to first know the empirical time series of the particular financial indicators. From financial statements it was calculated ten-year (1998 – 2009) time series with quarter indicators.[5] It is generally recommended to use maximal likelihood approach for estimation of the model parameters. However, since our data set consists of only few values, we will follow the generalized method of moments. In order to get the future distribution of particular financial indicators x_i, we will assume n = 300 000 independent scenarios for variance gamma distribution,

[4] www.federalreserve.gov, www.failedbankreporter.com

[5] The basic descriptive statistics are depicted in Appendix B.

$x_i \in VG(\nu; \theta, \vartheta)$.[6] From empirical correlation matrix, see Appendix C, it is obvious that there are some dependencies among the indicators. It is necessary to capture these dependencies. To do that, according to section 4, it was used the multivariate Gaussian copula. Applying this copula on the particular marginal distributions we get the quarter progress of the indicators with required dependencies. Afterwards by applying the simulated indicators into the model we easily get the probability distributions of PDs of analyzed banks. Figure 1 and Table 2 show the results.

čsob		KB		GE Money bank	
E(PD)	10,53%	E(PD)	18,17%	E(PD)	13,09%
stdev(PD)	0,175116	stdev(PD)	0,276025	stdev(PD)	0,196736

Table 2 The estimation of the future mean values and st. deviation of PDs

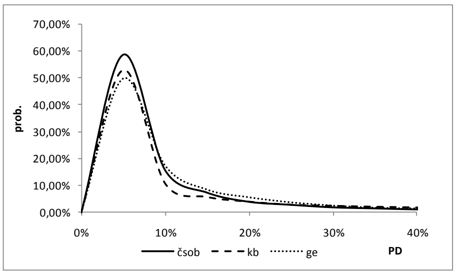


Figure 1 Estimation of future PD for ČSOB, KB and GE

From results it is obvious that risk of default of the analyzed banks is on relatively low level and all these three banks show a financial safety. It's hard to say what threshold of PD could cause the problems for the future, but from these results we can conclude that the financial market in the Czech Republic is more stable and better regulated compared to American one..

5.4 Stress testing

In this part the stress test on the impact of the extreme GDP downs on the analyzed banks stability is performed. The results can be observed in the following figure.

ΔGDP	0%	-25%	-50%	-100%	-150%	ΔGDP	0%	-25%	-50%	-100%	-150%	ΔGDP	0%	-25%	-50%	-100%	-150%
GDP	0,007	0,00525	0,0035	0	-0,0035	GDP	0,007	0,00525	0,0035	0	-0,0035	GDP	0,007	0,00525	0,0035	0	-0,0035
E (PD)	7,29%	9,04%	11,14%	16,51%	23,61%	E (PD)	16,62%	18,74%	21,04%	26,19%	32,03%	E (PD)	9,99%	12,21%	14,80%	21,16%	29,12%
stdev (PD)	0,091012	0,104748	0,119421	0,15025	0,179877	stdev (PD)	0,251824	0,26519	0,278118	0,301827	0,321309	stdev (PD)	0,119181	0,135036	0,151354	0,183607	0,211753

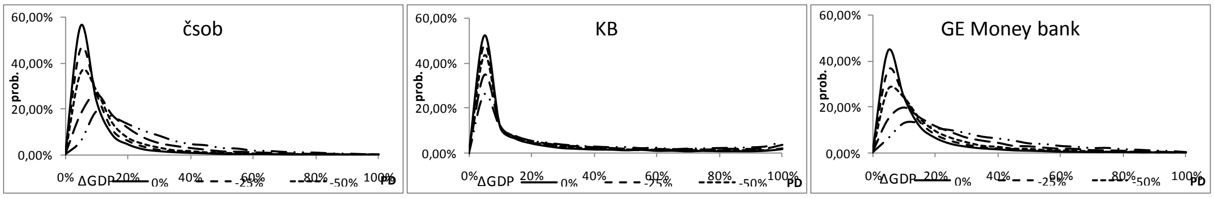


Figure 2 Impact of the GDP fluctuation on the PDs distribution

From results we can say that analyzed Czech banks are able to survive relative big recession. But, although all banks appear to be relatively healthy, there is still chance that "a financial crisis" will occur, at least in terms of probability.

6 Conclusion

The main purpose of the estimation of probability of default consists in its usage in the risk management, valuation of the credit derivatives, estimation of the creditworthiness of the borrowers and estimations of the banks' capital adequacy. Incorrect estimation of the PD can lead to the false valuation of the risk and consequently to the financial problems of the particular company. In this paper we have tried to estimate the distribution of a next stage PD of three Czech banks when the PD is supposed to be determined by the evolution of financial indicators through estimated logit model.

To estimate the probability distributions of the PDs of these banks we modeled particular financial indicators by means of any random process. The VG were chosen for simulation these indicators. Advantage of this

[6] We can see estimated parameters in Appendix B.

approach is the possibility to model the higher moments of the probability distribution. Next necessary step was to capture the dependencies among the particular financial indicators. The multivariate Gaussian copula was used to do that. From graphical results we can say that all of the analyzed banks are relatively financially stable.

At the conclusion the stress test on the impact of the extreme GDP downs on the analyzed banks stability was performed. It is obvious that analyzed banks are financial healthy with ability to survive also bigger negative movement of economy.

Acknowledgements

The paper is based on research activities sponsored through SGS research project No. SP2011/166. The support is greatly acknowledged.

References

[1] Altman, E.: *The Z-Metrics Methodology For Estimating Company Credit Ratings And Default Risk Probabilities*. RiskMetrics Group, 2010.
[2] Altman, E.: Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, September, 1968, 589-609.
[3] Cherubini, G., Luciano, E., Vecchiato, W.: *Copula Methods in Finance*. Wiley, 2004.
[4] Cont, R., Tankov, P.: *Financial Modelling with Jump Processes*. 2nd Printing, Chapman & Hall/CRC Press, London, 2004.
[5] Gurný, P., Gurný, M.: Comparison of the credit scoring models on PD estimation of US banks. *Mathematical Methods in Economics*. University of South Bohemia in České Budějovice, 2010.
[6] Karminsky, A., Peresetsky, A. Models for Moody's bank rating, *BOFIT Discussion Papers 17*, 2007.
[7] Nelsen, R. B.: *An Introduction to Copulas*. 2nd ed. Springer, 2006.
[8] Rank, J.: *Copulas. From theory to application in finance.* Risk books, 2007.
[9] Sato, K.: *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, 1999.
[10] Schoutens, W.: *Lévy Processes in Finance: Pricing Financial Derivates*. Wiley, 2003.

Appendix A

	growth of GDP	*interest rate*	*unemployment rate*
I.08	2,5%	3,66%	5,0%
II.08	2,1%	3,89%	5,3%
III.08	0,0%	3,86%	6,0%
IV.08	-1,9%	3,25%	6,9%
I.09	-3,3%	2,74%	8,2%
II.09	-3,8%	3,31%	9,3%
III.09	-2,7%	3,52%	9,7%
IV.09	0,2%	3,46%	10,0%
I.10	2,4%	3,72%	9,7%
II.10	3,0%	3,49%	9,6%
III.10	3,2%	2,79%	9,6%
IV.10	2,8%	2,86%	9,6%

Source: Eurostat, OECD Key Economic Indicators (KEI) Database

Appendix B

		YAEA	ROAA	PL GL	GDP
ČSOB	mean	0,0608	0,0110	0,0230	0,0070
	variance	0,0004	0,0000	0,0001	0,0001
	smodch	0,0209	0,0044	0,0097	0,0092
	skewness	0,7863	0,6608	1,5164	-1,8156
	kurtosis	3,9945	3,2969	6,4973	10,7607
KB	mean	0,0569	0,0197	0,0558	
	variance	0,0006	0,0002	0,0006	
	smodch	0,0240	0,0143	0,0246	
	skewness	1,2327	0,0880	0,4051	
	kurtosis	5,4396	3,8364	4,2010	
GE BANK	mean	0,0480	0,0308	0,0639	
	variance	0,0002	0,0002	0,0005	
	smodch	0,0139	0,0129	0,0221	
	skewness	1,2968	0,4792	0,4736	
	kurtosis	6,2185	3,4358	3,6464	

Basic characteristics of the probability distribution

		YAEA	ROAA	PL GL	GDP
ČSOB	Θ	0,0319	0,0085	0,0111	-0,00331
	σ	0,0154	0,0038	0,0045	0,00804
	v	0,1955	0,0689	0,5973	1,82573
KB	Θ	0,0274	0,0015	0,0181	
	σ	0,0157	0,0143	0,0457	
	v	0,4466	0,2771	0,3623	
GE BANK	Θ	0,0098	0,0259	0,0226	
	σ	0,0112	0,0104	0,0202	
	v	0,6826	0,0913	0,1629	

Estimated parameters of VG process

Appendix C

		čsob			KB			GE Money bank			
		ROAA	PL GL	YAEA	ROAA	PL GL	YAEA	ROAA	PL GL	YAEA	GDP
čsob	ROAA	1									
	PL GL	0,269831	1								
	YAEA	0,305974	0,530397	1							
KB	ROAA	-0,19264	-0,17073	-0,43499	1						
	PL GL	-0,08968	-0,06088	0,023422	-0,19202	1					
	YAEA	0,149201	0,464849	0,486332	-0,36433	0,156724	1				
GE Mone	ROAA	-0,22718	-0,03515	-0,27826	0,340256	-0,27858	-0,19421	1			
	PL GL	-0,26626	-0,20485	-0,50191	0,552744	-0,17992	-0,4759	0,420786	1		
	YAEA	0,096687	0,115396	0,023808	0,132243	0,11959	0,468029	-0,05606	0,085365	1	
	GDP	-2,4E-16	4,74E-16	-8,6E-16	1,62E-16	-1,9E-16	-7,9E-16	-2,1E-16	7,3E-17	-1,4E-16	1

Pearson correlation of financial indicators over 1998-2009

Comparing the core model of the Czech macroeconomy with benchmark univariate models

Jana Hanclova[1]

Abstract. The purpose of this paper is to compare the modelling strategy including a structural cointegrating *VAR* model with the in-sample fit of the individual equations in the core long-run structural model for the Czech macroeconomy using data over the period 1999Q1 – 2010Q4. We compare *ECM* specifications with a set of benchmark univariate time series representations using *ARMA(p,q)* specifications applied to the first differences of each of the nine core endogenous variables in turn (six domestic variables and three foreign variables).
The basic macroeconomic framework is a core small open economy model consisting of five long-run relationships based on production technology and output determination, arbitrage conditions, long-run solvency requirements and accounting identities and stock-flow relations. The first requirement in the construction of the benchmark model is the selection of an *a priori* maximum lag order for the autoregressive and moving average processes. We examine 25 different combinations and select our preferred benchmark model on the basis of the Akaike information criterion and the Schwarz Bayesian criterion. We compare the results for estimation and selection of univariate *ARMA* models using the AIC, SBC and adjusted R^2 statistics for each of the nine endogenous variables. We evaluate the estimated *CVAR* model and the preferred benchmark *ARMA* model at the end of this paper.

Keywords: *ECM* specifications, *ARMA* model, Czech macroeconomy, small open economy.

JEL Classification: C52, B22
AMS Classification: 62H12

1 Introduction

The aim of this presented paper is to compare the modelling possibilities of the elementary macroeconomic quantities using univariate *autoregressive moving average (ARMA) models* with the modelling of long-run equilibrium structural relations – using *the vector error correction model (VECM)* with restrictions. The preferred *ARMA* model is selected on the basis of the AIC and SBC information criteria minimalisation and diagnostic checking. The results are also relevant to describe the investigated models from the point of view of the adjusted coefficient of determination.

In section 2 we briefly summarize a framework of the macroeconometric modelling. The third section is devoted to the definition and process of estimation of two basic model groups – *ARMA(p,q)* models and cointegrating *VEC* models. The further one, fourth section deals with the empirical results. Based on the data analysis *ARMA* models and *VECM* are estimated for the small verified economy of the Czech Republic within the European Union environment in 1999Q1- 2010Q4. The final part compares these two groups of models from the point of view of their relevance but also from the point of view of the estimated coefficients and residual element in the equations of macroeconomic quantities.

2 Macroeconometric modelling

At present, there are two elementary methodological approaches to the modern macroeconometric modelling of economies – *Dynamic Stochastic General Equilibrium models (DSGE)* and *Cointegrated Vector Autoregression models (CVAR)*.

DSGE models follow and develop two significant waves of thinking - models of Real Business Cycle (RBC) and new Keynesian models (see [8]). *DSGE* models continue in quantitative economics with a strong theoretical

[1] VSB-Technical University of Ostrava, Faculty of Economics, Department of Mathematical Methods in Economics, Sokolska tr. 33, 701 21 Ostrava, Czech republic, jana.hanclova@vsb.cz.

background despite the effort to enrich them by standard statistical testing processes. These models are used by creators of macroeconomic policy mainly as a supportive tool for the analysis of the economic policy impacts on the economy in question.

In the structural models of *Cowles commission* the economic theory was used to identify the systems of equations and for the regression estimation the theoretical (a priori) limitations were applied. Jakon Marschak made widening on so called *probability approach* which evaluates the known data based on pre-defined stochastic model. In the first half of sixties in 20th century the methodology of London School of Economics (LSE) so called *British approach* was defined. This approach defined itself on contrary to the economics as logically-deductive science and demanded empirical testing of the economic theories with regard to the principal disagreement between equilibrium theory and mostly non-equilibrium data of a real word.

The other option of macroeconometric modelling is a use of vector autoregression (*VAR*) and mainly structural *VAR*. Models *CVAR* follow up modelling of a complex dynamic system the features of which respect real data (see [11] and [12]). In the article [9] King and others dealt with the cointegrating relations between consumption, investment, output and nominal interest rates and real monetary balance (see also [13] and [14]). Restrictions were the implications of RBC, Fisher equation theory and monetary demand equation. Anthony Garratt and others focus on structural *CVAR* model for small open economy of the U.K. in their book called *Global and National Macroeconometric Modelling* [6]. The model includes five long-run equilibrium relations derived from the economic theory and thus having a structural feature. When deriving these relations the authors took into account neoclassical production function, arbitrage conditions, solvency conditions and conditions of portfolio equilibrium. In the model following long-term relations occur: purchasing power parity in a relative version (*PPP*), parity of domestic and foreign interest rates representing Fischer´s effect (*IRP*), a gap of domestic and foreign product (*OG*). A long-term requirement on solvency and equilibrium of portfolio of actives is reflected in monetary demand (*MD*) and Fischer´s inflation parity represents Fischer´s domestic effect (*FIP*).

The basic macroeconomic core model includes six domestic variables and three foreign variables (Eurostat [4]). Domestic variables comprise: P_t - the domestic producer prices (manufacturing), PR_t - harmonised consumer prices, E_t - bilateral exchange rate CZK/EUR, Y_t - real per capita GDP (EUR), R_t - nominal interest rate as interbank 1Y middle rate, M_t - real per capita domestic monetary aggregate *M2* . Foreign variables for EU25 involve: YS_t - real per capita GDP, RS_t - 1Y interest rate as Euribor 1Y- offered rate, PS_t - producer prices.

Currently, we can express a core model in a form of five following equations

$$PPP: \quad p_t - ps_t - e_t = b_{10} + b_{11}t + \xi_{1,t+1} \tag{1}$$

$$MD: \quad my_t = b_{20} + b_{21}t + \beta_{22} \cdot r_t + \beta_{24} \cdot y_t + \xi_{2,t+1} \tag{2}$$

$$OG: \quad y_t - ys_t = b_{30} + b_{31}t + \xi_{3,t+1} \tag{3}$$

$$IRP: \quad r_t - rs_t = b_{40} + b_{41}t + \xi_{4,t+1} \tag{4}$$

$$FIP: \quad r_t - dpr_t = b_{50} + b_{51} + \xi_{5,t+1} \tag{5}$$

where $p_t = \ln(P_t), ps_t = \ln(PS_t), pr_t = \ln(PR_t), dpr = \Delta pr_t, e_t = \ln(E_t), y_t = \ln(Y_t), ys_t = \ln(YS_t), my_t = \ln(M_t / Y_t),$

$r_t = \ln(1+(R_t /100)), rs_t = \ln(1+(RS_t /100)), pps_t = ps_t - p_t .$

The long-run structural modelling strategy we follow was suggested by Garratt, Lee, Pesaran and Shin [6] and they applied it to the U.K. Other modifications were made for Germany [15] by Schneider, Chen and Frohn and also, for Switzerland [2] by Assenmacher-Wesche and Pesaran.

3 Econometric methods

In a following part of the paper we focus on a brief definition of two econometric approaches – estimation of individual dynamic equations of stationary time series using *ARMA* models and estimation using *CVAR* models.

3.1 Mixed Autoregressive Moving Average processes

An *ARMA* (*Autoregressive Moving Average*) model is a standard building block of univariate time series econometrics (see Hamilton [7]). In this section we shortly describe how to estimate the time series data generating process with *ARMA* model (see [1] or [10]). We will follow the Box-Jenkins methodology. *ARMA(p,q)* is process that includes both autoregressive and moving average terms of the orders *p* and *q* and it is described as

$$(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p)z_t=c+(1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q)\varepsilon_t. \tag{6}$$

In order the *ARMA(p,q)* process would be stationary equation roots $(1-\phi_1 L-\phi_2 L^2-\cdots-\phi_p L^p)=0$ have to lie outside a unit circle. In order the *ARMA(p,q)* process would be invertible equation roots $(1+\theta_1 L+\theta_2 L^2+\cdots+\theta_q L^q)=0$ have to lie outside a unit circle.

ARMA models are non-linear due to the *MA* terms and must be estimated using non-linear least squares. Non-linear least squares provide consistent and asymptotically normal coefficient estimates.

We can use *information criteria* to select the best *ARMA(p,q)* model and also the most common measure of goodness of fit is adjusted $R^2\,(R^2_{ADJ})$. We will use two the most frequently information criteria- *the Akaike information criterion (AIC)* and *the Schwarz Bayes information criterion (SBC).* To select the best *ARMA* model the value of the information criteria is to be minimized. The AIC is biased towards choosing an overparametrized model. The SBC is at least asymptotically consistent.

3.2 Cointegrating Vector Autoregression and Vector Error Correction Models

The estimation of the core long-run model involves an unrestricted *VAR* model. The *VAR* system can be written

$$\boldsymbol{A}(L)\boldsymbol{z}_t=\boldsymbol{u}_t, \tag{7}$$

where the matrix polynomial in the lag operator, *L*, *A(L)* has degree *k* and leading matrix equal to the identity matrix, reflecting the reduced-form nature of the system. The elements of u_t are correlated, that is $E(u_t u_t^{'})=\Omega$ is not diagonal, and Sims argued that is useful to transform them to orthogonal form to see the distinct patterns of movement of the system. The *VAR* system (8) can be rearranged as:

$$\boldsymbol{A}^*(L)\Delta\boldsymbol{z}_t=-\boldsymbol{\Pi}\boldsymbol{z}_{t-1}+\boldsymbol{u}_t, \tag{8}$$

where $\boldsymbol{\Pi}=\boldsymbol{A}(1)$ and the degree of $\boldsymbol{A}^*(L)=k-1$. If the elements of $\boldsymbol{z}_t$ are I(1) and cointegrated with rank $(\boldsymbol{\Pi})=r,\ \ 0<r<n,$ then $\boldsymbol{\Pi}=\boldsymbol{\alpha\beta}'$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are *(n x r)* matrices of rank *r*, giving the *VECM* representation

$$\boldsymbol{A}^*(L)\Delta\boldsymbol{z}_t=-\boldsymbol{\alpha\beta}'\boldsymbol{z}_{t-1}+\boldsymbol{u}_t, \tag{9}$$

Exact *r* identification of $\boldsymbol{\beta}$ requires *r* restrictions on each of the *r* cointegrating vectors (columns of $\boldsymbol{\beta}$), of which one is a normalization restriction and the *(r-1)* restrictions satisfy the identification rank condition. In the differenced (stationary) variables $\Delta\boldsymbol{z}_t=\boldsymbol{D}(L)\boldsymbol{u}_t$ the matrix ***D(1)*** of long-run multipliers. Reduced form *VECM(p)* can be written as

$$\Delta\boldsymbol{z}_t=\boldsymbol{\alpha\beta}'\boldsymbol{z}_{t-1}+\boldsymbol{d}_t\boldsymbol{\mu}+\boldsymbol{\Gamma}\left(L\right)\Delta\boldsymbol{z}_{t-1}+\boldsymbol{u}_t, \tag{10}$$

where $\boldsymbol{\Gamma}\left(L\right)$ is a matrix polynomial of degree *p* in the lag operator and $\boldsymbol{d}_t$ is the deterministic term. Estimation of the parameters of the model (11) was carried out using the Johansen's method (see Kočenda and Černý [10]).

4 Comparing the core model with benchmark univariate models

In this part of the paper we focus on data analysis, estimation and diagnostic testing of *VECM* and *ARMA* models using Eviews [5]. In a conclusion we compare the *VECM* specification with a set of univariate time series representations *ARMA(p,q)*.

The VECM for our Czech economy is described in equation (10), where $\boldsymbol{z}_t=(e_t, r_t, y_t, dpr_t, y_t, pps_t, my_t, rs_t, ys_t)'$. The data used in the applied study are quarterly, seasonally adjusted series covering the period 1999Q1-2010Q4 for the Czech economy.

In the first part of the data analysis we pre-test the variables for unit roots using ADF, PP and KPSS tests. We can conclude that all the endogenous variables are I(1) at the 5% level of significance.

4.1 VECM

The next step for the *CVAR* estimation is to select the order of the underlying unrestricted *VAR* in eight endogenous variables. We selected *VAR(2)* with regard on Wald test and also the LR test with SBC criterion and also by reason of the length of the time series and also for comparison of our results with other studies and we prefer over-estimation rather than under-estimation and four lags are too many for reason of loss of degrees of freedom. Next, having established the appropriate order of the VAR model, Johansen's cointegration tests are carried out using the trace and the maximum eigenvalue statistics. According to unit root tests and also graphs we assume an intercept with trend in the cointegrating relationship and a linear data trend. In this case we can indicate five cointegrating equations at the 5% level of significance, especially by the trace statistics.

The estimated long-run relationships for the Czech economy, incorporating 38 restrictions suggested by the theory and with *t-statistics* in [], are:

$$PPP: p_t - ps_t - e_t = -4.92 + \underset{[18.3]}{0.007}\, t + \xi_{1,t+1} \tag{11}$$

$$MD: my_t = 22.4 - \underset{[-6.56]}{0.023} t - \underset{[-19.7]}{7.02} \cdot r_t + \underset{[28.7]}{3.29} \cdot y_t + \xi_{2,t+1} \tag{12}$$

$$OG: y_t - ys_t = -1.31 + \underset{[8.5]}{0.0048} \cdot t + \xi_{3,t+1} \tag{13}$$

$$IRP: r_t - rs_t = 0.0043 + \underset{[0.021]}{0.000007} \cdot t + \xi_{4,t+1} \tag{14}$$

$$FIP: r_t - dpr_t = 0.064 - \underset{[-5.13]}{0.00085} \cdot t + \xi_{5,t+1} \tag{15}$$

Equation Equation (11) describes *PPP* relations and does not reject the context for the core model. The convergence of the Czech economy becomes evident by the average quarterly decrease in the real exchange rate by about 0.007%. Money market equilibrium is presented by equation (12). The quarterly long-run rate of decrease in money stock per capita and per capita Czech output is about 0.023% per quarter. The third long-run output relationship, given by (13), describes the average long-run growth rate in the Czech and EU25 economies. We estimate the increase of the output gap at about 0.005% per quarter. Next, equation (14) includes the interest rate parity condition. This includes the intercept, which can be interpreted as the deterministic component of the risk premium associated with bond and foreign exchange uncertainties. Its value is estimated at 0.0043 implying a risk premium of approximately 0.43% . Finally, the fifth equation (15) defines the FIP relationship, where the constant implies the average long-run Czech real interest rate is about 6.6%.

The results also confirm satisfactory diagnostic statistics of the estimated *VECM(1).* The assumption of normally distributed errors is tested by the chi-square distributed Jarque-Bera statistics. This assumption is rejected only in the equation of the EU25 output with respect to skewness. The LM tests are employed as a check of the residual serial correlation. We cannot reject the hypothesis of no serial autocorrelation of the first order in all equations at the 5% level of significance. The presence of autoregressive conditional heteroscedasticity is rejected in all equations. We can also detect a relatively high level of explanatory power for each evaluation with R^2_{ADJ}.

4.2 ARMA model

Further on we estimate *ARMA(p,q)* specifications applied to the first differences of each of $e_t, r_t,\ dpr_t, y_t, pps_t, my_t, rs_t, ys_t$ variables according to the equation (6). These benchmark models follow the Box-Jenkins methodology.

The first requirement in the construction of the benchmark model is selection of an *a priori* maximum lag order for the autoregressive and moving average processes (*p, q).* We choose 4, in light of the quarterly nature of the data, the number of available observation (48 observations for the sample period 1999Q4- 2010Q4) and considering that the degree of serial correlation in the first difference of the macrovariables is not very high. We examine the full set of *ARMA (p,q)* combinations that are spanned by all *p=0,1,…,4* and *q=0,1,…,4*, providing 25 different combinations. Our preferred benchmark models are then selected on the AIC and SBC criteria. AIC is a more appropriate selection criterion if the aim is to select the best approximating model in the information-theoretical sense. The theoretical ground for the use of SBC is in the case of models involving unit roots and cointegration that has not been fully developed.

The results of the estimation and selection of the univariate *ARMA* models are summarised in Table 1 providing also details of the AIC, SBC and adjusted R^2 *statistics* calculated for different models estimated for each the first difference of our eight endogenous macrovariables. The first column of this table is related to the unrestricted *ARMA(4,4)* specification for each variables, the second column shows our preferred benchmark *ARMA* model chosen by AIC, the third column is related to selected *ARMA* model by SBC. Residuals for a correctly specified *ARMA* model that captures the data generating process well should be white noise. We also provide residual diagnostics for the *ARMA* models chosen by AIC and SBC. We tested for serial correlation, normality, heteroscedasticity. Serial correlation Lagrange multiplier (LM) tests belong to the class of asymptotic tests and the null hypothesis is that there is no serial correlation up to lag order *p*. We used the *F*-statistics and also $Obs*R^2$ statistics. The serial correlation results for each selected model in the second and third column of Table 1 strongly reject the null of no serial correlation. We also employ White's heteroskedasticity test under the null hypothesis of homoscedasticity in the residuals of each equation. EViews reports two test statistics – the *F*-statistics, $Obs*R^2$ statistics. All two statistics reject the null hypothesis of homoscedasticity only for equation of difference of foreign output in model *ARMA(1,0).* The normality tests of residuals include the Jarque-Bera statistics. We reject normal distribution at 5% level only in *ARMA(4, 0)* model of differentiated foreign interest rate (*drs*) and also in v *ARMA(1, 0)* model of differentiated foreign output (*dys*). The main reason was a negative skewness and high kurtosis. All selected *ARMA* processes according to the AIC a SBC criteria fulfilled conditions of stability and invertibility. ARMA models in equations of variables *de* a *dmy* presented low R^2_{ADJ}.

		ARMA(p, q) model			*VECM*			*ARMA(p, q)* model			*VECM*
Variable		**(4,4)**	**AIC**	**SBC**	***(p)***	***Variable***		**(4,4)**	**AIC**	**SBC**	***(p)***
de	***AIC***	-4.796	-4.772	-4.772	**-5.447**	***dpps***	***AIC***	-7.590	**-7.454**	-7.289	-7.276
	SBC	-4.431	-4.693	-4.693	**-4.901**		***SBC***	-7.225	**-7.211**	-7.171	-6.731
	R^2_{ADJ}	0.265	0.077	0.077	**0.667**		R^2_{ADJ}	0.420	**0.298**	0.077 !!	0.205
	$(\hat{p},\hat{q})$	(4, 4)	(1, 0)	(1, 0)	x		$(\hat{p},\hat{q})$	(4, 4)	(4, 1)	(1, 1)	x
dr	***AIC***	-8.282	-8.577	-8.577	**-9.159**	***dmy***	***AIC***	-5.334	**-5.489**	-5.442	-5.301
	SBC	-7.917	-8.298	-8.298	**-8.613**		***SBC***	-4.969	**-5.291**	-5.403	-4.845
	R^2_{ADJ}	0.143	0.329	0.329	**0.767**		R^2_{ADJ}	0.109	0.139	0 !!!	**0.154**
	$(\hat{p},\hat{q})$	(4, 4)	(2, 4)	(0, 4)	x		$(\hat{p},\hat{q})$	(4, 4)	(2, 2)	(0, 0)	x
ddpr	***AIC***	-7.324	**-7,529**	**-7.529**	-7.395	***drs***	***AIC***	-8.923	-8.331	-8.331	**-8.610**
	SBC	-6.958	**-7.290**	**-7.290**	-6.849		***SBC***	-8.558	**-8.128**	**-8.128**	-8.065
	R^2_{ADJ}	0.333	**0.401**	**0.401**	0.381		R^2_{ADJ}	0.686	0.388	0.388	**0.589**
	$(\hat{p},\hat{q})$	(4, 4)	(2, 3)	(2, 3)	x		$(\hat{p},\hat{q})$	(4, 4)	(4, 0)	(4. 0)	x
dy	***AIC***	-6.645	-6.877	-6.823	**-9.159**	***dys***	***AIC***	-7.930	-7.940	-7.940	**-8.281**
	SBC	-6.280	-6.636	-6.744	**-8.613**		***SBC***	-7.565	**-7.861**	**-7.861**	-7.735
	R^2_{ADJ}	0.307	0.406	0.300	**0.767**		R^2_{ADJ}	0.594	0.547	0.547	**0.733**
	$(\hat{p},\hat{q})$	(4, 4)	(3, 2)	(1, 0)	x		$(\hat{p},\hat{q})$	(4, 4)	(1, 0)	(1, 0)	x

Table 1 Estimation and selection of the univariate *ARMA* models and *VECM* estimation

4.3 Comparing the core model with benchmark univariate models

The last column of our Table 1 introduces the error correction estimation of the core model discussed in section 4.1. Comparisons across the *ARMA*_AIC, *ARMA*_SBC and *VECM* columns show that the error correction model outperform the preferred *ARMA(p,q) model* for 5/8 of variables in terms of AIC and for 3/8 of endogenous variables in terms of SBC. The selected benchmark *ARMA* model was preferred for the first difference of the variables – domestic inflation (*dpr*), difference between domestic and foreign producer prices (*dpps*) and also monetary demand (*dmy*) in case of AIC (i.e if the aim is selected the best approximating model in the information-theoretical sense). In case we would like to meet certain other regularity conditions (i.e. a consistent model), we used SBC criterion which prefers *ARMA* models also for variables of foreign output (*dys*) and foreign interest rates (*drs*). From the point of data consistency with the estimated model (i.e. R^2_{ADJ}) *VECM* for 5/8 of variables is totally preferred. For example this *VECM(1)* explains as much as 66.7% of the total variation in the first difference of exchange rate (*de*) and preferred benchmark *ARMA(0, 4)* model only 7.7%. The similar situation is for equation of domestic output. On contrary the benchmark *ARMA(2, 3)* model for domestic inflation differentiation (*ddpr)* resp. *ARMA(4, 1)* model for change in difference of domestic and foreign producer price (*dpps*) explains 40.1% resp. 20.5 % of total variation but the long-run structural error correction specification only 38.1% resp. 20.5%.

5 Conclusions

The conclusion to be drawn for these results is that the long-run structural error correction model does indeed perform well in comparison to univariate time series models chosen according to our AIC, mainly for changes of variables exchange rate (*de*), domestic and foreign interest rate (*dr, drs*), domestic and foreign output (*dy, dys*). In case of SBC a situation is opposite and in 5/8 selected *ARMA(p,q)* model is prefered. It concerns equation of differentiation of domestic inflation (*ddpr)*, difference of domestic and foreign producer prices (*ddps*), monetary demand (*dmy*), foreign interest rate (*drs*) and foreign output (*dys*). The SBC statistics places greater weight on parsimony in model selection and this is reflected by the fact that relatively simple models are chosen in the third column of Table 1. Also relevance of the estimated R^2_{ADJ} *'s* is in favour of the error correction model in 6/8 of equations. This study was completed by the evaluation through *ARMA* and *VEC* models. These results are not published due to the publication limits of this paper.

Acknowledgements

This research work was supported by the grant no. 402/08/1015 (Macroeconomic Models of the Czech Economy and Economies of the other EU Countries) of the Czech Science Foundation.

References

[1] Arlt, J. and Arltlová, M. *Ekonomické časové řady.* Grada Publishing, Praha, 2007.
[2] Assenmacher-Weschke, K., Pesaran, M. H. A VECX Model of the Swiss Economy. *CESifo Working Paper,* No. 2 281. University of Cambridge, Cambridge.
[3] Chocholatá, M. Purchasing Power Parity and Cointegration:Evidence from Latvia and Slovakia. *Ekonomický časopis* **57**(2009), 344–358.
[4] Eurostat . *Eurostat database.* [Online], [cit. 15. 10. 2010]. Dostupnýz www: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database >.
[5] EViews. *EViews 7 User'Guide II.* Quantitative Micro Software, Irvine, 2009.
[6] Garratt, A., Lee, K., Pesaran, M. H. and Shin, Y. *Global and National Macroeconometric Modelling. A Long-Run Structural Approach.* Oxford University Press, Oxford, 2006.
[7] Hamilton, J. D. *Time series analysis.* Princeton University Press, Princeton, 1994.
[8] Hančlová, J. a kol. *Makroekonometrické modelování české ekonomiky a vybraných ekonomik EU.* EkF VŠB-TUO, Ostrava, 2010.
[9] King, R.G., Plosser, Ch. I., Stock, J. H., Watson, M. W. Stochastic Fluctuations in the United States 1921-1941. *American Economic review,* **81**(4),1991, 819-840.
[10] Kočenda, E. and Černý, A. *Elements of time series econometrics: an applied approach.* Karolinum Press, Charles University in Prague, 2007.
[11] Lukáčik, M. Makroekonomické modelovanie a VECM. In *AIESA - budovanie spoločnosti založenej na vedomostiach - zborník.* Vydavateľstvo EKONÓM, Bratislava, 2010.
[12] Lukáčik, M., Lukáčiková, A. and Szomolányi, K. The long-run determinants of money demand in Slovakia. In *Quantitative methods in economics: multiple criteria decision making XIII.* IURA, Bratislava, 2006.
[13] Mariel, P., Orbe, S. and Rodríguez, C. The knowledge-capital model of FDI: a time varying coefficients approach. *Scottish Journal of Political Economy*, **56** (2009), 196–212.
[14] Pánková, V.: FDI and economic growth – Central Europe. In: *Proceedings of 23rd International Conference - Mathematical Methods in Economics.* University of Hradec Králové, Hradec Králové, 2005, 255-258.
Schneider, E., Chen, P. and Frohn, J. A long-run structural macroeconometric model for Germany: an empirical note. *Economics:The Open-Access, Open-Assessment E-Journal,* 2 (2008).

Efficiency and Ownership of the Czech Firms

Jan Hanousek[1], Evžen Kočenda[2], Michal Mašika[3]

Abstract. In the present paper we analyze evolution of firm financial efficiency in the Czech Republic. Using a large panel consisting of more than 400 thousands Czech firm/years we study whether firms utilize fully their resources, how firm financial efficiency evolves over time, and how it is determined by an ownership structure. We employ a panel version of a stochastic production frontier model in the period 1996-2007 with time invariant efficiency. In that we differentiate among various degrees of ownership concentration and their domestic or foreign origin. In a two-stage set-up we estimate the degree of firms' inefficiency and then we estimate the effect of ownership structure on the distance from the efficiency frontier. Our results support the hypothesis of the concentration and foreign ownership being positively related to the financial efficiency.

Keywords: financial efficiency, ownership structure, panel data.

JEL Classification: G34
AMS Classification: 91B28

1 Introduction

The economic reforms of 1990s in Central and Eastern Europe (CEE) were aimed at creating a competitive market economies and increased efficiency of enterprises by means of firm restructuring, privatization and supporting institutional reforms. Still, as we show later in this section, there is a lack of reliable empirical evidence on medium and long-term firms' efficiency and its determinants in post-transition economies in the CEE region (Estrin et al., 2009). We fill the gap in the literature by analyzing financial efficiency of the Czech firms and how this efficiency is determined by ownership structures. We employ stochastic production frontier model and use a unique firm-level panel data of more than 400 thousand firm/year observations for period 1996-2007. Our results are in line with theoretical predictions of the concentration and foreign ownership being positively related to the financial efficiency (Blomström et al. 2001) and shed light on many other subtleties of how ownership affects firm's efficiency.

In this paper we advance the literature by addressing systematically the issues related to the efficiency effects of ownership and eliminate earlier shortcomings. In particular, we develop a more systematic analytical framework for evaluating the financial efficiency effect of the domestic versus foreign ownership, as well as the effect of various degrees of the ownership concentration.

2 Modeling Strategy

2.1 Theoretical Background

In our analysis we employ a stochastic production possibility frontier approach introduced by Aigner et al. (1977), Meeusen and van den Broeck (1977) and further adapted for panel data by Khumbhakar (1990) and Battese and Coelli (1995). The method measures the technical inefficiency under single output production. More important, the methodology helps to explain firm level differences in efficiency as a function of number of explanatory variable as opposed to estimating the average efficiency relative to "best practice" for number of sectors.

The methodology of the stochastic frontier develops in the following way. A firm has a production function

$$y_i = f(x_i;\beta)$$

[1] Charles University and Academy of Sciences CERGE-EI, Politickych veznu 7 P.O. Box 882 111 21 Prague Czech Republic, jan.hanousek @cerge-ei.cz.
[2] Charles University and Academy of Sciences CERGE-EI, Politickych veznu 7 P.O. Box 882 111 21 Prague Czech Republic, evzen.kocenda@cerge-ei.cz.
[3] Ludwig-Maximillians-Universität München, Department of Economics, Akademiestr. 1, 80799 Munich, Germany, michal.masika@lrz.uni-muenchen.de.

that defines a technological link between inputs (x) and resulting output (y) under an assumption that production is conducted in an efficient manner. Due to some degree of inefficiency a firm potentially produces less that it might and its production function is

$$y_i = f(x_i;\beta)\cdot TE_i$$

Firm's technical efficiency TE_i represents a ration of observed output to maximum feasible output and lies within an interval (0,1]; TE_i is considered to be nonnegative since firm's output is assumed to be positive. If $TE_i = 1$ then a firm employs all inputs efficiently and achieves an optimal output. If TE_i is smaller than one then a firm experiences a degree of inefficiency in its production. Further, two assumptions are made. One, efficiency is a stochastic variable with a distribution common to all firms and can be written as $TE_i = exp\ \{-u_{it}\}$; since $0 < TE_i \leq 1$, then $u_{it} \geq 0$. Two, firm's output is also subject to various random shocks that encompass anything from bad weather to unexpected luck and these effects are denoted as exp (v_{it}). Thus, the production function is further expanded to

$$y_{it} = f(x_{it};\beta)\cdot \exp(-u_{it})\cdot \exp(v_{it})$$

After taking the natural log of both sides we obtain

$$\ln y_{it} = \beta_0 + \sum_{j=1}^{k}\beta_{jit}\ln x_{it} + v_{it} - u_{it} \qquad (1)$$

In this general specification v_{it} is a pure noise component and two-sided normally distributed variable, while u_{it} is the nonnegative technical inefficiency component. Both terms form a compound error term with an a priori unknown distribution. The model is estimated by maximum likelihood assuming log-quadratic production function which encompasses the Cobb-Douglas specification and is a less restrictive one.

2.2 Empirical Approach – First Stage

The Cobb-Douglas function assumes that the input elasticities and returns of scale are constant, and that the elasticities of substitution are equal to one. From the empirical estimation perspective both assumptions are linked to the evidence that industries within one-digit NACE division differ with respect to capital-intensity, labor-intensity or technology-intensity (Laafia, 2002). Therefore, we follow the mainstream of the literature and consider interacting parameters of the Cobb-Douglas production function with 2-digit NACE industries. As a result, in the specification below we consider different parameters of the Cobb-Douglas function for each 2-digit NACE sector and this way we account for specifics of a given sector. Formally, our model of the financial efficiency frontier of I firms (i = 1,…,I) in J two-digit NACE sectors (j=1,…,J) over T time periods (t = 1,…,t) is specified as follows:

$$\ln y_{it} = \sum_{j=1\ldots J}\left[\beta_{0j} + \beta_{1j}\ln c_{it} + \beta_{2j}\ln l_{it}\right]\cdot ID_{itj} + v_{it} - u_{it} \qquad (2)$$

In the specification (2) ln y_{it} is the natural log of the value of production of firm i at time t, measured as firm turnover. Then, ln c_{it} is the natural log of the capital of each firm measured as working capital, and ln l_{it} is the natural log of the firm's labor, measured as the staff costs; β_0 is a common intercept for all firms. The working capital is the optimal proxy for the capital for our financial efficiency analysis. ID_{ijt} represents a vector of dummy variables to associate each firm with a specific industry sector j it operates in. By the construction of the model we interact the dummy variables for each of 45 two-digit NACE industries with both inputs (capital and labor) to control for industry specific effects. In addition, we divide these industry sectors into 6 basic groups based on different degrees of technology and knowledge intensity they represent (details available upon a request). The specification (2) is based on assumption that technical efficiency does not vary over time [we have consider also this option, but it was rejected by ML tests].

2.3 Empirical Approach – Second Stage

Ownership structures have been identified in numerous relevant studies as a key determinant of firm's performance (see Estrin et al., 2009 for a general overview and Hanousek et al., 2007a, 2009 for specific results related to the Czech firms). Therefore, in the second stage we model how is firm's efficiency (u_i) determined by its ownership structure: (u_i) = f (ownership structure); formal model is introduced later in this section. Specifically, we aim at answering the following questions formulated as hypotheses.

Hypothesis 1. *Majority owners reduce firm's inefficiency.*

Hypothesis 2. *Majority owners when being confronted with strong minority owners reduce firm's inefficiency more compared to uncontested majority owners.*

Hypothesis 3. *Minority and dispersed ownership reduce firm's inefficiency.*

Hypothesis 4. *Minority owners under presence of highly dispersed remaining ownership reduce firm's inefficiency.*

Hypothesis 5. *Foreign owners reduce firm's inefficiency.*

Hypothesis 6. *There exists a convergence in contribution to firm's efficiency between domestic and foreign owners.*

We aim to test the above hypotheses by employing a model that links firm efficiency with its ownership structure. The model for each year (period t) is specified as follows:

$$u_i = \alpha + \sum_{j=1}^{J} \delta_j OWN_i^j \qquad \text{for all } i = 1, \ldots N \tag{3}$$

Ownership structure (OWN_i^j) is defined for each firm i to distinguish specific ownership category j. We distinguish ownership of the private firms between domestic and foreign based on the exact knowledge of the owner's origin. In there is a missing information on the owner's domicile we introduce a special category of the "unknown" domicile. To this effect we consider the categories of domestic, foreign, and unknown domicile owners. From our data we can also distinguish an extent of the ownership concentration along with the extent of control over a firm. Following the country and legal-specific approach of Hanousek et al (2007a) we construct ownership categories to distinguish majority owner (stake above 50%), monitoring owner, minority owner (stake above 33%), and dispersed ownership. We elaborate more on the ownership categories in the data section.

3 Data

We develop a model to examine the impact of the ownership structure on the firm financial efficiency in the Czech Republic. We employ firm-level unbalanced panel data for the period 1996-2007 from the database Amadeus. Depending on a specific year we have firm-level balance sheet data (turnover, working capital, and staff costs) for 3,818 to 87,268 firms. As these are multiproduct firms we are unable to obtain exact information about quantities (input, output) connected with a production process of each of firm's product. For this reason we follow a standard approach in the literature and employ financial variables from the firms' balance sheets (see Coelli et al., 2005 for an overview). We further combine the balance-sheet data with ownership data obtained from the databases Amadeus, Aspekt, and Čekia. Altogether we work with a unique firm-level panel data of more than 400 thousand firm/year observations for the period 1996-2007. Detailed descriptive statistics are available upon a request.

In the first stage we derive firm efficiency based on a two-input (capital, labor) Cobb-Douglas production specification introduced in section 2. Given the space, we do not present the first stage results (available upon request). In order to capture different effects across sector-specific intensities we follow approach of Laafia (2002) who divided industries into different sectors based on their technology and knowledge intensity. This approach is based on the Eurostat official methodology assigning industries in manufacturing and services into s different groups to reflect different degrees of technology and knowledge they represent[4].

In the second stage we examine the impact of the ownership structure on estimated efficiency. We define the ownership variables to reflect different concentration thresholds based on the country's legal rules. Depending on their stakes, different blockholders have under the Czech law different opportunities to influence corporate governance. In particular, the law provides important rights of ownership and control to owners with *majority* ownership (more than 50 percent of shares), *monitored majority ownership* (there is a majority ownership and at the same time there exist also minority owner(s) with stakes higher than 10 per cent), *blocking minority ownership* (more than 33 percent but not more than 50 percent of shares), *controlling minority ownership* (an owner holds a stake in a firm that greater than 10 % and this stake is greater than the sum of the stakes of all remaining stakes that could be identified), *combined controlling minority ownership* (two owners whose combined stake

[4] In manufacturing industries we have four groups: high-technology; medium-high-technology; medium-low-technology; and low-technology. In services industries we have 5 groups: knowledge-intensive services (KIS); high-tech KIS; market KIS (that excludes financial intermediation and high-tech services); less knowledge-intensive services (LKIS); and market LKIS. Best results in terms of the highest turnover achieve medium-high and low technology firms in manufacturing sector, and LKIS and market LKIS firms in service sector.

exceeds 50%), a *legal minority* ownership (at least 10 but not more than 33 percent of shares), and finally *dispersed ownership* (if an owner holds a stake in a firm that is smaller than 10%). All ownership categories are exclusively defined and they are also distinguished for domestic and foreign owners, as well as those without know domicile.

4 Empirical Results

In Tables 1 and 2 we present our results that distinguish ownership effects depending on economic sectors firms operate in. We distinguish four manufacturing sectors based on technology levels (Table 1) and five service sectors based on knowledge intensity they represent (Table 2); as mentioned in section 3 this division strictly adheres to the methodology of the Eurostat.

Ownership category	Technology			
	High	**Medium-high**	**Medium-low**	**Low**
Majority foreign	0.14[a] (0.007)	0.017[a] (0.008)	0.031[a] (0.009)	0.022[a] (0.005)
Majority domestic	0.138[a] (0.007)	0.045[a] (0.007)	0.04[a] (0.006)	0.019[a] (0.005)
Majority unknown	0.20[b] (0.007)	0.094 (0.010)	0.097 (0.011)	0.046[a] (0.007)
Monitored majority foreign	0.142[a] (0.028)	0.018[a] (0.027)	0.04[c] (0.039)	0.087 (0.025)
Monitored majority domestic	0.119[a] (0.012)	0.017[a] (0.012)	0.021[a] (0.011)	0.022[a] (0.008)
Monitored majority unknown	0.089 (0.013)	0.088 (0.022)	0.126 (0.024)	0.09 (0.013)
Minority foreign	0.02[a] (0.019)	0.034[a] (0.020)	0.057[a] (0.020)	0.018[a] (0.015)
Minority domestic	0.023[a] (0.010)	0.027[a] (0.011)	0.039[a] (0.008)	0.019[a] (0.007)
Controlling minority foreign	0.013[a] (0.027)	0.048 (0.038)	0.053[c] (0.032)	0.089 (0.027)
Controlling minority domestic	0.012[a] (0.008)	0.018[a] (0.011)	0.022[a] (0.008)	0.013[a] (0.006)
Controlling minority domestic	0.009[b] (0.049)	0.003[a] (0.033)	0.001[a] (0.037)	0.003[b] (0.038)
Dispersed foreign	0.125 (0.036)	0.125 (0.045)	0.103 (0.041)	0.089 (0.027)
Dispersed domestic	0.071[a] (0.014)	0.06[a] (0.016)	0.056[a] (0.010)	0.041[a] (0.010)
Dispersed unknown	0.094 (0.019)	0.035[a] (0.021)	0.044[a] (0.020)	0.076 (0.017)
Unknown ownership (1996)	0.216[a] (0.001)	0.22[a] (0.002)	0.218[a] (0.002)	0.156[a] (0.001)
R-squared	0.023	0.042	0.039	0.026
N	24,890	10,439	11,608	19,732

Table 1 Ownership effects in manufacturing industries: Sectors by EUROSTAT; all years. Note: [a], [b], [c] denote the statistical significance at 1%, 5%, and 10% levels respectively.

Results for firms operating in services are presented in Table 2. On average owners of firms belonging to sectors of the less knowledge-intensive services (LKIS) and market LKIS are best conducive to firms' efficiency. On contrary, firms in sectors of the KIS and market KIS exhibit exceptionally poor results that are witnessed by rather high coefficients. Finally, firms in high-tech KIS do quite well with majority and monitored majority foreign owners driving the best efficiency results while other categories offer similar degree of efficiency irrespective of the owner's domicile.

Ownership category	Knowledge-intensive services (KIS)	High-tech KIS	Market KIS (1)	Less Knowledge-intensive Services (LKIS)	Market services less KIS
Majority foreign	0.216[a]	0.091[a]	0.248[a]	0.084[a]	0.084[a]
	(0.007)	(0.015)	(0.008)	(0.005)	(0.005)
Majority domestic	0.185[a]	0.062[a]	0.209[a]	0.108[a]	0.111[a]
	(0.005)	(0.013)	(0.006)	(0.004)	(0.004)
Majority unknown	0.316[a]	0.222	0.326[a]	0.185[a]	0.188[a]
	(0.005)	(0.016)	(0.006)	(0.004)	(0.004)
Monitored majority foreign	0.308	0.049[a]	0.417	0.146[a]	0.149[a]
	(0.030)	(0.052)	(0.038)	(0.022)	(0.022)
Monitored majority domestic	0.132[a]	0.106[a]	0.141[a]	0.057[a]	0.06[a]
	(0.010)	(0.023)	(0.011)	(0.008)	(0.008)
Monitored majority unknown	0.267[a]	0.109[a]	0.293[a]	0.108[a]	0.199
	(0.010)	(0.031)	(0.012)	(0.007)	(0.007)
Minority foreign	0.21[a]	0.048[a]	0.241[a]	0.161[a]	0.16[a]
	(0.015)	(0.044)	(0.018)	(0.012)	(0.012)
Minority domestic	0.196[a]	0.097[a]	0.196[a]	0.082[a]	0.086[a]
	(0.009)	(0.027)	(0.010)	(0.007)	(0.007)
Controlling minority foreign	0.365	0.044[c]	0.389	0.204[a]	0.207
	(0.028)	(0.092)	(0.030)	(0.019)	(0.020)
Controlling minority domestic	0.075[a]	0.05[a]	0.086[a]	0.042[a]	0.041[a]
	(0.008)	(0.015)	(0.009)	(0.005)	(0.005)
Minority domestic (33)	0.126[a]	0.003[c]	0.029[a]	0.154[a]	0.155
	(0.038)	(0.123)	(0.051)	(0.036)	(0.036)
Minority foreign (10)	0.292	0.22	0.371	0.002	0.002
	(0.294)	(0.280)	.	(0.115)	(0.115)
Dispersed foreign	0.484[a]	0.367[b]	0.471[a]	0.139[a]	0.144[b]
	(0.032)	(0.076)	(0.037)	(0.024)	(0.025)
Dispersed domestic	0.226[a]	0.101[b]	0.263[a]	0.182[a]	0.183[b]
	(0.014)	(0.049)	(0.016)	(0.011)	(0.012)
Dispersed unknown	0.26[a]	0.092[a]	0.288[a]	0.216[a]	0.215
	(0.014)	(0.034)	(0.016)	(0.010)	(0.010)
Unknown ownership (1996)	0.354[a]	0.207[a]	0.371[a]	0.207[a]	0.208[a]
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)
R-squared	0.031	0.036	0.027	0.019	0.019
N	141,31	10,108	120,331	154,882	156,089

Table 2 Ownership effects in service sectors: Sectors by EUROSTAT; all years. Note: [a], [b], [c] denote the statistical significance at 1%, 5%, and 10% levels respectively.

5 Conclusion

We analyze evolution of the financial efficiency in Czech firms during the period 1996-2007 and how the financial efficiency is determined by ownership structure. We provide evidence that ownership structure matters quite a lot and indicate numerous detailed results. Highly concentrated ownership is consistently beneficial to firm's efficiency and this finding is in favor of agency theory. Not surprisingly, dispersed ownership is the least preferable option.

On top of the above more or less expected results we show that a simple majority is not necessarily the best structure to affect financial efficiency. When a majority owner is monitored by a strong minority owner (monitored majority) this structure is more conducive to firm's efficiency than a pure majority. Further, we find that cooperative coalitions of minority owners that allow for control in a firm bring superior results. Minority owners who share control in a firm may end-up in rivalry position that is not conducive to efficiency. However, our evidence points at the act that minority owners do co-operate and improve financial efficiency of their firms.

We also show that financial efficiency is higher in less technology demanding and less knowledge intensive firms. This finding may question to some extent advancements of the Czech economy. On other hand we find that after 2004 financial efficiency converges between domestic and foreign (majority) owners. This is certainly a positive feature hinting at improved management and corporate governance in Czech firms.

References

[1] Aigner, D., Lovell, C., and Schmidt, P.: Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics* **6** (1977), 21–37.
[2] Aitken, B. J. and Harrison, A. E.: Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review* **89** (1999), 605–618.
[3] Arnold, J. M. and Javorcik, B. S.: Gifted kids or pushy parents? Foreign direct investment and plant productivity in Indonesia. *Journal of International Economics* **79** (2009), 42–53.
[4] Battese, G., and Coelli, T.: A model for technical inefficiency effect in a stochastic frontier production function. *Empirical Economics* **20** (1995), 325–332.
[5] Bernard, A. B. and Durlauf, S. N.: Convergence in International Output. *Journal of Applied Econometrics* **10** (1995), 97–108.
[6] Blomström, M.: Labor productivity differences between foreign and domestic firms in Mexico. *World Development* **16** (1988), 1295–1298.
[7] Blomström, M., Globerman, S., and Kokko A.: *The Determinants of Host Country Spillovers from Foreign Direct Investment.* Basingstoke: Palgrave, 2001.
[8] Brown, J. D., Earle, J. S and Telegdy, Á.: The productivity Effects of Privatization: Longitudinal Estimates from Hungary, Romania, Russia, and Ukraine. *Journal of Political Economy* **114** (2006), 61–99.
[9] Coelli, T. J.: Recent Developments in Frontier Modelling and Efficiency Measurement. *Australian Journal of Agricultual Economics* **39** (2005), 219–45.
[10] Cornwell, C., Schmidt, P. and Sickles, R. C.: Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels. *Journal of Econometrics* **46** (1990), 185–200.
[11] Claessens, S. and Djankov S.: Ownership Concentration and Corporate Performance in the Czech Republic. *Journal of Comparative Economics* **27** (1999), 498–513.
[12] Crinò, R.: Wages, Skills and Integration in Poland, Hungary and the Czech Republic: an Industry-Level Analysis. *Transition Studies Review* **12** (2005), 432–445.
[13] Demsetz, H.: The Structure of Ownership and The Theory of the Firm. *Journal of Law and Economics* **16** (1983), 375–390.
[14] Demsetz, H., and Lehn, K.: The Structure of Corporate Ownership: Causes and Consequences. *Journal of Political Economy* **93** (1985), 1153–1177.
[15] Djankov, S. and Murrell P.: Enterprise Restructuring in Transition: A Quantitative Survey. *Journal of Economic Literature* **40** (2002), 739–792.
[16] Estrin, S., Hanousek, J., Kočenda, E., Svejnar, J.: Effects of Privatization and Ownership in Transition Economies. *Journal of Economic Literature* **47** (2009), 699–728.
[17] Gugler, K.: Corporate Governance, Dividend Payout Policy, and the Interrelation between Dividends, R&D, and Capital Investment. *Journal of Banking and Finance* **27** (2010), 1297–1321.
[18] Hanousek, J., Kočenda, E. and Svejnar, J.: Origin and concentration: Corporate ownership, control and performance in firms after privatization. *Economics of Transition* **15** (2007), 1–31.
[19] Hanousek, J., Kočenda, E., Ondko, P.: The Banking Sector in New EU Member Countries: A Sectoral Financial Flows Analysis. *Czech Journal of Economics and Finance* **57** (2007), 200–224.
[20] Kim, C., Mauer, D. C. and Sherman A. E.: The determinants of corporate liquidity: theory and evidence**.** *Journal of Financial and Quantitative Analysis* **33** (1998), 335–359.
[21] Konings, J., Repkin, A.: How efficient are firms in transition countries? Firm-level evidence from Bulgaria and Romania. *Transition Economics* **1839** (1998), 1–26.
[22] Krueger, A. B. and Summers L. H.: Efficiency Wages and the Inter-Industry Wage Structure. *Econometrica* **56** (1998), 259–293.
[23] Khumbhakar, S.C.: Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency. *Journal of Econometrics* **46** (1990), 201–211.
[24] Laafia, I.: Employment in High Tech and Knowledge Intensive Sectors in the EU Continues to Grow in 2001. *Eurostat, Statistics in Focus: Science and Technology* **9** (2002), 1–8.
[25] Meeusen, W., van den Broeck, J.: Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* **18** (1977), 435–444.
[26] Melitz, M. J.: The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica* **71** (2003), 1695–1725.
[27] Shleifer, A., and Vishny R. W.: A Survey of Corporate Governance. *Journal of Finance* **52** (1997), 737–83.
[28] Thomsen, S., and T. Pedersen T.: Industry and ownership structure. *International Review of Law and Economics* **18** (1998), 386–404.

The mathematic modeling of process economization of natural resources

Simona Hašková[1], Pavel Kolář[2]

Abstract. The economic view of resources (to which crude oil belongs) distinguishes between the term of a "natural resource" and an "economic resource". The "natural" resource becomes economically interesting only if fulfilling the following two conditions:
There are technologies that enable mining and the utilization of natural resource;
There are prices and conditions that enable a meaningful usage of the resource.
The availability of condition two is being checked by the projects´ financial analyses of the exploration of natural resources and mining. The main task is to find out what kind of methods is available for use in exploration of the oil deposit and what strategy of follow-up mining to choose so that the expected NPV of the whole process (the exploration and mining) is the highest one. If the expected NPV is positive, the natural resource localized in the deposit justifies the economic resource and is therefore usable in practice.
In the contribution we draw on a typical decision tree of a project that analyzes exploration and mining. We identify the basic subjective factors (for example various types of options or price and maturity of mining rights), the absence of which, or on the other hand, high price and inappropriate timing deprives a natural resource of "economization". The mathematic models, that are effective tools for directing the transfer of the natural resources to the economical ones, derive from the detailed analyses of the relevant substructure of a decision tree. The interpretation is accompanied by a specific example of the exploration and mining analyses of an oilfield.

Keywords: Natural resource, economic resource, net present value (NPV), expected NPV, decision tree, option of a project termination, option of mining termination, oil well exploration method, seismographic exploration method.

JEL Classification: C02 C51 C58
AMS Classification: 91G06

1 Introduction

In 2009 the consortium of energy companies at the head of the American company Anadarko Petroleum confirmed a giant oil strike along the African coast. The oil area stretches from the Sierra Leone to Ghana. The relatively optimistic outlook was based on results of the exploration well Velus near the Sierra Leone shore, which is about 1 100 kilometers away from Ghana's predicted oil field.[3] At present the companies are working on several projects within which they want to start crude oil drilling. Recently there was news in the press about the discovery of a natural gas field at the European coast [8].

What the above mentioned news reports have in common is the information about the discovery of a natural resource. However, drilling the crude oil starts only if the field is **interesting from an economic point of view**. The first of the two assumptions of the "economic attractiveness" (the existence of technologies that enable drilling and utilization of the natural resource) mentioned in the abstract is evidently fulfilled in this case. Both of the resources have very attractive utilization (actually, they are essential for today's society). The potential threat of the global energy crises contributed significantly to the development of technologies available from various sites. According to the opinions of experts there is hardly anything else to improve in this regard.[4] That is why this assumption does not hinder the transformation of natural resources to economic recourses very much [6].

[1] VŠE Praha/Fakulta managementu JH, katedra managementu podnikatelské sféry, Jarošovská 1117/II, 377 01 Jindřichův Hradec, haskovas@post.cz.
[2] VŠE Praha/Fakulta managementu JH, katedra managementu podnikatelské sféry, Jarošovská 1117/II, 377 01 Jindřichův Hradec, kolar@fm.vse.cz.
[3] Original resource: The report in the Lidové noviny, September 2009 (amended).
[4] In the case of oil we have in mind the presence of oil in the liquid state. This oil was a natural resource in the time of pharaohs as well, but drilling technologies neither existed nor was the usages of it known. The natural resource is today even oil in the form of insoluble saturated clay. This natural resource is much more abundant

The situation is different in the case of the second assumption the "economic attractiveness" that concerns the price adequacy and general unspecified "conditions". Search for the natural resources fields for their exploitation is now in the hands of multinational companies. Deposits occur mostly in places where nobody has been looking so far, often in the countries of the "Third world". The governments of these states are increasingly realizing the strategic significance of raw oil. The local government perception of the wealth that is stored in oil resources reflects in their requirements applied when determining the conditions for permitting drilling. The significant role played in these conditions is in particular the price and mode of payment assigned drilling rights. The future of the oil field depends upon the agreement or disagreement on this issue and therefore, whether the resource is used for the benefit of all. To achieve the ubiquitous agreement contributes not only knowledge of how these factors translate into economic aspects of the project, but also the choice of survey methods and site selection strategies of drilling. The aim of the paper is to illustrate the crucial aspects related to this problem.

2 The decision tree of the problem

The contribution is based on the results of solving a large-scale case study entitled "The decision making between the use of traditional oil-well exploration method and the seismographic exploration method of a newly discovered oil field" solved in [3] in connection with the said finding of oil along the African coast. The model of the decision making problem about the optimal method choice, which is solved within the decision on the acceptance or rejection of the project, is the decision tree. All known information is reflected in the structure and in the evaluation of the individual leaves and derived expectations.[5] The result of many complex calculations and considerations is the completely specified decision tree in figure 1, on which we build our further considerations.

From the root node R of the decision tree in figure 1 two branches appear. The oil-well exploration branch leads upwards[6] and the seismographic exploration leads to the right side.[7]

As figure 1 show, it is not counted with oil drilling from the depths greater than 3000 feet. The numbers in the leaves of the tree indicate the NPV in millions USD (in case of the exploratory drilling its cost is already accounted to them), from which the unknown TP values (prices of the drilling rights) are deducted. In the right branch there is also X deducted, which in the model represents costs of the seismographic exploration (these are not included in NPV figures).

From the work [7] point of view we can regard the situational nodes as the elementary lotteries or compound lotteries (in case of the S_0 node). This economic approach in the form of NPV was studied in detail in [4]. The decision tree then can be viewed as a model of a certain strategic game: every application of option cuts off the part of the tree and every combination of application of options forms a strategy process. The optimal strategy is

---

than conventional oil fields but the lack of the cheap oil drilling technologies for extracting oil from the clay makes it economically unattractive.

[5] These are the mining conditions including information about the exploration field methods being considered, techno economic parameters of available drilling rigs and other mining equipment, information concerning the occurrence of oil in comparable locations, estimates of resource abundance, expectations about demand and prices, etc.

[6] The node V_i, i = 1, 2, 3 describes the situation of an exploratory well at a depth i measured in thousands of feet. The number of the horizontal edge that comes out of it gives the conditional probability that the well remains dry under the assumption that the well was dry even above 1000 feet. The opposite case (the oil is hit at depth i in thousands of feet) describes the edge upward. The situational node to which the edge enters distinguishes its output edges between the appearances of small or large quantities of oil (large quantity is twice as probable). The highest decision-making nodes to which these edges direct introduce the option (the right to choose) to abandon oil drilling in case the oil production is unprofitable. Only the presence of oil at a depth of 1000 feet means that the oil production can not be unprofitable and therefore the option is not needed at the V_1 node. On the contrary, low-lying decision-making nodes embedded in between the situational nodes V_i represent the options for abandonment of the project already in the exploration phase. The application of any of these options would occur at an oil-well exploration depth, where the hope to offset the marginal cost of further oil-well drilling by expected yield of extracted oil would fade.

[7] The node S_0 shows the possible outcomes of the seismographic exploration: We can expect oil at a depth of 1000 feet with a probability of 70 % (the situation S_1). With 10 % probability there is no oil at a depth under 3000 feet (which results in a loss of seismographic exploration costs; unlike the oil-well exploration method where, in the same case, the loss of oil-well exploration costs would increase the price of already purchased drilling rights). With the same probability (10 %) the oil can be expected at depths of 2000 or 3000 feet. The situations S_2 and S_3 are analogies to the situations at V_2 and V_3. However, unlike the situation at V_1, the options have to be included in the output edges of the situation S_1; the high price of drilling rights can cause a loss in oil production.

the one that offers the highest resulting expected NPV. The aim of the game is to find this strategy. And how to play the game let us see below.

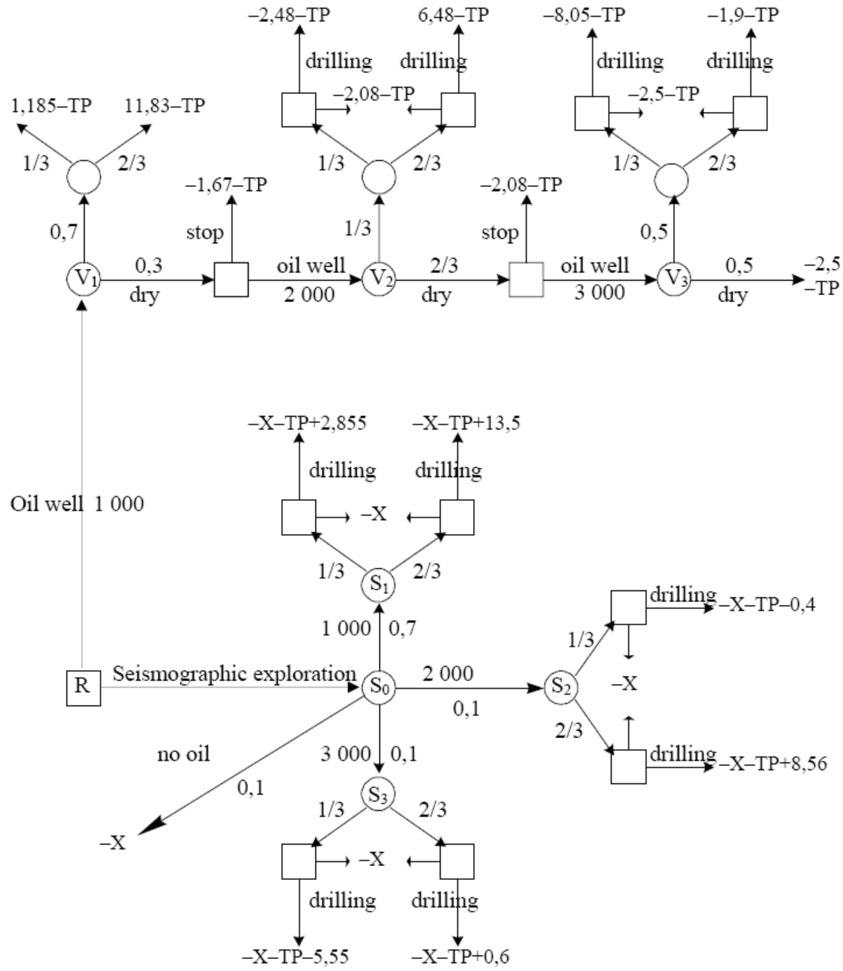


Figure 1 The simplified model of a decision tree problem

3 The effect of the drilling rights price on the exploration method choice and the drilling strategy

In this part we explore the impact of a parameter TP (price of the drilling rights) on the formation of an optimal game strategy. Every consideration of exploration methods will be analyzed separately; the optimal strategies and unbiased prognoses of NPV at the situational nodes of the tree are derived backward from the end to the beginning (that is from the outermost leaves to the root).

3.1 The drilling strategy after the oil-well exploration

The method of oil-well exploration requires purchasing the drilling rights before starting the oil-drilling (that is when we are not sure if we will ever come across the oil and therefore if there will be any oil to drill). The failure here means the loss of the total oil-well costs and cost of drilling rights. In figure 2 we can see upper branch (sub tree) of the decision tree from figure 1. Crossing out the edges indicate the possibilities that were not selected by the application or non-application of the options. Figures shown in italics at the situational nodes are the unbiased prognoses of NPV in the corresponding situation.

The optimal strategy for exploratory drilling and possible follow-up drilling can be easily read from figure 2 (at a depth of 1000 feet to drill regardless of the abundance of resources, at a depth of 2000 feet to drill only when there is a "lot" of oil and not to drill to a depth of 3000 feet). The dependence of the unbiased prognosis of NPV ($E[NPV]_1$) on TP of this strategy is given by $E[NPV]_1 = 5{,}74 - TP$, which was re-derived at the V_1 root of the sub tree. This strategy is independent of TP to the value TP = 5,74. The exploration oil-well method will not be used at TP > 5,74.

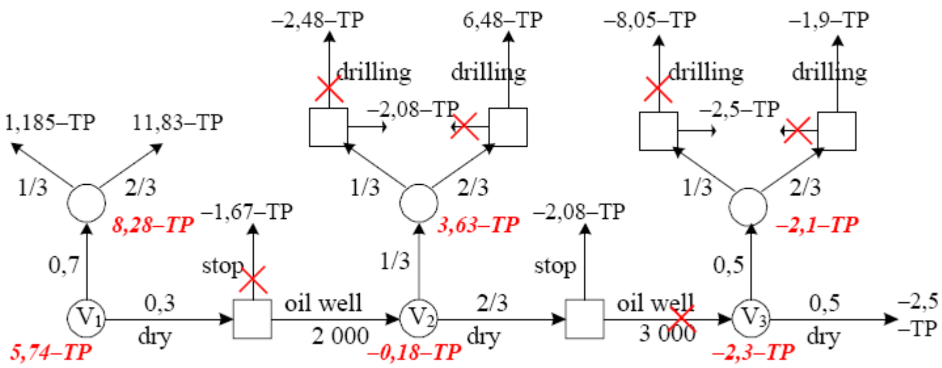


Figure 2 The tree-cut option "the oil-well exploration"

3.2 The drilling strategy after the seismographic exploration

Unlike the oil-well exploration, it is possible to purchase the drilling rights after the completion of the seismographic exploration; that is when we know that there is oil to be produced and we decide on starting the oil extraction. The failure here means the loss of X value (the seismographic exploratory costs). The decision on the use of the seismographic exploration method is based on the sub tree with the S_0 root (right branch) of the decision tree in figure 1. From the evaluated leaves of this sub tree it is obvious that as in case of oil-well exploration, low abundance of resource means that oil will never be drilled at a depth greater than 1000 feet. Whether oil will be drilled from a low abundant resource at a depth of 1000 feet or highly abundant resource from other depths depends solely on the TP price. Within the decision-making about the oil production the seismographic exploration costs of X are the sunk costs, therefore it does not affect the application or non-application of options. Around the prices TP $\in$ { 0,6, 2,855, 8,56 a 13,5 mil. USD} it leads to changes in the application of these options, which alters drilling strategies. Therefore it is necessary to examine the strategies separately at five successive intervals $0 \leq TP < 0{,}6$, $0{,}6 \leq TP < 2{,}855$, $2{,}855 \leq TP < 8{,}56$, $8{,}56 \leq TP < 13{,}5$ a $TP \geq 13{,}5$.
The examination result of the interval $0 \leq TP < 0{,}6$ can be seen in figure 3.

We can see that the optimal strategy in this case differs from the oil drilling strategy in figure 2 only in drilling the abundant oil resource even from a depth of 3000 feet. The predicted NPV of this strategy (the root S_0) takes on the value $E[NPV]_0 = -X - 0{,}83 \cdot TP + 7{,}576$.

The transition to the neighboring higher interval **0,6 ≤ TP < 2,855** means that the oil drilling at a depth of 3000 feet would be added to already marked loss oil drillings in figure 3. In the structure of the sub tree it would come to a change of cutting the edge in the right decision node placed under the situational node S_3 (application of the option "not to drill"). As a result the drilling strategy would fully coincide with the drilling strategy after the oil-well exploration (figure 2). It would also come to a change of the evaluation of the node S_3 to −X, which would alter the prognosis of NPV at the root S_0 to $E[NPV]_0 = -X - 0{,}77 \cdot TP + 7{,}536$.

The described consequences of the transition to the neighboring higher interval values of TP can be generalized and applied to other transitions:

- By the transition from the interval $0{,}6 \leq TP < 2{,}855$ to the interval **2,855 ≤ TP < 8,56** comes to an application of the option "not to drill" also to the oil drilling from the poor oil resource at a depth of 1000 feet (the decision node at the top left). This strategy change alters the evaluation of the node S_1 to $-X - 2 \cdot TP / 3 + 9$ and the NPV prognoses at the root S_0 to $E[NPV]_0 = -X - 0{,}53 \cdot TP + 6{,}871$;
- Compared to the previous case by the transition from the interval $2{,}855 \leq TP < 8{,}56$ to the interval **8,56 ≤ TP < 13,5** comes to an application of the option "not to drill" also at the oil drilling from the rich oil resource in a depth of 2000 feet (the decision root placed under the node S_2). This strategy change alters the evaluation of the node S_2 to –X and the NPV prognoses at the root S_0 to $E[NPV]_0 = -X - 0{,}47 \cdot TP + 6{,}3$;
- The transition to the interval **TP ≥ 13,5** enforces an application of the option that was the only one non applied in the previous case (the decision node at the top right), which blocks any possibility of oil drilling. All situational nodes (including S_0) then take on the value of – X and the seismographic exploration will not be realized. It does not come to the transformation of a natural resource to an economical resource.

The results of the above analyses are incorporated in a simple geometric model (see figure 4), for which multiple use can be found in practice.

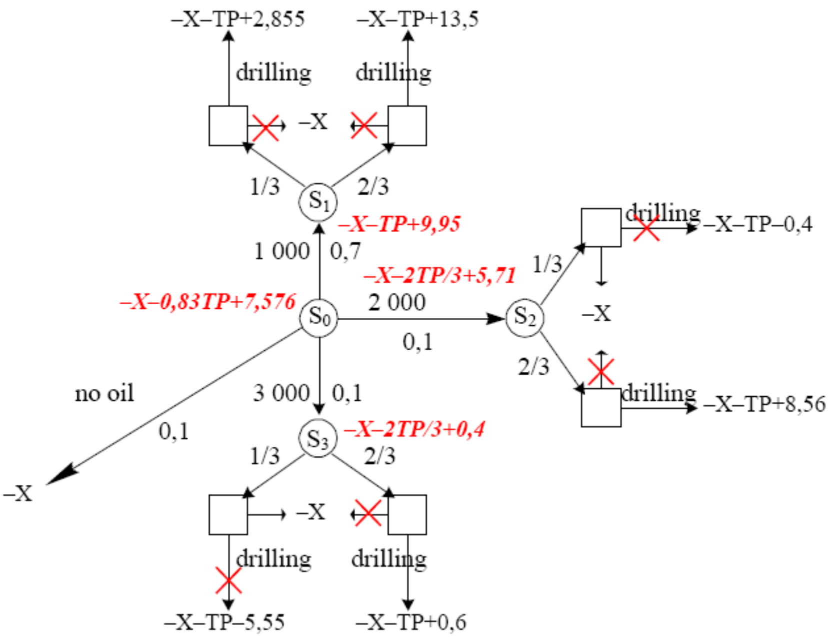


Figure 3 The tree-cut option “the seismographic exploration” at $0 \leq TP < 0{,}6$

4 The model dependence of the predicted NPV on the price of drilling rights (TP)

If we combine the partial sequences of the dependency of the unbiased NPV prognoses, which are obtained in the section 3.2, over the individual intervals of TP values, we get the course of this dependence over the entire range of TP values. This dependency in case of X = 0 (neglecting the seismographic exploration costs) is illustrated by the solid line in figure 4. By taking the seismographic exploration costs of X into account the line is moved in a parallel way about – X value vertically. An example of such a line is a broken line, plotted also in figure 4, moved about 2,141 downwards. It illustrates the course of dependence for the value X = 2,141 million dollars.

Apart from the two mentioned lines there is a dashed line plotted for the oil-well exploration in figure 4. Comparing both of the courses we can see that on the entire range of TP values the course of dependence decreases more slowly in case of the seismographic exploration than in case of oil-well exploration. However, within the seismographic exploration we see that with the growth of TP values the decrease of NPV values slows down. Therefore, when the both courses intersect somewhere (in our example it occurs when the price TP = 1,5 mil. USD) it means: to the left of the intersection the oil-well exploration method would be more advantageous and to the right of the intersection the seismographic exploration method would be better. The important information is the zero points of courses indicating the limit boundaries of usability of both examined methods. We can see that the price TP = 5,74 is the limit of usability of oil-well exploration, whereas the limit of usability of the seismographic exploration in the amount of X = 2,141 mil. lies somewhere near 8,6 mil. USD.

It should be emphasized that the model (as well as the decision tree, from which it was derived) is based on the knowledge of the situation before the oil field exploration. It is therefore a useful tool for only the first step in deciding on the entire project; when it comes to a decision whether or not to be concerned with the oil-field exploration or to reject the project. The project is rejected if the price of drilling rights is on the right side of the usability limit boundaries of both methods. Otherwise the project will be accepted and the model decides on the use of the oil-field exploration method.

The fact that the project was accepted and the oil-field exploration started would not mean that the natural resource became the economic resource. This would happen only if the oil drilling started. A decision on the oil production is based on the exploration results. Suppose that at the price of TP = 3 the seismographic exploration finds little abundant resource at a depth of 1000 feet (the tree branch protruding from the node S_1 to the left). In this case the oil production is loss making and the natural resource remains unused. None of the interested parties (owner of the resource or people interested in oil drilling) gain anything of it; on the contrary, both lose. This may be the motivation for the negotiation on the TP price reduction, which in this case prevents the transformation of the natural resource to the economic resource. If it comes to a price reduction by 0,5 on the basis of the negotiation, oil will be produced. A natural resource becomes an economic resource and both parties get richer.

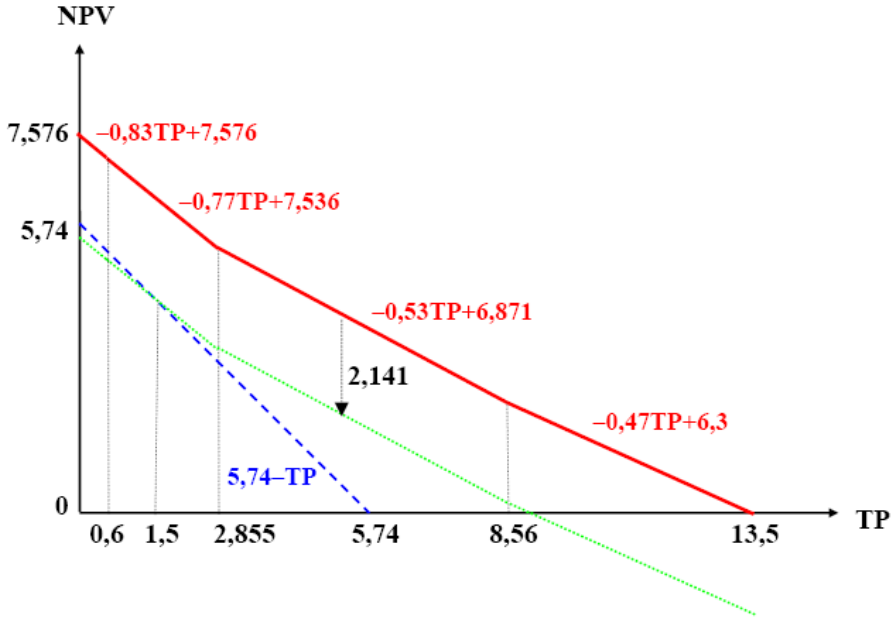


Figure 4 The graphic representation of the model

However, in the same situation it may become that the seismographic exploration does not express clearly to the richness of the resource (the possibility of an abundant resource is still in play) and the owner of a resource will not tend to reduce the TP price because of his pursuit of a fair profit distribution. Then other information can help to clarify the matter. However, this may not be available for free. By now the problem of setting an appropriate price of information comes into play. It is based on an argument that says that the price of information should not outweigh the benefit, which the information provides. In this context the work [5] deals with pricing of information based on Shannon's conception of uncertainty.

5 Conclusion

In our analyses we have seen that in case of the seismographic exploration the low cost of drilling rights made the oil an economic resource at a depth of 3000 feet and how increases in its price made it gradually economically unattractive in depths of 2000 or even 1000 feet. We have seen how the analyses of relevant substructures of the decision tree, which we conducted, allowed us to construct useful tools to detect and eliminate the causes and factors that prevent a meaningful application of natural resources. We have indicated the possibilities how can the resulting information contribute to an agreement on the TP price and thereby to remove barriers that hinder from the use of natural resources. Therefore the analyses can be considered as an useful tool that can help to discover new ”economic resources“.

References

[1] Brealey, R. A., Myers, S. C.: *Teorie a praxe firemních financí*. Computer Press, Praha 2000, ISBN 80-7226-189-4.
[2] Frank, R. H.: *The Return of the Economic Naturalist*. Virgin Books 2009, ISBN 9780753519660.
[3] Hašková, S.: *Základy teorie a praxe manažerských propočtů*. Diplomová práce FM VŠE, 2010.
[4] Hašková, S., Kolář, P.: *A contribution to a mathematic modeling of the behavior of private and institutionalized investors as observed in practice*. Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010, ISBN 978-80-7394-218-2.
[5] Hašková, S.: *The wealth maximization criterion & the utility maximization criterion*. Sborník příspěvků z vědecké konference doktorandů, studentů magisterského studia a mladých vědeckých pracovníků. Vysoká škola ekonomická v Praze, Nakladatelství Oeconomica – Jindřichův Hradec 2011.
[6] Hyne, N. J.: *Geology for petroleum exploration, drilling and production*. Pennwell Books, 1984, OSTI ID: 61066790.
[7] Varian, H. R.: *Microeconomic Analysis*. W.W.Norton & Company, Inc., 1992, ISBN 0- 393-95735-7.
[8] http://www.ordons.com/europe/western-europe/806-repsol-makes-two-new-oil-discoveries-offshore-spain.html

Comparison of the Multicriteria Decision Making and Client Clusters on the Retail Core Banking Services Market

Martina Hedvičáková[1], Alena Pozdílková[2], Ivan Soukal[3]

Abstract. The Czech Republic retail banking market is characterized by significant information asymmetry regarding the bank products and services. Banks provide complex and nontransparent tariffs. The analyses are based on more than 20 thousand of clients of 21 monitored banks offering 45 accounts and packages.
At the first part of the paper there are bank fees described using the client clusters characterized by certain usage pattern. There is a client decision based on the multi-criteria decision making at the second part of the paper. At first there are specified some criteria, their types and their independence is discussed. Then weights will be set and various multi-criteria decision methods will be applied. The first used treatment is one of the most widely known multi-criteria decision method - Saaty's AHP method. Then, as a basic approach pair-wise comparison is used and other methods are applied. At the end of the paper the possible application of uncertainty will be discussed. This application stems from inaccurate client answers. This question will be solved using the fuzzy sets, linguistic variable and linguistic scale that allows us to process the inaccurate answers or the answers when the client made just a guess which of the criteria is the key one.

Keywords: Bank, services, client, clusters, multi-criteria decision making, standardized weights.

JEL Classification: G21, C38, C10
AMS Classification: 62C86

1 Introduction

This paper is focused on the retail core banking services market (thereinafter only as RCBS abbreviation), which concern a great majority of consumers in European Union market (thereinafter only as EU abbreviation). These services include account administration, cash utilization, realizations of money transfers and additional services. Various studies assigned, or elaborated, by European Committee or European Union Directorate-General for Health and Consumers Protection [1], [2], [3] show the fundamental problem in this market from the point of view of the consumer. This problem is low transparency and very difficult product comparison of offers of individual banks with respect to client costs (generally included in the term bank charges), which restricts the principle of the invisible hand of the market. Opacity of offers of individual banks and impossibility to compare them, easily and clearly, is one of the key manifestations of asymmetry of information in the market. This problem was observed across the EU. [4]

As the very response in the Czech Republic (thereinafter only as CZ abbreviation) to this situation the web-based RCBS expert system called The Bank charge calculator (thereinafter only as Calculator) was created as an independent project of the RCBS market focused web bankovnipoplatky.com. This paper analyses Calculator respondent's data using cluster analysis based on k-means algorithm. Clusters were computed using statistical software IBM PASW 18 (formerly known as SPSS). [4]

2 The RCBS Calculator

The primary goal of the RCBS Calculator is advised the best product of the customer based on RCBS usage (what services are demanded and in what quantity). Knowledge base of the Calculator contains the tariff data of 12 banks (more that 98% of the RCBS market in the CZ) and their 45 accounts. Client only fills the detailed form concerning the usage of specific services – 52 questions in total (25 questions with attached sub questions

[1] University of Hradec Králové, Faculty of Informatics, Dept. of economics, Rokitanského 62, Hradec Králové, 500 03, Czech Republic, e-mail: martina.hedvicakova@uhk.cz
[2] University of Hradec Králové, Faculty of Informatics, Dept. of information technology, Rokitanského 62, Hradec Králové, 500 03, Czech Republic, e-mail: alena.pozdilkova@uhk.cz
[3] University of Hradec Králové, Faculty of Informatics, Dept. of economics, Rokitanského 62, Hradec Králové, 500 03, Czech Republic, e-mail: ivan.soukal@uhk.cz

and three additional questions). All fills are saved so the data for this study were acquired from 20 000 respondents who used the form of the Calculator in the time period from 12th October 2009 to 30th April 2011. The main tool for further help to choose the optimal cost account requires the identification of the most important RCBS client profiles. Because of specific respondent data gathering there has to be limited the population for which the client profiles will be identified. The client profiles in this paper are computed for and from e-banking client population (non-e-banking respondents were excluded from analysis). Analysis method of K-means was chosen by recommendation of [5], [6], [7] due to number of respondents and variable scale. [4]

From the marketing research point of view there are gathered data:

A. Multivariate – there has been monitored 53 variable concerning RCBS usage, 2 system variables for respondent identification and 45 variables containing the calculated costs for each of monitored RCBS product,

B. Primary – data were gathered directly from the client,

C. Subjective – data came from respondent himself, respectively it is his or hers subjective judge.

Due to specific data gathering process the data analysis outcome cannot be applied on the whole CZ population. Although the ratio of client with active e-banking and clients without electronic access to the account is 4:1 [7] and it will rise year by year we cannot include non-e-banking clients in our research. So the analysis and results concern e-banking population only. Still it is major share of the RCBS population. [4]

3 Cluster analysis results

Now there will be described client clusters. Within framework of this clustering, 6 943 respondents per 21 variables were analysed using the listwise method. Data validation of this treatment used more strict data filtering to guarantee that source data won't include extreme clients this paper described before. Exclusion of extreme clients allowed usage of source [8] method based on 5 clusters. [4]

Table 1. Centroid values for each cluster, source: own research

Variable/cluster	1	2	4	6
Domestic ATM withdrawal, own bank	3,1	2,5	2,8	2,1
Domestic ATM withdrawal, foreign bank	1,0	,4	,8	,4
Abroad ATM withdrawal, own bank	,4	,1	,1	,1
Abroad ATM withdrawal, foreign bank	,4	,1	,1	,1
Incoming payment from foreign bank	3,1	1,7	2,2	1,4
Incoming payment from own bank	2,3	,8	,7	,7
Direct payments to own bank at desk	,1	1,1	,0	,1
Direct payments to own bank Internet	3,6	,5	1,5	,8
Direct payments to foreign bank at desk	,1	1,7	,0	,1
Direct payments to foreign bank Internet	4,4	,7	4,0	1,3
Standing orders to own bank at desk	,0	1,6	,0	,1
Standing orders to own bank Internet	2,5	,2	,8	,3
Standing orders to foreign bank at desk	,0	2,7	,0	,2
Standing orders to foreign bank Internet	3,0	,3	3,0	,6
Encashment to own bank at desk	,0	,8	,0	,1
Encashment to own bank Internet	1,3	,1	,2	,1
Encashment to foreign bank at desk	,0	1,2	,0	,1
Encashment to foreign bank Internet	1,6	,1	1,1	,2
Cash deposit at desk	,6	,5	,2	,3
Cash withdrawal at desk	,3	,5	,1	,2
Cash back	,3	,1	,1	,1

In the table 1 there are shown values of cluster centroid for each of a typical client behavior pattern. The graph 1 then shows members count of the individual clusters. There will be described in detail 4 clusters [4]:

- The average client, 39,3% – cluster 3 is major group of the e-banking client population. It shares common frequency of ATM withdrawals with the others clusters (approximately 3 times from client's own bank and once from foreign bank). Typical for this client is preference of electronic banking usage both for direct payments, standing orders and encashment. Usage of at the desk services is sparse. Usage of these services is consisted of cash deposit, cash withdrawal only once or twice per year (this interpretation can be reversed as in the previous analysis, that is one from ten clients from this group uses an at desk cash withdrawal once per month).

- The active client, 16,3% – cluster 1 is a group of the more active clients, where, compared to the average client, the frequency of incoming payments is 2 times higher. Usage of services direct payments to own bank, cash ATM withdrawal from foreign bank, cash deposit or withdrawal and standing orders to own bank is 3 times higher. Concerning other services, this profile is similar to the average client and this client also shares with the average client the preference of the communication channel of e-banking. This client also has the highest frequency of ATM use abroad, although it is only 3 withdrawals per year on average.
- The average client with “at the desk services” preference 8,3% – cluster 2 is smaller than the cluster with desk preference clients from previous clustering. RCBS usage frequency of money transfers and incoming payments are very similar to the average client profile, the difference is that realization almost always occurs at the desk. This difference from the average client cluster can be noticed e.g. at almost 5 times higher frequency of cash deposits and withdrawals at the desk.
- The passive client, 35,9% – cluster 5 includes clients with low frequencies of monthly usage of all the monitored services. It can be noticed e.g. on services of money transfers and incoming payments, where this client profile receives only 2 payments per month and carries out only 2 direct payments and one standing order. All transfer services are done via internet. Compared to the average client profile, this cluster also has three times lower month frequency of ATM withdrawals from own bank and frequency of ATM withdrawal from foreign bank is lower by half. This client could be called low-cost client.

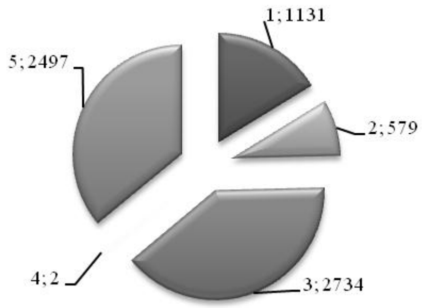


Figure 1 Shares of computed clusters,
source: own research

This treatment allowed cluster analysis to achieve better differentiation. It allowed the k-means algorithm to separate passive clients from the cluster of average clients. In the first clustering, passive clients had an effect up the profile of the average client. The major cluster (almost 70%) then has been split up to two smaller clusters that can be described in greater detail. Cluster member count is almost equal (the difference passive-average client is only approximately 5%). It is true that statistically this split did not have strong impact on average, but from the marketing point of view, it is a crucial to know the difference between the so-called low-cost client and average client with, however, 2–3 times higher use of services. Compared to the first treatment active client cluster’s share is here lower by 4%. The main reason is more strict data validation. The second treatment also raised the share of the client profile using mainly desk operations but an increase was less than 1%. The main benefit of the second more detailed treatment lies in the separation of passive clients from those average ones. [4], [9]

4 Multicriteria decision making

The second way, how to solve the problem described in previous chapters is application of Multicriteria decision making. Criteria properties and their weights will be described at the first part of the chapter. This knowledge will be applied on the profile of mean client and on one of the clusters (the mainstream one).

Of course we assume crisp data without uncertainty.

4.1 Criteria for multicriteria decision making

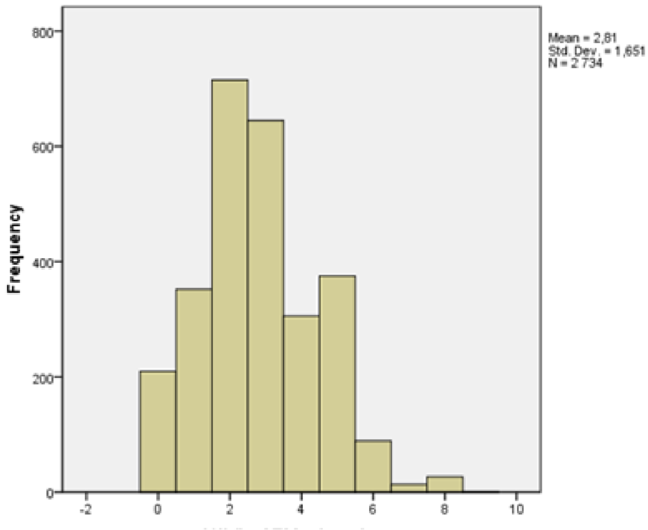
At first we have to describe used criteria. In this multicriteria decision making will be used the same main 21 criteria as in previous chapters – the criteria will be described in the next chapter.

Clearly all of these criteria are quantitative criteria.

The important condition, that has to be met, is independence of all these criteria. The independence was proved using correlation matrix (all pair correlations are zero or close to zero) and therefore independence of all used criteria is assumed.

4.2 Criteria weights

For the determination of the weights of basic methods there were used method of pair-wise comparisons, Saaty's AHP method and compensation method. The best results we got using the Metfessel's allocation method, in which statistical data and histograms for all criteria were used. We can see distribution in histograms of all values and their frequency, what helps us to determine weights.



Graph1 Histogram for Domestic ATM withdrawal in own bank

The weights are obviously standardized:

$$v_j \geq 0, j = 1, \dots m, \sum_{j=1}^{m} v_j = 1$$

The weights of the criteria are described in following table:

criterion	weight
Domestic ATM withdrawal, own bank	0,12
Domestic ATM withdrawal, foreign bank	0,12
Abroad ATM withdrawal, own bank	0,02
Abroad ATM withdrawal, foreign bank	0,03
Incoming payment from foreign bank	0,10
Incoming payment from own bank	0,10
Direct payments to own bank at desk	0,02
Direct payments to own bank Internet	0,07
Direct payments to foreign bank at desk	0,02
Direct payments to foreign bank Internet	0,07
Standing orders to own bank at desk	0,02
Standing orders to own bank Internet	0,07

Standing orders to foreign bank at desk	0,02
Standing orders to foreign bank Internet	0,07
Encashment to own bank at desk	0,01
Encashment to own bank Internet	0,03
Encashment to foreign bank at desk	0,01
Encashment to foreign bank Internet	0,03
Cash deposit at desk	0,03
Cash withdrawal at desk	0,03
Cash back	0,01

Table 2 Weights of criteria

4.3 Results of multicriteria decision making for the mean client

Now we describe results for the average client using the statistic data (average value of all data for all criteria was computed). Overall evaluation of variant will be computed as follows:

$$u(x) = \sum_{j=1}^{m} v_j \, u_j(x)$$

where v_j are standardized weights of criteria and $u_j(x)$ are evaluations of partial objectives (in this case this value is an average value in j-th criterion).

Overall evaluation is of course related to one variant.

We get overall evaluation of variant for the average client using these concrete values:

$$u(x) = 0{,}12 * 2{,}61 + 0{,}12 * 0{,}76 + 0{,}02 * 0{,}16 + 0{,}03 * 0{,}16 + 0{,}10 * 2{,}16 + 0{,}10 * 1{,}13 + 0{,}02 * 0{,}17$$
$$+0{,}07 * 1{,}67 + 0{,}02 * 0{,}21 + 0{,}07 * 2{,}82 + 0{,}02 * 0{,}17 + 0{,}07 * 0{,}97 + 0{,}02 * 0{,}30 + 0{,}07 * 2{,}03$$
$$+0{,}01 * 0{,}09 + 0{,}03 * 0{,}40 + 0{,}01 * 0{,}14 + 0{,}03 * 0{,}89 + 0{,}03 * 0{,}44 + 0{,}03 * 0{,}26 + 0{,}01 * 0{,}15$$
$$\boldsymbol{u(x) = 1{,}346}$$

4.4 Results of multicriteria decision making for the mean of the mainstream cluster

In this chapter we describe results for an average client in the mainstream cluster (mainstream cluster was described in one of the previous parts).

The computation is similar as in the previous chapter. We get overall evaluation of variant for an average client in the mainstream cluster using the specific values:

$$u(x) = 0{,}12 * 2{,}81 + 0{,}12 * 0{,}77 + 0{,}02 * 0{,}07 + 0{,}03 * 0{,}08 + 0{,}10 * 2{,}15 + 0{,}10 * 0{,}72 + 0{,}02 * 0{,}01$$
$$+0{,}07 * 1{,}53 + 0{,}02 * 0{,}03 + 0{,}07 * 4{,}00 + 0{,}02 * 0{,}01 + 0{,}07 * 0{,}79 + 0{,}02 * 0{,}02 + 0{,}07 * 3{,}03$$
$$+0{,}01 * 0{,}01 + 0{,}03 * 0{,}23 + 0{,}01 * 0{,}01 + 0{,}03 * 1{,}14 + 0{,}03 * 0{,}21 + 0{,}03 * 0{,}08 + 0{,}01 * 0{,}13$$
$$\boldsymbol{u(x) = 1{,}427}$$

4.5 Possible further expansion – uncertainty

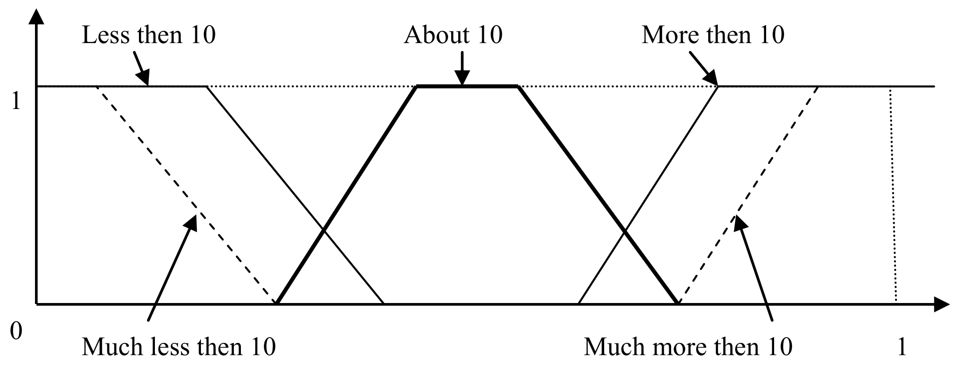
The decision problem was solved assuming the crisp data in previous chapters. But all clients (respondents) didn’t answer exactly, so in many real problems is contained a certain degree of uncertainty. For example, the client determined the criterion value of 10, which may not be exactly 10. The client for example meant that the value is between 8 and 12 and therefore he or she said 10.

This uncertainty can be solved with fuzzy sets that describe the answers to questions that are not exact.

Fuzzy sets (or fuzzy numbers, which are special case of fuzzy sets) can represent uncertain values (numbers) in our problem.

This problematic can be extended using linguistic variable and extended linguistic scale. This application helps us to better describe answers (because customers in their answers use natural language).

Next graph illustrates the value that is not exactly 10.



Graph 2 Application of the extended linguistic scale

5 Conclusion

The Czech Republic retail banking market is characterized by significant information asymmetry regarding the bank products and services. Banks provide complex and nontransparent tariffs. Using non-aggregated data concerning RCBS e-banking clients' month usage there was computed k-means cluster analysis. The treatments of clustering included of 6 943 respondents. The major cluster is the average client (39,3% share) by lower activity by 50 %, Internet preference is the same. This preference is also common for active client (16,3 % share). This cluster differs from major population by 2–3 times higher activity. The last group of clients (8,3 % share) showed at the desk preference and activity similar to the average client. The results can be used as a tool for further research of the price level/client's costs level of RCBS. The second possible use is as a tool for swift RCBS provider comparison when there are computed certain costs for each of retail account monitored by Calculator project. [4], [9]

The multicriteria decision making was used in the second part of the article. At first the standardized weights of criteria were computed and then the multicriteria decision making was applied on the mean client and on the mainstream cluster. Uncertainty is described at the end of the article as possible expandability, which can be extended in more directions. For example using fuzzy sets theory or linguistic scale offer better description of uncertainty in client answers.

Acknowledgements

This paper is written in the frame of the project of specific research "Retail banking adverse selection", original project name "Adverzní výběr v prostředí retailového bankovnictví", project number 2105, funded by Czech Republic Ministry of Education, Youth and Sport.

References

[1] European Union Directorate-General for Health and Consumers Protection. SEPA monitoring study. 2008, [online], URL: http://ec.europa.eu/consumers/rights/docs/SEPA_monitoring_study.pdf, cit. 22.8.2010

[2] European Union Directorate-General for Health and Consumers Protection. Retail financial services to the consumer markets scorecards. 2009, [online], URL: http://ec.europa.eu/consumers/rights/docs/swd_retail_fin_services_en.pdf, cit. 22.8.2010

[3] Commission of the European communities. Study on the Data collection for prices of current accounts provided to consumers. 2009, [online], URL: http://ec.europa.eu/consumers/strategy/docs/prices_current_accounts_report_en.pdf, cit. 24.8.2010

[4] Soukal, I., Hedvicakova, M.: Retail core banking services e-banking client cluster identification. In *Procedia Computer Science Journal,* volume 3. Published by Elsevier. 2010. p. 1205-1210, ISSN: 1877-0509

[5] Gordon, A.D.: Classification, 2nd edition, Chapman & Hall/CRC, London, 1999

[6] Brian, S. et al., Cluster analysis, 4th edition, Arnold Hodder headline group, London, 2001.

[7] Hebák, P.: Vícerozměrné staistické metody, 3. díl, (translated as Multivariate statistical methods, 3rd volume), Infromatorium, Prague, 2005

[8] Soukal, I.: Directive 2007/64/ES legal harmonization impacts on CBS client's costs in the ČR, Proceedings of the conference Hradec economical days 2010, Hradec Králové, 2008

Econometric Systems of Simultaneous Equations in Life Insurance

Radek Hendrych[1]

Abstract. In the present work, we deal with econometric systems of (linear) simultaneous equations. We apply this approach to a dynamic model of relevant relationships within a life insurance company. Concretely, we focus our attention on an econometric analysis of financial flows in a life insurance company operating in the Czech market. We assemble the linear econometric model and estimate its parameters using the usual three-stage least squares method. Selected results will be interpreted from the economic point of view. In this context, we also show a possibility of a residual bootstrap procedure which can be very useful here. For example, we construct bootstrap confidence intervals because in this specific situation (we operate with only a short data range) we cannot rely on the classical approximations by the normal distribution.

Keywords: econometric models, econometric systems of simultaneous equations, model of a life insurance company, residual bootstrap.

JEL classification: C30
AMS classification: 91G70

1 Introduction

Econometric systems of (linear) simultaneous equations are a special case of multivariate econometric models which allow us to work simultaneously with more than one dependent variable. For this reason, we can realize the essence of studied economic phenomena more precisely.

This approach can be appropriately applied in the context of (internal) models of relationships between different variables within a life insurance company. It has an extensive range of applications, e.g. predictions or generating scenarios of future developments (see [1] for more specifications).

2 Econometric Model of the Life Insurance Company

In this section, we deal with an econometric analysis of financial flows in a life insurance company operating in the Czech market. Concretely, we consider these variables: N_t - the number of the new life policies in year t (pieces), I_t - the technical interest rate in year t (% p.a.), IEF_t - the efficient interest rate in year t (% p.a.), i.e. the rate of return on investment activities of the insurance company, G_t - the relative growth coefficient of the average sum insured per one policy (e.g. $G_t = 0.01$ means the 1 % annual growth of the average sum insured when one compares the situation in year t with the one in year $t-1$, % p.a.), EX_t - the number of policies finished in year t (pieces), $PORT_t$ - the insurance portfolio in year t (pieces), i.e. the number of insurance policies, P_t - the insurance premium written in year t (in thousands of CZK), CS_t - the indemnity in year t (including surrenders and claims of policies, in thousands of CZK), V_t - the technical reserves in year t (in thousands of CZK), EAC_t - the acqusition expenses in year t (in thousands of CZK), EAD_t - the administrative expenses in year t (in thousands of CZK), RE_t - the reinsurance result in year t (in thousands of CZK), $PROF_t$ - the investment income of the insurance company in year t (in thousands of CZK), $PROFS_t$ - the share in profit of the insurance company in year t (in thousands of CZK) and R_t - the insurance result in year t (in thousands of CZK), $t = 1, \ldots, 11$. The complete data can be viewed on author's web page (`www.karlin.mff.cuni.cz/~hendrych`).

Now, define several other variables to better reflect the true dynamic economic relationships:

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech republic, e-mail: hendrych@karlin.mff.cuni.cz

- NG_t, the variable N_t adjusted by the average sum insured,

$$NG_t = \sqrt{1+G_t}\prod_{i=1}^{t-1}(1+G_i)N_t = \kappa_t N_t; \tag{1}$$

- $PORTG_t$, the insurance portfolio adjusted by the average sum insured,

$$PORTG_t = \kappa_t \cdot \frac{PORT_t + PORT_{t-1}}{2}; \tag{2}$$

- VI_t, the technical reserves adjusted by the technical interest rate,

$$VI_t = (1+I_t)\cdot\frac{V_t + V_{t-1}}{2}; \tag{3}$$

- $VIEF_t$, the technical reserves adjusted by the efficient interest rate,

$$VIEF_t = (1+IEF_t)\cdot\frac{V_t + V_{t-1}}{2}. \tag{4}$$

It is clear that variables $PORT$ and R are given by following equations, i.e. identities,

$$\begin{aligned} PORT_t &= PORT_{t-1} + N_t - EX_t, && (5)\\ R_t &= (P_t + RE_t + PROF_t) - && (6)\\ &- \left(CS_t + (V_t - V_{t-1}) + EAC_t + EAD_t + PROFS_t\right). \end{aligned}$$

The last identity (6) expresses the insurance result in year t as the total revenues $P_t + RE_t + PROF_t$ less the total costs $CS_t + (V_t - V_{t-1}) + EAC_t + EAD_t + PROFS_t$.

Next, proceed to formulate the econometric system of simultaneous equations. We have this model:

$$\begin{aligned} P_t &= a_1 + a_2 PORTG_t + \varepsilon_t^P, && (7)\\ CS_t &= b_1 + b_2 PORTG_t + b_3 V_{t-1} + b_4 EX_t + \varepsilon_t^{CS}, && (8)\\ V_t &= c_1 + c_2 V_{t-1} + c_3 P_t + c_4 CS_t + \varepsilon_t^V, && (9)\\ EAC_t &= d_1 + d_2 PORTG_t + d_3 NG_t + d_4 P_t + \varepsilon_t^{EAC}, && (10)\\ EAD_t &= e_1 + e_2 PORTG_t + e_3 NG_t + e_4 CS_t + \varepsilon_t^{EAD}, && (11)\\ RE_t &= f_1 + f_2 PORTG_t + f_3 NG_t + f_4 EX_t + \varepsilon_t^{RE}, && (12)\\ PROF_t &= g_1 + g_2(VIEF_t - VI_t) + g_3 P_t + g_4 CS_t + \varepsilon_t^{PROF}, && (13)\\ PROFS_t &= h_1 + h_2 PROF_t + \varepsilon_t^{PROFS}. && (14) \end{aligned}$$

This dynamic system of econometric equations originally contains six identities (1) - (6) and eight stochastic equations (7) - (14). The intercept and variables I, IEF, EX, G and N are assumed (strictly) exogenous variables and for this reason variables $PORT$, NG and $PORTG$, respectively, defined by equations (1), (2) and (5) are also strictly exogenous, i.e. these variables enter the system from outside. The lagged variable V_{-1} is supposed to be predetermined, i.e. it is determined by the system in time $t-1$. Note that the each equation in (7) - (14) satisfies the necessary condition for the identification in the framework of the econometric system of simultaneous equations (see [5] for the definition and more details).

2.1 Model Estimation and Interpretation

To estimate unknown parameters of previous stochastic simultaneous equations, we use the *three-stage least squares method*. This estimation technique is very common from the practical point of view and offers a set of suitable properties. It has shown that the parameter estimates are consistent, asymptotically normal distributed and asymptotically efficient (see [2] for more details).

For instance, the first estimated equation (7) is given by

$$\hat{P}_t = -13366.255 + 2.29916 \cdot PORTG_t$$

and the standard deviations of the estimated parameters a_1 and a_2 are 32166.17 and 0.175986, respectively. It follows from the just mentioned relationship that the average insurance premium is about 2299.16 CZK per one insurance policy and increases with the growth of the average sum insured.

In the eighth estimated equation (14), we obtain

$$\widehat{PROFS}_t = -6080.967 + 0.92877 \cdot PROF_t$$

and the standard deviations of the estimated parameters h_1 and h_2 are 4084.174 and 0.041221, respectively. From the economic point of view, it can be seen that the insurance company assigns their clients 92.88 % its investment income on average. In the Figure 1, the comparison of the variables P and $PROFS$ with their estimates is consulted. All parameter estimates are available on author's web page (see above).

The Hausman test of the model specification was accomplished to verify the structure of stochastic equations (7) - (14). It validates that all exogenous variables of the system (i.e. strictly exogenous and predetermined) are uncorrelated at any one time with the error terms (see [2] or [5]). The test statistics is calculated, $W = 1.73276$, it has 28 degrees of freedom and the p-value is virtually equals to one, i.e. the appropriateness of the model in the presented sense is not rejected.

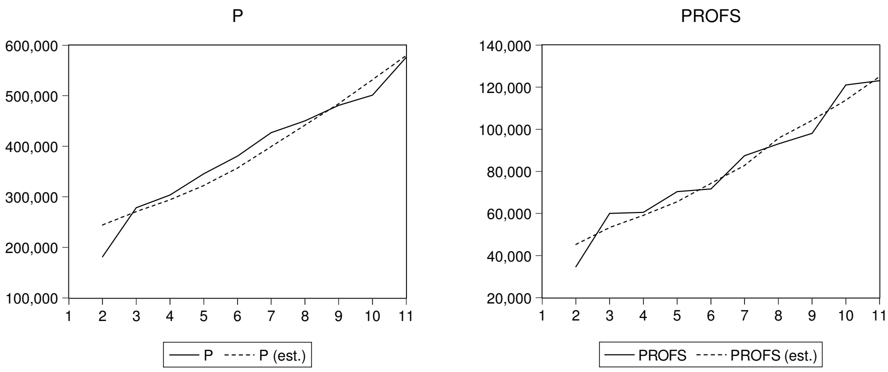


Figure 1: The variables P and $PROFS$ (solid curves) with their estimates (dashed curves).

2.2 Residual Bootstrap

If the distribution of the error terms (i.e. the disturbances) of some econometric system of simultaneous equations under considerations is unknown, so-called *bootstrap methods* may be applied to investigate the distributions of function of multiple time series or stochastic processes (see [4] or [3] for more details and the references given there). Especially, a residual based bootstrap is used in this context.

Suppose *the dynamic econometric system of simultaneous equations* in the general structural form

$$\mathbf{y}_t'\mathbf{\Gamma} + \sum_{l=1}^{p} \mathbf{y}_{t-l}'\mathbf{\Phi}_l + \mathbf{x}_t'\mathbf{B} + \boldsymbol{\varepsilon}_t' = \mathbf{0}, \qquad t = p+1, \ldots, T, \quad p \in \mathbb{N}, \tag{15}$$

where $\mathbf{y}_t$ denotes the vector of m endogenous variables, $\mathbf{y}_{t-k}$ is the vector of m lagged endogenous variables (i.e. the predetermined variables) and $\mathbf{x}_t$ is the vector of k strictly exogenous variables. Matrices $\mathbf{\Gamma}$ $(m \times m)$, $\mathbf{\Phi}_l$ $(m \times m)$ and $\mathbf{B}$ $(k \times m)$ contain the (unknown) parameters of the model and $\boldsymbol{\varepsilon}_t$ denotes the vector of m error terms. Compare with the model (7) - (14).

In the framework of the model (15), we commonly assume:

- $\mathsf{var}(\boldsymbol{\varepsilon}_t) = \mathsf{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \tilde{\mathbf{\Sigma}}$ and $\mathsf{cov}(\boldsymbol{\varepsilon}_s, \boldsymbol{\varepsilon}_t) = \mathsf{E}(\boldsymbol{\varepsilon}_s\boldsymbol{\varepsilon}_t') = \mathbf{0}$, $s \neq t$, and its asymptotic version $\text{plim}(1/T)\mathbf{E}\mathbf{E}' = \tilde{\mathbf{\Sigma}}$;
- $\mathsf{E}(\mathbf{x}_t\boldsymbol{\varepsilon}_t') = \mathsf{E}(\mathbf{x}_t)\mathsf{E}(\boldsymbol{\varepsilon}_t') = \mathbf{0}$ and its asymptotic version $\text{plim}(1/T)\mathbf{X}\mathbf{E}' = \mathbf{0}$;
- $\mathsf{E}(\mathbf{x}_t\mathbf{x}_t') = \mathbf{Q}$, $\mathbf{Q}$ is a finite regular matrix $(k \times k)$, and its asymptotic version $\text{plim}(1/T)\mathbf{X}\mathbf{X}' = \mathbf{Q}$;

where $\mathbf{X} = (\mathbf{x}_{p+1}, \dots, \mathbf{x}_T)$, $\mathbf{E} = (\boldsymbol{\varepsilon}_{p+1}, \dots, \boldsymbol{\varepsilon}_T)$ and plim denotes the probability limit operator as $T \to \infty$. We also suppose that $\boldsymbol{\Gamma}$ is an invertible matrix. If we multiply (15) from the right by $\boldsymbol{\Gamma}^{-1}$, we obtain so-called *the reduced form* of the econometric system of simultaneous equations.

In the context of using the residual bootstrap, [4] presents several stronger assumptions, e.g. $\{(\mathbf{x}_t, \boldsymbol{\varepsilon}_t)\}_t$ are i.i.d. random vectors which components have finite fourth moments, $\{\mathbf{y}_t\}_t$ is weakly stationary and a stability of parameter matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}_l$, respectively, is required (compare with [3]).

The residual based bootstrap procedure help us to find an estimator of a quantity of our interest, say $q := q(\boldsymbol{\Gamma}, \mathbf{B}, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p)$. If the values of strictly exogenous variables $\mathbf{x}_{p+1}, \dots, \mathbf{x}_T$ and variables $\mathbf{y}_1, \dots, \mathbf{y}_T$ are given, it proceeds as follows:

1. The residuals are estimated, i.e.

$$\hat{\boldsymbol{\varepsilon}}_t = -\mathbf{y}_t \hat{\boldsymbol{\Gamma}} - \sum_{l=1}^{p} \mathbf{y}_{t-l} \hat{\boldsymbol{\Phi}}_l - \mathbf{x}_t \hat{\mathbf{B}}, \qquad t = p+1, \dots, T,$$

where $\hat{\boldsymbol{\Gamma}}$, $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\Phi}}_1, \dots, \hat{\boldsymbol{\Phi}}_l$ are consistent estimates of parameter matrices (e.g. 3SLS estimates). These calculated residuals are centered. We get the sequence $\hat{\boldsymbol{\varepsilon}}_{p+1} - \overline{\boldsymbol{\varepsilon}}, \dots, \hat{\boldsymbol{\varepsilon}}_T - \overline{\boldsymbol{\varepsilon}}$, $\overline{\boldsymbol{\varepsilon}} = \frac{1}{T} \sum_{t=p+1}^{T} \hat{\boldsymbol{\varepsilon}}_t$.

2. The bootstrap residuals $\boldsymbol{\varepsilon}^*_{p+1}, \dots, \boldsymbol{\varepsilon}^*_T$ are then obtained by randomly drawing with replacement from the sequence of centered residuals. The bootstrap time series are computed recursively as

$$\mathbf{y}^*_t = -\sum_{l=1}^{p} \mathbf{y}^*_{t-l} \hat{\boldsymbol{\Phi}}_l \hat{\boldsymbol{\Gamma}}^{-1} - \mathbf{x}_t \hat{\mathbf{B}} \hat{\boldsymbol{\Gamma}}^{-1} - \boldsymbol{\varepsilon}^*_t \hat{\boldsymbol{\Gamma}}^{-1}, \qquad t = p+1, \dots, T,$$

where $\mathbf{y}^*_1 = \mathbf{y}_1, \dots, \mathbf{y}^*_p = \mathbf{y}_p$.

3. Based on the bootstrap time series, the parameter matrices $\boldsymbol{\Gamma}$, $\mathbf{B}$ and $\boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p$ are consistently reestimated. The bootstrap version of the statistic q, i.e. $\hat{q}^*$, is then calculated.

4. The second and third step are repeated B times, where $B \in \mathbb{N}$ is a large number.

After this procedure, the values $\hat{q}^*_1, \dots, \hat{q}^*_B$ are given. Define $\overline{\hat{q}^*}$, the bootstrap estimator of q, as follows

$$\overline{\hat{q}^*} = \frac{1}{B} \sum_{i=1}^{B} \hat{q}^*_i.$$

In this contribution, we deal with the construction of the bootstrap studentized confidence intervals. Suppose the studentized statistics

$$R = \frac{q - \hat{q}}{\sqrt{\widehat{\mathsf{var}}\,(\hat{q})}}$$

and its bootstrap counterpart

$$R^* = \frac{\hat{q}^* - \hat{q}}{\sqrt{\widehat{\mathsf{var}}\,(\hat{q}^*)}},$$

where the variances are estimated using standard statistical techniques.

If H is the distribution function of the statistics R and γ_p is its p-quantile, the $(1-\alpha)\%$ confidence interval for q, $\alpha \in (0,1)$, is defined as follows

$$CI_\alpha = \left(\hat{q} - \gamma_{(1-\alpha/2)} \sqrt{\widehat{\mathsf{var}}\,(\hat{q})}, \hat{q} - \gamma_{\alpha/2} \sqrt{\widehat{\mathsf{var}}\,(\hat{q})} \right). \tag{16}$$

The $(1-\alpha)\%$ *bootstrap (studentized) confidence interval* for q, $\alpha \in (0,1)$, is given by

$$CI^*_\alpha = \left(\hat{q} - \gamma^*_{(1-\alpha/2)} \sqrt{\widehat{\mathsf{var}}\,(\hat{q})}, \hat{q} - \gamma^*_{\alpha/2} \sqrt{\widehat{\mathsf{var}}\,(\hat{q})} \right), \tag{17}$$

where γ^*_p is the p-quantile of the distribution function H^* of the statistics R^*.

Note that $\sqrt{T}(\hat{q}-q)$ and $\sqrt{T}(\hat{q}^*-\hat{q})$ converge as $T \to \infty$ to the same limit distribution under suitable conditions. Therefore CI_α^* has the correct size asymptotically, i.e. $\mathsf{P}\left(q \in CI_\alpha^*\right) \to (1-\alpha)$ as $T \to \infty$, under general assumptions (see [6] for more details and references given there).

As mentioned earlier, this paper compares and contrasts these two approaches to constructing the different confidence intervals in the econometric model of financial flows in the insurance company. In the framework of the model (7) - (14) and its 3SLS estimation, 95% asymptotic (using the quantiles of the standardized normal distribution) and bootstrap confidence interval for unknown parameters are considered, i.e. intervals (16) and (17), respectively. The practical computational solutions have been implemented by a programme procedure in the EViews software (version 6.0) with the choice $B = 10\,000$.

For instance, in the first estimated equation (7) we obtain

$$CI_{0.05}(a_1) = (-76410.80, 49678.29), \quad CI_{0.05}(a_2) = (1.954231, 2.644085),$$

and

$$CI^*_{0.05}(a_1) = (-88074.82, 87470.76), \quad CI^*_{0.05}(a_2) = (1.778534, 2.718313).$$

For the parameters in the eighth equation (14), we get

$$CI_{0.05}(h_1) = (-14085.80, 1923.88), \quad CI_{0.05}(h_2) = (0.847980, 1.009562),$$

and

$$CI^*_{0.05}(h_1) = (-18371.82, 4803.17), \quad CI^*_{0.05}(h_2) = (0.818929, 1.051391).$$

In this special case, we operate with only the short data range and for this reason we cannot fully rely on the classical approximations by the normal distribution as shown in Figures 2 and 3. The bootstrap confidence intervals are then more reasonable.

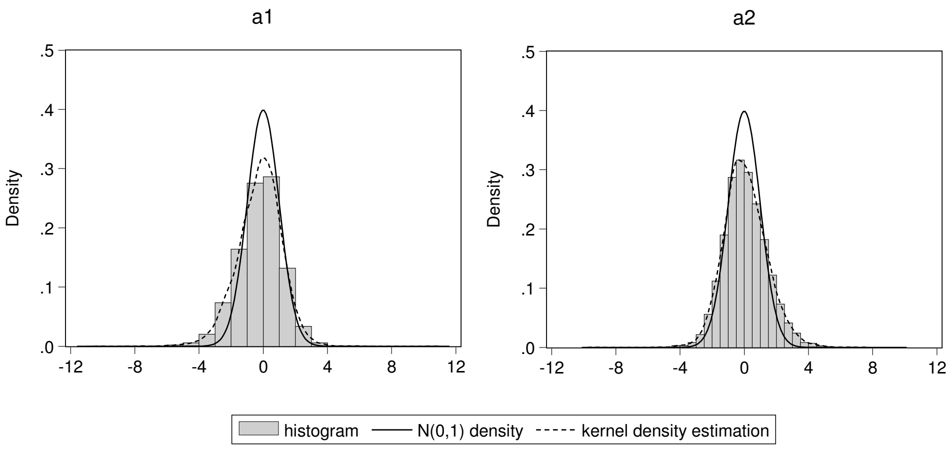


Figure 2: The distribution of $\{R_i^*\}_{i=1}^B$ is compared with the $\mathsf{N}(0,1)$ distribution (parameters a_1, a_2).

3 Conclusion

In the present work, we dealt with the econometric model of financial flows in the life insurance company operating in the Czech market. We assembled the econometric model of linear simultaneous equations to realize the (internal) relationships in the insurance company. It was shown that the estimated model equations provide us the reasonable and useful interpretation from the economic point of view.

The general residual based bootstrap technique for econometric models was introduced and it was applied to compare the relevant confidence intervals for unknown parameters.

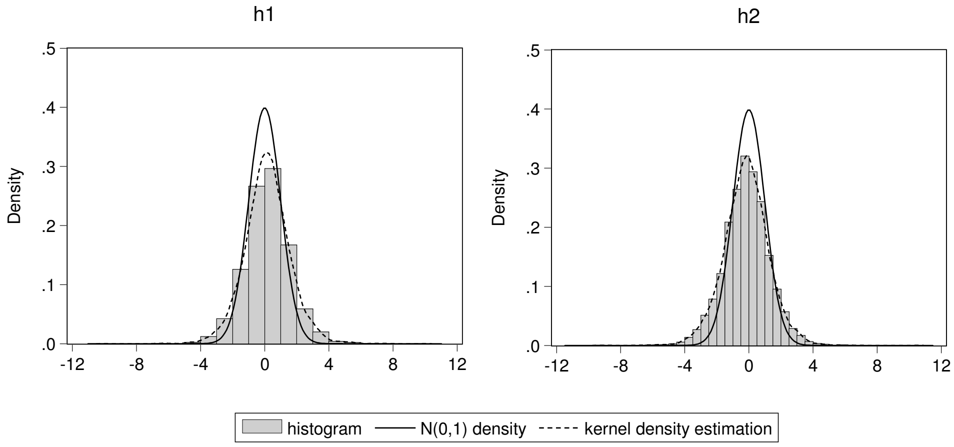


Figure 3: The distribution of $\{R_i^*\}_{i=1}^B$ is compared with the $\mathsf{N}(0,1)$ distribution (parameters h_1, h_2).

Acknowledgements

The work was supported by the grant SVV 261315/2011. The author also wishes to express his thanks to Prof. RNDr. Tomáš Cipra, DrSc. for several helpful comments.

References

[1] Cipra, T.: Econometric Analysis of Cash-flows in a life insurance company, *Pojistné rozpravy* **3** (1998), 66–72.

[2] Dhrymes, P. J.: *Topics in Advanced Econometrics: Linear and Nonlinear Simultaneous Equations.* Springer, New York, 1994.

[3] Faire, R. C.: *Estimating How the Macroeconomy Works.* Harvard University Press, Cambridge (MA, USA), 2004.

[4] Freedman, D. A: On Bootstrapping Two-stage Least-squares Estimates in Stationary Linear Model, *The Annals of Statistics* **12** (1984), 827–842.

[5] Greene, W. H.: *Econometric Analysis.* Prentice Hall, New Jersey, 2003.

[6] Lütkepohl, H.: *New Introduction to Multiple Time Series Analysis.* Springer, New York, 2005.

Estimating Output Gap in the Czech Republic

Pavel Herber[1]

Abstract. This contribution deals with an estimation of output gap in the Czech economy using a New Keynesian DSGE model. In this model the output gap is defined as a deviation of actual output from its flexible-price equilibrium level. The flexible-price equilibrium output is economically efficient and maximizes household welfare. These estimates are thus useful indicators for monetary policy.

The model is estimated using Bayesian methods. The model data fit and prior sensitivity is also carried out. The results indicate that the model is able to describe Czech data quite well. The estimated output gap is more volatile than traditional measure (HP filter), but the periods of booms and recessions are almost the same.

Keywords: output gap, flexible-price equilibrium, New Keynesian DSGE model,

JEL classification: E01, E12,

1 Introduction

The output gap is an important indicator of overall macroeconomic conditions and useful measure for monetary policy. The positive output gap indicates inflationary pressure and provides a signal to the monetary authority that policy may need to be tightened. The output gap is defined as deviation of actual from potential output. Unfortunately there is not any strict definition of potential output. Moreover potential output is not observable and researchers have to rely on its estimates. According to Kiley [5] output gap can be for instance defined as the deviation of output from its long-run stochastic trend (i.e., the Beveridge-Nelson cycle), the deviation of output from the level consistent with current technologies and normal utilization of capital and labor input (i.e., the production-function approach) or the deviation of output from flexible-price output (i.e., its natural rate). All above mentioned definitions can be incorporated to the New Keynesian DSGE model.

This contribution focuses on the last definition and estimates the output gap for the Czech economy using Bayesian techniques. Afterwards estimated output gap is compared to traditional gap estimated by Hodrick-Prescott filter. The model based output gap is little bit more volatile, but it illustrates Czech business cycle quite well. The periods of growth and fall are the same for both estimates. Finally the robustness of results is investigated. The model data fit and prior sensitivity is carried out and the results indicate that the model is able to describe Czech data pretty well.

The rest of the paper is organized as follows. The next section provides a brief description of the basic New Keynesian DSGE model which is used for analysis. Section 3 describes used data and presents the results. Final section concludes this contribution.

2 The model

Used DSGE model follows the paper of Justiniano and Preston [4]. The model consists of three representative agents – households, firms and monetary authority. A representative household maximizes its expected utility function with respect to its budget constraint and chooses optimal level of consumption and labour. Consumption contains a habit formation factor and is divided between domestic and imported goods. Domestic producers maximize their profits and optimize their behaviour by setting prices

[1]Masaryk University, Faculty of Economics and Administrations, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: pherber@mail.muni.cz

of its production in a Calvo style. Retail firms import foreign goods and have a small degree of pricing power. The monetary policy of a central bank is represented by a Taylor rule. It adjusts the interest rate with respect to the output gap and deviation of inflation from its steady-state value. Finally the foreign economy is represented by a closed version of the model.

The log-linearized system consists from following equations:

$$(1-\alpha)c_t = y_t - \alpha\eta(2-\alpha)s_t - \alpha\eta\psi_{F,t} - \alpha y_t^* \tag{1}$$

$$\Delta s_t = \pi_{F,t} - \pi_{H,t} \tag{2}$$

$$q_t = \psi_{F,t} + (1-\alpha)s_t \tag{3}$$

$$\pi_{H,t} = \beta\pi_{H,t+1} + \frac{(1-\theta_H)(1-\beta\theta_H)}{\theta_H}mc_t \tag{4}$$

$$mc_t = \varphi y_t - (1+\varphi)\varepsilon_{a,t} + \alpha s_t + \frac{\sigma}{(1-h)}(c_t - hc_{t-1}) \tag{5}$$

$$\pi_{F,t} = \beta\pi_{F,t+1} + \frac{(1-\theta_F)(1-\beta\theta_F)}{\theta_F}\psi_{F,t} \tag{6}$$

$$(c_t - hc_{t-1}) = (y_t^* - hy_{t-1}^*) + \frac{1-h}{\sigma}((1-\alpha)s_t + \psi_{F,t}) + \frac{1-h}{\sigma}(\varepsilon_{g,t} - \varepsilon_{g,t}^*) \tag{7}$$

$$(i_t - E_t\pi_{t+1}) - (i_t^* - E_t\pi_{t+1}^*) = E_t[\Delta q_{t+1}] + \varepsilon_{s,t} \tag{8}$$

$$i_t = \rho_i i_{t-1} + (1-\rho_i)(\psi_\pi\pi_t + \psi_{gap}gap_t) + \varepsilon_{i,t} \tag{9}$$

$$\pi_t = \pi_{H,t} + \alpha\Delta s_t \tag{10}$$

$$y_t^* - hy_{t-1}^* = E_t(y_{t+1}^* - hy_t^*) - \frac{1-h}{\sigma}(i_t^* - E_t\pi_{t+1}^*) + \frac{1-h}{\sigma}(\varepsilon_{g,t}^* - E_t\varepsilon_{g,t+1}^*) \tag{11}$$

$$\pi_t^* = \beta E_t\pi_{t+1}^* + \frac{(1-\theta^*)(1-\beta\theta^*)}{\theta^*}mc_t^* \tag{12}$$

$$mc_t^* = \varphi y_t^* - (1+\varphi)\varepsilon_{a,t}^* + \frac{\sigma}{(1-h)}(y_t^* - hy_{t-1}^*) \tag{13}$$

$$i_t^* = \rho_i^* i_{t-1}^* + (1-\rho_i^*)(\psi_\pi^*\pi_t^* + \psi_{gap}^* gap_t^*) + \varepsilon_{i,t}^* \tag{14}$$

$$\varepsilon_{a,t} = \rho_a\varepsilon_{a,t-1} + \mu_{a,t} \tag{15}$$

$$\varepsilon_{g,t} = \rho_g\varepsilon_{g,t-1} + \mu_{g,t} \tag{16}$$

$$\varepsilon_{s,t} = \rho_s\varepsilon_{s,t-1} + \mu_{s,t} \tag{17}$$

$$\varepsilon_{a,t}^* = \rho_a^*\varepsilon_{a,t-1}^* + \mu_{a,t}^* \tag{18}$$

$$\varepsilon_{g,t}^* = \rho_g^*\varepsilon_{g,t-1}^* + \mu_{g,t}^* \tag{19}$$

Variables and parameters with an asterisk relate to the foreign economy. Equation (1) is market clearing condition, equation (2) defines terms of trade. The relationship among the real exchange rate, law of one price gap and terms of trade is given by equation (3). Equation (4) represents a New Keynesian Phillips curve for domestic producers, where real marginal cost are defined by equation (5). Equation (6) represents a New Keynesian Phillips curve for importers, equation (7) expresses a link between foreign and domestic consumption levels. The uncovered interest rate parity condition is given by equation (8). Equation (9) represents a Taylor rule and equation (10) defines CPI inflation. The foreign economy is characterized by equations (11) – (14). The interpretation of these equations is the same as for corresponding equations for home economy. The dynamics of the model is driven by seven shocks. UIP shock and preference and productivity shocks in home and foreign economy are assumed to follow first-order autoregressive processes. Monetary policy shocks are assumed to be white noise.

In this model, the flexible price output is derived. The flexible price equilibrium is characterized by the absence of nominal rigidities. The flexible-price equilibrium output is economically efficient and maximizes household welfare. This level of output is therefore suitable as a target for monetary policy.[1] Flexible prices implies a constant mark-up and thus $mc_t = 0$. On the other hand the law of one price implies $\psi_{F,t} = 0$. By substitution to the corresponding equations and after some algebra, the following equations are obtained:

$$y_t^f = \frac{(1+\varphi)}{\varphi}(\varepsilon_{a,t} - \varepsilon_{a,t}^*) - \frac{1}{\varphi}s_t^f - \frac{1}{\varphi}(\varepsilon_{g,t} - \varepsilon_{g,t}^*) + y_t^{*f} \tag{20}$$

[1] For more details see Woodford [8].

$$
\begin{aligned}
s_t^f =& \frac{\sigma(1+\varphi)}{\varphi(1-h)(1-\alpha)^2+\sigma(\varphi\alpha\eta(2-\alpha)+1)}\Big((\varepsilon_{a,t}-h)-(\varepsilon^*_{a,t}-h\varepsilon^*_{a,t-1})\Big)+ \\
&+\frac{\sigma h(\varphi\alpha\eta(2-\alpha)+1)}{\varphi(1-h)(1-\alpha)^2+\sigma(\varphi\alpha\eta(2-\alpha)+1)}s^f_{t-1}- \\
&-\frac{\sigma+(1-h)(1-\alpha)}{\varphi(1-h)(1-\alpha)^2+\sigma(\varphi\alpha\eta(2-\alpha)+1)}(\varepsilon_{g,t}-\varepsilon^*_{g,t})+ \\
&+\frac{h\sigma}{\varphi(1-h)(1-\alpha)^2+\sigma(\varphi\alpha\eta(2-\alpha)+1)}(\varepsilon_{g,t-1}-\varepsilon^*_{g,t-1})
\end{aligned}
\tag{21}
$$

$$
y_t^{*f} = \left(\frac{(1-h)(1+\varphi)}{\varphi(1-h)+\sigma}\right)\varepsilon^*_{a,t}+\left(\frac{h\sigma}{\varphi(1-h)+\sigma}\right)y^{*f}_{t-1} \tag{22}
$$

Equations (20) and (22) define potential output and equation (21) represents flexible price terms of trade. Finally the output gap is defined as

$$
gap_t = y_t - y_t^f, \tag{23}
$$

$$
gap_t^* = y_t^* - y_t^{*f}, \tag{24}
$$

which measures the percentage deviation of the actual output from the flexible-price equilibrium output.

3 Estimation

3.1 Data and estimation techniques

A data set for the Czech Republic and the Euro area 12 from the first quarter 2000 to the third quarter 2010 is used for estimation. The data comes from the Eurostat database. Observable variables (for each country) are macroeconomic productivity (real GDP per worker), harmonized index of consumer prices, nominal interest rate and real exchange rate. All variables, except for interest rates, are seasonally adjusted and expressed as log changes. The nominal interest rates and inflation rates are annualized. Finally all variables are demeaned.

The model is estimated using Bayesian method. The likelihood function and unobserved variables are computed via the Kalman filter and the posterior distributions of the model parameters are computed by the Metropolis-Hastings algorithm. All computations are performed in Dynare[2] toolbox for Matlab (see Juillard [3]). More details of the estimation procedure can be found in Hamilton [1] and Koop [6].

3.2 Estimation results

Two parameters are calibrated. The discount factor β is set to 0.99, which implies an annual steady-state real interest rate of 4 per cent. The share of imported goods α is calibrated according to Musil [7] to the value of 0.7. The rest of parameters are estimated.

Table 1 presents the posterior estimates of parameters and 90% confidence intervals. Behaviour of a representative household is characterized by habit parameter h, inverse elasticity of interteporal substitution σ and by inverse elasticity of labour supply φ. The posterior estimates of these parameters are 0.59, 0.59 and 1.41. The current consumption is thus substantially influenced by the consumption in the last period and elasticity of labour supply is quite low, only 0.71.

The estimate of Calvo parameter θ_H is 0.68 and thus an average duration of price contracts for domestic producers is little bit more than 3 quarters. The estimated Calvo parameters for importers and foreign producers are little bit higher, θ_F is 0.70 and θ^* is 0.71. The average duration of price contracts is therefore 3.3 quarters and 3.5 quarters respectively. The posterior mean of elasticity of substitution between domestic and foreign goods η is 0.26.

The reaction function of the Czech National Bank is based on the parameters ρ_i, ψ_y and ψ_π. The estimates are following – ρ_i is 0.93, ψ_y is 0.22 and ψ_π is 1.29, which indicates a high stress on the interest

[2]Dynare verison 4.1.3 was used for estimation, available at: http://www.dynare.org/.

rate smoothing. On the other hand the Europe Central Bank focuses more on economic situation. The smoothing parameter is only 0.83 and weights on the output gap and inflation are higher, ψ_y^* is 0.34 and ψ_π^* is 1.3.

The estimates of autoregressive coefficients and standard deviations of shocks are higher in the Czech economy. The higher value of these parameters implies more persistent shocks and higher probability of larger shocks which is probably caused by the higher openness of the Czech economy.

Parameter	Prior mean	Prior s.d.	Post. mean	90% HPDI
h	0.7	0.1	0.5921	[0.4664; 0.7179]
σ	1	0.4	0.5913	[0.3072; 0.8821]
φ	1.2	0.2	1.4057	[1.1408; 1.6631]
η	1	0.4	0.2581	[0.0927; 0.4057]
θ_H	0.7	0.1	0.6801	[0.6051; 0.7738]
θ_F	0.7	0.1	0.7008	[0.2591; 0.7964]
θ^*	0.7	0.1	0.7120	[0.6403; 0.7862]
ρ_i	0.7	0.1	0.9296	[0.9093; 0.9506]
ρ_i^*	0.7	0.1	0.8307	[0.7917; 0.8725]
ψ_y	0.3	0.1	0.2205	[0.1235; 0.3139]
ψ_y^*	0.3	0.1	0.3407	[0.1808; 0.4937]
ψ_π	1.3	0.1	1.2920	[1.1320; 1.4514]
ψ_π^*	1.3	0.1	1.2989	[1.1343; 1.4594]
ρ_a	0.7	0.1	0.8440	[0.7583; 0.9294]
ρ_a^*	0.7	0.1	0.7285	[0.5709; 0.8933]
ρ_g	0.7	0.1	0.8910	[0.8161; 0.9628]
ρ_g^*	0.7	0.1	0.6837	[0.5709; 0.7951]
ρ_s	0.7	0.1	0.8066	[0.7033; 0.9112]
σ_a	0.1	∞	3.3099	[1.6274; 4.9439]
σ_a^*	0.1	∞	1.1651	[0.7020; 1.6167]
σ_g	0.1	∞	7.7784	[4.9995; 10.4488]
σ_g^*	0.1	∞	1.3163	[0.8416; 1.7815]
σ_s	0.1	∞	0.3722	[0.1453; 0.6001]
σ_i	0.1	∞	0.0967	[0.0768; 0.1165]
σ_i^*	0.1	∞	0.1171	[0.0924; 0.1414]

Table 1: Posterior estimates of parameters and innovations

Figure 1 plots the posterior mean of smoothed estimate of the output gap and its 90% confidence interval. The trajectory is compared with the gap obtained using Hodrick-Prescott filter with smoothing parameter equal to 10000[3]. The correlation between these two measures is only 0.52, but both measures indicate the same periods of growth and fall of GDP. The output gap had been positive during the early 2000s, but then dropped to be negative. The output gap reached the highest level in the middle of 2008 and after the begining of economic crisis dramaticly dropped down. Since 2010 the output gap has been slightly recovering.

An important advantage of DSGE model based output gap is that it enable to detect contributions of each shock. As can be seen from the shock decomposition the main sources of the output gap movements are demand disturbances. On the other hand monetary shocks caused only the small part of volatility and thus the differences between the actual and potential output are given by existence of staggered prices.

[3]The smoothing parameter is set according to the CNB estimate of the output gap (see Inflation report, avaible at: http://www.cnb.cz/).

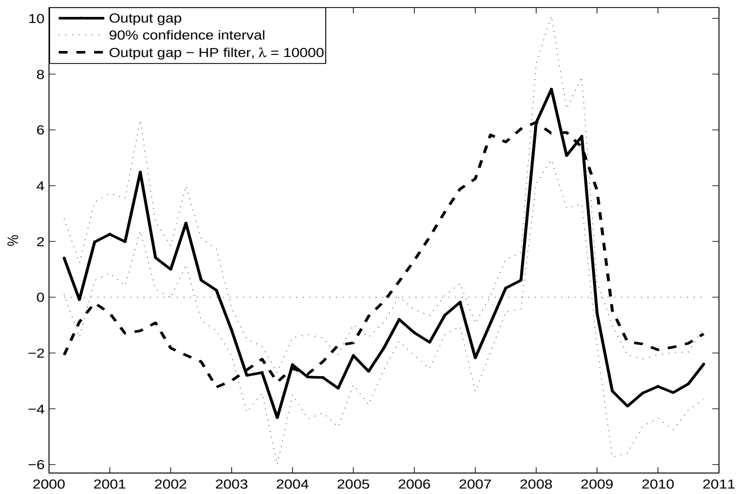


Figure 1: Output gap

3.3 Model evaluation

The estimated output gap strongly depends on the values of structural parameters and shocks. Their estimates may be influenced by the choice of prior hyperparameters and thus it is very important to investigate the prior sensitivity. The model was estimated using two sets of priors. In the case of loose prior the standard deviations has been doubled. In the case of noninformative prior the uniform distributions were used. Some parameters are not well identified, especially parameters in Taylor rule, but this is a common problem in DSGE models. However the trajectories of output gap obtained from these models are very similar, so the results are not very influenced by the choice of prior.

Finally the ability of the model to replicate the data is explored. Table 2 presents standard deviations and first and second order autocorrelations computed from the data and moments predicted by the model. The model little bit overestimates volatilities for all variables except for foreign interest rate. All computed autocorrelations have the same signs except for the first order autocorrelation of real exchange rate growth. The differences between moments computed from the data and from the model are not very high and therefore the model is able to describe Czech data quite well.

	Data	Model		Data	Model		Data	Model
σ_y	0.9184	1.1713	$\mathrm{ACF}(y,1)$	0.3920	0.3536	$\mathrm{ACF}(y,2)$	0.1510	0.0765
σ_{y^*}	0.6563	0.7431	$\mathrm{ACF}(y^*,1)$	0.2152	0.3014	$\mathrm{ACF}(y^*,2)$	-0.0894	-0.0262
σ_r	1.2980	1.3998	$\mathrm{ACF}(r,1)$	0.8979	0.9431	$\mathrm{ACF}(r,2)$	0.7654	0.8589
σ_{r^*}	1.3117	1.1954	$\mathrm{ACF}(r^*,1)$	0.9017	0.8539	$\mathrm{ACF}(r^*,2)$	0.7200	0.6557
σ_π	2.5556	3.1011	$\mathrm{ACF}(\pi,1)$	0.4714	0.5376	$\mathrm{ACF}(\pi,2)$	0.3009	0.2605
σ_{π^*}	1.3496	1.6652	$\mathrm{ACF}(\pi^*,1)$	0.3552	0.2049	$\mathrm{ACF}(\pi^*,2)$	0.0303	0.0156
σ_q	2.6201	2.7986	$\mathrm{ACF}(q,1)$	0.2952	-0.0094	$\mathrm{ACF}(q,2)$	-0.1336	-0.0346

Table 2: Statistical moments from the data and from the model

4 Conclusion

This contribution estimated output gap for the Czech economy using New Keynesian DSGE model. Unlike the estimates presented by Herber and Němec [2] the open economy model was used. In this model the output gap is defined similarly as a deviation of actual output from its flexible-price equilibrium level. The output gap is a useful measure for welfare and important indicator of macroeconomic conditions. Estimated output gap describes Czech business cycle quite well which differs from the finding of Herber and Němec [2]. The main driving forces of output gap are demand disturbances and the differences between actual and potential output are given mainly by stagerred prices which prevent more efficient allocation.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009.

References

[1] Hamilton, J. D.: *Time Series Analysis.* Princeton University Press, Princeton, New Jersey, 1994.

[2] Herber, P., and Němec, D.: Estimating Output Gap in the Czech Republic: DSGE Approach. In: *Proceedings of the 27th International Conference Mathematical Methods in Economics 2009* (H. Brožová, and R. Kvasnička, eds.), Czech University of Life Sciences Prague, Prague, 2009, 117–124.

[3] Juillard, M.: Dynare: A program for the resolution and simulation of dynamic models with forward variables through the use of a relaxation algorithm. CEPREMAP Working Papers (Couverture Orange), 1996, no. 9602.

[4] Justiniano A., and Preston B.: Small Open Economy DSGE Models: Specification, Estimation and Model Fit, unpublished, 2004, Available at: http://cama.anu.edu.au/Seminar%20Papers/Bruce_Preston_seminar_1.7.05.pdf

[5] Kiley, M. T.: Output gaps. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System, 2010, no. 27.

[6] Koop, G.: *Bayesian Econometrics.* John Wiley & Sons, 2003.

[7] Musil, K.: International Growth Rule Model: New Approach to the Foreign Sector of the Open Economy. Phd Thesis, Masaryk University, Brno, 2009.

[8] Woodford M.: *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton, NJ, 2003.

Financial accelerator in estimated model: application to the Czech economy

Miroslav Hloušek[1]

Abstract. The paper presents results of estimation of New Keynesian DSGE model with financial accelerator. The model is estimated using Bayesian methods on data of Czech economy. The models with and without financial accelerator are assessed how they fit the data using statistical moments and Bayesian factor. This exercise shows surprising result that the model without financial accelerator performs on data much better than model with it.

Keywords: financial accelerator, external finance premium, DSGE model, Bayesian estimation

JEL classification: E370

1 Introduction

The financial frictions are hot topic in current DSGE modeling. They are incorporated into models because they can amplify and propagate shocks into the real part of the economy. This issue is extremely attractive in the context of recent financial crises. This paper reacts to it and uses financial accelerator mechanism suggested by Bernanke et al.(1999) incorporated into the model from Christensen and Dib (2008). The model is adjusted in several dimensions and estimated using Bayesian techniques on Czech data. The model is then evaluated how it fits the data and is compared to model without financial accelerator. Results show that the model with financial accelerator does not perform on data better. Contribution of the financial accelerator mechanism for explaining business fluctuations is thus limited, at least for the Czech economy.

The rest of the paper is organized as follows. Section 2 presents main parts of the model, Section 3 briefly describes data and estimation technique. Results of the estimation, data fit of the models and dynamical properties are discussed in Section 4. Final section concludes with prospects for further research.

2 Model

The model is borrowed from Christensen and Dib (2008) and adjusted in several dimensions. The model includes frictions in the form of financial accelerator following Bernanke et al. (1999). There are several types of economic agents in the model but I will describe in detail only the most important ones.

Households live forever, they work, consume and save (supply funds to financial intermediary). Their optimization problem is quite standard and leads to usual first-order conditions.

Entrepreneurs are key part of the model. They manage firms that produce wholesale goods and borrow to finance the capital that is used in the production. Entrepreneurs are risk neutral and has finite horizon for planning purposes. There is a probability ν that the entrepreneur will survive until next period, so his expected lifetime is $1/(1-\nu)$. [2] They finance capital acquisition partly by net worth n_{t+1} and partly by borrowing $b_{t+1} = q_t k_{t+1} - n_{t+1}$ from financial intermediary. Financial intermediary obtains its funds from households and faces opportunity cost of funds R_t which is the economy's riskless gross rate of return. At the end of period t, entrepreneur decides to purchase capital k_{t+1} that will be used in the next period. The real price of capital is q_t, thus the cost of the purchased capital is $q_t k_{t+1}$.

[1]Faculty of Economics and Administration, Masaryk University, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: hlousek@econ.muni.cz

[2]New entrepreneurs are born and come into the market so that the fraction of entrepreneurs is constant.

The entrepreneur's demand for capital depends on the expected marginal return and the expected marginal external financing cost at t+1. The expected gross return on capital, $E_t f_{t+1}$, is given by

$$E_t f_{t+1} = E_t \left[\frac{z_{t+1} + (1-\delta) q_{t+1}}{q_t} \right] \tag{1}$$

where δ is the capital depreciation rate and z_{t+1} is the marginal product of capital at $t+1$ and $(1-\delta)q_{t+1}$ is the value of one unit of capital used in $t+1$. The optimal contract between the lender (financial intermediary) and the entrepreneur implies an external finance premium, $s(.)$, that depends on the entrepreneur's leverage ratio (net worth over capital investment).

$$s(.) = s\left(\frac{n_{t+1}}{q_t k_{t+1}} \right) \tag{2}$$

with $s'(.) > 0$ and s(1) = 1. As $n_{t+1}/q_t k_{t+1}$ falls, the borrower relies on uncollateralized borrowing (higher leverage) to larger extent funding of the project which rises cost of borrowing. The entrepreneur thus want to equate the return to capital to the marginal cost of external finance which is premium for external funds plus real opportunity costs represented by risk-free interest rate.

$$E_t f_{t+1} = E_t \left[s(.) R_t / \pi_{t+1} \right] \tag{3}$$

Log-linearized equation for the external funds rate is:

$$\hat{f}_{t+1} = \hat{R}_t - \psi(\hat{q}_t + \hat{k}_{t+1} - \hat{n}_{t+1}) \tag{4}$$

where ψ is elasticity of the external finance premium with respect to a change in the leverage position of entrepreneurs. Aggregate entrepreneurial net worth evolves according to

$$n_{t+1} = \nu v_t + (1-\nu) g_t \tag{5}$$

where v_t is the net worth of surviving entrepreneurs net of borrowing costs carried over from the previous period, $(1-\nu)$ is the share of new entrepreneurs entering the economy and g_t is the transfer that newly entering entrepreneurs receive from entrepreneurs who die and depart from the scene.

Entrepreneurs use capital k_t and labor h_t to produce output y_t following CRS production function. They sell their output on a perfectly competitive market.

Capital producers use a linear technology to produce capital goods. They use fraction of final goods from retailers as investment goods that are combined with existing capital stock to produce new capital goods k_{t+1}. Capital producers are subject to quadratic capital adjustment costs. Their optimization problem consists of choosing quantity of investment i_t to maximize their profits. It produces standard Tobin's Q equation that relates the price of the capital to the marginal adjustment costs.

The retail sector is used only to introduce nominal rigidity into model economy. Retailers purchase the wholesale goods from entrepreneurs (at price equal to the entrepreneur's nominal marginal cost) and differentiate them at no cost. Then they sell these differentiated retail goods in a monopolistically competitive market. This market structure allows them to set price of their goods. This mechanism is modeled a la Calvo (1983). Optimization problem is standard and leads to conventional New Keynesian Phillips curve.

Central bank adjusts the nominal interest rate R_t in response to deviations of inflation π_t and output y_t from equilibrium and partly responds to development of interest rate.

3 Data and estimation

The steady state of the model is computed and model equations are log-linearized around symmetric steady state. Solution of the system can be obtained using e.g. Blanchard and Kahn (1980) procedure. The model parameters are then estimated by Bayesian techniques. It combines maximal-likelihood with some prior information to get posterior distribution of the parameters. Specifically, posterior inference was obtained by Random Walk Chain Metropolis-Hastings algorithm which generated 500,000 draws from the posterior distribution. They were computed in two chains with 250,000 replications each, 50 % of replications were discarded so as to avoid influence of initial conditions. MCMC diagnostics were used for verification of the algorithm. All the computations were carried out using Dynare toolbox (Adjemian et al., 2011) in Matlab software.

The model is estimated using data for following model variables: output, investment, consumption, nominal interest rate, and inflation. Time series are obtained from Czech Statistical Office and Czech National Bank and cover time period 1996:Q1 – 2010:Q4. Specifically, output is gross domestic product (GDP), investment is gross fixed capital formation, consumption is measured by expenditure of households, interest rate is represented by 3M Pribor and inflation rate is q-on-q change of consumer price index (CPI). Data on output, investment and consumption are expressed per worker using data of total employment. All time series used for estimation are expressed as deviation from trend which is estimated by Hodrick-Prescott filter (with $\lambda = 1600$).

4 Results of estimation

Some model parameters are calibrated. Priors for the estimated parameters are set according to ratios from data or other empirical studies for the Czech economy. Table 1 shows results of Bayesian estimation. The key parameter related to financial accelerator mechanism is ψ – it expresses elasticity of the external finance premium with respect to firm's leverage. Value of 0.028 is smaller than the prior mean that was set to 0.05. This value is often used in calibrated versions of the model (see e.g. Bernanke et al., 1999). Also Christensen and Dib (2008) got lower value for this parameter, 0.042. However, with 90 % probability our estimate is different form zero. The capital adjustment cost parameter χ is estimated to

		Prior distribution			Posterior distribution		
Parameter	Interpretation	Density	Mean	Std	Mean	5 %	95 %
Structural parameters							
ψ	fin. accelerator	beta	0.05	0.10	0.0288	0.0209	0.0366
χ	adjustment cost	beta	0.50	0.10	0.7260	0.6481	0.8091
α	capital share	beta	0.33	0.10	0.4352	0.2777	0.5856
ϕ	Calvo parameter	beta	0.66	0.10	0.1842	0.1432	0.2259
Monetary policy rule							
ρ_r	int. rate smoothing	beta	0.80	0.10	0.2836	0.1996	0.3724
ρ_π	weitht to inflation	norm	1.50	0.10	1.4499	1.2874	1.6272
ρ_y	weight to output	norm	0.125	0.05	0.1010	0.0227	0.1792
Shock persistence							
ρ_a	technology	beta	0.70	0.10	0.7159	0.6134	0.8114
ρ_e	preference	beta	0.70	0.10	0.4486	0.3546	0.5467
ρ_x	investement	beta	0.70	0.10	0.4531	0.3325	0.5729
ρ_{cp}	cost push	beta	0.70	0.10	0.7634	0.6446	0.8799
Standard deviation of shocks							
ϵ_r	monetary	invg	0.01	Inf	0.0382	0.0304	0.0452
ϵ_e	preference	invg	0.01	Inf	0.0166	0.0136	0.0195
ϵ_a	technology	invg	0.01	Inf	0.0173	0.0099	0.0283
ϵ_x	investment	invg	0.01	Inf	0.0299	0.0238	0.0358
ϵ_{cp}	cost push	invg	0.01	Inf	0.0174	0.0020	0.0555
ϵ_g	government	invg	0.01	Inf	0.0912	0.0769	0.1050

Table 1: Bayesian estimation results

0.726, much higher than the prior setting and commonly used values, see again Bernanke et al. (1999). However, Meier and Müller (2006) report also quite high estimate of this parameter, 0.65. Capital share in the production function, α is also higher than the prior mean but is in line with empirical studies for the Czech economy, see e.g. Hloušek (2007). Estimated value of Calvo parameter ϕ is 0.1842 which is implausibly low. It implies average duration of price contracts 1.2 quarters. Hloušek (2010) also found out significant flexibility of domestic prices in the Czech economy, but his estimate of Calvo parameter is 0.40, still higher than in this model. Monetary policy rule shows very low interest rate smoothing, parameter ϱ_r is only 0.283. It is in stark contrast with other estimated models for the Czech economy, see again Hloušek (2010) or Musil (2008). Weights to inflation and output gap in monetary policy rule are quite standard. The most persistent shocks are cost push shock and technology shocks, $\rho_a = 0.7169$ and $\rho_{cp} = 0.7634$, respectively. Looking at volatility of the shocks, government spending shock and monetary shock are the most volatile ($\epsilon_g = 0.0912$ and $\epsilon_r = 0.0382$).

This benchmark model (referred to as FA model) is further compared to model without financial accelerator. That model is obtained when the parameter ψ is set to zero and other parameters are again estimated. This reduced version is referred to as noFA model. There are not large differences between estimated parameters of these models. However, the models are compared how they fit the data.

4.1 Assessing of fit

This subsection compares moments from the data to model implied moments from FA and noFA model. Moments of the interest are standard deviations and autocorrelations.

Variable	Data		FA model		noFA model	
	Std	Std rel. to y_t	Std	Std rel. to y_t	Std	Std rel. to y_t
y_t	2.26	1.00	4.16	1.00	1.93	1.00
i_t	3.29	1.45	9.87	2.37	4.66	2.42
c_t	1.28	0.57	4.62	1.11	2.54	1.32
π_t	3.57	1.58	3.55	0.85	3.52	1.83
R_t	2.20	0.97	1.30	0.31	1.45	0.75

Table 2: Standard deviations and relative volatilities

Table 2 shows standard deviations of output, investment, consumption, inflation and nominal interest rate calculated from data and from data series generated by models. The third, fifth and seventh column shows volatilities relative to output. Looking at absolute values of standard deviations, both models match volatility of inflation perfectly. Consumption and investment is less volatile while interest rate is more volatile in the data than in the models. Generally, noFA model is closer to data than FA model. Regarding relative volatilities, both models capture the fact that investment is more volatile than output but both overstate the size of volatility. Similarly, both models fail to match smaller volatility of consumption relative to output. In case of volatility of inflation and interest rate noFA model performs better than FA model.

Figure 1 plots unconditional autocorrelation of the data and autocorrelations of time series obtained by model simulations. Moments for data were calculated from estimation of reduced VAR(1) process. It is evident that model without financial accelerator does a good job in matching autocorrelations

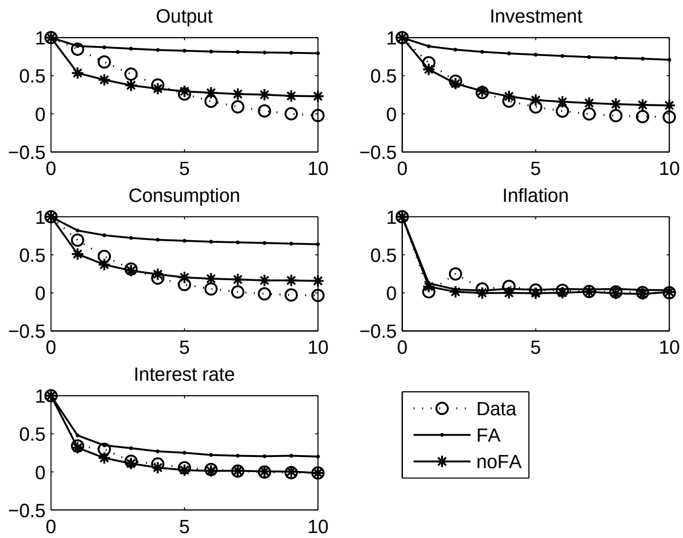


Figure 1: Autocorrelations

for all variables. Only for consumption and output it is less correlated than data at one and two (or three) quarters horizon and more correlated at six and more quarters horizon. Higher autocorrelation for inflation for two lags is not matched by any of the models. Outcome of FA model is quite poor. Variables are very persistent and descent only slowly which is at odd with the data.

4.2 Dynamical properties of the model

Dynamical properties of the model can be studied by impulse response functions and variance decomposition. Figure 2 plots behavior of several endogenous variables in reaction to investment efficiency shock. Solid line is used for FA model and dashed line for noFA model. Positive investment efficiency shock means that final (consumption) goods are more effectively turned into investment goods. This shock increases investment, worked hours and hence output. The price of efficiency unit of capital, q_t, falls which lowers the return on capital and net worth. The resulting rise of external finance premium increases the cost of funding of investment projects which brings investment back to equilibrium. Interest rate rises as central bank reacts to high inflation and output which contributes that the economy returns to steady-state. For model without financial accelerator, the initial reaction of real variables (such as output, investment or consumption) is slightly lower and subsequent response is more persistent than for model with it. However, the distinction between these models is negligible and the question whether the financial accelerator amplifies or dampens shocks to the economy can not be resolved based on these results.

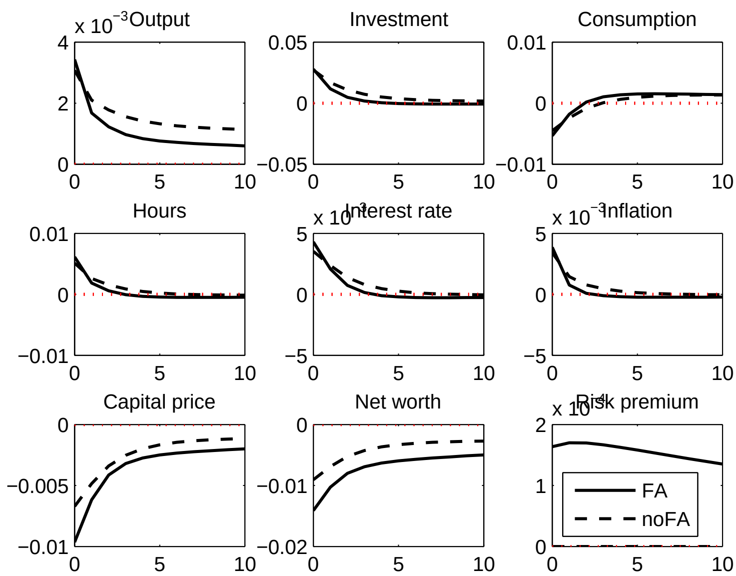


Figure 2: Impulse responses to investment efficiency shock

Variation of model variables can be decomposed into contribution of each exogenous shock using variance decomposition. Table 3 shows forecast error variance of the variables in the long run horizon. Even if the volatility of monetary policy shocks were not largest, their contribution to behavior of model variables is enormous. They explain from 62 to 88 per cent of the variation of selected variables. On the other hand, cost push shocks are quite unimportant. Investment efficiency shocks explain only 12 % of variation of investments and also partly contributes to variation of interest rate. Technology shocks contribute to behavior of output and consumption by 12 % and 14 % respectively. Contribution of preference shocks is relatively minor and government spending shocks which have largest volatility only partly explain variation of consumption and, to smaller extent, of other variables.

Finally, Bayesian approach to estimation can be used for quantitative comparison of models data fit. Specifically, Bayesian factor (BF) which is based on calculated marginal data densities expresses how many times one model is more probable than the other. Bayesian factor of noFA/FA models is 86716.29 which indicates decisive evidence in favor of model without financial accelerator.[3]

[3] See DeJong and Dave (2007) for ranges of BF and its interpretation.

Variable	Percentage owing to the type of shock					
	Investment	Technology	Preference	Monetary	Cost push	Government
y_t	2.2	12.0	1.6	76.9	0.1	7.2
i_t	12.0	3.3	0.9	80.1	0.0	3.6
c_t	4.5	14.1	6.5	62.4	0.1	12.4
π_t	1.0	3.0	0.9	88.1	0.0	6.9
R_t	10.5	5.7	1.8	74.7	0.1	7.2

Table 3: Variance decomposition

5 Conclusion

This paper presented results of estimation of New Keynesian model with financial accelerator on Czech data. Financial accelerator is usually incorporated into models because of better data fit and for the reason that it amplifies and propagates shocks into the economy. The results for the Czech economy are not in accordance with this practice. Even if the model with financial accelerator was successfully estimated, its ability to explain data is not satisfactory. The restricted model (without financial accelerator) fits the data better. Next, the issue regarding propagation of the shocks into the economy can not be resolved based on current results. Also some estimated parameters do not have plausible and clearly interpretable values. The reason can be that the model omits some important channels. It is closed economy model and includes only one type of nominal rigidity. The topic for further research is extension of the model by foreign sector and other nominal and real rigidities.

Acknowledgements

This work is supported by MŠMT project Research centers 1M0524.

References

[1] Adjemian, S., Bastani, H., Juillard, M., Mihoubi, F., Perendia, G., Ratto, M. and Villemot S. (2011), Dynare: Reference Manual, Version 4, *Dynare Working Papers,* 1, Cepremap.

[2] Bernanke, B.S., Gertler, M., and Gilchrist S. (1999) The Financial Accelerator in a Quantitative Business Cycle Framework, In J.B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics,* vol. 1C. North-Holland Publishing Co.

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

[4] Christensen, I., Dib, A. (2008) The financial accelerator in an estimated New Keynesian model, *Review of Economic Dynamics,* 11, 155178.

[5] DeJong, D. N., Dave, C. (2007) *Structural macroeconometrics,* Princeton: Princeton University Press.

[6] Hloušek, M. (2007) Dual approach to growth accounting – application for the Czech Republic. In *Mathematical Methods in Economics 2007.* Ostrava: Faculty of Economics.

[7] Hloušek, M. (2010) *Nominal rigidities and wage-price dynamics in estimated DSGE model: application for the Czech economy,* PhD Thesis, Masaryk University.

[8] Meier, A., Müller, G.J. (2006) Fleshing out the monetary transmission mechanism: output composition and the role of financial frictions. *Journal of Money, Credit, Banking* 38, 1999–2133.

[9] Musil, K. (2009) *International Growth Rule Model: New Approach to the Foreign Sector of the Open Economy,* PhD Thesis, Masaryk University.

[10] Wieland, V., Cwik, T., Mller, G. J., Schmidt, S. and Wolters, M. *A New Comparative Approach to Macroeconomic Modeling and Policy Analysis,* Working Paper, Goethe University of Frankfurt, May 2011 (revised version).

Unified approach utilizing entropy to measure operational complexity of supplier-customer system – case studies

Jiří Hofman[1], Ladislav Lukáš[2]

Abstract. Paper discusses unified approach based upon entropy for quantitative measurement of operational complexity of company supplier-customer relations. We assume a problem-oriented database exists, which contains detailed records of all product forecasts, orders and deliveries both in quantity and time, scheduled and realized, too. Basic DB pre-processing detects the most important flow variations, e.g. order – forecast, delivery – order, and actual production – scheduled one, provided set of states monitoring various flow variations having been defined. The variation states are classified considering their severity with respect to management control decisions. After consistency checks of collected data, histograms and empirical distribution functions are constructed. Finally, the entropy, information-theoretic measure of supplier-customer operational complexity, is calculated by Mathematica module developed. Results of supplier-customer system analysis from selected Czech SME are presented in detail.

Keywords: Business economics, supplier-customer systems, firm performance, complexity measures, information and entropy.

JEL Classification: C63, C81, L25, M21
AMS Classification: 62B10, 91F99, 94A17

1 Introduction

There is well-known that business economics considers two types of supplier-customer system complexity – a structural complexity and an operational one. Opposite to structural complexity, which prefers a static concept, the operational complexity is concerned with dynamic one. There is a lot of literature available presenting various approaches and aspects thereon, see e.g. [3], [5]. In principle, a measurement and analysis of operational complexity should express behavioral uncertainty of the system with respect to a specified level of its control. We assume that such data are available in a company management information system (MIS), and they should trace in detail all possible types of uncertainty of information within and across companies, e.g. replenishment time disturbances, deviations of material in/out flows, etc.

2 Theoretical background

There is well-known that information theory provides instruments of quantifying complexity in general. Among others, the most popular is Shannon's information-theoretic measure and corresponding entropy. It provides quantitative measure of the expected amount of information required to describe the state of a system. Since we are going to concern with empirical observations, the probabilistic framework seems a proper one. We refer [2] for more details, too. In general, one assumes a trial which issues an event A_i belonging to given complete set of mutually disjunctive events $\{A_1,...,A_N\}$ having probabilities $p_i = \mathrm{P}(A_i)$, $i=1,...,N$, which satisfy the equation $p_1+...+p_N = 1$, because of set of events completeness. Making a large number of independent trials n, we get ratios $n_i = n(A_i)/n$ approaching p_i, i.e. $n_i \approx np_i$, $i=1,...,N$, where $n(A_i)$ denotes the number of occurrences of event A_i within such n independent trials. Having probabilities p_i one can calculate a quantity called *entropy*, sometimes entropy of information, since it is related in information theory to the most effective binary coding of each individual trial in average. The entropy is traditionally denoted I and given by following expression

$$I = -\sum_{i=1}^{N} p_i \log_2(p_i). \tag{1}$$

[1] Univ.of West Bohemia, Faculty of Econ., Dept. of Business Administation and Management., Husova 11, 30614 Pilsen, hofman@kpm.zcu.cz.
[2] Univ.of West Bohemia, Faculty of Econ., Dept. of Econ. and Quantitative Methods, Husova 11, 30614 Pilsen, lukasl@kem.zcu.cz.

It's worth to note that the quantity I, more precisely $I(p_1,...,p_N)$, depends upon particular event probability distribution. So, if the events $\{A_1,...,A_N\}$ obey an uniform distribution, i.e. having $p_i = 1/N$, for all i, we get directly

$$I(p_1,...,p_N) = -\sum_{i=1}^{N} p_i \log_2(p_i) = I_u = -\sum_{i=1}^{N} (1/N) \log_2(1/N) = \log_2(N). \quad (2)$$

It is also the greatest value, which is possible to get for a system, which states are completely described by mutually disjoint events $\{A_1,...,A_N\}$ with probabilities $(p_1,...,p_N)$. Further, due to our purposes to use entropy (1) for quantitative analysis of supplier-consumer system, we temporarily leave the traditional notation and adopt another one, which is more popular in physics, in particular H for I, and H_u for I_u, eventhough with same formulas.

3 Operational complexity of supplier-customer system

There is well known that from management science point of view a supplier-customer system belongs to broad theory of inventory control. With a supplier-consumer system basic scheme sketched on Table 1, any deviations in quantity of goods delivered, time gaps between actual supply times and ordering ones, and other quantities used to monitor various replenishment deviations are representable. We refer to [2] and [5] for details.

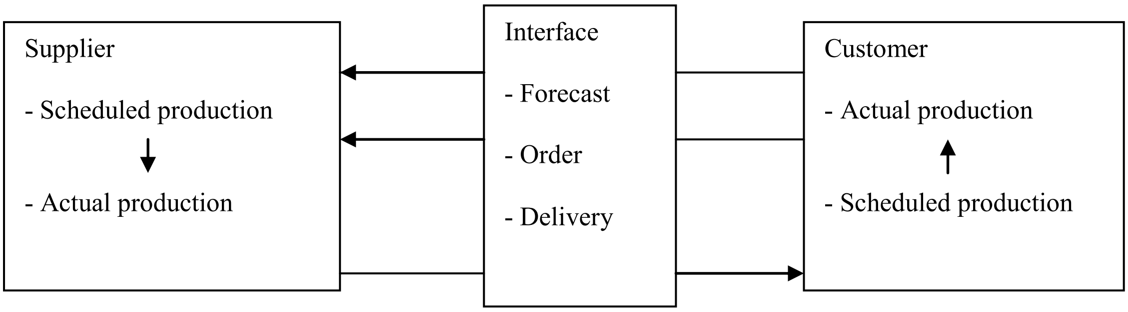


Table 1 Basic scheme of supplier-customer system

The essence of monitoring supplier-customer system containing set of products $\{P_1, ..., P_n\}$ in order to evaluate its operational complexity on the base of entropy consists of two types of variables being related to quantity and time of each product P_j, which occur at supplier, customer and/or at the interface side. In general, we are able to define a lot of variables being denoted systematically ${}_{(a,b)}Q_i$ and ${}_{(a,b)}T_i$, i=1,...,n, where a stands for a side (s – supplier, i – interface, c – customer) and b denotes a type of production, e.g. scheduled, actual, forecast, etc., following [2]. Since we are looking mainly for flow variations, we are to consider differences, e.g. (Order – Forecast), (Delivery – Order), (Actual production – Scheduled production), etc. Hence, such quantities are expressible in general form as: ${}_{(a,b)}Q_i - {}_{(u,v)}Q_i$ and ${}_{(a,b)}T_i - {}_{(u,v)}T_i$, just with case-dependent and properly selected prefix tuplets (a,b) and (u,v). However, natural and crucial question arises how to get all probabilities introduced. We know the probability estimation depends on the specific supplier-customer system investigated. So far, from theoretical point of view we conclude that all necessary probabilities will be estimated from data collected by monitoring the supplier-customer system and stored in a proper data base.

4 Problem-oriented database and case studies

Case-collected data sheets having been excerpted either by SQL processing of reports generated by MIS or manually in the simplest case, are processed first by EnComP1.java program, which serves three main purposes to – check data consistency, build up DB, and generate output files for further processing. The program EnComP2.java makes graphical outputs. However, these are optionally produced by EnComP2mma.nb, too, the program written in Mathematica (Wolfram Research, Inc.) with purposes to create additional graphical outputs, to calculate EDF (empirical distribution function), and to evaluate entropy.

The EDF is calculated from raw data $\{y_k\}$, k=1,..,K, which contains all observations available including repeating values, of a random variable Y. The Y makes a theoretical framework for any Q- or T-flow observed. The procedure has two steps: i) sort and scale $\{y_k\}$ by affine map in order to get $\{x_k\}$ of a random variable X identically distributed as Y, but having dom(X)=[0,1], and extract all the repeating values form $\{x_k\}$ in order to get strictly increasing subset $\{x_i\}$, i=1,..,N, $0 \le x_1 < x_2 < .. < x_N \le 1$, with frequencies $\{f_i\}$, i=1,..,N of repeated values, ii) calculate EDF $F(\xi) = P(\xi < x)$, $x \in \{x_i\}$, with dom(F(.)) = range(F(.)) = [0,1]. Empirical frequency function gives relative frequencies $\{p_i\}=\{f_i/K\}$, i=1,..,N. In general framework, these values are systematically used for calculation entropy H by (1), and H_u by (2).

For the paper we have selected four SME-ranked firms, called anonymously FA, FB, FC, and FD in order to keep business secrets, with different business orientations: FA – bulding engineering, FB – fashion shop, FC – mechanical engineering, and FD – transportation engineering. In a similar anonymous way we denote suppliers, e.g. S1fA and S2fA, which denote two different suppliers of company FA. We concern exclusively with time flow variations given by $T_s - T_v$, thus representing a time gap between order issue time T_v and receipt time T_s of product delivery. Such quantity is rather important in practice, and is called lead time in the inventory theory.

Case studies:

First, we start with the FA company. The corresponding data were collected from 2008 till 2010, and fetched manually into the problem-oriented database. More details are given in [1], where variations both in time and quantity are presented for company the most important products (concrete, solid bricks, masonry mortars, plasters and building blocks). However, we have concern in this paper with quantity $T_s - T_v$, only.

The Figures 1-5 present typical outputs from programs EnComP2mma.nb and mme11HJLLplot.nb, which provides graphical representation of results. Eventhough we show results of two products only (concrete and solid brick), we can see how difficult would be to make managerial decision as to preferencing one of suppliers S1fA and S2fA. On the contrary, see the Figure 5, the table of values H, H_u and $h = H/H_u$, and the plot in particular, makes it very easy – there is evident that S2fA outperforms S1fA. Of course, keeping in mind that we have analysed quantity $T_s - T_v$, only.

The second and the third company, i.e. FB – fasion shop, and FC – mechanical engineering co., represent very different companies as to their prouct lines. In both cases, we have analysed the quantity $T_s - T_v$, again, and we have investigated the role of two main suppliers, denoted S1fB, S2fB, and S1fC, S2fC, respectively. Because of the limited space, we present final results only – Figures 6 and 7, which are similar to Figure 5. The products $\{P_1, P_2, P_3\}$ of company FB are {blouses, dresses, skirts} for ladies, while at the company FC there are mechanical components for assembling {tanks, masts, exchangers}. The Figure 6 shows, that outperformace of S2fB is achieved in blouses and dresses, only. On the contrary, the Figure 7 shows monotone outperformace of the supplier S1fC in all thre products investigated.

The last company FD selected, the transportation company, is as to its size the biggest one among FA, FB, FC, and FD. It is well to be identified on the top of SME size-ranks. The company has its MIS and the corresponding data were excerpted from huge MS-Excel file. It contains more than 42 thousand of records in total.

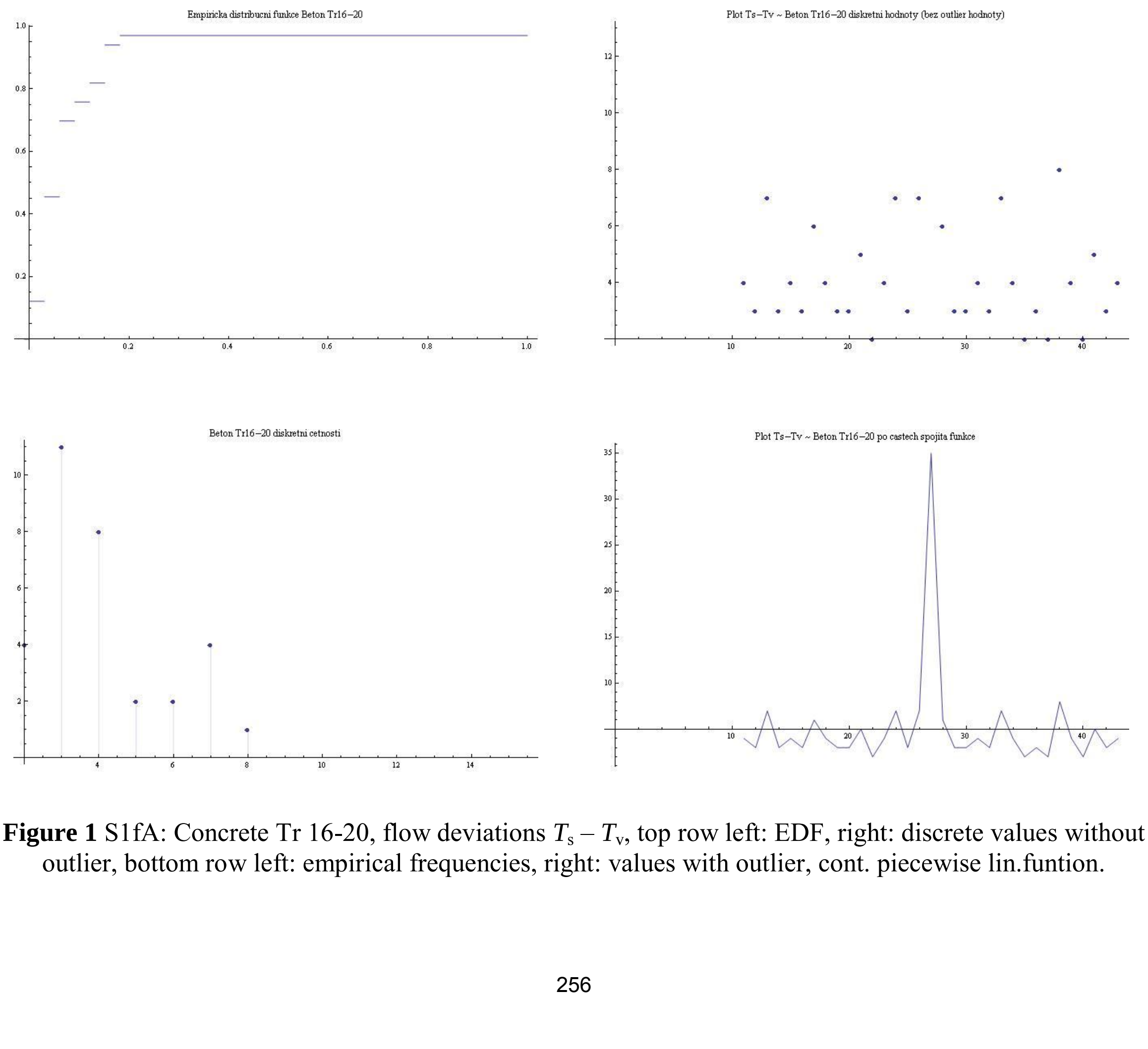


Figure 1 S1fA: Concrete Tr 16-20, flow deviations $T_s - T_v$, top row left: EDF, right: discrete values without outlier, bottom row left: empirical frequencies, right: values with outlier, cont. piecewise lin.funtion.

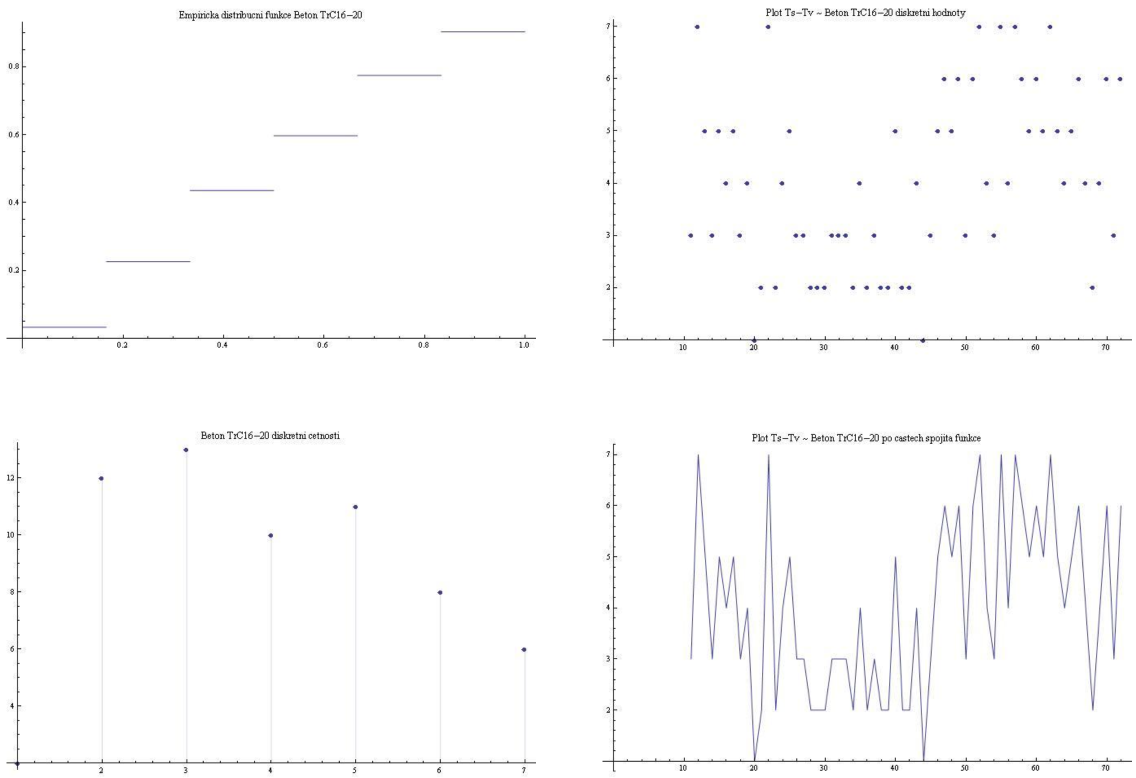


Figure 2 S2fA: Concrete TrC 16-20, flow deviations $T_s - T_v$, top row left: EDF, right: discrete values without outlier, bottom row left: empirical frequencies, right: values with outlier, cont. piecewise lin.funtion.

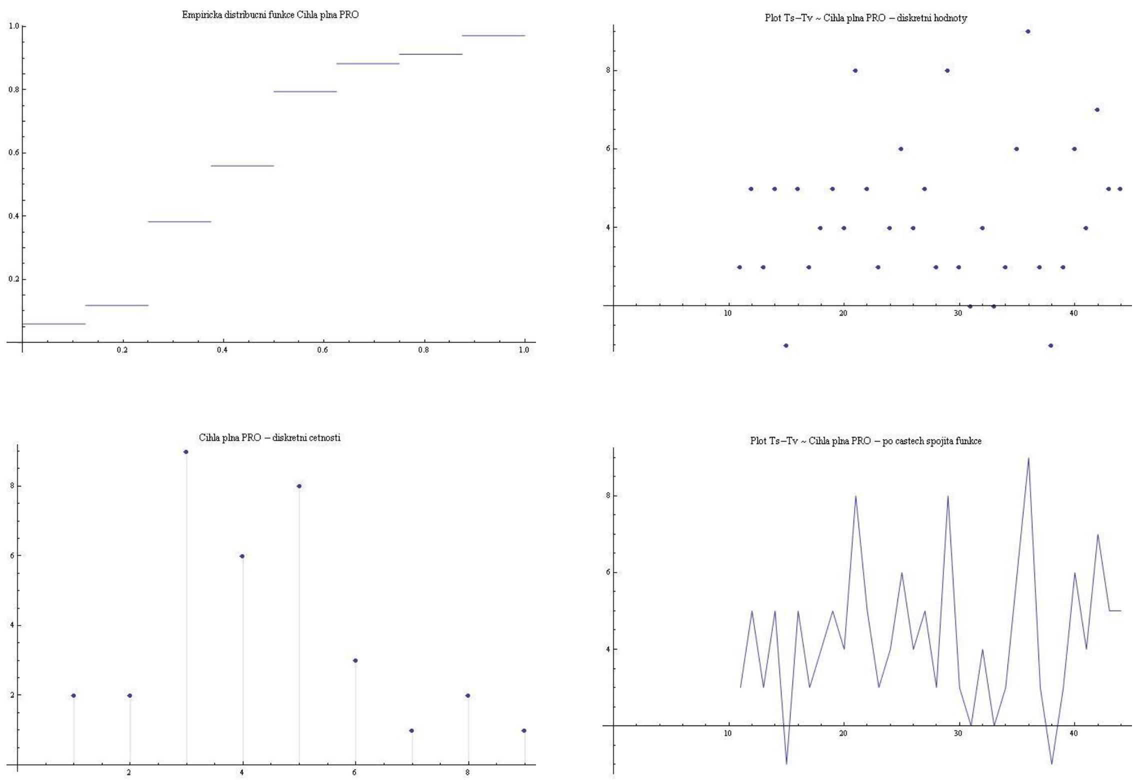


Figure 3 S1fA: Solid brick CP290x140x65, flow deviations $T_s - T_v$, top row left: EDF, right: discrete values without outlier, bottom row left: empirical frequencies, right: values with outlier, cont. piecewise lin.funtion.

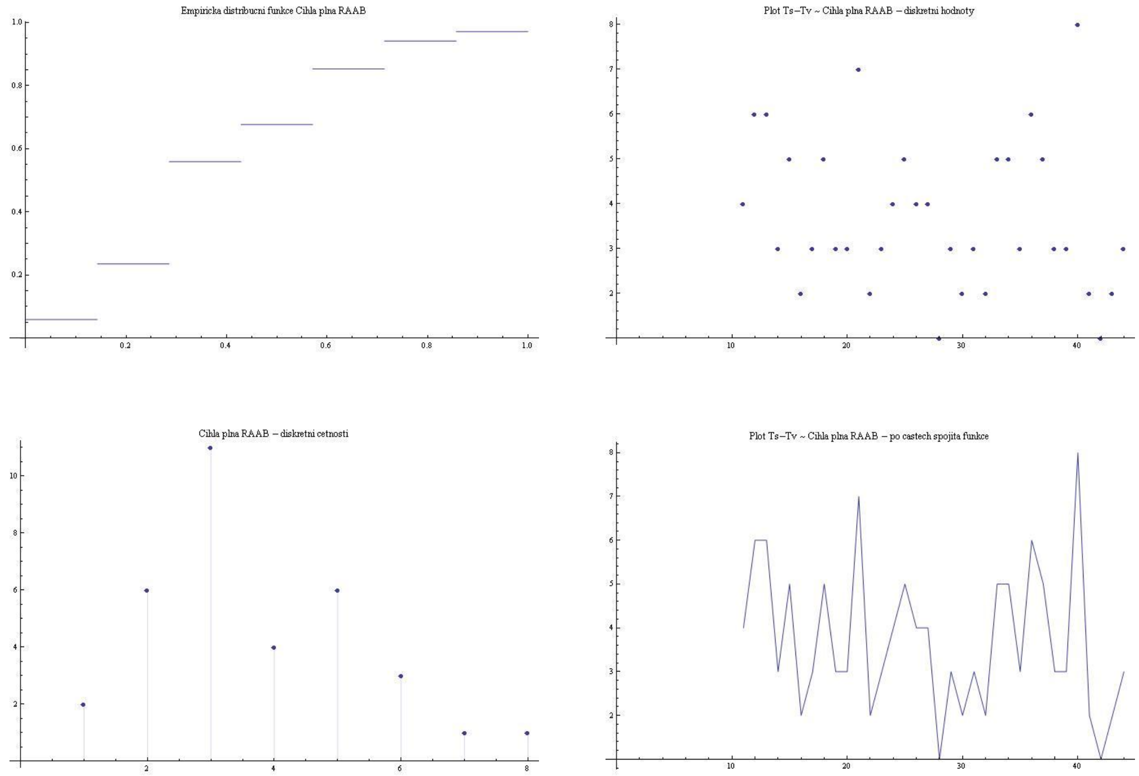


Figure 4 S2fA: Solid brick CP290x140x65, flow deviations $T_s - T_v$, top row left: EDF, right: discrete values without outlier, bottom row left: empirical frequencies, right: values with outlier, cont. piecewise lin.funtion.

Supplier: product	H	H_u	$h = H/H_u$
S1fA: Concrete	2.55792	5.04439	0.507082
S2fA: Concrete	2.66536	5.9542	0.447643
S1fA: Solid brick	2.76999	5.08746	0.544474
S2fA: Solid brick	2.62193	5.08746	0.51537
S1fA: Building block	2.8014	4.52356	0.619292
S2fA: Building block	2.93795	4.9542	0.593023

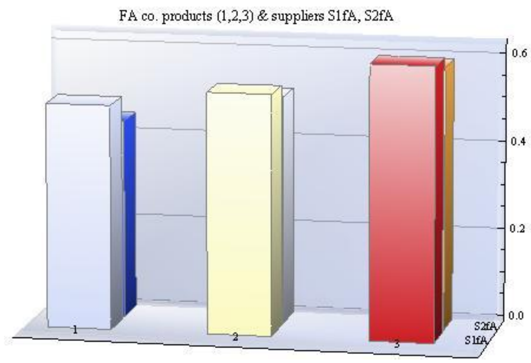


Figure 5 Left: table of calculated values H, H_u and h for S1fA and S2fA, right: h bar chart plot.

Supplier: product	H	H_u	$h = H/H_u$
S1fB: Blouses	1.96692	4.0000	0.491729
S2fB: Blouses	1.22791	6.04439	0.203148
S1fB: Dresses	1.85475	4.45943	0.415917
S2fB: Dresses	0.932112	4.52356	0.206057
S1fB: Skirts	1.22791	6.04439	0.203148
S2fB: Skirts	1.3392	6.37504	0.210069

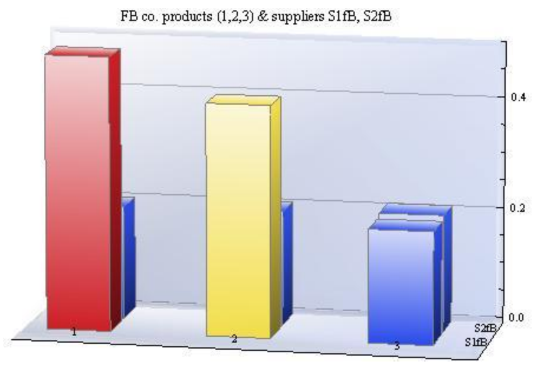


Figure 6 Left: table of calculated values H, H_u and h for S1fB and S2fB, right: h bar chart plot.

Supplier: product	H	H_u	$h = H/H_u$
S1fC: Tank G100	2.31212	7.29462	0.316962
S2fC: Tank G100	2.69223	7.29462	0.369071
S1fC: Mast LTA	3.24267	6.45943	0.502005
S2fC: Mast LTA	3.34846	6.45943	0.518383
S1fC: Exchanger P12	2.31212	7.29462	0.316962
S2fC: Exchanger P12	2.52078	7.29462	0.345568

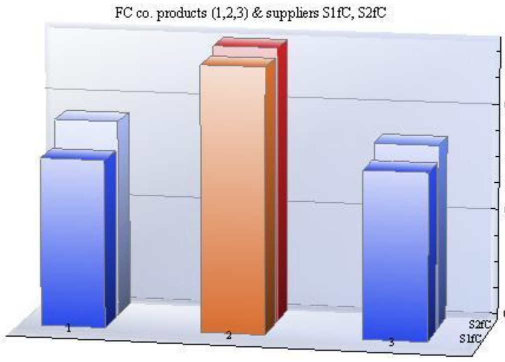


Figure 7 Left: table of calculated values H, H_u and h for S1fC and S2fC, right: h bar chart plot.

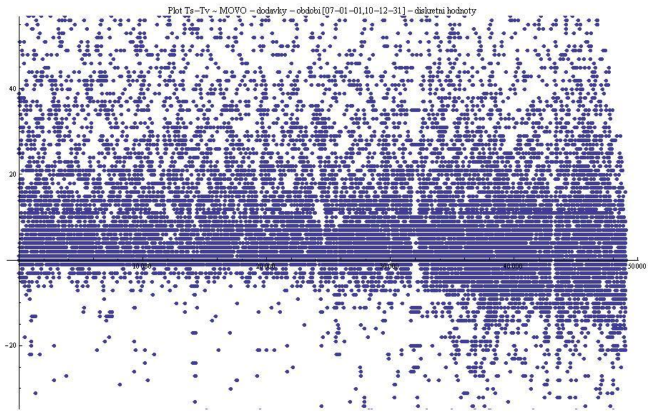


Figure 8 Fractal character of $T_s - T_v$, all deliveries from 2007-01-01 till 2010-12-31.

5 Conclusion

The measure of operational complexity based upon entropy provides versatile instrument for supplier-customer system analysis. In case studies presented, we applied unified approach for analysis of lead time variations of product-line main products at four SME-ranked firms. The results are briefly discussed, and possibilities of their direct application for managerial decision making are outlined, too. Evidently, from both managerial and theoretical point of view there is still a lot of work to do. In particular, collection and processing of data, probability estimation of all mutually disjunctive states in specific supplier-customer system considered, and last but not least an accumulation of experience with different kinds of application. The research thereon is ongoing.

Acknowledgements

The second author acknowledges funding by the project LC06075 of the Ministry of Education, Czech Republic.

References

[1] Častoralová, K.: Measurement of supplier-customer system complexity using quantitative measures in TECTON, Ltd (in czech). Plzeň: ZČU/FEK, Bachelor thesis, 2011, pp.72.

[2] Hofman, J., Lukáš, L.: Application of entropy for supplier-customer system complexity analysis - case studies. *Proc. of the 28th Int.Conf. on Mathematical Methods in Economics* 2010, Part I, Publisher Univ. of South Bohemia in České Budějovice, České Budějovice, 2010, pp. 238-243, ISBN 978-80-7394-218-2.

[3] Lau, A.H.L., Lau H-S.: Some two-echelon supply-chain games improving from deterministic-symmetric-information to stochastic-unsymmetric-information. EJOR **161** (2005), 203-223.

[4] Sivadasan, S., Efstathiou J., Calinescu A., Huaccho Huatuco L.: Advances on measuring the operational complexity of supplier-customer systems. EJOR **171** (2006), pp. 208-226.

[5] Tomek, G., Vávrová, V.: Operation, Operative or Operational Production Management (in czech). *E+M Ekonomie a Management* (4/2009), 46-53.

A comparison of smoothing filter based on fuzzy transform and Nadaraya-Watson estimators

Michal Holčapek[1], Tomáš Tichý[2]

Abstract. Data smoothing is an important step within a data processing allowing one to stress the most important patterns. In literature we can find many different smoothing techniques and filter types. Recently, Holapek and Tich (2010, 2011) suggested smoothing filters based on fuzzy transform approach introduced by Perfilieva (2004). For this purpose, a generalization of the concept of fuzzy partition was suggested and the smoothing filter was defined as a combination of the direct discrete fuzzy transform and a slightly modified inverse continuous fuzzy transform. In this paper we compare the proposed filter with the Nadaraya-Watson estimator. We provide an approximative relation of both filters by an optimal parameter selection.

Keywords: Fuzzy transform, Nonparametric regression, Nadaraya-Watson estimator, FT-smoothing filter estimator.

JEL classification: G17
AMS classification: 62G08

1 Introduction

The fuzzy transform (shortly, F-transform) is a simple approximation technique proposed by Perfilieva in [9] (see also [8]) based on fuzzy partitioning of a closed real interval into fuzzy subsets. The F-transform is introduced and investigated in a continuous and discrete design context. In both designs, a (continuous or finite) function defined on a closed interval $[a, b]$ is transformed, using fuzzy sets (basic functions) that form a fuzzy partition of $[a, b]$, to a finite number of real numbers called the components of F-transform. This type of F-transform is called the direct F-transform. An inverse F-transform assigns to F-transform components a continuous function in the continuous design and a finite function in the discrete design, which is an approximation of the original function.

In [3], we presented an application of the F-transform to the non-parametric derivation of a probability density function (PDF) from a data sample. More precisely, we introduced an FT-smoothing filter which is the combination of the discrete direct and the continuous inverse F-transform and derived an optimization of parameters of uniform partitions to obtain the best estimation of PDF with respect to the integrated square error (ISE). In comparison with the results obtained by the Parzen window estimator (for a survey, see [10]), the FT-smoothing filter provides very satisfactory estimates of PDFs. Moreover, similarly to the vector quantization based on Parzen windows, or the finite Gaussian mixture, the FT-smoothing filter decreases significantly the model complexity being proportional for Parzen windows to the number of sample data which can lead to the memory storage problem. A disadvantage of the FT-smoothing filter is an overfitting for a smaller bandwidth of basic functions which is among others caused by a limitation of the original definition of fuzzy partition. This fact motivated us to generalize the original idea of the fuzzy partition and the corresponding definition of FT-smoothing filter in [5]. Following the Stefanini's idea presented in [12] (or in the recent paper [13]) we introduced a fuzzy r-partition as a special case of a fuzzy r-cover, where r referees to the sum of membership values over all basic functions for any point of a given interval. Note that the original Perfilieva's definition of the fuzzy partition can be obtained by putting $r = 1$ and the Stefanini's definition supposing that $2r$ is the number of active basic functions.[1] The FT-smoothing filter is defined analogously to the inverse $\mathrm{F}^{(\mathrm{r})}$-transform in [12, 13].

[1]University of Ostrava/Institute for Research and Applications of Fuzzy Modeling, University of Ostrava, 30. dubna 22, 701 03 Ostrava 1, Czech Republic, e-mail: nmichal.holcapek@osu.cz

[2]Faculty of Economics VŠB-TU Ostrava/Department of Finance, Sokolská 33, 701 21 Ostrava 1, Czech Republic , e-mail: tomas.tichy@vsb.cz

[1]That means a number of basic functions giving a non zero value in the evaluation of the inverse F-transform.

In this contribution, we shall present an interesting approximative relations between the FT-smoothing filter as an estimator and the Nadaraya-Watson (NW) estimator. This relation is derived from the theoretical results on the asymptotic behavior of both estimators, namely, on the comparison of the optimal bandwidth of kernel and basal fuzzy set with respect to the asymptotic mean square error (AMSE).

The paper is structured as follows. After Introduction, we briefly summarize the important notions about the fuzzy r-partition. The third section is devoted to the definitions of NW and FT-smoothing filter estimators. Furthermore, the approximative relation between both estimators is provided. The last section is a practical comparison of both estimators assuming financial data.

2 Fuzzy r-partitions

We shall use $\mathbb{R}$ to denote the set of all natural, integer and real numbers, respectively. A *fuzzy set* on $\mathbb{R}$ is a function $A : \mathbb{R} \to [0, 1]$. We shall say that a fuzzy set A is *empty*, if $A(x) = 0$ holds for any $x \in \mathbb{R}$. Further, we shall use $\mathrm{Ker}(A)$ to denote the set of all $x \in \mathbb{R}$ for which $A(x) = 1$. The set $\mathrm{Ker}(A)$ is called the *kernel* of A. We shall say that a fuzzy set is *convex*, if $A(\lambda x + (1 - \lambda)y) \geq \min(A(x), A(y))$ for any $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$, *continuous*, if $A(x)$ is a continuous function in the common sense, and *normal*, if $\mathrm{Ker}(A) = \{x\}$ for a suitable $x \in \mathbb{R}$. In this paper, we shall suppose that each fuzzy set is continuous, convex and has the non-empty kernel.[2] Note that the choice of the fuzzy sets shapes is motivated by the shapes of fuzzy sets used in the original definition of fuzzy partition [9] and this choice seems to be natural.

Definition 1 (Fuzzy r-cover)**.** Let $r \geq 1$ be a real number. A *fuzzy r-cover of* $[a, b]$ is a collection $\mathcal{A} = \{A_i \mid i = 1, \ldots, n\}$ of non-empty fuzzy sets that satisfies $\sum_{i=1}^{n} A_i(x) \geq r$ for any $x \in [a, b]$.

Using the concept of fuzzy r-cover we can naturally generalize the concept of fuzzy partition of an interval $[a, b]$. We shall use $A \upharpoonright [a, b]$ to denote the restriction of a function A to $[a, b]$.

Definition 2 (Fuzzy r-partition)**.** Let $r \geq 1$ be a real value. A collection $\mathcal{A} = \{A_i \mid i = 1, \ldots, k\}$ of non-empty fuzzy sets is called a *fuzzy r-partition of* $[a, b]$, if there exists a fuzzy r-cover $\mathcal{B} = \{B_i \mid i = 1, \ldots, k\}$ of $[a, b]$ such that

1. $A_i = B_i \upharpoonright [a, b]$ for any $i = 1, \ldots, n$,
2. $\sum_{i=1}^{n} A_i(x) = r$ for any $x \in [a, b]$.

The fuzzy sets of $\mathcal{A}$ are called *basic functions*.

A *basal fuzzy set* is a continuous, convex, normal fuzzy set $S : \mathbb{R} \to [0, 1]$ such that $S(x) = S(-x)$ for any $x \in \mathbb{R}$ and

$$\int_{-\infty}^{\infty} x^2 S(x) dx < \infty. \tag{1}$$

Note that this equality enables us to significantly simplify the description of e.g. the Bias, Var, MSE, or AMSE, that will be investigated in the subsection devoted to the statistical analysis of the FT-smoothing filter. Let S be a basal fuzzy set and $h > 0$ be a real number. On can check easily that also $S_h(x) = S(x/h)$ is a basal fuzzy set. The value h is called the *bandwidth*.

Definition 3. Let $\mathcal{T} = \{t_i \mid i = 1, \ldots, n\}$ be an ordered set of nodes from $\mathbb{R}$ with $t_i < t_{i+1}$ for any $i = 1, \ldots, n$ and $\mathcal{S} = \{S^{(i)} \mid i = 1, \ldots, n\}$ be a set of basal fuzzy sets. A *fuzzy r-partition of* $[a, b]$ *determined by* $(\mathcal{T}, \mathcal{S})$ is a fuzzy r-partition $\mathcal{A} = \{A_i \mid i = 1, \ldots, n\}$ of $[a, b]$ such that

$$A_i(x) = S^{(i)}(x - t_i) \tag{2}$$

holds for any $x \in [a, b]$ and $i = 1, \ldots, n$. We shall say that a fuzzy r-partition of $[a, b]$ determined by $(\mathcal{T}, \mathcal{S})$ is *uniform*, if $S = S^{(i)}$ for any $i = 1, \ldots, n$ and $t_{i+1} - t_i = u$ for any consecutive nodes t_{i+1} and t_i in $\mathcal{T}$.

[2]Note that if $\mathrm{Supp}(A) = \mathrm{cl}\{x \in \mathbb{R} \mid A(x) > 0\}$, where cl is the closure operator, then the considered fuzzy sets are fuzzy intervals or, accepting the assumption of the normality, fuzzy numbers (see [6, 1]).

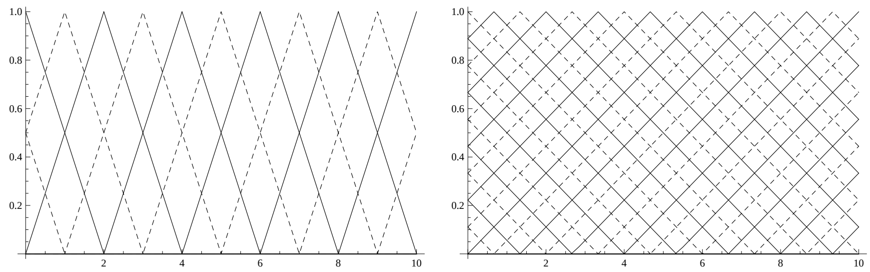


Figure 1: Triangle uniform fuzzy 2-partition for $h = 2$ (left) and 9-partition for $h = 6$ (right).

Example 1 (Triangle fuzzy r-partition)**.** The triangle basal fuzzy set is given by

$$S_h(x) = \max\left(0, \frac{h - |x|}{h}\right). \tag{3}$$

On Fig. 1, a uniform fuzzy 2-partition of $[0, 10]$ with the bandwidth $h = 2$ and a uniform fuzzy 9-partition of $[0, 10]$ with the bandwidth $h = 6$. One can note that both fuzzy r-partitions are obtained as a combination of two fuzzy r-partitions (pictured by the normal and dashed lines) with lower values of r.

3 Theoretical comparison of the NW and FT-smoothing filter estimators

For our comparison using asymptotic behavior of both estimators, we shall consider a fixed design. Recall that a *fixed design* consists of $x_1, \ldots, x_n$ which are ordered non-random numbers. We shall say that a fixed design is *equally spaced*, if $x_{j+1} - x_j$ is constant for all j. Sometimes, a fixed equally spaced design has a form $x_j = \frac{j}{n}$ for $j = 0, \ldots, n$. For the fixed design the response variables are assumed to satisfy

$$Y_j = g(x_j) + \varepsilon_j, \quad j = 1, \ldots, n, \tag{4}$$

where g is a (non-random) function and ε_j is a random variable representing the error in x_j having $\mathrm{E}(\varepsilon_j) = 0$, $\mathrm{Var}(\varepsilon_j) < \infty$ and $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$, respectively. In our investigation, we restrict ourselves to the common choice $\mathrm{Var}(\varepsilon_j) = \sigma^2$ for any $j = 1, \ldots, n$ (a homoscedastic model).

3.1 Nadaraya-Watson estimator

Let us recall that the aim of the kernel-based nonparametric regression is to estimate the unknown function g. The Nadaraya-Watson (NW) estimator of g at a point x is defined by

$$\hat{g}_{\mathrm{NW}}(x) = \frac{\sum_{j=1}^{n} Y_j K(x_j - x)}{\sum_{j=1}^{n} K(x_j - x)}, \tag{5}$$

where Y_j is expressed by (4) and K is a kernel (see e.g. [7, 10, 11, 14]). Recall that a *kernel* is a continuous, symmetric, unimodal function $K : \mathbb{R} \to [0, \infty)$ having the following properties:

$$\int_{-\infty}^{\infty} K(x)dx = 1 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2 K(x)dx < \infty. \tag{6}$$

Obviously, each kernel is a symmetric density function with a single mode. Typical examples of kernels are the Uniform, Triangular, Cosine, Epanechnikov or Gaussian density functions (see e.g. [2, 11]). Let $h > 0$ be a real number and K be a kernel, then $K_h(x) = 1/hK(x/h)$ is again a kernel, where h is the bandwidth of K (see [14]).

The optimal bandwidth of a kernel K with respect to the asymptotic mean square error (AMSE) is

$$h_{\mathrm{AMSE}}^{\mathrm{NW}} = \left(\frac{\sigma^2(b-a)}{ng''(x)} C(K)\right)^{\frac{1}{5}}, \tag{7}$$

where $C(K) = R(K)/\mu_2(K)^2$ with $\mu_2(K) = \int_{-1}^{1} z^2 K(z)dz$ and $R(K) = \int_{-1}^{1} K^2(z)dz$ may be understood as a characterization of the kernel K (see [2, 14]).

3.2 FT-smoothing filter estimator

Let $\mathcal{A} = \{A_i \mid i = 1, \dots, n\}$ be a fuzzy r-partition of $[a, b]$ determined by $(\mathcal{T}, \mathcal{S})$. We shall say that a set $X = \{x_j \mid j = 1, \dots, n\}$ of reals is *sufficiently dense with respect to* $\mathcal{A}$, if for each $A_i \in \mathcal{A}$ there exists at least one node $x_j \in X$ such that $A_i(x_j) > 0$.

Definition 4. Let $(x_1, Y_1), \dots, (x_n, Y_n)$ define a finite random function, where $Y_1 \dots, Y_n$ are given by (4) and $X = \{x_1, \dots, x_n\} \subseteq [a, b]$, and $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ be a uniform fuzzy r-partition of $[a, b]$ determined by $(\mathcal{T}, S)$ such that X is sufficiently dense with respect to $\mathcal{A}$. We shall say that a collection of random variables $\{\Phi_i \mid i = 1, \dots, k\}$ is the *direct discrete stochastic F-transform of the finite random function defined by* $(x_1, Y_1), \dots, (x_n, Y_n)$ *with respect to* $\mathcal{A}$, if

$$\Phi_i = \frac{\sum_{j=1}^{n} Y_j S(x_j - t_i)}{\sum_{j=1}^{n} S(x_j - t_i)}. \tag{8}$$

Definition 5. Let $x_1 < \cdots < x_n$ be a sequence of nodes of $[a, b]$, $\mathcal{A}$ be a fuzzy r-partition of $[a, b]$ determined by $(\mathcal{T}, \mathcal{S})$, $X = \{x_1, \dots, x_n\}$ be sufficiently dense with respect to $\mathcal{A}$ and $Y_j = g(x_j) + \varepsilon_j$ for $j = 1, \dots, n$. The FT-smoothing filter estimator of the function g at the point x is given by

$$\hat{g}_{\mathrm{FT}}(x) = \frac{1}{r} \sum_{i=1}^{k} \Phi_i A_i(x), \tag{9}$$

where Φ_i are the stochastic F-transform components of the random function given by $(x_1, Y_1), \dots, (x_n, Y_n)$.

For our investigation, the following relation between basal fuzzy sets and kernels is fundamental.

Proposition 1. *There is a one-to-one correspondence between the classes of all basal fuzzy sets and all kernels.*

Comparing the NW estimator with the formula in (8) for the stochastic F-transform components, one can see the similarity. More precisely, we may state the following proposition.

Proposition 2. *Let* $\mathcal{A} = \{A_i \mid i = 1, \dots, k\}$ *be a uniform fuzzy r-partition of* $[a, b]$ *determined by* $(\mathcal{T}, S)$. *If* $Y_1, \dots, Y_n$ *are random variables defined by (4) and* $\hat{g}_{\mathrm{NW}}$ *is the NW estimator with*

$$K(x) = \frac{S(x)}{\int_{-\infty}^{+\infty} S(x)dx}, \tag{10}$$

then $\Phi_i = \hat{g}_{\mathrm{NW}}(t_i)$ *for any* $t_i \in \mathcal{T}$.

Using the above propositions, we proved in [5] (see also [4]) the asymptotic bias and variance of $\hat{g}_{\mathrm{FT}}(x)$. A consequence of these expressions is the optimal bandwidth of a basal fuzzy set S with respect to the asymptotic mean square error

$$h_{\mathrm{AMSE}}^{\mathrm{FT}} = \left(\frac{k^2 u^2 \sigma^2}{4ng''(x)^2 (b-a)} C(K) \right)^{\frac{1}{5}}$$

where K is related to S by (10), $C(K) = R(K)/\mu_2(K)^2$, $u = t_{i+1} - t_i$ and k denotes the number of basic functions obtained by the kernel K. Since $t_i \notin [a, b]$ for some $i = 1, \dots, k$, we can simply deduce $ku \geq b - a$. One can note that $ku \approx b - a$ for small h and

$$h_{\mathrm{AMSE}}^{\mathrm{FT}} \approx \left(\frac{(b-a)\sigma^2}{4ng''(x)^2} C(K) \right)^{\frac{1}{5}}.$$

Comparing this result with that in (7) provided by the NW estimator, we obtain an approximated equality

$$h_{\mathrm{AMSE}}^{\mathrm{FT}} \approx 0.76 h_{\mathrm{AMSE}}^{\mathrm{NW}}, \tag{11}$$

where $0,76 \approx \sqrt[5]{1/4}$. Thus, a lower asymptotic bandwidth for the FT-smoothing filter estimator is needed to obtained an optimal model of an unknown function which would correspond to the NW model.

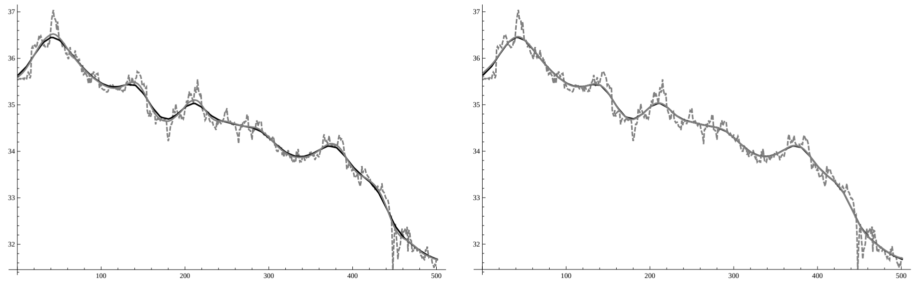


Figure 2: Comparison of the resulted functions for $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left) and $h^{\mathrm{NW}} = 1/0.76h^{\mathrm{FT}}$ (right) with $h^{\mathrm{FT}} = 20$ and $r = 2$.

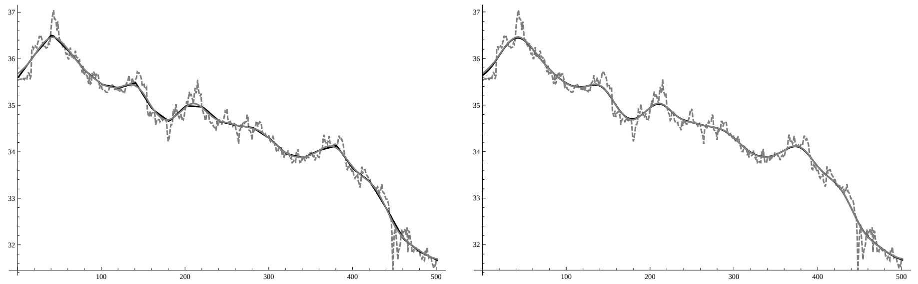


Figure 3: Comparison of the smoothness of resulted functions for $h^{\mathrm{NW}} = 1/0.76h^{\mathrm{FT}}$ with $h^{\mathrm{FT}} = 20$ and $r = 1$ (left) and $r = 8$ (right)

4 Practical comparison of the NW and FT-smoothing filter estimators

For a simple comparison, we chose data sets consisting of 500 and 75 daily quotations of CZK/EUR exchange rate randomly selected from the original time series covering the last eight years. On Fig. 2, one can see a comparison of the resulted functions obtained by the NW estimator (grey line) and by the FT-smoothing filter estimator (black line), when we use $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left) and $1/0.76h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (right) with $h^{\mathrm{FT}} = 20$ and $r = 2$. Although, a small number of h is not here supposed, the correction of h^{FT} against h^{NW} works well. On Fig. 3, we demonstrate the smoothing property which is clearly dependent on the level of r for the fixed values $h^{\mathrm{FT}} = 20$ and $h^{\mathrm{NW}} = 20/0.76$. The both approaches give us practically identical resulted functions. Note that 26 components are used and only 2 basic functions are active (giving a non-zero value) for $r = 1$ to find the values of the smoothed function. For $r = 8$, the number of components and basic functions used for the calculation is naturally much greater, namely, 215 components and at most 16 basic functions. On the other hand, this is still less than the number of values over which is the kernel active, namely, 40 values are used to evaluate the function values using NW estimator. On Fig. 4, we present in the first column further smoothed functions by the NW estimator (grey line) and the FT-smoothing filter (black line) assuming the identical bandwidth $h^{\mathrm{FT}} = 4, 8$ with $r = h^{\mathrm{FT}}$. The results assuming the correction on the bandwidth for the NW estimator introduced in (11) are then presented in the second column. Again both approaches provide similar behavior with respect to the same bandwidth (the FT-smoothing filter estimator gives more smoothed functions) and become nearly identical after the proposed correction. Moreover, the number of active values for the kernel is still greater than the number of basic functions used for the evaluation of the FT-smoothing filter estimator. Summarizing our observation, the FT-smoothing filter can be advantageously used in cases when larger numbers of data are considered to reduce the model complexity of the NW estimator, but to retain the quality of estimates.

Acknowledgements

The support provided by the Institutional Research Plan MSM 6198898701 and VSB-TU Ostrava under the SGS project SP2011/7 is kindly announced.

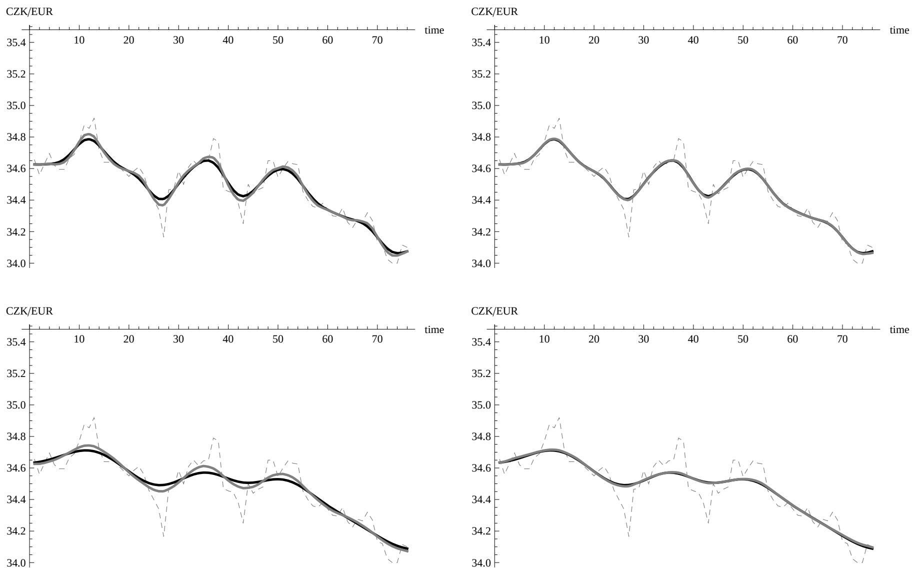


Figure 4: Smoothed functions for financial data consisting of 75 daily quotations of CZK/EUR exchange rate - $h^{\mathrm{NW}} = h^{\mathrm{FT}}$ (left column) and $h^{\mathrm{NW}} = 1/0.76 h^{\mathrm{FT}}$ (right column), $h^{\mathrm{FT}} = 4, 8$

References

[1] D. Dubois and H. Prade. Operations on fuzzy numbers. *Int. J. Syst. Sci.*, 9:613–626, 1978.

[2] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications.* Chapman&Hall/CRC Monographs on Statistics & Applied Probability, 1996.

[3] M. Holčapek and T. Tichý. A probability density function estimation using F-transform. *Kybernetika*, 46(3):447– 458, 2010.

[4] M. Holčapek and T. Tichý. Smoothing of financial data by a filter based on fuzzy transform. In *Proceedings of MME 2010*, pages 244–249, České Budějovice, Czech Republic, 2010.

[5] M. Holčapek and T. Tichý. A Smoothing Filter Based on Fuzzy Transform. *to appear in Fuzzy sets syst.*

[6] G.J. Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, New Jersey, 1995.

[7] A. Pagan and A. Ullah. *Nonparametric Econometrics.* Cambridge University Press, New York, 1999.

[8] I. Perfilieva. Fuzzy transforms. Peters, James F. (ed.) et al., Transactions on Rough Sets II. Rough sets and fuzzy sets. Berlin: Springer. Lecture Notes in Computer Science 3135. Journal Subline, 63-81 (2004), 2004.

[9] I. Perfilieva. Fuzzy transforms: Theory and applications. *Fuzzy sets syst.*, 157(8):993–1023, 2006.

[10] B. W. Silverman. *Density estimation for statistics and data analysis.* Chapman & Hall/CRC, Lodon, 1986.

[11] J. S. Simonoff. *Smoothing methods in statistics.* Springer-Verlag, New York, 1996.

[12] L. Stefanini. Fuzzy transform and smooth function. In *Proceedings of IFSA/EUSFLAT 2009*, pages 579–584, Lisbon, Portugal, 2009.

[13] L. Stefanini. F-transform with parametric generalized fuzzy partitions. *Fuzzy sets syst.*, doi:10.1016/j.fss.2011.01.017

[14] M.P. Wand and M.C. Jones. *Kernel Smoothing.* Chapman&Hall/CRC Monographs on Statistics & Applied Probability, London, 1995.

Fuzzy classification systems and their applications

Pavel Holeček[1], Jana Talašová[2], Jan Stoklasa[3]

Abstract. In both science and practice, it is common to work with classes of objects, defined by verbally specified values of the object's characteristics. These linguistic expressions can be interpreted as values of linguistic variables representing the characteristics of interest. The classes may be denoted by numbers, usually integers. A fuzzy classification system can then be described by means of a rule base in the following way. On the left-hand side of each rule, there is a combination of values of linguistic variables that defines a particular class. The integer-valued variable, which identifies the same class, is on the right-hand side. For any object described by values of its characteristics, its assignment to a class is determined by correspondence between these values and the left-hand sides of the rules. The output of the fuzzy classification system depends on whether the numeric class identifiers form a nominal, ordinal, or even cardinal scale. In case of a nominal scale, there is no sense in applying any arithmetic operations on class identifiers. The result is processed into information whether the object can be clearly classified or not. If the numeric class identifiers form a uniform cardinal evaluating scale, it is meaningful to perform calculations whose result can be the location of an object in between two classes. In both these cases, a verbal description of the result is desirable. The solution to this problem will be illustrated on two examples of academic staff performance evaluation.

Keywords: fuzzy classification, fuzzy rule base, academic staff evaluation.

JEL Classification: C44
AMS Classification: 91B74

1 Introduction

In the real world, the following problem is commonplace. A set of classes, defined by verbally specified values of characteristics, is given. Moreover, an object described by values of these characteristics is available. We want to determine which of these classes the object belongs to. Because the verbal expressions used for description of the classes can be interpreted as values of linguistic variables, it is natural to describe the classes by means of a fuzzy rule base. The fuzzy rule base can be either derived from data or it can be set expertly.

Many papers were written on the topic of fuzzy classification. The book [4] gives a broad overview. Various methods and their performance are compared e.g. in [2] and [3]. However, the vast majority of authors focus mainly on deriving fuzzy rules from data (e.g. in [5], [6], [7], and [13]), rather than on the classification itself. In this paper, we will restrain ourselves to situations where a fuzzy rule base is already known. Also, we will discuss various cases in connection to the nature of the problem, and suitable methods for choosing the most appropriate class for the object. The theory will be accompanied by real-world examples.

Two applications of fuzzy classification will be shown, both stemming from the academic area. In the first one, academic staff members will be classified according to their specialization. Three classes will be used – *Researcher*, *Teacher* and *Nonspecific*. The result of this classification can be used in HR management, so that the academic staff members could dedicate their time and effort in that area which appeals to them the most. In the second example, the academic staff members will be divided into classes according to their overall performance [10]. The overall performance is calculated from the performance of an academic staff member in the area of pedagogical activities and R&D. For solving both of the problems the FuzzME software (http://fuzzme.wz.cz) was used. The evaluation model that contains both of these fuzzy classifications is currently being applied at the Palacky University in Olomouc.

[1] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, pavel.holecek@upol.cz.
[2] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz.
[3] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jan.stoklasa@upol.cz

2 Preliminaries

Fundamentals of the fuzzy set theory (introduced in [11]) are described in detail, e.g., in [1]. A fuzzy set A on a universal set X is characterized by its membership function $A : X \to [0,1]$.

A linguistic variable (see [12]) is defined as a quintuple (X, $T(X)$, U, G, M), where X is a name of the variable, $T(X)$ is a set of its linguistic values (linguistic terms), U is an universe on which the mathematical meanings of the linguistic terms are modeled, G is a syntactic rule for generating linguistic terms from $T(X)$, and M is a semantic rule which to every linguistic term $\mathcal{A} \in T(X)$ assigns its meaning $A = M(\mathcal{A})$ which is a fuzzy set on U. In this chapter, the linguistic term $\mathcal{A}$ will be distinguished from its mathematical meaning A, which is a fuzzy set, by means of a different font. In real-life applications, the universe U is usually a closed interval of real numbers, i.e. $U = [a,b]$, and meanings of linguistic terms are fuzzy numbers on U.

A linguistic scale (see [9]) is a special case of a linguistic variable. A linguistic scale offers simplified description of a continuous real variable with values on $[a,b]$ by specifying a finite number of linguistic values modeled by fuzzy numbers on $[a,b]$. We say that a linguistic variable (X, $T(X)$, $[a,b]$, M, G), where $T(X) = \{\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_s\}$, is a linguistic scale on $[a,b]$ if the fuzzy numbers $T_1, T_2, ..., T_s$, representing meanings of its linguistic values, form a fuzzy scale on $[a,b]$. A fuzzy scale on $[a,b]$ is defined as a set of fuzzy numbers $T_1, T_2, ..., T_s$ on this interval that form a fuzzy partition on the interval, i.e., for all $x \in [a,b]$ it holds that $\sum_{i=1}^{s} T_i(x) = 1$, and the T's are indexed according to their ordering.

3 Fuzzy classification

In the following text, the classes will be described by fuzzy rules. On the left-hand side of a rule, there are linguistic variables together with their linguistic values specifying the class of interest. On the right-hand side, there is an integer variable whose value denotes the number of the class. It is possible to describe one class by multiple rules. Under these circumstances, we will be solving the problem of classification of an object described by crisp values of its characteristics, which are on the left-hand sides of the rules.

Let C = {1,…,k}, k $\in N$, be a set of numeric identifiers of the classes of interest. A fuzzy classification system can then be described by means of a fuzzy rule base in the following form:

If F_1 is $\mathcal{A}_{11}$ and ... and F_m is $\mathcal{A}_{1m}$, then class D_1.
If F_1 is $\mathcal{A}_{21}$ and ... and F_m is $\mathcal{A}_{2m}$, then class D_2.
....
If F_1 is $\mathcal{A}_{n1}$ and ... and F_m is $\mathcal{A}_{nm}$, then class D_n,

where for i = 1,…, n, j = 1,…, m,

- $\left(F_j, T(F_j), \left[p_j, q_j\right], M_j, G_j\right)$ are linguistic scales for the individual features,
- $\mathcal{A}_{ij} \in T(F_j)$ are their linguistic values and $A_{ij} = M_j(\mathcal{A}_{ij})$ are fuzzy numbers representing their meanings,
- $D_i \in \{1, ..., k\}$ are the class identifiers.

In the general case, we will assume that the integers on the right-hand sides of the rules form only a nominal scale. Although the above-mentioned fuzzy rule base is similar to the one used in the Sugeno fuzzy controller [1], in the general case the Sugeno fuzzy inference algorithm cannot be used for fuzzy classification. In the next chapter, fuzzy inference algorithms convenient for this case will be described. The result of classification is the class which is the most appropriate for the object, or alternatively, the information that the object cannot be classified unambiguously. This type of fuzzy classification will be illustrated on the example of determining the type of an academic staff member.

If the classification is basically an evaluation, it makes sense to assume that the integers, which identify the classes, form an ordinal, or even a cardinal scale. It is meaningful to consider the object as lying between two neighboring classes. Moreover, in the case of a cardinal evaluation scale, it is also meaningful to calculate the location of the objects in relation to these classes and the Sugeno inference approach [1] can be used. The advantage of using the tools of linguistic fuzzy modeling in both mentioned cases is that the results of classification can be interpreted verbally in a natural way.

4 Applications in the area of academic staff performance evaluation

The two types of fuzzy classification mentioned in the previous section will be illustrated by examples originating from the system of academic staff performance evaluation that is applied at the Palacký University in Olomouc, Czech Republic.

The data on working activities and achieved results of academic staff members are gathered annually into an information system. Particular activities have their scores assigned. The scores reflect the importance and/or the time demands of the activities. Academic staff members are evaluated according to their total scores in two main areas – pedagogical activities, and research and development. Their working position is taken into consideration, too. Due to the used methodology of R&D output evaluation (methodology currently valid in the Czech Republic), the evaluating scales for both these areas differ. This was taken into consideration when describing the evaluations by linguistic scales [10].

In the first example, the evaluations of performance in both areas will be used to classify academic staff members according to their specialization (*Teacher*, *Researcher*, *Nonspecific*). In the second example, the overall evaluation of academic staff members, which makes use of linguistically defined classes, will be described.

4.1 Determining the type of an academic staff member

In this application, academic staff members are classified according to their specialization into one of three types – *Teacher*, *Researcher,* and *Nonspecific*. The knowledge of the type of an academic staff member can be used in human resource management at the university. If an academic staff member performs significantly well in one area and has not-so-good results in the other, then their supervisor can allow them more space to focus on that area of activities for which they are better suited. To calculate the overall evaluation, a different rule base can be used than in the case of a *Nonspecific* academic staff member.

The classification of academic staff members is based on their evaluation in the area of pedagogical activities and R&D. The designed fuzzy rule base is shown in Figure 1.

Academic staff member type		Research and Development Performance				
		Very low	Low	Standard	High	Extreme
Pedagogical activities performance	Very low	Nonspecific	Nonspecific	Nonspecific	Researcher	Researcher
	Low	Nonspecific	Nonspecific	Nonspecific	Researcher	Researcher
	Standard	Nonspecific	Nonspecific	Nonspecific	Nonspecific	Researcher
	High	Teacher	Teacher	Nonspecific	Nonspecific	Nonspecific
	Extreme	Teacher	Teacher	Teacher	Nonspecific	Nonspecific

Figure 1. Fuzzy rule base used for determining the type

Two classification algorithms were tested – *Single Winner* and *Voting by Multiple Fuzzy Rules* [3].

Single Winner

In the S*ingle Winner Method* [3], the classification of an object is done as follows. Let us suppose that an object is described by values of its characteristics, i.e. by the real numbers $a_1,\dots, a_m$. First, the degrees of correspondence h_i, i=1,…,n, between the inputs and the i-th rule are calculated:

$$h_i = A_{i1}(a_1)\cdot A_{i2}(a_2)\cdot\ldots\cdot A_{im}(a_m),\ i=1,\ldots,n.$$

Next, the number of votes is calculated for each class as follows:

$$v_T = \max_{i=1,\ldots,n}\{h_i \mid D_i = T\}, T=1,\ldots,k.$$

The resulting class T^* for a given object is the one with the maximum value of v_T, i.e. the one for which it holds that

$$v_{T^*} = \max_{T=1,\ldots,k}\{v_T\}.$$

Voting by Multiple Fuzzy Rules

In case of *Voting by Multiple Fuzzy Rules* [3], the degrees of correspondence h_i are calculated in the same way as with the Single Winner:

$$h_i = A_{i1}(a_1)\cdot A_{i2}(a_2)\cdot \ldots \cdot A_{im}(a_m),\ i = 1,\ldots,n$$

Next, the number of votes is calculated for each class as the sum of degrees of correspondence pertaining to those rules that voted for this class:

$$v_T = \sum_{\substack{i=1,\\ D_i = T}}^{n} h_i, T = 1,\ldots,k$$

The resulting class T^* is again the one with the maximum value of v_T. Figure 2 compares the results obtained by both methods. Each dot represents an object. Its position is determined by the evaluation in the area of pedagogical activities (x axis) and R&D (y axis). The resulting classes are differentiated by the color of dots. It can be seen that the border between classes is smoother for *Voting by Multiple Fuzzy Rules*.

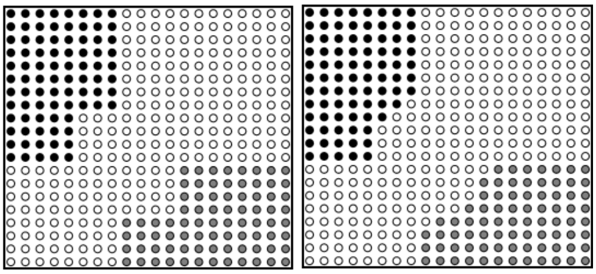


Figure 2. Results obtained by *Single Winner* (left) and *Voting by Multiple Fuzzy Rules* (right)

Objects that cannot be classified

In some cases an object cannot be classified unambiguously. First, as the classes of interest generally need not cover the whole input space, the object need not be a member of any of these classes. Second, the object can belong to several classes with similar membership degrees.

In the above example, academic staff members were classified into one of three classes that were covering the whole space. The optimum class was determined by the largest number of votes; if there were more such classes, it was possible to select any of them. So far, we have not studied how reliable the assignment of the class to an academic worker was. This will be discussed in the following text and a method for identifying unclassifiable academic staff members will be proposed.

The boundary between classifiable and unclassifiable objects can be set by means of two parameters:

1. **Minimal required support for the winner class** – if the number of votes for the winner class T^* is lower than a specified value, the object does not seem to belong to any of the classes. Therefore, it should be marked as unclassifiable.
2. **Minimal required distinctiveness of the winner** – if this condition is not satisfied, it means that the classification is ambiguous and the object belongs to more than one class in similar degrees. The distinctiveness of the winner CF is defined as a real number on [0, 1]. Let us assume that the number of votes for the winner class V is greater than zero. Then CF can be calculated in the following way:

   $$CF = 1 - (V'/V),$$

   where $V = \max_{i=1,\ldots,s}\{v_i\}$ is the number of votes for the winning class T^* and $V' = \max_{\substack{i=1,\ldots,s\\ i\neq C^*}}\{v_i\}$ is the number of votes for the second best fitting class.

Figure 3 shows the results for various values of the minimal required distinctiveness CF_{min}. The unclassifiable objects are shown as crosses. For example, let us assume three academic staff members with different evaluations in the area of pedagogical activities and R&D. The results of the fuzzy classification with the *Single Winner* algorithm are summarized in the Table 1. The *Voting by Multiple Fuzzy Rules* algorithm would give similar results in this case.

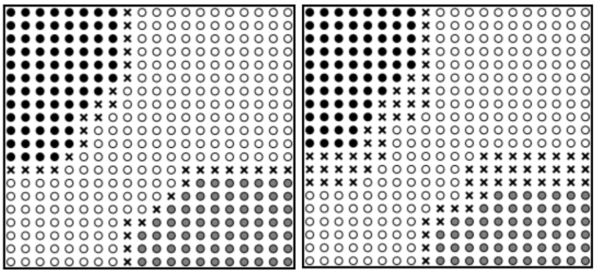


Figure 3. Results for CF_{min}= 0.4 (left) and CF_{min}= 0.7 (right).

Pedagogical activities	R&D	Proposed class	Distinctiveness
0.9 (Standard)	2.8 (Extreme)	Researcher	0.75
2.1 (Extreme)	0.5 (Low)	Teacher	1
1.3 (High)	2.2 (High)	Nonspecific	0.83

Table 1 Sample results of the academic staff members' fuzzy classification

4.2 Classifying academic staff members according to their performance

In this application, several performance classes are defined for academic staff members [10]. Again, the classification is based on their evaluations in the areas of pedagogical activities and R&D. The fuzzy rule base used is shown in Figure 4.

Overall performence of an academic staff member		Research and Development Performance				
		Very low	Low	Standard	High	Extreme
Pedagogical activities performance	Very low	Unsatisfactory	Unsatisfactory	Substandard	Standard	Very Good
	Low	Unsatisfactory	Unsatisfactory	Substandard	Very Good	Excellent
	Standard	Substandard	Substandard	Standard	Very Good	Excellent
	High	Standard	Very Good	Very Good	Excellent	Excellent
	Extreme	Very Good	Excellent	Excellent	Excellent	Excellent

Figure 4. Fuzzy rule base used for classification according to performance

Contrary to the previous example, the indicators of classes form a cardinal (evaluating) scale. The numeric values 0, 0.5, 1, 1.5, and 2 represent typical evaluations for classes denoted linguistically as *unsatisfactory, substandard, standard, very good*, and *excellent*. These real numbers lie in the kernels of values of the fuzzy scale depicted in Figure 5.

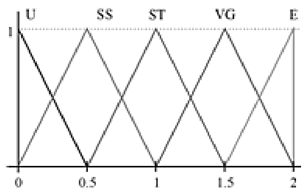


Figure 5. The linguistic fuzzy scale used for performance classes.

The evaluations will be calculated as follows. First, the rule base described in Figure 4 will be created, with the difference that on the right-hand sides of the rules there will be real numbers, which are the representatives of the fuzzy classes that express the overall performance of an academic worker. Then, the Sugeno inference algorithm [8] will be used for crisp evaluations in the area of pedagogical activities (*pa*) and R&D (*rd*). In this way, a crisp value of the overall evaluation (*eval*(*pa*,*rd*)) will be calculated. This procedure can be expressed by the following formula:

$$eval(pa, rd) = \frac{\sum_{j=1}^{n} A_{j1}(pa) \cdot A_{j2}(rd) \cdot ev_j}{\sum_{j=1}^{n} A_{j1}(pa) \cdot A_{j2}(rd)} = \sum_{j=1}^{n} A_{j1}(pa) \cdot A_{j2}(rd) \cdot ev_j \ ,$$

where A_{j1} is the fuzzy number representing the meaning of the linguistic term describing evaluation in the pedagogical area in rule j, $j=1, ...,n$, A_{j2} is the fuzzy number representing the meaning of the linguistic term describing evaluation in the area of *R&D* in rule j, $j=1, ..., n$, and ev_j is the real number representing the most typical value of the linguistic term describing the resulting class D_j in rule j , j=1, ..., n; ev_j lies in the kernel of the respective triangular fuzzy number.

From numeric evaluation we will proceed to its linguistic description, which, in the context of academic staff members, is more suitable. For that purpose, we will make use of the linguistic scale in Figure 5. If *eval*(*pa, rd*) = D_j, for some j=1,..,n, then the academic staff member fully belongs to the class with the characteristic value D_j and the linguistic interpretation of the result is clearly given by the corresponding term. Otherwise, it belongs to two neighboring classes which are the closest, i.e. where the value *eval*(*pa, rd*) belongs with a non-zero membership degree. Membership degrees of *eval*(*pa, rd*) to these two classes-are used for the linguistic description of the resulting evaluation. For example, a possible result can be that the overall performance of a given academic staff member is *70 % standard and 30 % very good*.

5 Conclusion

On a real example from the human resource management in the academic area, two types of fuzzy classification were described in this paper. The first type represents an identification problem: it is necessary to decide to which of the classes that are described verbally by the fuzzy values of their characteristics does a given object belongs (or alternatively, to decide that the object cannot be classified). There are no relationships among the classes; their identifiers form a nominal scale. In the second case, the fuzzy classification can be viewed as a certain type of evaluation; the class identifiers form a cardinal evaluating scale. Both of the mentioned cases (identification and evaluation) represent the typical problems, where the fuzzy classification is applied in the practice. The further research will be focused on suitable fuzzy classification algorithms for different types of evaluating scales and, specifically, on the application of the linguistic fuzzy modeling tools for linguistic interpretation of the fuzzy classification results. A case of fuzzy classification of objects with verbally defined characteristics values will be also studied.

Acknowledgements

The research is supported by the grant *PrF_2011_022 Mathematical Models and Structures* by IGA of the Palacky University in Olomouc.

References

[1] Dubois, D., and Prade, H. (Eds.): *Fundamentals of Fuzzy Sets. The Handbook of Fuzzy Sets Series*. Kluwer Academic Publishers, Boston-London-Dordrecht, 2000.

[2] Ishibuchi, H. and Murata, T., and Gen, M.: Performance evaluation of fuzzy rule-based classification systems obtained by multi-objective genetic algorithms, *Computers & Industrial Engineering,* Volume 35, Issues 3-4, 1998, pp. 575-578.

[3] Ishibuchi, H., Nakashima T., and Morisawa T.: Voting in fuzzy rule-based systems for pattern classification problems, *Fuzzy Sets and Systems*, Volume 103, Issue 2, Soft Computing for Pattern Recognition, 1999, pp. 223-238.

[4] Kuncheva, L. I..: Fuzzy classifier design. Springer-Verlag, Heidelberg, 2000.

[5] Lei, Z., and Ren-hou, L.: Designing of classifiers based on immune principles and fuzzy rules, Information Sciences, Volume 178, Issue 7, 2008, pp. 1836-1847.

[6] Li, T.-H., Guo N. R., and Kuo C.-L.: Design of adaptive fuzzy model for classification problem, *Engineering Applications of Artificial Intelligence*, Volume 18, Issue 3, 2005, pp. 297-306.

[7] Roubos, J. A., Setnes M., and Abonyi J.: Learning fuzzy classification rules from labeled data, Information Sciences, Volume 150, Issues 1-2, 2003, 77-93.

[8] Sugeno, M.: An introductory survey on fuzzy control. *Information Sciences*, 1985, 36:59-83.

[9] Talašová, J.: *Fuzzy methods of multiple-criteria evaluation and decision making* (in Czech). VUP, Olomouc. ISBN 80-244-0614-4, 2003.

[10] Talašová, J., and Stoklasa, J.:. Fuzzy approach to academic staff performance evaluation. Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010, September 2010, České Budějovice, Czech Republic, p. 621-626, ISBN 978-80-7394-218-2.

[11] Zadeh, L. A.: Fuzzy Sets. *Inform. Control,* 8, pp. 338-353, 1965.

[12] Zadeh, L. A.: The concept of linguistic variable and its application to approximate reasoning. *Information sciences*, Part 1: 8 (1975), 199-249, Part 2: 8 (1975), 301-357, Part 3: 9 (1975), 43-80, 1975.

[13] Zhou, E., and Khotanzad, A.: Fuzzy classifier design using genetic algorithms, *Pattern Recognition*, Volume 40, Issue 12, 2007, pp. 3401-3414.

Possibility of strictly Pareto efficient consumption allocation in a general equilibrium in an oligopolistic economy

Milan Horniaček
Institute of Public Policy and Economics, Faculty of Social and Economic Sciences
Comenius University in Bratislava
E-mail:milan.horniacek@fses.uniba.sk

Abstract: We analyze an abstract infinite horizon general equilibrium model of an oligopolistic economy. It is an extensive form non-cooperative game. A strict strong perfect general equilibrium (SSPGE) is the applied equilibrium concept. SSPGE requires that there does not exist a coalition of players that can weakly Pareto improve the vector of continuation payoffs of its members in some subgame by a coordinated deviation. We identify a class of SSPGEs, in which, in each subgame, the vector of consumers' equilibrium payoffs is strictly Pareto efficient.

Keywords: coalitions, general equilibrium, oligopoly, Pareto efficiency, strong perfect equilibrium.
JEL Classification: C73, D43, D51.
AMS Classification: 91A-06.

1 Introduction

Oligopoly is a typical feature of many markets in the current world. Nevertheless, standard general equilibrium models (i.e., Arrow - Debreu - Hahn type of models; see, for example,[2] or [1] for description of these models) follow the paradigm of perfect competition. The same holds also for computable general equilibrium models focusing on typical oligopolistic industries (see [5] for a recent example). Thus, general equilibrium models that incorporate the interaction between firms in oligopolies are needed. So far, the paper by Dierker and Grodal [3] is the most comprehensive result in this direction. As the title of their paper suggests, they pay a great attention to the issue of the objective of the firm in an oligopolistic industry. This objective should take into account not only the effect of the prices charged by the firm on its profit but also their effect on the shareholders' expenditures on consumer goods.

General equilibrium models of imperfect competition developed so far are static or (at least) have a finite time horizon. Nevertheless, the interdependence between the firms in the oligopolistic industries (and the behavior taking into account such an interdependence) fully manifests itself when the firms have infinite time horizon. (See [4] for various infinite horizon oligopoly models that have much richer sets of equilibria than finite horizon oligopoly models.) Therefore, in the present paper, we construct an abstract infinite horizon discrete time general equilibrium model with oligopolistic markets and infinitely lived consumers. This model has the form of an infinite horizon non-cooperative game with discounting of single period payoffs. The model is abstract because we do not specify all fundamentals of the economy that generate the analyzed extensive form game.

A strict strong perfect general equilibrium (henceforth, SSPGE) is the equilibrium concept applied to the model. It requires that there does not exist a coalition of players that can weakly Pareto improve the vector of continuation payoffs of its members in some subgame by a coordinated deviation. It is a refinement of Rubinstein's [6] concept of a strong perfect equilibrium. We identify a class of SSPGEs, in which, in each subgame, the vector of consumers' equilibrium average discounted utilities is strictly Pareto efficient.

Taking into account space limitations, we describe only the features of the model that are crucial for the proof of our result and we only outline the proof, formulating it partially in a verbal form, without introducing a symbol for each variable in the model.[1]

2 Model

The model is identified with the infinite horizon discrete time extensive form non-cooperative game $\Gamma = \left\langle K, H, (H_k)_{k\in K}, (\gamma_k)_{k\in K}\right\rangle$ with observable actions of players.

$$K = I \cup J \cup B \cup L \cup \{G\} \cup \{F\}$$

is the finite set of infinitely lived players. I is the set of consumers, J is the set of firms, B is the set of commercial banks (henceforth, only banks), L is the set of labor unions,G is the government, and F is the central bank. H is the set of (both terminal and non-terminal) histories. We denote the set of terminal histories by H_∞.[2] For each $k \in K$, H_k is the set of non-terminal histories after which player k takes an action and γ_k is k's payoff function in Γ defined on the set of terminal histories. (Instead of specifying the player correspondence, assigning to each non-terminal history the set of players who take an action after it, we specify directly $H_k, k \in K$.)

There are $\varkappa$ consumer goods in the model. For each consumer good $n \in \{1, ..., \varkappa\}$, $\chi_n > 0$ is the upper bound on its output, $p_n^{\min} > 0$ is the lower bound on its price, and $p_n^{\max} > p_n^{\min}$ is the upper bound on its price. These bounds are the same for each period. We set $X = \prod_{n=1}^{\varkappa} [0, \chi_n]$ and $P = \prod_{n=1}^{\varkappa} \left[p_n^{\min}, p_n^{\max}\right]$.

We denote by A the set of all feasible (with respect to available inputs and technologies) consumption allocations (i.e., sequences of single period consumption vectors) in the model. Of course, $A \subset \left(X^{\#(I)}\right)^\infty$. We assume that A is nonempty, compact and convex.

Trading between firms is based on single period contracts specifying the traded quantity and the unit price. Borrowing by firms, consumers, other banks,

[1] We give a more detailed description of the model, together with the proof of the sufficient condition of the existence of an SSPGE in it, in our (so far) unpublished manuscript "Strict strong perfect general equilibrium." Since September 1, 2011, the latter manuscript is available at http://www.fses.uniba.sk/?id=2290.

[2] A terminal history is a history (i.e., a record of past actions taken by the players) after which no player takes an action. In our infinite horizon model, the set of terminal histories coincides with the set of infinite histories.

and the government from banks is based on single period contracts specifying the borrowed amount and the interest rate. Agreements between labor unions and firms are based on single period contracts specifying wage rate and the minimum number of employees for each type of labor service. Employment is based on single period contracts between firms and consumers specifying the provided amount of each type of labor service. A buyer and a seller make a contract proposal simultaneously. A contract is concluded if and only if their proposals coincide.

In each period, the government collects income taxes and sets the income tax rate common for consumers, firms, and banks. It pays unemployment benefits and social subsidies to consumers, which are not taxed. The government can borrow from any bank. Any bank can borrow from another bank. In each period, the central bank sets the discount rate, at which banks can borrow reserves, and the reserve requirement ratio (common for all types of accounts).

Firms and consumers keep all their money in bank accounts. The government keeps all its money in the account in the central bank. Obtained loans are first deposited in a bank account. There is the upper bound on money supply, which is the same for each period. There are upper bounds on prices, wage rates, interest rates, and income tax rate that are the same for each period. Consumers have only endowments of labor services. There is the upper bound on aggregate endowment of each labor services that is the same for each period. In order to simplify and shorten the paper, we disregard durable consumer goods. Consumers' endowments of labor services and firms' technologies determine upper bounds on production of each good across all periods.

The firms and banks are limited liability companies. (Despite this, with some abuse of terminology, we use the term "dividends.") Their shareholders do not bear their losses. Shareholders cannot change firms' and banks' payoff functions described below. Shares cannot be traded.

Payoff functions γ_k, $k \in K$, have the following properties. For each player, his single period utility function is the same in every period. For each $k \in K$, γ_k assigns to every terminal history k's average discounted single period payoff. All players use the same discount factor $\delta \in (0, 1)$.

Consumers' single period payoffs functions are their single period utility functions. The latter functions represent locally non-satiated single period preferences. They are strictly concave and homogeneous of degree one. Only consumed quantities of consumer goods (including services) are arguments of single period utility functions. Leisure is not included among their arguments. (Nevertheless, it can be taken into account indirectly. Consumed quantities of goods and services can reflect time available for consuming them.) For each $i \in I$, $u_i : X \to \mathbb{R}_+$ is his single period utility function and $e_i : P \times \mathbb{R}_+ \to \mathbb{R}_+$ is his single period expenditure function. We have $u_i(0) = 0$ for each $i \in I$.

Consumers are special players in our model. They behave strategically with respect to conclusion of labor and credit contracts. Nevertheless, given their net income in each period, obtained loans, interest payments, and received interest, they choose their consumption stream on the basis of maximization of their average discounted utilities. Since their single period utility functions are strictly concave, their consumption streams are uniquely determined.

A usable single period income of consumer $i \in I$ in period $t \in \mathbb{N}$ is i's net income in period t plus his withdrawals from bank accounts in period t in the excess of his net income minus the sum from his net income that he does not withdraw from his bank accounts in period t. It is the sum of money that i uses for purchases of consumer goods in period t. We work with usable single period incomes of shareholders of firms and banks (members of labor unions) in the definition of firms' and banks' (labor unions') single period payoff functions. Such an approach takes into account that consumers' purchasing opportunities in each period are constrained by their purchasing plans for other periods.

In the definition of single period payoff functions of firms, banks, and labor unions, we take actual consumption vectors of all consumers as fixed. The single period payoff of each firm and bank equals the ratio of the sum of its shareholders' usable single period incomes to the market value of the sum of its shareholders' fixed consumption vectors. (This formulation is in the spirit of definition in [3, p. 266], but it allows the computation of sum of discounted single period payoffs.) The single period payoff of each labor union equals the ratio of the sum of its members' single period usable incomes to the market value of the sum of its members' fixed consumption vectors. The government's single period payoff function is strictly increasing in real gross domestic product (henceforth, GDP) and strictly decreasing in unemployment rate and the government's debt. The central bank's single period payoff function is strictly increasing in real GDP and strictly decreasing in the absolute value of the rate of inflation (computed using GDP deflator).

We restrict attention to pure strategies in Γ. A pure strategy of player $k \in K$ is a function that assigns to each history $h \in H_k$ one element of the set of his feasible vectors of actions after h. First of all, it assigns to elements of H_k a vector of proposals of contracts that k makes to the other players and (with the exception of the central bank) his decisions on withdrawals from and deposits to his bank accounts. Besides this, a pure strategy of a consumer $i \in I$ assigns to some elements of H_i a consumption bundle that he buys (thus, a pure strategy does not prevent him from buying a positive quantity of each consumer good), a pure strategy of firm $j \in J$ producing consumer goods assigns to some elements of H_j prices that it charges to consumers for its products, a pure strategy of bank $b \in B$ assigns to some elements of H_b interest rates for deposits, the government's pure strategy assigns to some elements of H_G the income tax rate and the levels of unemployment benefits, the central bank's pure strategy assigns to some elements of H_F the discount rate and the reserve requirement ratio.

For each $k \in K$, let S_k be the set of pure strategies of player k in Γ. It assigns to each $h \in H_k$ one of k's possible actions after h. We set $S = \prod_{k \in K} S_k$ and for each coalition $C \in 2^K \setminus \{\varnothing\}$, $S_C = \prod_{k \in C} S_k$. For every $k \in K$, $\pi_k : S \to \mathbb{R}$ is k's payoff function in Γ defined on the set of profiles of pure strategies. It is defined by $\pi_k(s) = \gamma_k(\eta(s))$, where $\eta : S \to H_\infty$ is the outcome function.

For each $h \in H \backslash H_\infty$, $\Gamma_{(h)}$ is the subgame of Γ following history h. For every symbol defined for Γ its restriction to subgame $\Gamma_{(h)}$ is indicated by subscript (h).

We assume that S_k is a nonempty and compact space and function π_k is continuous for each $k \in K$. (With respect to the page limitations, we make assumptions about S_k and π_k instead of about the fundamentals of Γ.)

Definition 1 *A strategy profile $s^* \in S$ is an SSPGE of Γ if (i) there does not exist $h \in H \backslash H_\infty$, $C \in 2^K \backslash \{\varnothing, K\}$, and $s^{(C)} \in S_{C(h)}$ such that*

$\pi_{k(h)}\left(s^*_{-C(h)}, s^{(C)}\right) \geq \pi_{k(h)}\left(s^*_{(h)}\right)$ for each $k \in C$ with strict inequality for at least one member of C, and (ii) there does not exist $h \in H \backslash H_\infty$ and $s \in S_{(h)}$ such that $\pi_{k(h)}(s) \geq \pi_{k(h)}\left(s^*_{(h)}\right)$ for each $k \in K$ with strict inequality for at least one $k \in K$.

3 Possibility of the strict Pareto efficiency of a consumption stream in an SSPGE

For each $h \in H \backslash H_\infty$ let

$$S^+_{(h)} = \arg\max \left\{ \sum_{k \in L \cup J \cup B} \pi_{k(h)}(s) \mid s \in S_{(h)} \right\}. \tag{1}$$

Thus, each $s \in S^+_{(h)}$ maximizes the sum of payoffs of labor unions, firms, and banks in $\Gamma_{(h)}$. The maximization takes place with respect to the profile of players' pure strategies (and, hence, indirectly with respect to the values of players' decisions that are assigned to the non-terminal histories by their pure strategies).

Proposition 1 *Suppose that $s^* \in S$ is an SSPGE of Γ and for each $h \in H \backslash H_\infty$, it satisfies the following conditions (i) $s^*_{(h)} \in S^+_{(h)}$, (ii) in every period of $\Gamma_{(h)}$, the market for each consumer good is in equilibrium, (iii) for each consumer good, its price is the same in every period of $\Gamma_{(h)}$, (iv) the equilibrium consumption allocation in $\Gamma_{(h)}$ is an interior point of A, (v) for each $j \in J \cup B$, the sum of consumption vectors of j's shareholders is the same in every period of $\Gamma_{(h)}$, (vi) for each $\ell \in L$, the sum of consumption vectors of ℓ's members is the same in every period of $\Gamma_{(h)}$, (vii) there exists $\epsilon > 0$ such that in each period of $\Gamma_{(h)}$, the government's income from taxes exceeds the sum of its expenditures (on unemployment benefits, social subsidies, repayment of debt, and interest payments) by at least ϵ .Then, for each $h \in H \backslash H_\infty$, the vector of consumers' payoffs $\left(\pi_{i(h)}\left(s^*_{(h)}\right)\right)_{i \in I}$ is strictly Pareto efficient with respect to the set $\left\{\left(\pi_{i(h)}(s)\right)_{i \in I} \mid s \in S_{(h)}\right\}$.*

Proof. Take s^* satisfying the assumptions of Proposition 1 and $h \in H \backslash H_\infty$. Let

$$\left\{(x_i(t))_{i \in I}\right\}_{t \in \mathbb{N}} \in A$$

be the equilibrium consumption allocation in $\Gamma_{(h)}$ and $p \in P$ the vector of prices of consumer goods in each period of $\Gamma_{(h)}$. Suppose that the claim of Proposition 1 does not hold for h. Then there exists a consumption allocation $x^{+} \in A$ such that the vector $\left((1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}u_i\left(x_i^{+}(t)\right)\right)_{i\in I}$ weakly Pareto dominates

$$\left((1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}u_i\left(x_i(t)\right)\right)_{i\in I}=\left(\pi_{i(h)}\left(s^{*}_{(h)}\right)\right)_{i\in I}. \tag{2}$$

Since A is convex and consumers' single period utility functions are strictly concave, for $\lambda \in (0,1)$ arbitrarily close to one, the vector

$$v=\left((1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}u_i\left(\widetilde{x}_i(t)\right)\right)_{i\in I},$$

where $\widetilde{x}=\lambda x+(1-\lambda)x^{+}$, weakly Pareto dominates $\left(\pi_{i(h)}\left(s^{*}_{(h)}\right)\right)_{i\in I}$. For each $i\in I$ and every $t\in\mathbb{N}$, let $\triangle_i(t)=u_i(\widetilde{x}_i(t))-u_i(x_i(t))$. Then,

$$(1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}\triangle_i(t)>0$$

for each $i\in I$ with $v_i>\pi_{i(h)}(s^{*}(h))$ and

$$(1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}\triangle_i(t)=0$$

for each $i\in I$ with $v_i=\pi_{i(h)}(s^{*}(h))$. As consumers' single period utility functions are homogeneous of degree one, the minimal sum of average discounted expenditures on reaching the stream of single period utilities

$$\{u_i(x_i(t))+\triangle_i(t)\}_{t\in\mathbb{N}},$$

when the vector of prices of consumer goods is p in each period, is

$$(1-\delta)\,e_i(p,1)\sum_{t\in\mathbb{N}}\delta^{t-1}\left[u_i(x_i(t))+\triangle_i(t)\right]\geq(1-\delta)\sum_{t\in\mathbb{N}}\delta^{t-1}px_i(t),\forall i\in I, \tag{3}$$

with strict inequality if $v_i>\pi_{i(h)}(s^{*}(h))$. Taking into account condition (vii) of Proposition 1, when λ is close enough to one, the government can give (by reducing the income tax rate) to each $i\in I$, in every period t of $\Gamma_{(h)}$ in which it is needed, increased usable single period income enabling him to cover the minimum costs of achieving utility $u_i(x_i(t))+\triangle_i(t)$, without eliminating the government's budget surplus in any period of $\Gamma_{(h)}$. For each $i\in I$, let $z_i=\{z_i(t)\}_{t\in\mathbb{N}}$ be i's consumption stream that minimizes his average discounted costs of reaching utility stream $\{u_i(x_i(t))+\triangle_i(t)\}_{t\in\mathbb{N}}$, when the vector of prices of consumer goods in each period equals p. Taking into account condition (iv) of Proposition 1, $z=(z_i)_{i\in I}\in A$. Let $\widetilde{s}\in S_{(h)}$ generate sequences of consumers' usable single period incomes described above, consumption allocation z, and the vector of prices of consumer goods equal to p in each period of $\Gamma_{(h)}$. Then (for fixed consumption allocation x) $\widetilde{s}$ increases the sum of payoffs of players in $I\cup J\cup B$ in $\Gamma_{(h)}$ in comparison with $s^{*}_{(h)}$. This contradiction with condition (i) of Proposition 1 completes the proof. ■

4 Conclusion

We have indicated the possibility of strict Pareto efficiency of equilibrium consumption allocation in a general equilibrium in an oligopolistic economy. This suggests that strict Pareto efficiency of consumption allocation in a general competitive equilibrium cannot be a sufficient justification for advocating of antitrust policy. In Proposition 1, cooperation between firms, banks, and labor unions on equal grounds plays an important role. This indicates that strong labor unions need not prevent Pareto efficient allocation of resources.

Acknowledgement 1 *The research reported in the present paper was supported by the grant VEGA 1/0138/09.*

References

[1] Arrow, K.J., and Hahn, F.H.: *General competitive analysis.* Holden and Day, San Francisco, 1971.

[2] Debreu, G: *Theory of value.* Yale University Press, New Haven, 1959.

[3] Dierker, E., and Grodal, B.: The price normalization problem in imperfect competition and the objective of the firm. *Economic Theory* **14** (1999), 257-284.

[4] Frriedman, J.W.: *Oligopoly theory.* Cambridge University Press, Cambridge, 1983.

[5] Miťková, V.: Modeling the crisis in automobile industry - the applied CGE study. In: *Proceedings of the Mathematical Methods in Economics 2009* (Brožová, H., and Kvasnička, R., eds.). Czech University of Life Sciences, Faculty of Economics and Management, Prague, 2009, 221-226.

[6] Rubinstein, A.: Strong perfect equilibrium in supergames. *International Journal of Game Theory* **9** (1980), 1-12.

Using indicators of ecological stability in stochastic programming

Michal Houda[1]

Abstract. When building bigger constructions (transport, industrial, etc.), the EU law impose the so-called The EIA (Environmental Impact Assessment) process – evaluation of possible influences of the construction. All important constructions have to go through the process, prescribing obligatory the judgement of a series of factors that influence the environment and population health, grouped into several categories. Outputs of the EIA process are usually a set of recommendations and obligations to the investors and building companies in order to compensate the negative impacts of the constructions by additional arrangements.

In modelling phase, several uncertain parameters (e.g. future transport intensities) enter into the problem. Also many criteria of the EIA process are of subjective character, so modelling through utility functions (again not well defined) is proposed. In our contribution we develop an innovative approach to model the expenses devoted to obey the EIA rules by stochastic programming tools: especially, we represent uncertainty in parameters by their probabilistic distributions, and subjective utility function representing the ecological demands is modelled via so-called indicators of ecological stability. The model takes into account budget limitations, several legislative obligations (as pollution limits), and other ecological aspects influencing costs of transport constructions; the goal is to help choose, among possible compensating constructions and arrangements, the optimal ones. The resulting stochastic programming model could be seen as parallel to the usual V@R problem – type of problems used in finance.

Keywords: EIA process, indicator of ecological stability, stochastic programming, value-at-risk models.

JEL classification: C44
AMS classification: 90C15

1 Introduction

The growth of economics in whole has many visible et invisible manifestation in our surroundings, more or less distant. One of the very visible manifestation of these is number of constructions that grow every day at our nearness. This is not new phenomenon, at all; in fact it is old as the humanity itself. What is modern, in this context, is very strong emphasis on the impact of such development to the *environment*. Any of new big constructions, as transport line constructions (highways, railways), industrial or commercial areas, or other important constructions, cannot be nowadays realized without precise treatment of impacts (negative as positive) to the environment.

In [3] we already presented the whole legal framework of this treatment, let only briefly summarize it. By European Union law, every new construction has to pass through the process evaluating the impact of the construction to the environment. The process is divided into several phases, the most important one is so-called Environmental Impact Assessment (EIA). The output of this phase of the process is to evaluate if the negative impacts of the construction are still acceptable, in view of the society development, and (what is especially important in view of our present paper), to propose possible duties and arrangements to compensate these negative impacts. Unfortunately, the impacts are of very different nature and their evaluation is not simple at all. In [3], we already provided a description of the EIA categorization

[1]Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: houda@ef.jcu.cz

of possible impacts (see also [1]), so we will not go through it again, only summarize that it includes human healthy impacts (air pollution, noise pollution, social-economic factors) and environmental impacts (climate, water, landscape, natural resources, ecosystems, etc.).

Our present intention is to continue research already started therein: to develop stochastic programming models incorporating to the above decision process some uncertainty factors, as unknown transport intensities (on transport line constructions), efficiency of compensation, or subjectivity of some EIA criteria.

2 Model description

Denote $x \in X \subset \mathbb{R}^n$ the collection of the possible compensations and arrangements ready to be used in some construction. The values of the variable x can be discrete or continuous according to the nature of the arrangement. Let $\xi \in \Xi$ be the random vector representing uncertainty factors, with $\Xi \subset \mathbb{R}^s$ being the predefined support of ξ. From the computational perspective, it is necessary to consider the probability distribution of ξ to be known in advance.

The actual expenses of all arrangements are represented by the cost function $c : X \times \Xi \to \mathbb{R} : (x;\xi) \mapsto c(x;\xi)$. We make here the first simplification and suppose c to be a linear function, i. e., $c(x;\xi) = c^T x$ where $c \in \mathbb{R}^n$ are constant unit costs of the arrangements. Let B be the budget limit on these expenses.

All factors of subjective and evaluative character are represented by the utility function $u : X \times \Xi : (x;\xi) \mapsto u(x;\xi)$. The quantification of this function is difficult; later, we will use a function based on the indicators of ecological stability to represent the values of the function u.

We are know ready to formulate the optimization model with uncertainty. Instead of viewing the problem as (traditionally) cost-minimizing, we use a utility-maximizing view:

$$\text{maximize } u(x;\xi) \text{ subject to } c^T x \leq B, x \in X_0. \tag{1}$$

where $X_0 \subset X$ includes all deterministic constraints, technical parameters, and 0-1 type constraints. This is an optimization model with unknown uncertainty parameters, so we are looking for a probabilistic formulation incorporating available information on the random parameter ξ. Before do that we deal first with subjective utility function u. Similarly as in [3] we use so-called indicators of ecological stability.

Indicators of ecological stability is a set of well defined and quantified variables that evaluates the quality of environment. The overal set of these indicators is still growing (especially in case of indicators qualifying such things as landscape, efficiency, total prosperity of environment, and similar); the actual standard set, and propositions to new possible indicators are at the disposition of the reader on the homepage of the European Environmental Agency (EEA, see [2]).

3 β-V@R formulation of the problem

The main idea of our approach is to replace the subjective utility function by a weighted sum of objective indicators of ecological stability. Let $g : \mathbb{R}^n \times \Xi \to \mathbb{R}^G : (x;\xi) \mapsto g(x;\xi)$ be a function representing the values of the EEA indicators. Moreover, we introduce a new parameter L representing a required limit on values of g, and the weights $w \in [0;1]^G$ representing importance of each of indicators. Now we are ready to incorporate the probabilistic information on ξ provided by its probability distribution. We do it through imposing the probabilistic constraint in the form

$$\Pr\{w^T g(x;\xi) \geq L\} \geq \beta \tag{2}$$

where L is now considered as the decision variable, and $\beta \in (0;1)$ is a prescribed (sufficiently high) value of probability of exceeding the limit L.

The whole proposed formulation of the problem read:

$$\text{maximize } L \text{ subject to } \Pr\{w^T g(x;\xi) \geq L\} \geq \beta, c^T x \leq B, x \in X_0. \tag{3}$$

The formulation is seen parallel to the known β-V@R problems, thoroughly analyzed in finance optimization (see e. g. [4] and references therein). The only difference is that our left-hand side of probabilistic constraint does not represent losses (as in original formulation) but the positive effect of arrangements. In fact, this is not any real obstacle, and we can use without worries the methods developed by many authors for V@R problems.

3.1 Example

Consider an artificial example of building a segment of a (non-specified) highway. For simplicity we consider two indicators of ecological stability only:

- $1 - i_1 \ldots$ excedance of the air pollution limits (percents of area where the limits are exceeded);
- $-i_2 \ldots$ noise pollution (number of habitants exposed to heavy noise).

To share previously used notation we use this modified notation so that indicators i_1 and i_2 are represented by positive values.

Consider further three possible arrangements in mind to diminish the negative impacts of the building.

- $x_1 \ldots$ imposed speed limit of 80 km/h;
- $x_2 \ldots$ imposed speed limit of 110 km/h;
- $x_3 \ldots$ noise wall dimensions.

The variables x_1 and x_2 are binary (the speed limit is imposed or not), the variable x_3 is continuous and could represent the length of noise wall (with standard width and height), or the consumption of the material to build it.

Finally, the only random factor considered in the example is ξ, random transport intensity on the highway segment considered. Assume that the values of indicators depends linearly on x and ξ. (In fact, we expect that this is far from reality for real problems and is subject to the future research.) Then we could write

$$
\begin{aligned}
i_1 &= (a_{10} + a_{11}x_1 + a_{12}x_2)\xi \\
i_2 &= (a_{20} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3)\xi
\end{aligned}
$$

The parameters a_{11} and a_{12} represent the positive effect of the imposed speed limit to the air pollution; a_{21} and a_{22} the same to the noise pollution. Denote $x = (1, x_1, x_2)$, $a_0 = (a_{10}, a_{20})$, $A = (a_{ij})_{i,j=1,2}$. Consider further that the overall impact of the highway construction is measured by the weighted sums of indicators i_1 and i_2 with $w = (w_1, w_2)$ beeing the weights of the indicators. In such notation, the β-V@r model formulation of the problem reads:

$$
\text{maximize } L \tag{4}
$$

subject to

$$
\begin{aligned}
\Pr\big\{w_1 i_1 + w_2 i_2 = w^T(a_0 + Ax)\xi \geq L\big\} &\geq \beta \\
c^T x &\leq B \\
x_1 + x_2 &\leq 1 \\
x_{1,2} \in \{0,1\}, x_3 &\geq 0
\end{aligned}
$$

The resulting optimization program is the linear stochastic programming mixed-integer problem with probabilistic constraints, and it is solvable under some additional assumption on distribution of the random variable.

4 β-CV@R formulation

Inspired by [4] we now proceed to better formulation of the model. The β-CV@R model is, according to [4] formulated as the conditional expectation of losses above the amount L. Sharing the notation of the previous section, the β-CV@R formulation of our optimization problems reads

$$\text{to find } \phi_\beta(x) := \frac{1}{1-\beta}\,\mathbb{E}\big[w^T g(x;\xi) \mid w^T g(x;\xi) \le L_\beta(x)\big] \tag{5}$$

where

$$L_\beta(x) = \max\big\{L : \Pr\{w^T g(x;\xi) \ge L\} \ge \beta, c^T x \le B, x \in X_0\big\} \tag{6}$$

i. e., $L_\beta(x)$ is the optimal value of the β-V@R formulation of the problem.

Denote

$$F_\beta(x;L) = -L + \frac{1}{1-\beta}\,\mathbb{E}[L - w^T g(x;\xi)]^+ \tag{7}$$

where $[t]^+$ is positive part of t (i. e., $[t]^+ = t$ if $t \ge 0$ and zero otherwise).

Proposition 1. *As a function of L, $F_\beta(x;L)$ is convex, continuously differentiable and β-CV@R value ϕ_β of any $x \in X$ can be determined from the formula*

$$\phi_\beta(x) = \min_L F_\beta(x;L). \tag{8}$$

Proof. The proposition follows directly from Theorem 1 in [4]. □

The proposition enable simple numerical scheme to solve the β-CV@R formulation of our problem because the minimization of the convex differentiable function is numerically easy. Moreover, the explicit dependence on β-V@R value L_β (which computation may not be always easy) is discarded. We finally illustrate the new formulation on Example 3.1.

4.1 Example continued

The β-CV@R formulation of our example problem reads simply

$$\phi_\beta(x) = \frac{1}{1-\beta}\,\mathbb{E}\big[w^T(a_0 + Ax)\xi \mid w^T(a_0 + Ax)\xi \le L_\beta(x)\big] \tag{9}$$

with L_β being the optimal value of the problem (4). Using (7) and Proposition 1 we arrive at final problem formulation:

$$\text{minimize } -L + \frac{1}{1-\beta}\,\mathbb{E}[L - w^T(a_0 + Ax)\xi]^+ \tag{10}$$

subject to

$$\begin{aligned} c^T x \le B, &\quad x_1 + x_2 \le 1 \\ L \in \mathbb{R}, &\quad x_{1,2} \in \{0,1\}, x_3 \ge 0 \end{aligned}$$

If we have an approximation or estimate of the vector ξ at our hand (for example, done by sampling), various numerical techniques can be used to estimate value and solution of problem (10).

5 Conclusion

In this paper we presented a V@R motivated approach to solve problems coming from the environmental politics. We designed a probabilistic approach to model uncertainties and subjective factors in evaluating impacts of constructions to the environment, incorporating well-defined and objectively determined indicators of ecological stability. Our model are especially useful in that they obey assumptions on traditional financial V@R and CV@R models, so that open the door to many modelling and computational methods developed originally for these financial models.

Acknowledgements

This research was partially supported by the Czech Science Foundation under the project No. P402/10/0956.

References

[1] Council directive 85/337/EEC of 27 June 1985 on the assessment of the effects of certain public and private projects on the environment. *Official Journal of the European Union*, L 175, 05/07/1985:40–48, June 1985.

[2] EEA Core set of indicators [online]. European Environment Agency. Available from World Wide Web: <http://themes.eea.europa.eu/IMS/CSI>, February 2011 [cited 20 May 2011].

[3] Houda, M.: Environmental factors in optimization of traffic line construction expenses. In *Proceedings of the 24th International Conference Mathematical Methods in Economics 2010* (M. Houda and J. Friebelová, eds.), University of South Bohemia in Česke Budějovice, 2010, 262–267.

[4] Rockafellar, R. T., and Uryasev, S.: Optimization of conditional value-at-risk. *The Journal of Risk* **2** (2000), 21–41.

Exchange rate pass-through analysis for Czech Republic and Poland: implications towards eurozone accession

Roman Hušek[1], Tomáš Formánek[2]

Abstract. Exchange rate pass-through (ERPT) is a useful tool for describing properties of exchange rate shock-transmission into domestic prices. This type of analysis is of great relevance for small open economies such as Czech Republic and Poland, especially in situation of prospective ERM-II regime participation with subsequent eurozone accession. Existing pass-through studies for current and former eurozone candidate economies show diverse ERPT results, depending on the methodology used, as well as on objective economic circumstances in different countries. We analyze and compare the short-term and long-term pass-through ERPT behavior for Czech Republic and Poland. Our estimated long-term ERPT for Czech Republic is approximately 25% and just under 10% for Poland. Impacts from low-inflation and inflation targeting environment are considered, along with other implications, as the ERPT analysis may serve as a basis for improved predictions of monetary and real sector indicators for small open economies.

Keywords: Exchange rate pass-through (ERPT), inflation, VAR model, impulse-response function (IRF).

JEL Classification: C32, E31, E52, F33.

1 Introduction

Current Czech macroeconomic policy is not aimed at measures directly motivated by seeking and facilitating our eurozone accession. Poland, our only non-eurozone neighbor has recently gained some momentum on its way towards euro-adoption. Also, given the level and dynamics of our economic integration with eurozone countries, it is likely that within a few years the Czech government will (have to) be more active in this respect. Our prospective participation in the ERM-II regime would require tight exchange rate stability towards the euro, yet at the same time the catching-up process and faster relative macroeconomic development of the new EU member states exerts pressure towards exchange rate appreciation. This paper is designed to analyze and compare the short-term and long-term **exchange rate pass through** (ERPT) dynamics for the Czech Republic and Poland.

ERPT was traditionally defined as the percentage change in the local currency price of an imported good resulting from a 1% change in the nominal exchange rate between the exporting and importing countries. However, most of the recent studies on ERPT extend the pass-through definition to consumer prices and producer prices. Taking into account the **transmission mechanism** and pricing chain, we may expect the ERPT to be gradual and incomplete. For example, we may expect consumer prices to have lower and delayed response to an exchange rate shock, if compared with the response of import prices. Vector autoregression (VAR) methodology seems suitable for this type of analysis, allowing for construction of impulse-response functions (IRFs) where transmission processes are determined by variables' ordering as used for Cholesky decomposition. Section 2 of this paper provides brief literature review, section 3 outlines our methodological approach, section 4 consists of interpreted estimation results and section 5 concludes our paper.

2 Literature Review

There are many empirical analyses focusing on the estimation of ERPT behavior in developed countries, as well as numerous studies tackling the pass-through behavior for small open economies. Czech Republic and other new EU-members and eurozone candidates have also been studied from a broadly defined ERPT perspective, for example by Babetskaia-Kukharchuk [1], Darvas [3] and Coricelli et al. [2] as detailed in this section.

Babetskaia-Kukharchuk [1] examines the pass-through effect based upon evaluating numerous estimates of different VAR-model specifications and also by assessing the IRFs instability induced by alternative ordering of

[1] University of Economics, Department of Econometrics, 4. W. Churchill Sq., Prague 3, husek@vse.cz.
[2] University of Economics, Department of Econometrics, 4. W. Churchill Sq., Prague 3, formanek@vse.cz.

variables as Cholesky decomposition is performed. The ERPT estimation proper is than performed using a six-variable VAR model and according to a general concept used by McCarthy [9], who analyzes the pass-through using variables located along the pricing chain (distribution chain), which may be ordered as follows: import prices, producer prices and consumer prices. Peak of the short-term pass-through estimated response for Czech economy is 25 percent and the long-term (total) ERPT is approximately 30%. Also, some theoretical consideration is given to time stability and symmetry of price adjustment. As prices tend to be more rigid downwards, the ERPT responses to positive and negative exchange rate shocks could exhibit asymmetric behavior. Due to data availability and other empirical restrictions, asymmetric pass-through analyses are scarce.

Faruqee [4] provides economic explanation for **incomplete pass-through**, which is a prominent empirical feature among many economies, including eurozone members and candidates. Some of the major causes for low ERPT ratio include nominal rigidity, local currency pricing, market segmentation (mark-up changes, international price discrimination), local distribution costs and other offsetting factors. Those factors play diverse roles at different levels of the pricing chain and multiple disaggregated price level indicators are deemed necessary for this type of analysis. Darvas [3] proposes to decompose inflation changes into ERPT and price convergence effects, pointing out that changes in domestic price level are not necessarily induced only by changes in the exchange rate. This author also points out that ERPT may significantly vary over time. However, as the Czech Republic and Poland approach eurozone accession and the ERM-II transition regime, it is increasingly important to keep in mind the potential costs from ERPT induced inflation.

Coricelli et al. [2] point out that monetary policy arrangements have substantial influence on pass-through behavior. *Inflation targeting* and flexible exchange rate tend to result in significantly lower ERPT ratios when compared to other economies with imposed exchange rate regulation. Remarkably, Coricelli et al. [2] also indentify a complete ERPT for Slovenia and Hungary and somewhat lower pass-through for Czech Republic and Poland. This result may be examined in contrasts with some of the other published estimations, e.g. Gagnon and Ihrig [6] who analyze twenty industrialized economies and find a very low and declining ERPT with long-term pass-through averaging around 5 percent. Also, McCarthy [9] finds the eurozone ERPT inflation response as mostly insignificant.

In general, individual authors take different approaches to ERPT analysis and there is a variety of procedures and methods leading to pass-through estimations. Hence, we should not be surprised by the fact that individual ERPT estimates can differ fundamentally as they result from a wide spectrum of macroeconomic frameworks, econometric procedures, data aggregation, handling and transformation, etc. There is not a single, widely adopted and accepted model for ERPT measurement, which somewhat complicates the comparison of results among different authors and working papers as well as the selection of suitable model for our own analysis. Nevertheless, using aggregated data we have been able to construct and use a relatively simple and efficient three-variable VAR model that has unambiguous ordering of variables and a clear interpretation of IRFs, as described in the next section.

3 Methodology and Data

We use a **VAR model** approach to study and compare the exchange rate pass-through for the Czech Republic and Poland. The usual two-stage methodology is used. First, we examine the behavior of the selected time series by estimating the VAR model. Second, we construct impulse-response functions from the estimated VAR model in order to identify the pattern of (ERPT) behavior. This two-stage process is consistently performed for the Czech economy and for Poland, in order to obtain commensurable ERPT behavior that may be directly compared and interpreted accordingly. The VAR(2) model is specified as follows:

$$\begin{aligned}
n_t &= \beta_{10} + \beta_{11} n_{t-1} + \beta_{12} n_{t-2} + \beta_{13}\pi^{eu}_{t-1} + \beta_{14}\pi^{eu}_{t-2} + \beta_{15}\pi^{d}_{t-1} + \beta_{16}\pi^{d}_{t-2} + \beta_{17} t + u_{1t} \\
\pi^{eu}_t &= \beta_{20} + \beta_{21} n_{t-1} + \beta_{22} n_{t-2} + \beta_{23}\pi^{eu}_{t-1} + \beta_{24}\pi^{eu}_{t-2} + \beta_{25}\pi^{d}_{t-1} + \beta_{26}\pi^{d}_{t-2} + \beta_{27} t + u_{2t} \\
\pi^{d}_t &= \beta_{30} + \beta_{31} n_{t-1} + \beta_{32} n_{t-2} + \beta_{33}\pi^{eu}_{t-1} + \beta_{34}\pi^{eu}_{t-2} + \beta_{35}\pi^{d}_{t-1} + \beta_{36}\pi^{d}_{t-2} + \beta_{37} t + u_{3t}
\end{aligned}$$

where n_t is the nominal effective exchange rate; year on year relative differences are used,

π^{eu}_t - consumer prices index for the eurozone; year on year relative differences are used,

π^{d}_t - domestic consumer prices index; year on year relative differences are used,

t - trend,

u_{it} - seemingly unrelated random variables.

According to information criteria results we use a VAR(2) specification, which is widely used in macroeconometric models of the Czech economy (as discussed e.g. by Hušek, [8]). Linear trend was explicitly introduced to the VAR model, as our data exhibit integrated or near-integrated $I(1)$ behavior.

Considering Cholesky decompostion, VAR model variables are ordered according to a pricing chain conception implicating that prices are fixed in the very short run and therefore cannot react to an exchange rate shock immediately, whereas the exchange rate can and does respond to many exogenous shocks instantaneously (see e.g. Babetskaia-Kukharchuk [1]). It is often assumed that the exchange rate determination is a forward looking process which takes place at the asset market rather than at the goods market. Hence, orthogonalized innovations to the exchange rate depend only on the residuals from the first (exchange rate) equation and not on other variables of our model. Although the short run price rigidity restriction may be inappropriate for certain individual commodity prices, its imposition at the aggregated CPI level is applied and advocated by different authors (e.g. Faruqee, [4]).

As shown e.g. by Hušek [7], ordering of VAR model variables affects the Cholesky decomposition and thus has an impact on IRFs calculation, effectively imposing the theoretical pass-through restrictions and influencing the estimated ERPT behavior. The exchange rate variable order is indeed important in our model, as experiments with alternatively specified transmission mechanisms produced significant alterations to IRF behavior.

All data used in our model were obtained from the IMF's International Financial Statistics Online (IFSO) database. About half of the working papers referred to in section 2 use data expressed in logarithms, yet we have found little practical evidence supporting such approach for our model. Instead, we compare each quarter to its corresponding value from the previous year, using relative differences (t to t-4 percentage changes). Time span of the data covers the period from 2001:1Q to 2010:3Q and due to y-o-y differences being used, we lose observations from the first year. EViews and PcGive software were used for all estimations and IRFs construction.

Resembling the approach of Babetskaia-Kukharchuk [1], we had to replace import prices by foreign consumer prices (we use eurozone CPI) as a measure of preserving the influence of external macroeconomic environment while dealing with the shortage of adequate data on import prices. While assembling our model, we tested the inclusion of alternative inflation indicators at different levels of aggregation, such as producer prices index (PPI) and GDP deflator. It is assumed that ERPT is the same for all goods in a consumption basket and so the inclusion of e.g. PPI might help with describing pass-though properties along the pricing chain. However, given the volume and character of data observations, we find this approach unsuitable for ERPT modeling for the Czech Republic and Poland. Adding new endogenous variables to the VAR(2) model leads to unacceptable growth in IRF confidence limits and restrains interpretation and reliability of the calculated pass-through.

4 Estimation Results

Before the VAR model estimation proper, we conducted some preliminary tests on the data. ADF unit root tests suggest that most of our data series exhibit integrated or near integrated $I(1)$ behavior, even though the tests were applied to differenced data (relative y-o-y first differences) and therefore non-linear trend in the data could be present instead, as ADF tests may not distinguish between unit root and non-linear trend. Yet, the explicit introduction of trend as an exogenous variable to the model in order to tackle the deterministic constituent part of the trend factor from the underlying undifferenced observations has provided sufficient improvement in model performance and residuals' properties. Granger causality tests were performed with the expected result of eurozone CPI being independent on local variables from Czech Republic and Poland. We do not to change the specification of the VAR(2) model upon this intuitive and theoretically correct result, as π_t^{eu} remains a crucial and theoretically justified variable within our ERPT estimation model. F-tests on the significance of inclusion for each endogenous variable of the model confirm our specification at the 1% significance level. Similarly, F-tests on the significance of inclusion for each lag to the VAR(2) model confirm our specification at the 1% significance level. All tests and estimations were performed on a consistent basis for both the Czech Republic and Poland, in order to obtain comparable results. VAR estimation results for both economies are presented in table 1. As multicolinearity of predetermined variables in VAR models is a common factor, some of the individual regression coefficient estimates are not significant at the 5% significance level. On the other hand, the overall performance of our model is satisfactory for both economies included in this study. R^2 coefficients and their corresponding F-tests provide sufficient confidence for the main stage of our analysis, which is the **ERPT modeling through impulse-responses**.

Vector Autoregression Estimates Sample(adjusted): 2001:3 2010:3. Included observations: 37 after adjusting endpoints Standard errors in ( ) & t-statistics in [ ]						
	Czech Republic			**Poland**		
	n_t	π_t^{eu}	π_t^{d}	n_t	π_t^{eu}	π_t^{d}
n_{t-1}	**0,931**	**0,047**	**-0,051**	**0,968**	**0,022**	**-0,017**
	(0,182)	(0,026)	(0,049)	(0,232)	(0,016)	(0,022)
	[5,130]	[1,850]	[-1,045]	[4,178]	[1,426]	[-0,767]
n_{t-2}	**-0,453**	**-0,028**	**-0,061**	**-0,197**	**-0,014**	**-0,004**
	(0,191)	(0,027)	(0,052)	(0,253)	(0,017)	(0,024)
	[-2,371]	[-1,026]	[-1,181]	[-0,779]	[-0,818]	[-0,150]
π_{t-1}^{eu}	**1,330**	**0,764**	**1,483**	**0,141**	**1,011**	**0,018**
	(1,915)	(0,269)	(0,519)	(3,833)	(0,258)	(0,365)
	[0,695]	[2,838]	[2,859]	[0,037]	[3,912]	[0,050]
π_{t-2}^{eu}	**-2,736**	**-0,148**	**-0,701**	**-3,228**	**-0,354**	**0,187**
	(1,685)	(0,237)	(0,457)	(3,094)	(0,209)	(0,295)
	[-1,624]	[-0,627]	[-1,536]	[-1,043]	[-1,696]	[0,635]
π_{t-1}^{d}	**0,464**	**0,137**	**0,759**	**1,617**	**-0,122**	**1,213**
	(0,845)	(0,119)	(0,229)	(2,145)	(0,145)	(0,204)
	[0,550]	[1,154]	[3,316]	[0,754]	[-0,843]	[5,936]
π_{t-2}^{d}	**0,561**	**-0,174**	**-0,060**	**-1,468**	**0,033**	**-0,430**
	(0,769)	(0,108)	(0,208)	(2,129)	(0,144)	(0,203)
	[0,729]	[-1,613]	[-0,288]	[-0,689]	[0,228]	[-2,122]
Constant	**4,815**	**0,915**	**-1,002**	**7,183**	**1,024**	**-0,304**
	(3,188)	(0,448)	(0,864)	(8,315)	(0,561)	(0,792)
	[1,510]	[2,043]	[-1,160]	[0,864]	[1,827]	[-0,384]
Trend	**-0,099**	**-0,005**	**0,024**	**-0,047**	**-0,005**	**0,018**
	(0,059)	(0,008)	(0,016)	(0,128)	(0,009)	(0,012)
	[-1,675]	[-0,620]	[1,480]	[-0,371]	[-0,524]	[1,451]
R-squared	0,788	0,793	0,862	0,707	0,789	0,858
Adj. R-squared	0,737	0,743	0,829	0,636	0,738	0,824
Sum sq. resids	254,618	5,029	18,691	1124,982	5,114	10,206
S.E. equation	2,963	0,416	0,803	6,223	0,420	0,593
F-statistic	15,405	15,830	25,881	10,002	15,499	25,103
Log likelihood	-88,184	-15,580	-39,868	-115,671	-15,889	-28,673
Akaike AIC	5,199	1,275	2,587	6,685	1,291	1,982
Schwarz SC	5,547	1,623	2,936	7,033	1,640	2,331
Mean dependent	4,219	2,043	2,462	0,633	2,043	2,616
S.D. dependent	5,777	0,821	1,940	10,329	0,082	1,415

Table 1 Estimation results

Short-term ERPT is measured as an impulse-response function to exchange rate shock. Because software IRF outputs are normalized to one standard deviation shocks (using the standard error of the regression for the estimated equation), we apply linear transformation as to obtain a one percentage shock to the exchange rate n_t. Figure 1 shows the properties of short-term pass-through for Czech Republic and Poland. In both cases, IRFs show an incomplete ERPT behavior that fades away quickly. Short-term pass-through for the Czech economy peaks in the second quarter at approximately 10%. In Poland, short-term IRF values decline from the first period's level of a 3% ERPT. It is important to note that for all calculated IRF periods the ± 2 s.e. confidence intervals encompass zero values, making the interpretation of results somewhat restricted. This is a problem inherent to many IRFs calculated from real data VAR models (see Hušek [7]). Other authors (e.g. Babetskaia-Kukharchuk [1]) face the same reasoning limitations, as well as their results also show IRF oscillations that sometimes cross over into negative values, bringing the possibility of *price puzzle* into short-term pass-through behavior. Poland's lower ERPT may be attributed to more prominent local currency pricing and some offsetting factors within the CPI basket, such as higher consumer market competitiveness (full inflation targeting regime is used in both countries).

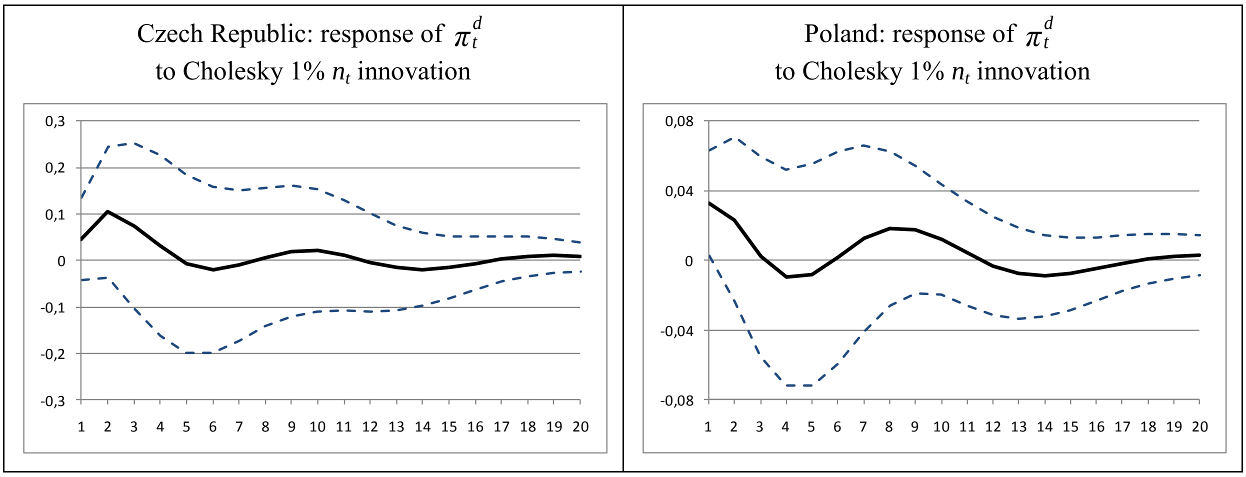


Figure 1 Short-term ERPT

Following the general concept of ERPT calculation, accumulated IRFs are frequently interpreted as the total or **long-term ERPT** (e.g. Faruqee [5]). Figure 2 shows the estimated long-term pass-through for Czech Republic and Poland. Presuming the five year-IRF (i.e. 20 quarters) period is sufficient to illustrate long-term ERPT behavior, we come to the conclusion that the total reaction to an exchange rate shock is approximately 25% for the Czech Republic and 8-10% for Poland. Again, data interpretation must take into account the confidence limits for the accumulated IRFs. Most of the exchange rate shock being transmitted (albeit incompletely) to inflation takes effect within the first 4 quarters.

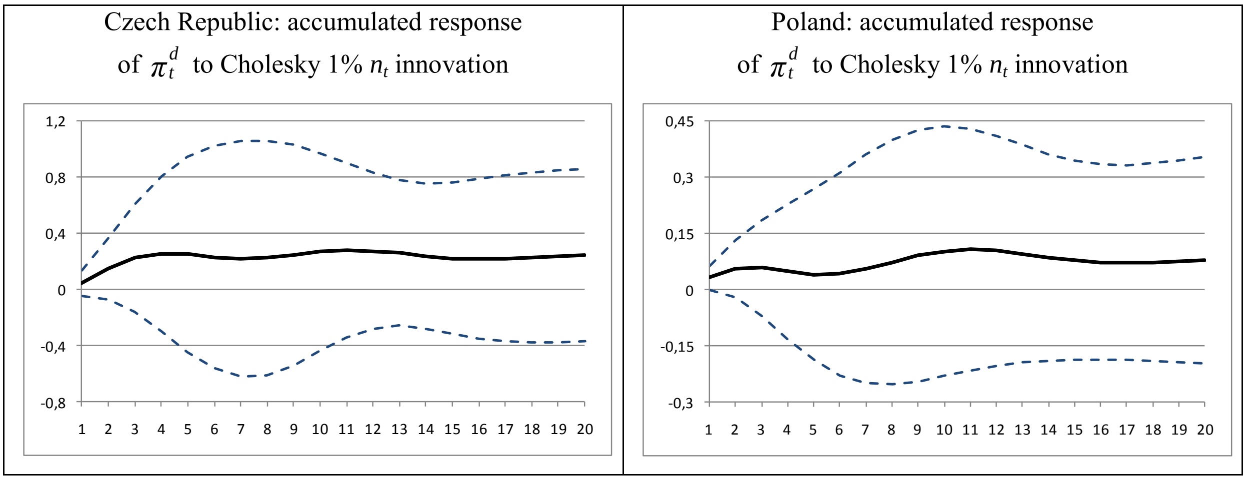


Figure 2 Long-term ERPT

The estimated long-term pass-through of 25% for the Czech Republic (acquired from the period 2001-2010) is in line with Babetskaia-Kukharchuk [1] who found a 30% ERPT for the period 1996-2006. Our results for Czech Republic and Poland also match the conclusions presented by Coricelli et al. [2] who identify inflation targeting and low-inflation environment as the main causes of low and declining ERPT.

5 Conclusions

This paper is focused on the study of exchange rate pass-through for Czech Republic and Poland. ERPT behavior describes the properties of exchange rate shock transmission into domestic prices. Using 2001-2010 quarterly data, we estimate a 3-variable VAR model for Czech Republic and Poland, upon which IRFs are constructed to assess the short-term and long-term pass-through of exchange rate shocks into domestic inflation. We find that ERPT for Czech Republic is higher, with short-term responses peaking around 10% and total (long-term) pass-through being 25%. We may interpret our results as follows: given a 1 percent nominal effective exchange rate shock, we expect the CPI inflation to rise by approximately 0,25% in the long run; according to our analysis, most of this response would take place within the first year. The pass-through for Poland is significantly lower as short-term response peaks at 3% and long-term ERPT stabilizes around 9%. Also, attention needs to be paid to different estimation and interpretation restrictions inherent in modeling ERPT through VAR models and IRFs.

Acknowledgements

The support from grant GAČR No. 402/09/0273 is gratefully acknowledged.

References

[1] Babetskaia-Kukharchuk, O.: Transmission of Exchange Rate Shocks into Domestic Inflation: The Case of the Czech Republic. *Czech Journal of Economics and Finance*, 2009, 59/2, 137–152.
[2] Coricelli, F., Jazbec, B. and Masten, I.: Exchange Rate Pass-Through in EMU Acceding Countries: Empirical Analysis and Policy Implications. *Journal of Banking and Finance*, 2006, Vol. 30/5, 1375-1391.
[3] Darvas. Z.: *Exchange Rate Pass-Through and Real Exchange Rate in EU Candidate Countries*. Economic Research Centre of the Deustche Bundesbank, Discussion Paper, 2001, 10/01.
[4] Faruqee, H.: *Exchange Rate Pass-Through in the Euro Area: The Role of Asymmetric Pricing Behavior*. IMF Working Paper, 2004, No. 04/14.
[5] Faruqee, H.: *Exchange Rate Pass-Through in the Euro Area*. IMF Staff Papers, 2006, Vol. 53, No. 1.
[6] Gagnon, J., Ihrig, J.: *Monetary Policy and Exchange Rate Pass-Through*. Board of Governors of the Federal Reserve System, International Finance Discussion Papers. 2002, No. 704 (July).
[7] Hušek, R.: *Ekonometrická analýza*. Praha: VŠE, Nakladatelství Oeconomica. 2007. ISBN 987-80-245-1300-3.
[8] Hušek, R.: *Aplikovaná ekonometrie – Teorie a praxe*. Praha: VŠE, Nakladatelství Oeconomica. 2009. ISBN 987-80-245-1623-3.
[9] McCarthy, J.: Pass-Through of Exchange Rates and Import Prices to Domestic Inflation in Some Industrialized Economies. *Eastern Economic Journal*, 2007, Vol. 33/4, 511-537.

Analysis of Exchange Rates Effects on the Main Czech Export Sectors in the Context of Real & Nominal Convergence

Roman Hušek[1], Michal Řičař[2]

Abstract. In the last decade the Czech Republic passed through a very interesting period of the economic development. A continuously increasing income for each citizen together with much more stable environment and an access to many opportunities not only in the borders but also in the new formed labor market of the EU and the world has been a strong impulse for further trade activities. On the one hand Czech firms have been allowed to use great market opportunities on the other hand they have been under a great pressure of foreign competitors. Further, to Czech exporters and importers has appeared a brand new problem connected with a floating exchange rate, a factor which has a strong power to influence revenues and costs of their activities hence their ability to survive in the long term view.
This paper aims to determine if revenues from the export of the main Czech exporters are significantly affected by the exchange rates of the main trade partners of the Czech Republic or there exist other and stronger factors which are able to compensate exchange rates volatility such as the real convergence of the CZ. For this purpose have been used econometric techniques in form of Vector Autoregressive (VAR) and Vector Error Correction (VEC) models.

Keywords: exchange rate, Czech export, export revenues, exchange rate effect, real convergence, nominal convergence.

JEL Classification: C10, C32, D21, F31

1 Introduction

A floating exchange rate in the modern economy is a component through which is being performed automatic compensations in different economic development across countries in a nominal price level, internationally demanded goods and services which reflect comprehensive and deep factors of the economic growth, such as productivity, capital facilities, quality of products etc. It could be said that a floating exchange rate is a healthy tool how to correct an economic distortion with respect to foreign economies in a short period of time and in the market way. Contrary, for export orientated firms a floating exchange rate is connected with a high risk which comes from different revenues and costs due to its strong volatility which is almost unpredictable. Modern companies use techniques such as VaR (Value at Risk) together with ARIMA and GARCH methods to determine whether and how they should be insured against future fluctuations in inflows and outflows. However, if an exchange rate is being appreciated systematically and continuously it is inevitable to react not only by speculations or with some kind of cover, e.g. forwards, futures, option and other market instruments.

The development of the Czech Republic from 1993 could be depicted as a continuously increasing performance in all sectors of the economy with linked growing standard of living for all inhabitants. This fact is closely connected with a foreign capital and investments which have come from more developed economies. Especially foreign direct investments (FDI) indisputably have helped to bring brand new technologies, know-how, new type of management, organization etc. which have caused rapid growth in productivity and efficiency. In contrast the market field has been filled up with capable competitors with foreign experience hence much higher knowledge about the market environment. Incoming FDI together with high level of foreign demand for new opportunities in the developing market have caused that the main exchange rates of the CZ appreciated by more than 35 % in the period 2000-2011 which has been huge challenge for domestic exporters.

In this context it looks necessary to examine whether and how these significant changes in the exchange rates have been reflected to the performance of the main Czech exporters captured by their revenues from export activities. For this purpose has been analyzed a wide and strong portfolio of export sectors together with composition of the main countries of foreign trade with CZ which is introduced in the second chapter. The third chapter examine consequences of the real and nominal convergence in the EU focused to the CZ. The fourth chapter introduced examined data and quantitative framework in a form of VAR and VEC thus could be examined if there exist feed-back effects, similar stochastic trends or rather one-way causality or non influence power of examined variables. The fifth chapter summarizes the results of this paper.

[1] University of Economics, Prague , W. Churchill Sq. 4, Czech Republic, husek@vse.cz.
[2] University of Economics, Prague , W. Churchill Sq. 4, Czech Republic, michal.ricar@vse.cz.

2 Main characteristics of the Czech export sector

The Czech Republic is characteristic with its wild portfolio of exported and imported goods. In 2010 was the GDP in the CZ 3 669 bil. of CZK at current prices where export of goods was 2 472 bil. and import was 2 338 bil. It is clear that consumption, investments (further development), employment, revenues etc. are closely connected with opened sectors. Table 1 shows the key sectors of the foreign trade of the CZ.

Industry Sector (NACE)		Export ration on total Export	Import Ration on Total Import
10	Manufacture of food products	2,75%	4,37%
20	Manufacture of chemicals and chemical products	4,39%	6,28%
22	Manufacture of rubber and plastic products	4,83%	4,66%
24	Manufacture of basic metals	4,29%	6,29%
25	Manufacture of fabricated metal products, except machinery	5,99%	4,18%
26	Manufacture of computer, electronic and optical products	16,26%	16,88%
27	Manufacture of electrical equipment	8,19%	6,16%
28	Manufacture of machinery and equipment n. e.c.	11,80%	8,21%
29	Manufacture of motor vehicles, trailers and semi-trailers	18,58%	9,24%
Sum		**77,09%**	**66,28%**

Table 1 Main Sectors of Export and Import in 2009
Source: Czech Statistical Office.

Export orientated sectors are primarily manufactures of vehicles, machinery and electronics which have the higher level of the exported than imported goods ratio. They **together form more than 65 % of total Czech export**; the rest of the manufacturers are focused on chemicals and agriculture products where the import ratio prevails over the export ratio.

From this perspective it is clear that an impact of exchange rate volatility to earnings of firms should be different across the sectors due to the fact that companies have to optimize their portfolios of revenues and costs at foreign currencies in various relations with respect to minimize their exposition to transaction risks.

A next important parameter is a composition of foreign countries and concrete exchange rates at which trade off is being made. Table 2 shows approximately 80 % of the export and the import divided by countries. Germany, Slovakia, Poland, France and the UK are the most important export partners, together form more than 42 % of total export. In contrast Germany, China, Japan and Russia are strategic partners for the import. The main trading currency is **the Euro** which **covers 79 % and 60 % of the export and the import trades** in the survey respectively. The Polish Zloty and the British Pound are the second and third most important currencies for export activities respectively. On the other hand the Chinese Yuan and the Japanese Jen are the main currencies for importers after the Euro.

Country	Currency	Export ration on Total Export	Import Ration on Total Import	Balance (mil. CZK)
Germany	EUR	33,43%	27,12%	162 182
Slovakia	EUR	9,31%	5,54%	84 374
Poland	PLZ	6,02%	6,53%	-2 707
France	EUR	5,83%	3,95%	43 360
United Kingdom	GBP	5,09%	2,21%	61 967
Austria	EUR	4,87%	3,72%	27 974
Italy	EUR	4,53%	4,47%	6 521
Netherlands	EUR	4,00%	3,42%	16 021
Hungary	HUF	2,64%	2,29%	9 968
Russia	RUB	2,40%	5,27%	-52 903
USA	USD	1,66%	2,16%	-7 641
China	CNY	0,77%	10,24%	-183 232
Japan	JNY	0,39%	3,21%	-54 428
Sum		**80,94%**	**80,13%**	**111 456**

Table 2 Main Countries of Export and Import in 2009
Source: Czech Statistical Office.

3 Real and nominal convergence of the EU countries

A transformation process in Middle and East Europe has serious consequences for the development of these economics. The European Union is wilder and stronger player in a political and economical field each year and incorporated and surrounded countries are caught by its massive development. This fact is closely linked with the real and nominal convergence of Middle and East Europe. Increasing foreign trade activities forcing countries to adapt their competitiveness, productivity, effectiveness mainly due to a FDI and connected increasing capital facilities thus they're achieving higher level of real output, this fact has been also closely examined by [8]. Contrary, West Europe has much higher price level calculated as purchasing power parity (PPP) which has been caused due to higher productivity in the past. This phenomenon has been generally explained by Ballassa-Samuelson effect. The authors [2], [7] have argued that a different level of productivity across countries causes that country with higher productivity has a higher price level but a nominal exchange rate doesn't reflect this level hence PPP is disturbed and more developed country can buy more products for cheaper prices in less developed country, vice versa. Main reason lies in the fact that international tradable goods have to have the same market price but non-tradable goods do not. Higher productivity in tradable sectors is connected with higher nominal wages without losing competitiveness for producers in the international field. But productivity in non-tradable sectors is almost the same across countries, e.g. services as a hairdresser's. Non-tradable sectors wouldn't work for low level wages if there were opportunities to transfer their production factors to tradable sectors with higher wages thus wages in more productive countries increase in non-tradable sectors also and inflation is rising up. Contrary a nominal exchange rate reflects only prices of tradable goods and their potential inflation which is immediately incorporated to a nominal exchange rate; otherwise there would be an arbitrary opportunity for lower priced products which would be immediately eliminated by international traders. Thus is explained how is possible that countries don't have PPP stable and equal to one in the long term. This fact has crucial consequences, **distortion from PPP means that a country with lower PPP has systematically underestimated a nominal exchange rate and wages which is very attractive for more developed countries**.

With the context of the real and nominal convergence of Middle and East Europe this conclusion has substantial impacts. If FDI and capital facilities constantly increasing, which is a case of the Czech Republic, thus productivity grows, which has been also shown in [5], then could be expected that nominal wages grow too but not only in tradable sectors but also in non-tradable. With this mechanism could be expected that PPP will be harmonized gradually. Figure 1 represents the real and nominal convergence of the EU27 in 2009 as a combination of a GDP per capita in PPP and a comparative price level.

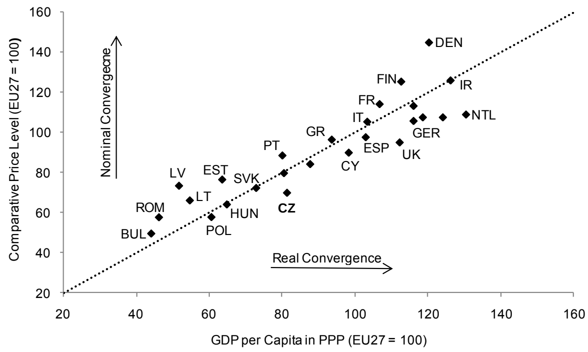


Figure 1 Real and Nominal Convergence in 2009
Source: Eurostat.

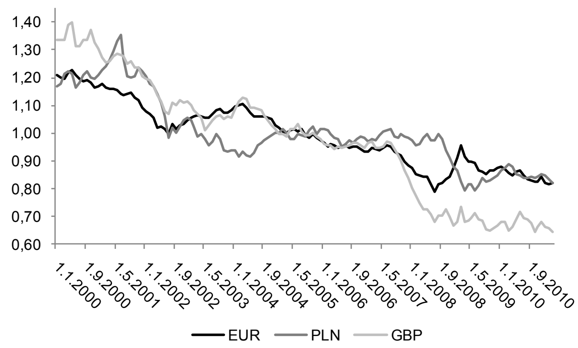


Figure 2 Main export exchange rates of the CZ
Source: The Czech National Bank.

It could be observed that the Czech Republic is underestimated in comparative price level regarding GDP per capita in PPP. More specifically, if the real development of the CZ is approx 81 % of the EU-27 then the nominal price level and the exchange rate convergence of the CZ should be around 81 % instead of 70 %, otherwise the nominal price level or the exchange rate is underestimated in the view of productivity and the EU-27. The Czech National Bank (CNB) applies inflation targeting coherently with the ECB, in other words, the nominal price level is almost the same as in the developed countries of the EU-27. Thus there exist only one channel through which PPP - in fact prices in the economy and the equality of the international purchasing power of the CZ - could be realized and that are the exchange rates of the CZ with its trade partners. Figure 2 demonstrates the main nominal export exchange rates related to their average level in 2005. Over the period **2000 – 2011 the exchange rates appreciated by more than 35 %,** as has been mentioned above. The conclusion is clear, **the nominal convergence is being achieved only through the exchange rates**, which has been further analyzed by [1] and [3]. This fact plays **crucial role for the Czech export sector which is continuously exposed to diminishing revenues from its export activities**.

More concretely, **Czech exporters could solve the long term appreciation trend in four ways hence eliminate an influence to their revenues**. (1) With **higher level of prices** of their goods on foreign markets. Total revenues in this case depend on demand elasticity which is related to quality of their products, marketing and affordable substitutes; on the very high competitive international market is this solution very improbable. (2) The second effect of appreciation could be partly positive due to **decreasing costs** of inputs from foreign countries but the most of costs are paid in the Crowns. Thus they are forced to decrease their costs through optimization of suppliers, wages, energy savings etc. These factors should increase operation efficiency. (3) The third way is in **increasing productivity of labor and capital through better capital facilities** which is possible due to FDI **and access to new technology equipment** on the international market. This leads to a total restructuring and optimization of a company and to a possibility to achieve an international competitive advantage which is crucial for a long survive of a company. Finally, (4) the fourth solution could be partially consequence of the third point, a possibility to **dispose with better and more competitive products which able to penetrate to new foreign markets thus a firm expends its activities and increases quantity of sold goods which eliminate decreasing price in a domestic currency**. In real business could be expected combinations of all mentioned factors; only this leads to the modern competitive company in the international field.

4 Exchange rate shocks and revenues of the export sectors

As has been described above, the real and nominal convergence of the Czech Republic is a strong phenomenon which systematically affects firms. Main question is whether and how the main exchange rates influence revenues of the export sectors. For this purpose could be used a well known VAR system together with numerous tests. Due to a VAR system it is possible to analyze shocks in appropriate variables with responses in examined variables, chosen export sectors. Thus could be analyzed if short term fluctuations have significant influence to performance of the export activities or whether they are able to protect themselves with better productivity and new instruments on the market.

4.1 Data and model details

Table 1 summarized sectors which have been analyzed in our VAR and VEC systems; variables are denoted regarding their numbers – M*10, M20, M22, M24, M25, M26, M27, M28, M29*. Whole spectrum of reaction functions could be found at the appendix of this document. Data has been obtained from the Czech Statistical Office as an index of revenues from exports divided by sectors. With respect to table 2 we have included the performance of the German (*GER*) and Slovak (*SK*) industry in the form of industrial production index from the Federal Statistical Office of Germany and the Statistical Office of the Slovak Republic respectively. Finally, the main export exchange rates – variables *EUR*, *PLN* and *GBP* – have been obtained from the Czech National Bank. Data has monthly frequency for the period 1/2002 - 3/2011 and related to the average of month 2005 (111 obs.). If seasonality has been proven then has been eliminated by the TRAMO/SEATS technique [4].

In the first step we have examined correlations between variables and we have tested if exist significant changes in various periods. This analysis has proven that the full length period has optimal characteristics without significant changes in correlations according the length of the period. The next step has tested an unit root by the ADF test which has proven that all time series are non-stationary; test with first differences has proven stacionarity for every time series. Then we have tested cointegration between *GER, SK, EUR, PLN* and *GBP* with chosen sector using the Johansen cointegration test and also here we have tested whether the length of the period has significant influence to a cointegration vector. If a cointegration between the sector and other variable(s) has been found we have estimated a non restricted VAC system including *GER, SK, EUR, PLN* and *GBP* for each sector separately with various lags depending on AIC, SC, HQ criteria and Wald test. Further, we have tested whether variables have significant coefficients to the sector and also if they are explained endogenously regarding whole system with Granger causality/Block exogeneity Wald test. Variables with non-significant coefficients and without Granger causality to the sector have been eliminated thus we have obtained a restricted VAR system. We have also preferred lag of a VAR system which represents the sector as an endogenous regarding whole system and the German and Slovak production index rather exogenous; we have assumed that single Czech sector doesn't have a power to influence whole production performance, especially in Germany, and also that the sector is connected to foreign countries rather than opposite way. If cointegration vector hasn't been found then we have used first differences and estimated VAR system and the same procedure with elimination of variables as in a VEC system has been used. During eliminations have been compared the information criteria.

If we denote a vector of all variables as $\boldsymbol{y}_t$ then a VAR system in lag p could be written as:

$$\mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\phi}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad (1)$$

where $\boldsymbol{\phi}_p$ is the matrix of parameters in lag *p*, $\boldsymbol{\varepsilon}_\mathrm{t}$ is *p*-dimensional Gaussian process of white noise with the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$. We estimate a VEC system from (1) as:

$$\Delta \mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{\mathrm{p}-1} \Delta \mathbf{y}_{t-p+1} + \boldsymbol{\Pi} \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_\mathrm{t}, \qquad (2)$$

where

$$\boldsymbol{\Gamma}_\mathrm{i} = -\mathbf{I}_l + \sum_{\mathrm{j}=1}^{\mathrm{i}} \boldsymbol{\Phi}_\mathrm{j} \ \text{ for } \ i = 1, \dots, p-1, \boldsymbol{\Pi} = -\left(\mathbf{I}_l - \boldsymbol{\Phi}_1 - \cdots - \boldsymbol{\Phi}_\mathrm{p}\right). \qquad (3)$$

It could be seen that the VEC system in the form (2) contains the matrix $\boldsymbol{\Pi}$ which incorporates long term influences which could be characterized as a common stochastic trend for connected variables and the matrix $\boldsymbol{\Gamma}_\mathrm{i}$ contains short term effects [6].

4.2 Impact of a shock in the exchange rates & the foreign economies to the Czech export sectors

Our analysis of each sector has shown that **the main explanatory power has the German index of industrial production** which has been significant for all sectors except 26 Manufacture of computer and electronic. **Other variables have occurred sporadically** and if so they have had a low explanatory power. Table 3 summarized all significant results. Cointegration vector has been found in 3 cases, sectors 27 – 29, and only with Germany. **This proves that the German industry is the main costumer for the main Czech exporters and in the long term significantly affects their performance**. The Slovak industry hasn't been found as an explanatory variable at any case which implies that the **development of the industrial sectors probably goes in a different trajectory**. Columns of the table with variables show at the first place the intensity of Cholesky one S.D. innovation (shock) from the first to tenth period in direction from the variable to the sector, at the second place is recorded variance decomposition from this shock for the same period. Thus could be seen the intensity of the shock for each sector in 10 months from the perspective of intensity and for every relevant variable in the system. For instance, when a shock in the first system occurs at the German industry then sector 10 reacts at the same time with the highest shock at level 2,24 and after 10 months the innovation from the first month is equal to -0,76, while the ratio from 0,12 to 0,35 corresponding to explained variance of the shock in sector 10 by the shock in the German industry; could be seen that innovation in the German equation has increasing power to explain the variance of the innovation of sector 10 which means that shocks in equations corresponding each other in increasing level in the time, in other words, future development of the German industry is being resonated to sector 10 steadily which proves their connection.

Industry	Sector	Coin.	Model	Lag	R^2	EUR	PLN	GBP	GER	SK
10	Manufacture of food products	-	VAR	3	0,37	-	-	-	2,24; -0,76 0,12; 0,35	-
20	Manufacture of chemicals and chemical products	-	VAR	3	0,24	0,32; -0,13 0,04; 0,08	-	-	1,62; -0,06 0,08; 0,18	-
22	Manufacture of rubber and plastic products	-	VAR	4	0,18	-	-	-	1,80; -0,09 0,28; 0,31	-
24	Manufacture of basic metals	-	VAR	2	0,12	-0,46; 0,01 0,02; 0,05	-	-	2,91; -0,09 0,15; 0,20	-
25	Manufacture of fabricated metal products, except machinery	-	VAR	5	0,28	-	-	-	3,40; -0,36 0,18; 0,30	-
26	Manufacture of computer, electronic and optical products	-	VAR	2	0,24	-	-0,92; 0,00 0,01; 0,04	-	-	-
27	Manufacture of electrical equipment	1	VEC	3	0,28	-	-	-	3,61; 5,41 0,04; 0,24	-
28	Manufacture of machinery and equipment n.e.c.	1	VEC	3	0,40	-	-	-	3,54; 2,33 0,28; 0,41	-
29	Manufacture of motor vehicles, trailers and semi-trailers	1	VEC	4	0,37	-	-	-	3,83; -1,03 0,29; 0,13	-

Table 3 Impact of the shock in the exchange rates and the foreign economies to the Czech export sectors
Source: Authors.

Table 3 reveals one very important conclusion, it could be said that **none of the main Czech export sector is significantly affected by the main exchange rates**. This could seem as a **confusing inference if we consider that the nominal convergence of the Czech Republic is being realized just through the exchange rates**. In contrast, it has been described that Czech firms have many new opportunities due to new markets and continuously incoming FDI also proving that the Czech market is an attractive environment for further development and has its future perspective. More concretely, **no effects of the exchange rates conclusively mean that Czech firms have to use last innovations, operational techniques, achieving higher productivity and efficiency, better marketing etc. thus they could realize constantly increasing revenues, sells and new costumers on foreign markets and simultaneously successfully face to continuing appreciation of the exchange rates** which is consequence of the nominal and real convergence. Thus real convergence, productivity is stronger factor than nominal and in the future we could expect that appreciation of the exchange rates will continue but also that Czech firms will be ready to new innovations and will be competitive enough to challenge new threats and opportunities. Thus the market of the Czech Republic and the level of living of Czech citizens will converge to more developed economies with implicit assumption that the development of the world economy will continue in optimal trajectory.

5 Conclusion

The paper has constituted a formal econometric assessment of the main exchange rates of the Czech Republic and the German and Slovak industry with impacts to revenues of the main Czech export sectors. Our main finding reveals that the performance of the Czech exporters wasn't being influenced by the main exchange rates of the Czech Republic in the period 2002/01 – 2011/03. In contrary it has been proven that the main export sectors are closely connected with the German industry which is the main costumer for Czech motor vehicles, trailers, semi products, machinery and electrical equipment, this fact has been shown by cointegration models and short term VAR models. The Slovak industry hasn't disposed with a relevant stochastic common trend and in the short term has been non-significant at models which imply that the development of the Czech and Slovak countries goes in a different direction.

In the context of the real and nominal convergence of the Czech Republic seems confusing that models haven't detected connections between the exchange rates and the export sectors regarding fact that the nominal exchange rates continuously appreciating from 2002. This phenomenon has been explained due to modern technologies, new styles of management, higher productivity and efficiency, marketing etc. as a consequence of the real convergence. Thus firms are able to consistently achieve lower product costs, otherwise they wouldn't be able to survive in the long term which is controversial claim regarding increasing revenues from a foreign trade. These findings lead us to the conclusion that the Czech market could expect a rising trajectory and trends associated with an increasing globalization effect.

Acknowledgements

We gratefully acknowledge the support of this project by the research grant GAČR 402/09/0273.

References

[1] Babecký, J., Bulíř, A., and Šmídková, K:. Sustainable Real Exchange Rates in the New EU Member States: Is FDI a Mixed Blessing? *European Economy Economic Papers* **368** (2009), 1-77. Available via the Internet: http://ec.europa.eu/economy_finance/publications/publication14297_en.pdf

[2] Ballassa, B.: The Purchasing-Power Parity Doctrine: A Reappraisal. *The Journal of Political Economy* **72** (1964), 584-596.

[3] Čihák, M., and Tomáš H.: Price Convergence in EU-Accession Countries: Evidence from the International Comparison. *Economie Internationale* **102** (2005), 59-82.

[4] Gómez, V., Maravall, A.: *TRAMO* (*Time Series Regression with ARIMA Noise, Missing Observations and Outliers) and SEATS (Signal Extraction in ARIMA Time Series): Instruction for the Users.* Madrid, 1997.

[5] Hunya, G., Holzner, M., and Wörz, J. How to Assess the Impact of FDI on an Economy. *OECD Concept Paper*, 2007. Available via the Internet: http://www.oecd.org/dataoecd/9/0/38937107.pdf

[6] Johansen, S.: *Likelihood-Based Inference in Cointegrated Vector Auto-regressive Models.* Oxford University Press, New York, 1995.

[7] Samuelson, P. A. : Facets of Balassa-Samuelson Thirty Years Later. *Review of International Economics* **2** (1994), 201–226.

[8] Sylwester, K.: Foreign Direct Investment, Growth and Income Inequality in Less Developed Countries. *International Review of Applied Economics* **19** (2005), 289–300.

Constant Harvesting in the Predator-Prey Model

Viktor Chrobok

Abstract The paper is focused on the Predator-Prey model modified in the case of harvesting one or both populations. Firstly there is given a short description of the basic model and its solution. Than the model is modified to provide a description of a constant harvesting. Solving this system requires a linearization, which was properly done and brought valuable results applicable even for the basic harvesting model. This model is generally diverging, but when just predator population is needed the oscillating character can be sustained.

Keywords: Predator-Prey model, Lotka-Volterra equation, ordinary differential equations, mathematical biology

JEL Classification: C62
AMS Classification: 92BXX

1 Introduction

The Predator-Prey model describes interactions between two species, where one of them is a prey to the other. Some examples of such populations are foxes and rabbits, lynxes and hares or various fish populations. The paper is focused on the case of introducing a third party to the system, which is harvesting one or both populations. The main assumption of this text is that this harvesting is done continuously. There will be given fundamental characteristics of the basic model and description of an addition of a constant harvesting element into the system.

The model is described by differential equations, which were firstly proposed independently by Alfred J. Lotka in 1925 and Vito Volterra in 1926. It is an useful tool for analysing development of biological systems in time, which is mainly the subject of mathematical biology. However when a harvesting modification is added, there arises also a problem from economist's point of view. Description of harvesting in the Predator-Prey model is currently quite rare and incomplete. Some mention about the topic can be found in [2], other resources covering this topic are currently unknown to the author of this paper.

2 The Basic Model

The Predator-Prey system is described by a following pair of first-order, non-linear, differential equations:

$$\frac{dx}{dt} = (a - by) * x \wedge \frac{dy}{dt} = (-m + nx) * y,$$

$$t \geq 0, x(0) = x_0, y(0) = y_0, \tag{1}$$

where x is the amount of the prey population, y is the amount of the predator population, t is time, a is a growth rate of the prey population, b shows how much of the prey population is killed by the predator population, m is a death rate of the predator population, n shows how much of the predator population is renewed by consuming the prey population. All parameters a, b, m, n are essentially positive constants and since x and y are representing the amount of prey and predator populations they are non-negative numbers.

It can be easily proved that for a positive number of both populations, there exists just one equilibrium point, $[x_e, y_e] = \left[\frac{m}{n}, \frac{a}{b}\right]$.The system given by (1) can be solved by a separation of variables to get the following formula:

$$\left(\frac{y}{y_0}\right)^a e^{-b(y-y_0)} = \left(\frac{x}{x_0}\right)^{-m} e^{n(x-x_0)}, \tag{2}$$

which implies that the system is not converging to the equilibrium, neither diverging from it, but oscillating; i.e. if $[x_0, y_0] \neq [x_e, y_e]$ then the system is periodically moving on a closed curve and gets back to $[x_0, y_0]$ after a specific period of time.

3 Graphical Representation of the Basic Model

Let's choose specific values of parameters described above and show a behaviour of the Predator-Prey system.

For the following figures were chosen these values of the parameters:

$$a = 5; b = 0.02; m = 10; n = 0.01; x_0 = 1000; y_0 = 500. \tag{3}$$

For plotting the results was used Maple v. 12, the stepsize for all figures in this paper was set to 0.01 time units. Those parameters of the model shall be used till the end of this paper, the differences in the initial values will be always especially stated. The figure showing the development of the system given by (1) with parameters and initial conditions given by (3) is plotted below. The arrows show the direction of the derivatives.

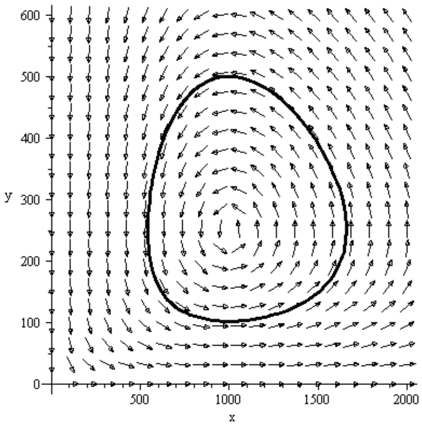


Figure 1. Basic Predator-Prey Model ([*x, y*] plot)

4 Constant Harvesting

This modification assumes that a harvester is always harvesting the same absolute amount of the harvested population. It can be illustrated on an example of the fisherman catching fish until the certain quota is fulfilled, but his harvestings are uniformly distributed in time. Modified equations are

$$\frac{dx}{dt} = (a - by)x - c \wedge \frac{dy}{dt} = (-m + nx)y - d, \tag{4}$$

where variable c is an absolute amount of prey harvested by a harvester and d is an absolute amount of predator harvested by a harvester (i.e. positive numbers).

The curves showing zero values of derivatives given by (4) are hyperbolas, which complicates finding of the equilibrium. However the equilibrium point exists (note that if just prey population is harvested and $c/a < m/n$ then the equilibrium does not exist) and is unique (specific prescription could be found in [1]).

4.1 Linearization

Since the stability of the system is not clear from known facts now a linearization of the system is needed. The process of linearizing is based on the Taylor's theorem and precisely done in [1], the following text will show just the most crucial steps of this procedure.

When apply Taylor near the equilibrium one gets a linear system given by:

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix} = \begin{bmatrix} \frac{c}{x_e} & -bx_e \\ \frac{dn}{nx_e - m} & nx_e - m \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} ax_e - 2c \\ -d - \frac{dnx_e}{nx_e - m} \end{bmatrix}. \tag{5}$$

For the stability of the system is necessary to know the eigenvalues of the homogenous system, which could be written as

$$\lambda_{1,2} = \frac{\frac{c}{x_e} - m + nx_e \pm \sqrt{\left(\frac{c}{x_e} + m - nx_e\right)^2 - 4\left(\frac{cm}{x_e} - cn + \frac{bdnx_e}{nx_e - m}\right)}}{2}. \tag{6}$$

Since the real part is always positive (note that it is zero for the basic system) the system is diverging from the equilibrium if it starts in any non-equilibrium point. The eigenvalues determine the solution of the homogenous system, when consider the equilibrium point as a particular solution one can get the following:

$$\begin{aligned} x(t) &= x_e + e^{\varphi t}[x_d \cos(\omega t) + (2i_1 y_d + 2i_2 x_d)\sin(\omega t)], \\ y(t) &= y_e + e^{\varphi t}[y_d \cos(\omega t) + (2i_3 y_d + 2i_4 x_d)\sin(\omega t)], \end{aligned} \tag{7}$$

where$x_d = x_0 - x_e$, $y_d = y_0 - y_e$ and i_1, i_2, i_3, $i_4$are generally real numbers, φ is the real part of λ and ω is the absolute value of the imaginary part of λ. Equations (7) imply that the speed of divergence is positively correlated with φ and that the system reaches extreme values with frequency ω.

The following plot shows the difference between the non-linearized (left side) and the linearized (right side) system:

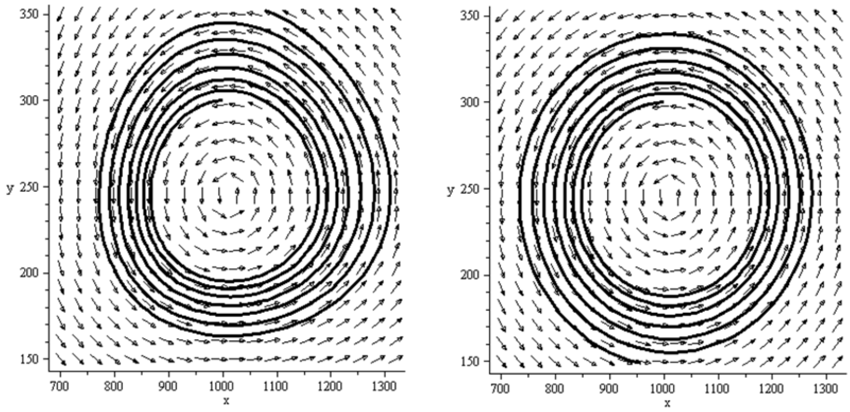


Figure 3 Constant Harvesting in the Predator-Prey Model ([x_0, y_0]=[100, 300])

4.2 Keeping the System Oscillating

Imagine that a harvester needs just predator population and that they do not care about the prey population. On the other hand they do not want the extinction of preys since this would lead to the extinction of the predators. Following the basic scheme of constant harvesting outlined before will lead to the extinction of both populations.

A possible solution of this problem could be not harvesting until the prey population reaches its maximum, then harvest prey till its minimum on the same cycle and put them separately to a place without any interactions. Now the prey could infinitely rise according to the model. A harvester could be continuously taking preys so that there is the same number of them all time and putting them to the original system. This would lead to the convergence of the system since it is the inverse operation to the harvesting. However the aim of the harvester is a constant harvesting of the predator population, not moving a system to the equilibrium. So they will just harvest predators as much as to keep the system oscillating, i.e. so that $\varphi = 0$.

Mathematically it means that identity

$$x_e = \frac{m+\sqrt{m^2+4cn}}{2n}, \quad (8)$$

where c stands for the number of preys added to the system, must be satisfied. After plugging for prey equilibrium value one could get the maximum number of predators available for harvesting in the oscillating system as a function of number of added preys:

$$d = c\frac{n}{b} * \frac{2a-m+\sqrt{m^2+4cn}}{m+\sqrt{m^2+4cn}}. \quad (8)$$

The function given by (10) is strictly increasing and asymptotically linear. It is strictly convex if $m > a$, strictly concave if $a > m$ and linear if $a = m$. This implies that a harvester could have increasing returns to scale in such system if the death rate of predators is lower than the growth rate of the prey population and diminishing returns to scale in the opposite situation. Moreover the following relationship holds when the system is purely oscillating:

$$\frac{c}{x_e} = \frac{d}{y_e}. \quad (9)$$

5 Conclusion

General linearization provides valuable information about the constant harvesting model's behaviour. We found out that the equilibrium point is unstable and that this model is diverging. The linearized system fits sufficiently close to the equilibrium, but not very well far from it, but even the non-linearized model would not provide usable results in the real world for the extreme values.

The thought of keeping the constant harvesting system oscillating when only predators are needed for a harvester, could be applied to various agriculture businesses. Derived relationships should provide a valuable help for any harvester dealing with such system. This modification is also a fruitful area for future research when incentives to harvest more or giving up of some present profit with a vision of higher profit in the future when some positive interest rate or finite horizon is added to the problem.

References

[1] Chrobok, V.: *Harvesting in the Predator-Prey Model*, VŠE, Praha 2009.

[2] Clark, C. W.: *Mathematical Bioeconomics: the optimal management of renewable resources*, John Wiley, New York 1976.

[3] Doucet, P., Sloet, P. B.: *Mathematical Modeling in the Life Sciences*, 2nd edition, Redwood Books, Trowbridge 1993.

[4] Hájková, V., John, O., Zelený, M.: *Matematika*, Karolinum, Praha, 2003.

[5] Lee, H. J., Schiesser, W. E.: *Ordinary and Partial Differential Equation Routines in C, C++, Fortran, Java, Maple, and MATLAB*, Chapman & Hall/CRC, London 2004.

[6] Pánková, V.: *Nelineární optimalizace pro ekonomy*, Professional Publishing, Praha 2003.

[7] Rektorys, K.: Přehled užité matematiky, SNTL, Praha 1973.

[8] Varian, H. R.: *Intermediate Microeconomics: a modern approach*, W. W. Norton & Company, New York 2006.

Evaluating the Efficient Market Hypothesis by means of isoquantile surfaces and the Hurst exponent

Kristýna Ivanková[1], Ladislav Krištoufek[2], Miloslav Vošvrda[3]

Abstract. This article extends our previous work on applications of isoquantile (formerly *isobar*) surfaces to market analysis. The approach is applied to lagged returns of selected stock market indices and compared to various estimations of the Hurst exponent.

We evaluate the Efficient Market Hypothesis by means of the two aforementioned approaches for the ASPI, BET, BUX, JSX, NASDAQ, PX and S&P500 indices. The more does a time series satisfy the EMH, the closer it resembles Brownian motion. In this case isoquantile surfaces form a circle and the Hurst exponent approaches $\frac{1}{2}$.

Keywords: isoquantile, isobar, Hurst exponent, Efficient Market Hypothesis, stock market index

1 Introduction

This article applies two approaches to evaluate the Efficient Market Hypothesis for selected stock market indices. The notion of an *efficient market* was introduced in Fama [6]. We study the particular version of EMH stating that returns of efficient stock market indices follow the behaviour of Brownian motion; as for this article, closeness to this ideal will be understood as a measure of efficiency.

The first section discusses the isoquantile approach introduced in [4] (originally called the *isobar* approach). For a fixed center point and a level u, an isoquantile encloses the u-th quantile of a multi-dimensional data distribution, mapping every direction from the center to a particular distance. The term "isoquantile" is used for both the mapping and its image (a continuous surface). Since isoquantiles of varying levels u continuously cover the whole domain, points in it can be ordered by comparing the levels of their (unique) encompassing isoquantiles.

The second author of [4] focused on practical estimation of isoquantile shapes. The article [9] considers estimation of the edge of the bounded support ($u \to 1$) using nonparametric regression and [10] extends this method for unbounded support using asymptotical location and isoquantiles.

In this article, we follow the approach of [9] and limit ourselves to *homothetic* isoquantiles (i.e. isoquantiles of varying levels are assumed to be scaled copies of each other). In our prior article [8] we've formulated two isoquantile shape estimation methods (differing in the chosen coordinate system) and applied them on the PX and NASDAQ indices.

The next section discusses the Hurst exponent [7, 5, 3], a measure of fractal dimension. It's commonly used in financial analysis as an indicator of long-term persistence of time series. We've chosen four commonly used methods for its estimation: detrended fluctuation analysis, rescaled range analysis, detrending moving average and height-height correlation analysis.

The third section is concerned with an application. We discuss the means to apply the approaches to a time series, summarize the tested indices and discuss the results.

[1]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Praha 1; Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08, Praha 8, e-mail: kristyna.ivankova@gmail.com
[2]Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, 110 00, Praha 1; Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08, Praha 8, e-mail: kristoufek@ies-prague.org
[3]Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08, Praha 8, e-mail: vosvrda@utia.cas.cz

2 Isoquantile surfaces

Isoquantiles require us to work in polar coordinates. Choosing the Euclidean space, the transformation of a non-zero vector $\mathbf{x} \in \mathbb{R}^d$ to generalized polar coordinates is given by

$$r = \|\mathbf{x}\|_2 , \quad \theta = \frac{\mathbf{x}}{\|\mathbf{x}\|_2},$$

where $\|\mathbf{x}\|_2$ is the Euclidean norm of the vector $\mathbf{x}$. Observe that the generalized angle θ lies on S^{d-1}, the sphere of unit radius in $\mathbb{R}^d$.

We'll use the definition of isoquantile as it appears in [4], page 2: For every $u \in (0, 1)$, the u-level isoquantile is defined as a mapping of a fixed θ to the value of the inverse distribution function of the Euclidean distance from the origin: $\theta \to F^{-1}_{R\,|\,\Theta}(u)$. The name "$u$-level isoquantile" will also be used interchangeably for the surface $S_u = F^{-1}_{R\,|\,\Theta}(u)$ determined by each θ with a fixed quantile u in the inverse of the conditional distribution function $F^{-1}_{R\,|\,\Theta}$.

We'll assume our sample to originate from the random variable $X = (R, \Theta)$. Assume continuity of the marginal density $f_\Theta(\theta)$, conditional density $f_{R\,|\,\Theta}(r\,|\,\theta)$ and the conditional distribution function $F_{R\,|\,\Theta}(r\,|\,\theta)$. The distribution function is assumed to be invertible, the introduced mapping continuous and strictly positive.

A rigorous definition of the introduced ordering is as follows. Consider a sample of n independent realizations of the random variable X, e.g. $X_i = (R_i, \Theta_i)$, $1 \le i \le n$. For every i there exists an unique u_i-level isoquantile containing the point X_i. Denoting $X_{i,n}$ the realizations ordered by their respective quantile values u_i, the maximum value is given by the point $X_{n,n}$ which belongs to the upper-level isoquantile with level $\max_{1\le i\le n} u_i$.

In practice, we'll assess the 1-level isoquantile on the grounds of the *asymptotical location property* as described in [10]. For large n, the furthest points from the origin lie near the $\frac{n-1}{n}$-level isoquantile. The 1-level isoquantile is then simply the edge of the bounded support.

Isoquantile estimation is performed by the non-parametric regression of [9, 10]. For the estimation we'll assume *homotheticity* of isoquantiles. The function $v(\theta)$ corresponds to the 1-level isoquantile and unambiguously describes the shape of all isoquantiles. The distribution of $\frac{\mathbf{x}}{v(\theta)}$ is spherically symmetric.

We estimate $v(\theta)$ using radial regression:

$$w(\theta) = E(R\,|\,\Theta = \theta) = c\, v(\theta),$$

The estimate of the expected value of R given $\Theta = \theta$ describes the shape of 1-level isoquantile up to a multiplicative constant. This constant is chosen in a way that the estimated expected value shape $\hat{w}(\theta)$ contains the whole data after scaling:

$$\hat{v}(\theta) = \frac{\hat{w}(\theta)}{\hat{c}}, \quad \text{where } 1/\hat{c} = \max_{1\le i\le n} \tfrac{R_i}{\hat{w}(\Theta_i)}$$

For practical estimation we'll use the two parametrizations introduced in [9] (hyperspherical) and [8] (unit sphere projection). For details and rationale see our previous work [8].

To enable quantitative comparison with the methods of the following section, we're introducing a numerical measure that preserves index ordering from the previously used visual assessment. This novel market efficiency measure is computed via Fourier analysis of the estimated isoquantile shape; due to size constraints, details will be revealed in an upcoming article.

3 Hurst exponent

For comparison, we present the results for another measure frequently used as a measure of market efficiency − the Hurst exponent H. The exponent $0 < H < 1$ is a characteristic measure of long-range dependence in the time series. For $H > 0.5$, the series is persistent, i.e. following a trend; while for $H < 0.5$, the series is anti-persistent, i.e. switching more frequently than a random series does. Therefore, a deviation from $H = 0.5$ indicates possible profitable trading as there are long-range correlations in the series. Out of many Hurst exponent estimators, we use the most popular ones for the financial series – detrended fluctuation analysis, rescaled range analysis, detrending moving average and height-height correlation analysis.

Rescaled range analysis (RS) is the most traditional of the methods, proposed by [7]. According to [14], the time series of length $2^{\upsilon_{max}}$ is divided into the sub-periods of length 2^{υ}. For each sub-period, the range R of the profile is calculated as well as the standard deviation of the increments S. The rescaled range R/S_{υ}, based on the average rescaled ranges for each sub-period with length υ, scales as $R/S_{\upsilon} \propto \upsilon^{H}$. In our application, we set $\upsilon_{min} = 4$ and $\upsilon_{max} = \log_2 T$.

Detrended fluctuation analysis (DFA), proposed by [13], is based on scaling of variance of the detrended series. In the procedure, the profile (cumulative return deviations from the average) of the time series of length T is divided into sub-periods of length s and for each sub-period, a linear fit $X_s(t)$ of the profile is estimated. A detrended signal $Y_s(t)$ is then constructed as $Y_s(t) = X(t) - X_s(t)$. Fluctuation $F^2_{DFA}(s)$, defined as an average mean squared error from the linear fit over all sub-periods of length s, scales as $F^2_{DFA}(s) \propto s^{2H}$ [11]. To get reasonable estimates of H, we set $s_{min} = 5$ and $s_{max} = T/5$.

Detrending moving average (DMA), proposed by [1], is based on a moving average filtering. For a set window size λ, we construct a centered moving average $\bar{X}_{\lambda}(t)$ for each data point $X(t)$ of the series. Similarly to DFA, fluctuations $F^2_{DMA}(\lambda)$, defined as the mean squared error of $X(t)$ from $\bar{X}_{\lambda}(t)$, scale as $F^2_{DMA}(\lambda) \propto \lambda^{2H}$. As the centered moving average is used, we set $\lambda_{min} = 3$ and $\lambda_{max} = 21$ with a step of 2.

Height-height correlation analysis (HHCA), proposed by [2], is based on a scaling of height-height correlation function of the series $X(t)$ with time resolution ν and $t = \nu, 2\nu, ..., \nu\lfloor \frac{T}{\nu} \rfloor$ (where $\lfloor\rfloor$ is a lower integer operator). Second-order height-height correlation function of $X(t)$ is then defined as $K_2(\tau) = \sum_{t=1}^{\lfloor T/\nu \rfloor} |X(t+\tau) - X(t)|^2 / \lfloor T/\nu \rfloor$ where time interval τ generally ranges between $\nu = \tau_{min}, \ldots, \tau_{max}$. $K_q(\tau)$ then scales as $K_q(\tau) \propto \tau^{2H}$. In the following, we set $\tau_{min} = 1$ and $\tau_{max} = 20$.

4 Evaluating the efficient market hypothesis – stock market indices

The efficient market hypothesis states that returns (closing−opening price) of market indices in efficient markets behave ideally like Brownian motion (see e.g. [12]). In practice, this assumption is violated mostly by the periodic structure (day, week, quarter, year) of agent behaviour. Further bias mostly reveals non-rational behaviour, non-zero information costs or delayed reactions. Our goal is to measure the efficiency of a market using both the isoquantile and Hurst exponent approach and to compare them.

Our data consists of weekly closing and opening prices for the past ten years (sample size around 500) obtained from the Reuters Wealth Manager service.

Firstly we'll describe the results for isoquantiles. The y-axis denotes the current value of stock market index returns, the x-axis denotes their lagged values. Under the efficient market hypothesis, the isoquantile shape for this configuration should be close to a circle (since Brownian motion is independent to itself when lagged). The results were computed for lags between one and sixteen weeks. Image 1 shows examples of various isoquantile shapes for the assessed stock market indices. Our previous work [8] contains complete depictions of 1–14-week lags for the PX and NASDAQ indices.

We've applied the methods on seven stock market indices. We'll shortly summarize them before presenting the results:

- The All Share Price Index: 241 Sri Lankan stocks of the Colombo Stock Exchange
- The BET Index: 10 Romanian stocks of the Bucharest Stock Exchange.
- The BUX Index: 13 Hungarian stocks of the Budapest Stock Exchange.
- The JSX Composite Index: 379 Indonesian stocks of the Indonesia Stock Exchange.
- The NASDAQ Composite Index is comprised of 2742 stocks of the NASDAQ Stock Market.
- The PX Index is comprised of 14 stocks of the Prague Stock Exchange (only five of which are Czech).
- S&P500: 500 stocks traded on NYSE or NASDAQ.

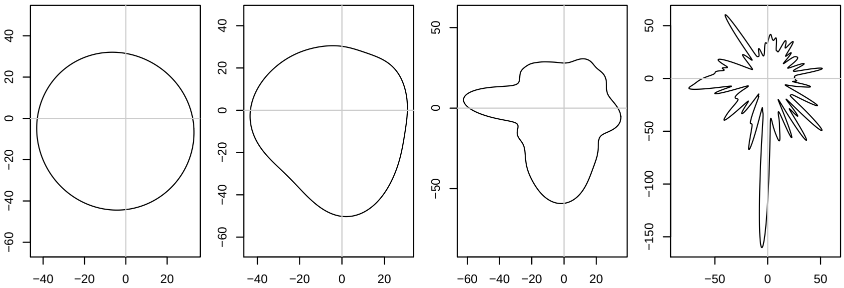


Figure 1: Examples of isobar shapes.

Isoquantile shapes for the All Share Price Index, NASDAQ Composite Index and S&P500 are very close to circles. Small deviations from the circle shape can be observed in ASPI (lags 1, 3 and 5), NASDAQ (lags 12, 13 and 16) and in S&P500 (lags 12, 13 and 16). Deviations in the 13-week lag can be explained by the expected quarterly periodicity of agent behaviour. Based on visual examination, the underlying markets of ASPI, NASDAQ and S&P500 may follow the efficient market hypothesis.

Isoquantile shapes for BET differ from circles in multiple lags (of 2, 3, 4, 11 and 13 weeks): the deviations are distinctive, which suggests short-time dependency in the data.

The isoquantile shapes of the PX Index, BUX and JSX Composite Index deviate from circles constantly: for PX it's the longer lags of 4, 7, and 9–15 weeks, for BUX it's 3 and 5–16 weeks. Isoquantiles for the JSX Composite Index don't resemble a circle for any lag. Observing a systematic deviation from independence between current values and lagged ones, we can postulate that the efficient market hypothesis doesn't apply to markets described by these indices.

The first parametrization prefers rounder shapes; isoquantiles resemble a circle more often. The second parametrization follows the data shape better.

We proceed by comparing both parametrizations with the four chosen methods of Hurst exponent estimation. Image 2 depicts both the isoquantile measures on horizontal axes plotted against Hurst exponent estimations on vertical axes. The hyperspherical isoquantile parametrization is shown in the left column, the projection parametrization in the right one. Rows subsequently depict the four Hurst exponent estimation methods in this order: rescaled range analysis (RS), detrended fluctuation analysis (DFA), detrending moving average (DMA) and height-height correlation analysis (HHCA). Each of the images additionally contains a linear fit.

The ideally effective market has isoquantile measure of zero and Hurst exponent equal to 0.5; the isoquantile measure increases with decreasing similarity to Brownian motion while the Hurst exponent approaches 1 for persistent time series. Both of these phenomena signify lower efficiency, so the measures should be positively correlated: this prediction holds for rescaled range analysis, detrended fluctuation analysis and height-height correlation analysis. The DFA method fulfills this prediction best (see e.g. [14]), followed by the RS and HHCA methods (DFA differs for NASDAQ having a lower value than S&P500, and for JSX being higher than PX). The most distant index from the linear fit is ASPI - according to the isoquantile approach it's among the most Brownian-like indices while the Hurst exponent shows pronounced persistence. This result is possibly caused by the fact that we've used short lags (1 to 16 weeks) in the isoquantile approach and the Hurst exponent measures long-term dependece. We interpret this by stating that ASPI shows dependence only in the long term. The results for DMA show almost no linear relationship between the Hurst exponent and the isobars measure. This difference is caused mainly by the low DMA estimation for JSX. Based on the construction of the methods, this indicates that JSX exhibits seasonal or cyclical behavior.

5 Conclusion

We've contrasted two approaches for studying the time dependence in time series data: the isoquantile approach (formerly called *isobar*) for short-time dependence (1–16 week lags) and the Hurst exponent for

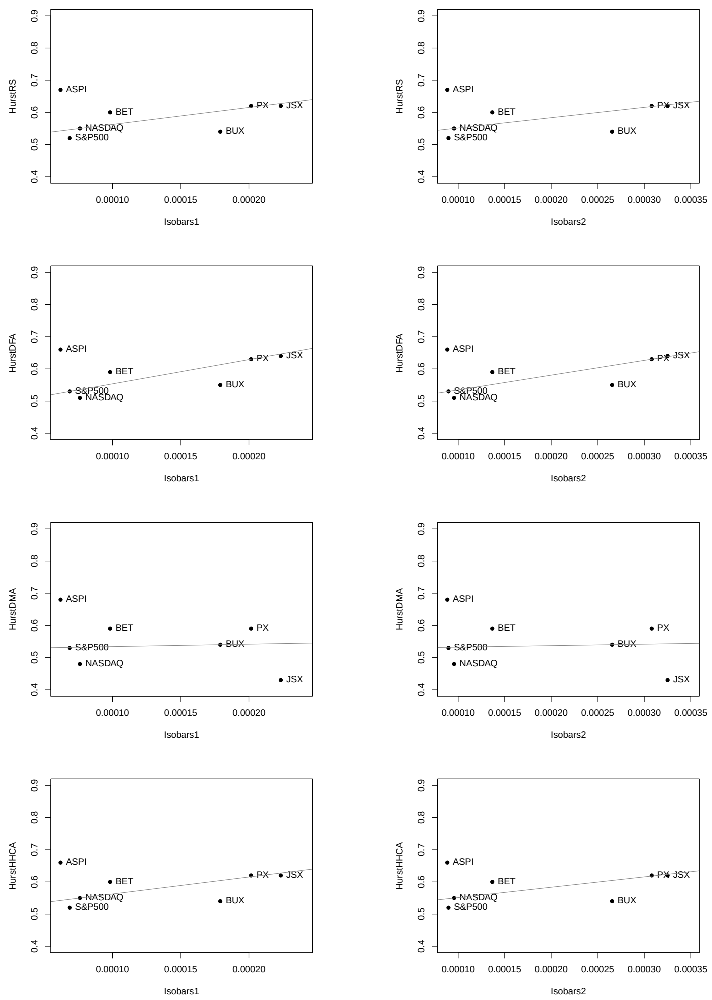


Figure 2: Isoquantile measures plotted against various Hurst exponent estimators.

long-time dependence. We've compared two parametrizations for isoquantiles (hyperspherical coordinates and unit sphere projection) and four methods for estimating the Hurst exponent (detrended fluctuation analysis, rescaled range analysis, detrending moving average and height-height correlation analysis). Using these methods we've tested the EMH for selected indices: the All Share Price Index, the NASDAQ Composite Index, S&P500, BET, PX Index, BUX and the JSX Composite Index. Since none of tested indices have shown strong anti-persistence, we've assumed a positive correlation between isoquantiles and the Hurst exponent – an assumption confirmed for three of four Hurst exponent estimation methods.

According to our results, the isoquantile approach and the Hurst exponent approach complement each other nicely, each focusing on a different dependency scale.

Acknowledgements

This work has been supported by the Czech grant agency Grant 402/09/H045, the Grant Agency of Charles University (GAUK) under project 118310 and project SVV 261 501.

References

[1] Alessio, E., Carbone, A., Castelli, G., and Frappietro, V.: Second-order moving average and scaling of stochastic time series. *European Physical Journal B* **27 (2)** (2002), 197–200.

[2] Barabasi, A.-L., Szepfalusy, P., and Vicsek, T.: Multifractal spectre of multi-affine functions. *Physica A* **178** (1991), 17–28.

[3] Cajueiro, D.O., and Tabak, B.M.: The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Physica A* **336/3–4** (2004), 521–537.

[4] Delcroix, M. F., and Jacob, P.: Stability of extreme value for a multidimensional sample. *Stat. Analyse Donées* **16** (1991), 1–21.

[5] Eom, C., Oh, G., and Jung, W.: Relationship between efficiency and predictability in stock price change. *Physica A* **387/22** (2008), 5511–5517.

[6] Fama, E.: Random Walks in Stock Market Prices. *Financial Analysts Journal 5* **21** (1965), 55–59.

[7] Hurst, H.E.: Long term storage capacity of reservoirs. *Transactions of the American Society of Engineers* **116** (1951), 770–808.

[8] Ivanková, K.: Application of isobars to stock market indices. *Mathematical Methods in Economics Proceedings* (2010).

[9] Jacob, P., and Suquet, C.: Regression and edge estimation. *Stat. Prob. Lett.* **27** (1996), 11–15.

[10] Jacob, P., and Suquet, C.: Regression and asymptotical location of a multivariate sample. *Stat. Prob. Lett.* **35** (1997), 173–179.

[11] Kantelhardt, J., Zschiegner, S., Koscielny-Bunde, E., Bunde, A., Havlin, S., and Stanley, E.: Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. *Physica A* **316(1-4)** (2002), 87–114.

[12] Osborne, M. F. M.: Brownian motion in the stock market. *Operations research* **March-April 1959**, 145–173.

[13] Peng, C., Buldyrev, S., Havlin, S., Simons, M., Stanley, H., and Goldberger, A: Mosaic organization of DNA nucleotides. *Physical Review E* **49/2** (1994), 1685–1689.

[14] Weron, R.: Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A* **312** (2002).

Efficiency evaluation of a branch network using DEA models

Josef Jablonský[1]

Abstract. The paper deals with efficiency analysis of branch network of one of the Czech firms. The analysis is based on two-stage DEA model. First stage of the analysis measures external efficiency where the main output is the number of transactions of the branch which is taken as one of the inputs in the second stage. This stage evaluates internal efficiency of the branch. Total efficiency of the branch is given by synthesizing both external and internal efficiencies. The system for efficiency evaluation is illustrated on real data set with 67 branches.

Keywords: data envelopment analysis, efficiency, two-stage model, multiple criteria decision making.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The paper is focused on efficiency analysis of branch network of a general firm operating using its branch selling and service network by means of data envelopment analysis (DEA) model. DEA models are relatively often applied in practice as a tool for efficiency analysis in different fields. There are described many applications in finance and insurance (efficiency evaluation of bank or insurance companies branches), health care (efficiency of hospitals or their parts), education (benchmarking of schools) or other non-profit sectors in literature. An extensive review of DEA application can be found e.g. in Emrouznejad et al. [5]. In the Czech Republic real-world studies using data envelopment analysis as main modelling tool are not frequent. Among others it is possible to mention studies presented in Dlouhy et al. [4] and Grmanova and Jablonsky [6].

Most of the applications of DEA use standard models formulated by Charnes et al. [2] and Banker et al. [1] or some of their modifications presented in following years. The application described in this paper deals with firm's branch network. The DEA model for its efficiency analysis can be inspired by many real-world DEA applications in efficiency evaluation of bank branches or network branches of firms in different sectors. The model presented in this paper is quite easy but general enough and can be applied for any firms with an extensive selling or service network (e.g. mobile operators, photo shops or any other selling and service branches). The model presented in this paper uses two-stage efficiency analysis. The first stage evaluates external efficiency while the second one deals with internal efficiency of the branches. The overall efficiency is given as the combination of both stages.

The paper is organized as follows. Section 2 presents formulation of standard DEA models including two-stage models. Section 3 presents an original model for efficiency evaluation of firm's branch network. The model is verified in Section 4 on the set of 67 branches of one of the Czech firms. Some concluding remarks are presented in final section of the paper.

2 DEA models for two-stage production processes

Let us suppose that the set of decision making units (DMUs) contains n elements. The DMUs are evaluated by m inputs and r outputs with input and output values x_{ij}, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$ and y_{kj}, $k = 1,2,\dots,r$, $j = 1,2,\dots,n$, respectively. The efficiency of the q-th DMU can be expressed as the weighted sum of outputs divided by the weighted sum of inputs with weights reflecting the importance of single inputs/outputs v_i, $i = 1,2,\dots,m$ and u_k, $k = 1,2,\dots,r$ as follows:

$$\theta_q = \frac{\sum_{k=1}^{r} u_k y_{kq}}{\sum_{i=1}^{m} v_i x_{iq}}. \tag{1}$$

[1] University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, W. Churchill Sq. 4, 13067 Praha 3, Czech Republic, jablon@vse.cz

Standard CCR input oriented DEA model formulated by Charnes et al. [2] consists in maximization of efficiency score (1) of the DMU$_q$ subject to constraints that efficiency scores of all other DMUs are lower or equal than 1. The linearized form of this model is as follows:

$$
\begin{aligned}
&\text{maximize} && \theta_q = \sum_{k=1}^{r} u_k y_{kq} \\
&\text{subject to} && \sum_{i=1}^{m} v_i x_{iq} = 1, && && (2)\\
& && \sum_{k=1}^{r} u_k y_{kj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, && j = 1,2,\ldots,n, \\
& && u_k, v_i \ge \varepsilon, && k = 1,2,\ldots,r,\ i = 1,2,\ldots,m.
\end{aligned}
$$

If the optimal value of the model (2) $\theta^*_q = 1$ then the DMU$_q$ is CCR efficient and it is lying on the CCR efficient frontier, otherwise the unit is not CCR efficient. The model (2) is often referenced as primal CCR model. Its dual form is sometimes more convenient and its mathematical model is as follows:

$$
\begin{aligned}
&\text{minimize} && \theta_q \\
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, && i = 1,2,...,m, && (3)\\
& && \sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, && k = 1,2,...,r, \\
& && \lambda_j \ge 0, && j = 1,2,\ldots,n,
\end{aligned}
$$

where λ_j, $j = 1,2,\ldots,n$ are weights of DMUs, s^-_i, $i = 1,2,\ldots,m$, and s^+_k, $k = 1,2,\ldots,r$ are slack (surplus) variables and θ_q is the efficiency score of the DMU$_q$ which expresses necessary reduction of inputs in order this unit becomes efficient.

The models (2) and (3) are CCR models with input orientation, i.e. they look for reduction of inputs in order to reach the efficient frontier. The output oriented modification of the presented models is straightforward. The BCC models under variable returns to scale assumptions originally presented by Banker et al. [1] extend the formulation (3) by convexity constraint $\sum_j \lambda_j = 1$.

The models (2) and (3) measure the relative efficiency of one-stage transformation of m inputs into r outputs. The transformation of inputs into final outputs can be taken as a two-stage process. The inputs of the first stage are transformed into outputs and all or at least some of these outputs are utilized as inputs of the second stage that are using for production of final outputs. Let us denote the input values of the first stage x_{ij}, $i = 1,2,\ldots,m$, $j = 1,2,\ldots,n$ and the output values of the first stage y_{ij}, $i = 1,2,\ldots,r$, $j = 1,2,\ldots,n$. Supposing that all outputs of the first stage are taken as inputs of the second stage and that the final output values are z_{ij}, $i = 1,2,\ldots,p$, $j = 1,2,\ldots,n$, the two-stage DEA model under constant returns to scale assumption can be formulated according to Chen et al. [3] as follows:

$$
\begin{aligned}
&\text{minimize} && \theta_q - \phi_q \\
&\text{subject to} && \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, && i = 1,2,...,m, && (4)\\
& && \sum_{j=1}^{n} y_{kj}\lambda_j \ge \tilde{y}_{kq}, && k = 1,2,...,r, \\
& && \sum_{j=1}^{n} y_{kj}\mu_j \le \tilde{y}_{kq}, && k = 1,2,...,r,
\end{aligned}
$$

$$\sum_{j=1}^{n} z_{lj}\mu_j \geq \phi_q z_{lq}, \qquad l = 1,2,...,p,$$
$$\theta_q \leq 1, \phi_q \geq 1,$$
$$\lambda_j \geq 0, \mu_j \geq 0, \qquad j = 1,2,\dots,n,$$

where λ_j and μ_j, $j = 1,2,\dots,n$, are weights of the DMUs in the first and second stage, θ_q and φ_q efficiency scores of the DMU$_q$ in the first and second stage and $\tilde{y}_{kq}$ are variables to be determined.

The DMU$_q$ is recognized as efficient according to the model (4) if the efficiency scores of both stages equal to 1, i.e. $\theta_q = 1$, $\varphi_q = 1$ and the optimal objective function value of the presented model is 0. The inefficient units can be ranked relatively by the following geometric average efficiency measure:

$$e_q = (\theta_q / \varphi_q)^{1/2} \tag{5}$$

Two stage models are quite popular and there are published many papers with this topic. Among the newest ones is the paper of Paradi et al. [10] presenting the model for efficiency evaluation of bank branches. This approach is similar to the model which is described in the following section.

3 General efficiency model for two-stage firm's branch network

Overall efficiency of branch network depends on many input and output factors. The main goal of branches consists in generating of financial contribution by realization of its business activities (mainly sales and service support). In order to reach this main goal two groups of efficiencies must be on an appropriate level. One of them is external (volume) efficiency which expresses the ability of the branch to transform its localized potential into corresponding number of customers visiting the branch and realizing a transaction (sales or service demand). That is why the main output influencing the external efficiency is the *number of transactions* or customers. These two dates are not identical but the number of transactions is easily available in operator's statements. There are many input factors influencing the external efficiency (the number of customers). Among them we use the following ones:

- *Operational expenses* (rental costs, wages and overheads) cover attractiveness of the locality of the branch and its size (greater branch has more employees and that is why higher wages).

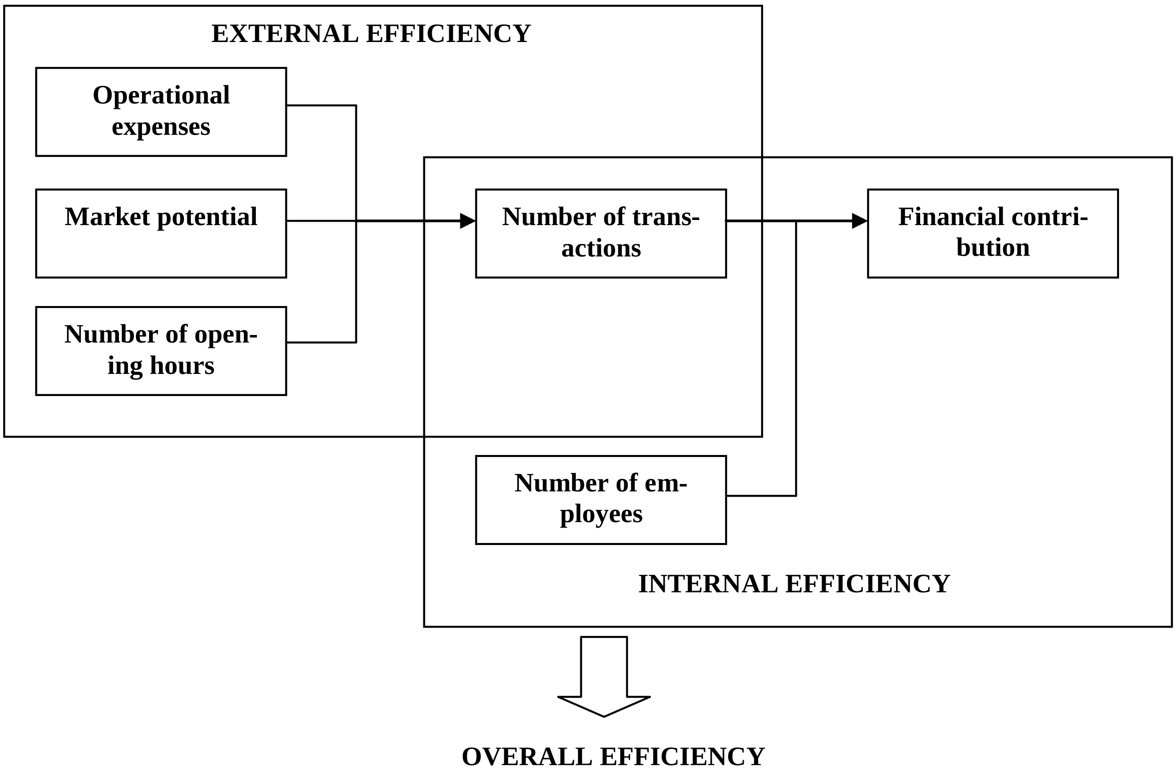


Figure 1 Two-stage efficiency model

- *Market potential* measures the potential of business area of the branch. It depends on the number of inhabitants of the area and the number of other branches of the same or competitive operator within the given area. In our study it is measured by the number of inhabitants per one branch within the area.
- *Number of business hours* per year of the branch is an important characteristic influencing total number of transactions.

Internal efficiency measures the ability of the branch to transform its potential into an appropriate financial contribution. In our model the main output of internal efficiency is financial contribution of the branch expressed in Czech currency unit (CZK). Second possible output is *ICCA score* which measures the satisfaction of customers. Finally we did not use this output in the model because of its low discrimination among the branches of the data set which was available. Two inputs are taking into account in the model. One of them – the number of transactions – is the output of external efficiency. The second one is the number of full time employees.

Overall efficiency is a combination of both particular efficiencies. The model for efficiency evaluation is presented in Figure 1 (see Sabata [11] for more details). The mathematical formulation is a slight modification of model (4). It looks as follows:

$$
\begin{aligned}
\text{minimize} \quad & \theta_q - \phi_q \\
\text{subject to} \quad & \sum_{j=1}^{n} x_{ij}\lambda_j \le \theta_q x_{iq}, \qquad i = 1, 2, 3, \\
& \sum_{j=1}^{n} y_j\lambda_j \ge \tilde{y}_q, \\
& \sum_{j=1}^{n} y_j\mu_j \le \tilde{y}_q, \\
& \sum_{j=1}^{n} t_j\mu_j \le t_q, \\
& \sum_{j=1}^{n} z_j\mu_j \ge \phi_q z_q, \\
& \theta_q \le 1, \phi_q \ge 1, \\
& \lambda_j \ge 0, \mu_j \ge 0, \qquad j = 1,2,\ldots,n,
\end{aligned}
\tag{6}
$$

where $\mathbf{y} = (y_1, y_2,\ldots, y_n)$ is the vector of only output of the first stage (number of transactions), $\mathbf{t} = (t_1, t_2,\ldots, t_n)$ is the vector of one of the inputs of the second stage (number of employees), and $\mathbf{z} = (z_1, z_2,\ldots, z_n)$ is the vector of only output of the second stage (financial contribution). The remaining symbols have the same meaning as in model (4).

4 Results and discussion

Efficiency measures the ability of the evaluated unit to transform the inputs into the outputs and using DEA models it is measured in a relative way by comparison to other units of the data set. Another view on the performance of the selling and service branches is their ability to create profit (profitability) which is one of the most important characteristics for shareholders. Profitability of the branch can be defined as the ratio (financial contribution – operational and other expenses)/financial contribution, and it is expressed in %.

Table 1 presents results of efficiency and profitability analysis based on the data set of 67 branches of one of the Czech firms. Due to the limited space of this paper Table 1 contains results for the best 20 branches identified by model (6) only. The columns of this table contain:

- External (θ_q) and internal (φ_q) efficiency score given by model (6) under variable returns to scale assumptions (the model is extended by convexity constraints - sum of λ and μ variables is 1).
- Overall efficiency given by (5).
- Profitability of the branches calculated from the source data set (not presented in this paper).

Rank	External	Internal Efficiency	Overall	Profitability
1	1,000	1,000	1,000	73%
2	1,000	1,000	1,000	53%
3	0,841	1,000	0,917	73%
4	1,000	1,244	0,896	73%
5	1,000	1,273	0,886	70%
6	0,851	1,113	0,874	71%
7	0,938	1,329	0,840	62%
8	1,000	1,435	0,835	66%
9	0,900	1,296	0,833	67%
10	0,925	1,340	0,831	64%
11	0,913	1,330	0,829	54%
12	0,911	1,337	0,826	64%
13	0,679	1,000	0,824	55%
14	0,922	1,401	0,811	62%
15	0,986	1,550	0,798	78%
16	1,000	1,582	0,795	70%
17	1,000	1,606	0,789	60%
18	0,737	1,186	0,788	56%
19	0,957	1,605	0,772	76%
20	0,883	1,516	0,764	70%

Table 1 Efficiency and profitability

Table 1 shows that the external efficiency score for first 20 units is on average much higher that their internal efficiency score. External score is greater than 0,9 for almost all branches listed in Table 1but their internal score (for comparison its reciprocal values) is much worse in typical cases. More detailed information about distribution of efficiency scores of all branches is given in Table 2. According to their efficiency scores and profitability the branches can be split into several classes. Best branches are those having overall efficiency and profitability better than upper quartile. Only 4 branches of 67 ones fulfil this condition. The worse branches are those with overall efficiency and profitability worse than lower quartile (6 branches of the data set). A deeper analysis is necessary for identification of main factors of inefficiencies of the particular branches. It is not possible within the given scope of the paper.

	External	Internal Efficiency	Overall	Profitability
Minimum	0,665	1,000	0,301	35%
Lower quartile	0,851	1,502	0,609	62%
Median	0,938	1,922	0,687	70%
Upper quartile	0,987	2,405	0,789	73%
Maximum	1,000	8,707	1,000	78%
Mean value	0,907	2,267	0,686	66,4%
Std. deviation	0,098	1,423	0,142	9,13

Table 2 Statistical characteristics of efficiency and profitability

5 Conclusions

The paper presents an original procedure for efficiency evaluation of branch network of one of the Czech firms. The procedure is based on a two-stage DEA model which evaluates internal and external efficiency of the branch. Relative external efficiency of all branches is quite high – many of them work on the efficient frontier and the worse one has its external efficiency score 0,665, i.e. this branch must reduce its inputs by one third approximately in order to reach efficient frontier. On the contrary relative internal efficiency has much higher variability (Table 2) – only a few branches are efficient and the worse branch must increase its financial contribution more than eight times in order to work efficiently. The analysis shows almost no relation between efficiencies (internal, external, overall) and profitability. There are some branches with very low overall efficiency but their profitability is on quite high level and branches with high efficiency and low probability on the contrary. An interesting conclusion is given by comparison of results of here presented results with firm's common practice. The presented model was created for one concrete firm but it can be used with minor modifications for any firms operating using their branch network.

Acknowledgements

Research is supported by the Grant Agency of the Czech Republic - project no. 402/09/0231, and by the Internal Grant Agency of the University of Economics Prague - project F4/18/2011.

References

[1] Banker, R.D., Charnes, A., and Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1984, 1078–1092.

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

[3] Chen, Y., Liang, L. and Zhu, J.: Equivalence in two-stage DEA approaches. *European Journal of Operational Research*, 193, 2009, 600-604.

[4] Dlouhy, M., Jablonsky, J. and Novosadova, I. : Using data envelopment analysis for efficiency evaluation of Czech hospitals. *Politicka ekonomie*, 1, 2007, 60-71.

[5] Emrouznejad, A., Parker, B. and Tavares, G.: Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Journal of Socio-Economic Planning Science*, 42(3), 2002, 151-157.

[6] Grmanova, E. and Jablonsky, J.: Efficiency analysis of Slovak and Czech insurance companies using data envelopment analysis models. *Ekonomicky casopis*, 57(9), 2009, 857-869.

[7] Jablonsky, J.: Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling*, 46, 2007, 1091–1098.

[8] Jablonsky, J. A two-stage AHP/DEA model for evaluation of efficiency. In: *Mathematical Methods in Economics 2009*, University of Life Sciences Press, Praha, 2009, 159-164.

[9] Liang, L., Li, Z.Q., Cook, W.D. and Zhu, J.: Data envelopment analysis efficiency in two-stage networks with feedback. *IIE Transactions*, 43(5), 2011, 309-322.

[10] Paradi, J.C., Rouatt, C. and Zhu, H.: Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*, 39(1), 2011, 99-109.

[11] Sabata, O.: *Efficiency evaluation of mobile operator's branch network using data envelopment analysis models* (in Czech). University of Economics, Prague, 2010.

[12] Zhu, J.: *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Kluwer Academic Publishers, Boston, 2002.

Propensity to criticalness in the PERT method, the expectation of time and distance of activities from project beginning

Marek Jakubík[1]

Abstract. Project management deals with the planning and managing of projects. One of the most popular and well known critical path method is Program Evaluation and Review Technique. In this method, activity durations are random variables, depending on the real random elements. Estimated duration of each task is known only with certain probability. Even this method is not quite perfect. It has some pitfalls, which can largely affect results of this method. Uncertainty of a task does not exactly determine whether a task is really critical, or whether it even provides a slack. Activities' propensity to criticalness depends not only on estimates of activity durations, but also on the distance of activity from project beginning, activity's duration, structure of a network diagram, project manager's experiences etc. Each activity is a part of one or more project network paths. Such activity – in case it is not the final one - can be quite long, FS dependent and still not critical because many alternative paths with various duration estimates can influence its criticality.

Keywords: propensity, criticalness, PERT method, activity duration

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Project Management presents a certain way of beforehand planning and realising variously difficult events, which need to be realised in a required deadline and with beforehand planned financial and source costs in order to meet determined goals. As [5] states and further elaborates, project management can be also briefly characterised as effective and efficient reaching significant changes. It concerns an important part of managerial decision-making process.

As [4] states and further elaborates, project management uses tasks solved by network analyses methods, such as the PERT method (Program Evaluation and Review Technique). Since we work with rateable data, these data are burdened with a statistical error. To reach significantly high probability we use such activities arrangement which ensures the meeting of the deadline of the whole project in time. For the PERT model calculation we use a qualified estimate of individual activities duration lengths.

A weak point is the fact that activity duration is not known accurately, but it is only given with a high probability. This is mentioned in [2]. Activity duration is a random quantity with a certain probability distribution. Therefore, even a pessimistic estimate of activity duration is a random quantity dependent on an expert's qualified estimate. However, such an estimate is subjective and hence activity criticalness is given with certain probability depending on the selection of all activity duration estimates. Within the total context the due date of the whole project can be influenced by individual activities duration estimates, and thus it is important to determine the criticalness of individual activities. Among indicators (activity duration estimate time and its criticalness) there is a mathematical relation which is used in practice for instance in heuristic approaches. The longer the task is, the more threatening it is for meeting the deadline of the whole project. The dispersion of activity duration depends on the beta distribution which is discussed in [3]. The pessimistic estimate of activity duration is one of its components. In practice the estimate is often exceeded. Therefore, the estimate influence on activity criticalness is often researched. The determination of the relation between pessimistic duration estimate and activity criticalness is the topic of this paper. The paper endeavours to find and describe a functional relation between a pessimistic duration estimate b_{ij} and the activity criticalness (i,j) using tests on a selected example.

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

2 Materials and methods

Since a qualified estimate of individual activities duration is composed of three indicators, an optimistic estimate of activity duration (i,j) a_{ij}, a pessimistic estimate of activity duration (i,j) b_{ij} and the most probable estimate of activity duration (i,j) m_{ij}, we will focus on one of them, i.e. a pessimistic estimate of activity duration (i,j) b_{ij} and its influence on the activity criticalness (i,j).

Using the indicators of activity duration we calculate basic statistical characteristics of the activity, or an expected duration t^e_{ij} and its dispersion $\sigma^{2(}t_{ij})$. Both characteristics are in line with corresponding beta distribution characteristics for duration.[2]

Using standard PERT calculations we calculate mean values and dispersions of all deadlines possible at the earliest and the latest for all activities and nodes and the determination of the so-called expected critical path. Based on a probability analysis of calculated parameters we can judge the probability of the development of time slack and further the probability of a particular critical path, the probability of meeting a planned deadline of the project etc. [1].

3 Results and discussion

During the testing, the activity duration in a pessimistic estimate b_{ij} changed and its influence on the activity criticalness was observed (i,j). Other estimates of duration remained constant and thus the influence was more remarkable. The dispersion of activities time parameters directly depends only on pessimistic estimate of activity duration, as shown in the example in Figure 2. For better observation of a pessimistic estimate influence, the so-called criticalness potential was established, which can be calculated as follows:

$$Pk_{ij}=(p_i/h_i)(b_{ij}/T) \tag{1}$$

Where

Pk_{ij} ... criticalness potential.
p_i... the probability of a critical path gateway through this node,
h_i ... number of tasks ending in the node,
b_{ij} ... pessimistic estimate of activity (i,j) duration,
T... total project duration

Further, the normalisation of a criticalness potential was carried out due to a percentage observation of the influence on the activity criticalness and for better orientation among activities. It is calculated as a share from the criticalness potential to the total sum of all criticalness potentials. The limit for the choice of criticalness potential from a normalised state is one over the number of activities in the whole project. Using the potential we observed the influence of a pessimistic estimate of activity duration on activity criticalness. If the criticalness potential normalisation is higher than the limit we receive the information that this activity is slightly more influential for the whole project duration than others. This serves as a quick survey of pessimistic estimate influence on possible activity criticalness. The higher the criticalness potential, the closer is the activity duration tie to the whole project duration, and hence its influence on its criticalness is increased.

The influence of a pessimistic estimate of activity duration on the activity criticalness (i,j) was tested by the model demonstrated in Figure 1. In the network of the tested model there are 17 activities and 12 nodes with node 1 being the initial node and node 12 being the terminal one. Based on particular formulas we calculated all possible basic statistical characteristics of the activities and also expected duration periods t^e_{ij} and their dispersions $\sigma^2(t_{ij})$.

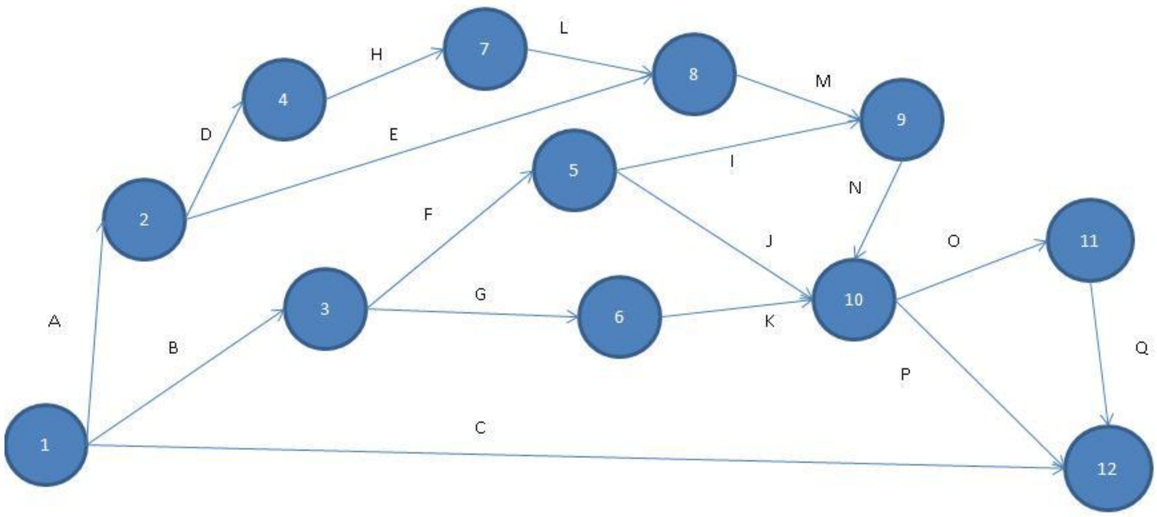


Figure 1 Network of the tested model

Activity	(i,j)	Optimistic estimate of activity duration(i,j) a_{ij}	Pessimistic estimate of activity duration (i,j) b_{ij}	Most probable estimate of activity duration i (i,j) m_{ij}	Expected duration time t^e_{ij}	Dispersion $\sigma^2(t_{ij})$	Criticalness potential normalisation	Criticalness potential
A	(1, 2)	1	20	1	4,166666667	10,0277778	0,226628895	0,5
B	(1, 3)	1	1	1	1	0	0,011331445	0,025
C	(1, 12)	1	30	1	5,833333333	23,3611111	0,339943343	0,75
D	(2, 4)	1	1	1	1	0	0,005665722	0,0125
E	(3, 8)	1	1	1	1	0	0,005665722	0,0125
F	(3, 5)	1	1	1	1	0	0,005665722	0,0125
G	(3, 6)	1	1	1	1	0	0,005665722	0,0125
H	(4, 7)	1	1	1	1	0	0,005665722	0,0125
I	(5, 9)	1	1	1	1	0	0,002832861	0,00625
J	(5, 10)	1	1	1	1	0	0,002832861	0,00625
K	(6, 10)	1	19	1	4	9	0,107648725	0,2375
L	(7, 8)	1	1	1	1	0	0,005665722	0,0125
M	(8, 9)	1	1	1	1	0	0,011331445	0,025
N	(9, 10)	1	1	1	1	0	0,014164306	0,03125
O	(10, 11)	1	1	1	1	0	0,011331445	0,025
P	(10, 12)	1	20	1	4,166666667	10,0277778	0,226628895	0,5
Q	(11, 12)	1	1	1	1	0	0,011331445	0,025
						Total:	1	2,20625

Table 1 Testing on the model - activities A, C, K, P

Table 1 presents the model tested on activities A, C, K and P. The criticalness potential selected all four activities; however, a critical path leads through nodes 1-2-4-7-8-9-10-12, which is demonstrated in Table 2. Hence only activities A and P are really critical, the other two are not on the critical path.

Nodes	Deadline mean value of the earliest activity node occurrence (i,j) T^0_i	Deadline mean value of the latest allowable activity node occurrence (i,j) T^1_i	Mean value of critical (interferential) reserve μ (R_i)
1	0	0	0
2	4,166666667	4,166666667	0
3	1	4,166666667	3,166666667
4	5,166666667	5,166666667	0
5	2	7,166666667	5,166666667
6	2	5,166666667	3,166666667
7	6,166666667	6,166666667	0
8	7,166666667	7,166666667	0
9	8,166666667	8,166666667	0
10	9,166666667	9,166666667	0
11	10,16666667	12,33333333	2,166666667
12	13,33333333	13,33333333	0

Table 2 Critical path – activities A, C, K, P

Activity	(i,j)	Optimistic estimate of activity duration(i,j) a_{ij}	Pessimistic estimate of activity duration (i,j) b_{ij}	Most probable estimate of activity duration i (i,j) m_{ij}	Expected duration time t^e_{ij}	Dispersion $\sigma^2(t_{ij})$	Criticalness potential normalisation	Criticalness potential
A	(1, 2)	1	1	1	1	0	0,019512195	0,041666667
B	(1, 3)	1	1	1	1	0	0,019512195	0,041666667
C	(1, 12)	1	40	1	7,5	42,25	0,780487805	1,666666667
D	(2, 4)	1	1	1	1	0	0,009756098	0,020833333
E	(3, 8)	1	1	1	1	0	0,009756098	0,020833333
F	(3, 5)	1	1	1	1	0	0,009756098	0,020833333
G	(3, 6)	1	1	1	1	0	0,009756098	0,020833333
H	(4, 7)	1	1	1	1	0	0,009756098	0,020833333
I	(5, 9)	1	1	1	1	0	0,004878049	0,010416667
J	(5, 10)	1	1	1	1	0	0,004878049	0,010416667
K	(6, 10)	1	1	1	1	0	0,009756098	0,020833333
L	(7, 8)	1	1	1	1	0	0,009756098	0,020833333
M	(8, 9)	1	1	1	1	0	0,019512195	0,041666667
N	(9, 10)	1	1	1	1	0	0,024390244	0,052083333
O	(10, 11)	1	1	1	1	0	0,019512195	0,041666667
P	(10, 12)	1	1	1	1	0	0,019512195	0,041666667
Q	(11, 12)	1	1	1	1	0	0,019512195	0,041666667
						Total:	1	2,135416667

Table 3 Testing on the model - activity C

Table 3 demonstrates the model being tested on activity C. The criticalness potential selected the activity C; however, the critical path is 1-2-4-7-8-9-10-11-12, which is demonstrated in Table 4. Therefore, activity C is not critical. As we can see, the normalised potentials are higher at the beginning and at the end of the graph even though the expected duration is the same for all, except for activity C. This is caused by the structure of the graph, because at the beginning of the graph there are not many variants through which the critical path could lead, and the same is true for the end of the graph. The potentials at the end of the graph are higher also due to the fact that three activities J,K and N terminate in one node – number 10 -, and thus the potential that the critical path will lead through this node is higher. Moreover, other activities have many alternative paths through which the critical path can lead.

Nodes	Deadline mean value of the earliest activity node occurrence (i,j) $T^0{}_i$	Deadline mean value of the latest allowable activity node occurrence (i,j) $T^1{}_i$	Mean value of critical (interferential) reserve $\mu\,(R_i)$
1	0	0	0
2	1	1	0
3	1	3	2
4	2	2	0
5	2	4	2
6	2	5	3
7	3	3	0
8	4	4	0
9	5	5	0
10	6	6	0
11	7	7	0
12	8	8	0

Table 4 Critical path - activity C

4 Conclusion

From the model situations it follows that the proposed relation between a pessimistic activity duration and its criticalness is not explicit because it depends on the structure of the net model. However, it can give us precise information through which part of the model a critical path will most probably lead. If we select one activity based on this relation, the fact that this activity is really critical cannot be justified. The solution is not accurate; however, by further arranging the relation and possible enlarging the relation, we could achieve better results. The application of the proposed solution may serve as a fast survey and orientation in the model, provided the relationship can be improved and described in a better way.

References

[1] Anderson, D. A., Sweeney, D. J., Williams, T. A.: Introduction to Management Science: *Quantitative Approaches to Decision Making*. West Publ., Minneapolis, 1994.

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

[3] Herreri´As-Velasco, J. M. et Al.: *Revisiting the PERT mean and variance*. European Journal of Operational Research Volume 205, 2010.

[4] Jablonský, J.: *Operační výzkum: Kvantitativní modely pro ekonomické rozhodování*. Professional Publising, Prague, 2002.

[5] Jonas, D.: *Empowering project portfolio managers: How management involvement impacts project portfolio management performance*. International Journal of Project Management Volume 28, 2010.

Lower Bound on Large p-Median Problem Using the Sequential Zone Adjustment Method

Jaroslav Janáček[1]

Abstract. This paper deals with an approximate approach to the p-median problem. Large instances of the p-median problem can be often met when a real public service system is designed and discomfort of its customers is minimized. Attempts at exact solving real-world instances of the problem using a location-allocation model often fail due to enormous computational time. An approximate approach, which reformulates the p-median problem to a covering problem, seems to be an admissible way how to master the computational complexity. Nevertheless, the faster computational process is paid for by a loss of accuracy. This contribution deals with a sequential method of the dividing point deployment to obtain a tighter lower bound of the solved p-median problem and to provide a user of the approximate approach with an estimation of accuracy.

JEL Classification: C61
AMS Classification: 90C06

1 Introduction

The p-median problem, which consists of selection of at most p center location at the network nodes so that the sum of network distances from each node to the nearest located center is minimum, constitutes a basic model for designing optimal structure of public service system [2], [5], [6], [7], [8]. To describe the p-median problem on a network, we denote J a set of served nodes, similarly, symbol I denotes a set of possible center locations. We will study here only following formulation of the p-median problem, where it is necessary to determine at most p locations from I so that sum of network distances from each element of J to the nearest location is minimum. The network distance between a possible location i and customer j from J is denoted as d_{ij} and a matrix of all distances between elements of I and J will be denoted as $\{d_{ij}\}$ The basic decisions in any solving process of the p-median problem concern location of centers at network nodes from the set I. To model these decisions at particular nodes, we introduce the zero-one variable $y_i \in \{0,1\}$, which takes the value of 1 if a center should be located at location i, and which takes the value of 0 otherwise. The p-median problem can be formulated as follows:

$$Minimize \qquad F(\mathbf{y})=\sum_{j\in J} min\{d_{ij}: i\in I, y_i=1\} \text{ subject to } \sum_{i\in I} y_i \le p, y_i \in \{0, 1\} \text{ for } i\in I \tag{1}$$

Attempts at exact solving real-world instances of the problem using a location-allocation model and some exact solving technique of mathematical programming often fail due to enormous computational time [4]. An approximate approach, which reformulates the p-median problem to a covering problem, seems to be an admissible way how to overcome the computational complexity. This approach uses an approximation of a common distance between a service center location and a customer by some of pre-determined distances. To improve the covering approach to p-median problem, we have developed a sequential method of the dividing point deployment. Nevertheless, the faster computational process is paid for by a loss of accuracy. This contribution deals with technique of obtaining a tight lower bound of the solved p-median problem to provide a user of the approximate approach with estimation of accuracy.

2 Formulation of the p-Median Problem as the Covering Problem

The presented approximate approach [3] is based on a relaxation of the assignment of a service center to a customer. The unique assignment of a service center to a customer is broadly used in location-allocation problems [4], [5] and it enables to specify the distance between a customer and assigned center. The presented approach approximates the distance between a customer and the nearest center unless the assignment must be specified. The range $<0, max\{ d_{ij}: i\in I, j\in J \}>$ of all possible distances is partitioned the into $r+1$ zones. The zones are separated by a finite ascending sequence of dividing points $D^1, D^2, \ldots, D^r$, where $0 = D_0 < D_1$ and $D_r < D_m = max\{ d_{ij}: i\in I, j\in J \}$. A zone k corresponds with the interval $(D_k, D_{k+1}>$, the starting zone (k=0) corresponds with the

[1] [1] Faculty of Management Science and Informatics, University of Žilina, 01026 Žilina, Slovak Republic, jaroslav.janacek@fri.uniza.sk.

interval (D_0, D_1> and so on, up to the zone r, which corresponds with the interval (D_r, D_m>. A length of the interval k is denoted by e_k for k= 0, …, r.

The covering-type model of the p-median problem is formulated using binary variables y_i and x_{jk} . A variable $y_i \in \{0,1\}$ takes the value of 1 if a center should be located at location i, otherwise it takes the value of 0. An auxiliary variable x_{jk} takes the value of 1 if the distance of the customer $j \in J$ from the nearest located center is greater than D_k and this variable takes the value of 0 otherwise. Then the expression $e_0\, x_{j0} + e_1 x_{j1} + e_2 x_{j2} + e_3 x_{j3} + \ldots + e_r x_{jr}$ is an upper approximation of d_{ij}. If the distance d_{ij} belongs to the interval (D_k, D_{k+1}>, it is estimated by upper bound D_{k+1} with a possible deviation e_k. Similarly to the covering model [3], we introduce zero-one constant a_{ij}^k for each triple $i, j, k,$ where $i \in I$, $j \in J$, and k=0, ..., r. The constant a_{ij}^k is equal to 1 if and only if the distance between the served customer j and the possible center location i is less or equal to D_k, otherwise a_{ij}^k is equal to 0. Then the mathematical description follows:

$$\text{Minimize} \qquad \sum_{j \in J} \sum_{k=0}^{r} e_k\, x_{jk} \tag{2}$$

$$\text{Subject to} \qquad x_{jk} + \sum_{i \in I} a_{ij}^k\, y_i \geq 1 \qquad \text{for } j \in J \text{ and } k=0, ..., r \tag{3}$$

$$\sum_{i \in I} y_i \leq p \tag{4}$$

$$x_{jk} \geq 0 \qquad j \in J \text{ and } k=0, ..., r \tag{5}$$

$$y_i \in \{0,1\} \qquad \text{for } i \in I. \tag{6}$$

In the above model, the objective function (2) gives the upper bound of the sum of the original distances. The constraints (3) ensure that the variables x_{jk} are allowed to take the value of 0, if there is at least one center located in radius D^k from the customer j. The constraint (4) limits the number of located centers by p. The constraints (5) and (6) are obligatory constraints. Even if only zero or one are feasible values of the variables x_{jk}, the associated obligatory constrains can be relaxed from the model due to partial integrality property concerning these variables. This fact makes the computational process easier.

3 The Distance Range Partitioning by the Sequential Method

The zone determination is based on an initial anticipating of a distance relevancy and its step-by-step improving. The distance relevancy n_h of a distance from the matrix $\{d_{ij}\}$ is in general derived from the frequency, with which the given distance d^h belongs to an optimal solution of the solved instance of the p-median problem. It means the frequency with which the distance d_{ij} is the shortest distance between a served customer and the nearest located center. At the start of the sequential zone adjustment method the relevancy n_h can be only estimated using a heuristic rule as no optimal solution of the solved instance is at disposal. We exploit the expression (7) to determine the initial value of the relevancy.

$$n_h = N_h e^{-d^h/T} \tag{7}$$

Symbol T denotes a positive parameter of the heuristic rule and N_h is the frequency of d^h in the matrix $\{d_{ij}\}$, where only p-1 biggest distances is discharged from each matrix column.

Determination of the current relevancy n_h is a key-stone of the presented zone adjustment method, which minimizes the deviation of the computed lower bound from an anticipated optimal solution. The dividing points associated with the current relevancy enable to formulate the covering problem (2)-(6) accordingly to [3] with below redefined coefficients e_k. Then, any integer-programming solver can obtain an approximate solution of the p-median problem. As concrete location of centers are part of the solution, the distances corresponding with the shortest distances between served customers and located centers can be found in the matrix $\{d_{ij}\}$. The frequency of the distance d^h updates the current value of relevancy n_h. Then the searching process of new dividing points determination and subsequent solving of the approximate p-median problem can be repeated.

Concerning the above-mentioned method for optimization of dividing points, the corresponding algorithm is based on the following analysis and the below-presented model.

As noted above, the elements of the distance matrix $\{d_{ij}\}$ form a finite ordered set of values $d^0 < d^1 < \ldots < d^m$, where $D_0 = d^0$ and $D_m = d^m$. Let us assume that a distance d between customer and the nearest located center belongs to an interval (D_k, D_{k+1}> given by a pair of succeeding dividing points. Let us denote $D_k^1, D_k^2, \ldots, D_k^{r(k)}$ the values of the sequence $d^0<d^1<\ldots<d^m$, which are greater than D_k and less than D_{k+1}, then the maximum deviation of d from the lower estimation D_k^1 is D_{k+1} - D_k^1.

As the variable x_{jk} from model (2)-(6) takes the value of 1 if the distance of the customer $j \in J$ from the nearest located center is greater than D_k and this variable takes the value of 0 otherwise, we can redefine the zone coefficients e_k in accordance to $e_0=D_0^1-D_0$ and $e_k=D_{k+1}^1$- D_k^1 for k=1,…, r. Then the expression $e_0 x_{j0} +e_1x_{j1}+ e_2x_{j2}+ e_3x_{j3}+\ldots+ e_rx_{jr}$ is a lower approximation of the d_{ij}, which corresponds to the nearest distance of the served node j from the nearest located center i.

If a relevancy n_h of each d^h is given, we could minimize the deviation using dividing points obtained by solving the following problem:

$$\text{Minimize} \qquad \sum_{t=0}^{m-1}\sum_{h=t}^{m-1} (d^{h+1} - d^{t+1})n_{h+1}z_{th} \tag{8}$$

$$\text{Subject to } z_{t,h+1} \le z_{th} \quad \text{for } t=0, \ldots, m-1 \text{ and } h=t, \ldots, m-1 \tag{9}$$

$$\sum_{t=0}^{h} z_{th}=1 \text{ for } h=0, \ldots, m-1 \tag{10}$$

$$\sum_{t=1}^{m-1} z_{tt}=r \tag{11}$$

$$z_{th}\in\{0, 1\} \quad \text{for } t=0, \ldots, m-1 \text{ and } h=t, \ldots, m \tag{12}$$

If the distance d^h belongs to the interval starting with a dividing point d^t then the decision variable z_{th} takes the value of 1. Link-up constraints (9) ensure that a distance d^{h+1}can belong to the interval starting with d^t only if each distance between $d^{h+1}$and d^t belongs to this interval. Constraints (10) assure that each distance d^h belongs to some interval and constraint (11) enables that only r dividing points will be chosen. After the problem (8)-(12) is solved, the nonzero values of z_{tt} indicate the distances, which correspond with dividing points.

4 Numerical Experiments

We performed a sequence of numerical experiments to test effectiveness of the suggested zone adjustment method for lower bound on of the p-median problem. Solved instances of the problem were obtained from OR-Lib set of the p-median problem instances [1]. To make the results more comparable, we preserve the original notation of the instances, e.g. pmed21, and if we modified an instance in the number of centers p to enlarge the set of benchmarks, then we denote the modified instance name by an asterisk, e.g. pm*d21. An individual experiment is organized so that the initial sequence of dividing points is obtained accordingly to the model (8)-(12), where the relevancy n_h of a distance d_h is computed in accordance to the formula (7) for T=1000. The sequential zone adjustment method starts from the initial sequence of dividing points, defines associated constants a^k_{ij} and e_k and solves the problem (2)-(6). Resulting values of y_i are used to obtain reduced sequence of the relevant distances d_h with their frequencies n_h as described in the previous section. As the optimal value of the objective function (2) is only lower bound of the associated value of the objective function (1), we use the formula (1) to obtain this associated real value.

This step is repeated for new sequences d_h and n_h until no improvement of (2) is obtained or the number of relevant distances drops below 20. The obtained results are reported in table 1 and they are separated in the following row-blocks.

Block “Desc” comprises a concise description of an instance, which includes name of the instance (“Name”), the number of customers (“Cust”), which is equal to the number of possible locations and the number p denoting the upper limit on the number of located centers. The denotation “Exact” indicates the row of the objective function values (1) for the individual instances obtained by the optimization environment XPRESS-IVE. The row-block “LB Gap” refers on the sequence of solutions of the problem (2)-(6) obtained for step by step improved set if the dividing points. In the rows denoted by Run1, Run2, etc., there are given the gaps of the solution values (1) from the exact solution in percents.

	Name	pmed5	pm*d5	pmed8	pm*d8	pmed12	pm*d12	pmed16	pm*d16
	Cust	100	100	200	200	300	300	400	400
Desc	*p*	33	10	20	100	10	100	5	128
	Exact	1434	3505	4459	933	6645	1706	8068	1750
	Run1	30.8	14.8	25.8	89.3	13.9	63.1	10.4	63.9
LB Gap	Run2	3.8	4.3	4.5	0.4	3.0	0.2	2.5	0.0
[%]	Run3				1.9		0.1	2.4	
	Run4				0.0				

Table 1 Gaps between the lower bounds and exact solutions

The experiments were performed using the optimization software Xpress-IVE. The associated code was run on a PC equipped with the Intel Core 2 6700 processor with parameters: 2.66 GHz and 3 GB RAM.

5 Conclusions

We present the sequential adjustment method, which enables to improve the lower bound on an optimal solution of the p-median problem. This method can be very useful, when an exact approach to the p-median problem fails due to enormous size of the associated model and an approximate approach must be used. The presented results of preliminary numerical experiments show that the approximate approach combined with the sequential method is a promising way to solving large instances of the p-median problem. As can be seen in the table that only two or three steps of the sequential method are necessary to obtain a lower bound, which differs from the optimal solution by less than five percent.

Acknowledgements

This work was supported by the research grant VEGA 1/0361/10 „Optimal design of public service systems under uncertainty“.

References

[1] Beasley J. E., 1990, OR Library: Distributing Test Problems by Electronic Mail, Journal of the Operational Research Society, 41(11), 1069-1072.

[2] Doerner, K.F., Gutjahr, W.J., Hartl, R.F., Karall, M.; Reimann, M.: Heuristic Solution of an Extended Double-Coverage Ambulance Location Problem for Austria. Central European Journal of Operations Research, Vol. 13, No 4, pp 325-340, (2005)

[3] Janáček, J.: Approximate Covering Models of Location Problems. In: Lecture Notes in Management Science. Proceedings of the 1st International Conference on Applied Operational Research – ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp 53-61

[4] Janáček, J. - Buzna, Ľ. (2008): An acceleration of Erlenkotter-Körkel's algoriths for uncapacitated facility location problem. In: Annals of Operations Research, Vol. 164, pp 97-109

[5] Janáček, J., Linda, B., Ritschelová, I. (2010): Optimization of Municipalities with Extended Competence Selection. Prager Economic Papers-Quarterly Journal of Economic Theory and Policy, Vol. 19, No 1, 2010, pp 21-34

[6] Janáčková, M., Szendreyová, A.: An impact of set of centres of distribution system design problem on computational time of solving algorithm. In: Mikulski J (ed). Advances in Transport Systems Telematics, Katowice, pp 109-112, (2006).

[7] Jánošíková, L.: Emergency Medical Service Planning. Communications –Scientific Letters of the University of Žilina, Vol. 9, No 2, pp 64-68, (2007).

[8] Marianov, V., Serra, D.: Location problems in the public sector. In: Drezner Z (ed.) et al. Facility location. Applications and theory. Berlin: Springer, pp 119-150, (2002).

Renewable Energy Sources and DEA Models

Jana Kalčevová[1]

Abstract. According to the directive of European Union the increased use of energy from renewable energy sources (RES) is necessary. There are five real possibilities of RES building-up in the Czech Republic. They are wind power stations for wind energy production, photovoltaic power stations based on obtaining of sun energy, the geothermal power stations deal with thermal energy of the Earth, places with large water areas or swift rivers are suitable for hydro-electric power stations and the last but not least possibility is to gain biomass energy from biomass. Each of the mentioned energy sources has its advantages and also disadvantages that can be evaluated by investor from the multicriteria point of view. Evaluation of investments into RES is a multiple criteria decision making problem with the evaluated units (energy sources) and with the multiple inputs (investments, recovery, etc.) and multiple outputs (expected output, profit, etc.). This problem can be solved by several modeling techniques. In this paper data envelopment analysis (DEA) models are used and methodology for evaluation of projects for RES building is introduced.

Keywords: renewable energy sources, multicriteria decision making, data envelopment analysis

JEL Classification: C44
AMS Classification: 90B50, 90C29, 90C90, 91B06

1 Introduction

In the last year the paper [7] focused on multicriteria evaluation of renewable energy sources (RES) was published, as well as [6]. The necessary consequence of increasing living standard in many countries over the years is the increase in energy use. Energy demand was covered by the use of fossil fuels (e.g. coal, petroleum, etc.) but their excessive use led to a decrease of these sources [11]. In addition, we cannot ignore the environmental impact from their use. These are reasons for effort to use and develop more environmentally friendly forms of energy.

The Czech Republic as a member of the European Union is bounded by the Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing Directives 2001/77/EC and 2003/30/EC. According to this document the increased use of energy from renewable sources is necessary. One of the appropriate and achievable objectives is 20 % target of the overall share of energy from renewable sources before 2020. This directive also specifies national targets. For the Czech Republic it implies 13% of the overall share of energy from renewable sources. In addition the Czech Republic tends to 15% share in 2030 and 30% share of energy from renewable sources in 2050 in accordance to the State energy conception.

There are five real possibilities of renewable energy sources building-up in the Czech Republic. The first ones are wind power stations for wind energy production. The second possibility is a photovoltaic power station based on obtaining of sun energy. The geothermal power station deals with thermal energy of the Earth. Places with large water areas or swift rivers are suitable for hydro-electric power stations. The last but not least possibility is to gain biomass energy from biomass. There exist several other alternative energy sources but such as energy of sea waves are not reachable in the Czech conditions.

Each of the mentioned energy sources has its advantages and also disadvantages. In [6] and [7] we evaluate them from the multicriteria point of view. It is a complex decision making problem with a small number of alternatives (five in the case of this paper) and many decision criteria. For the evaluation TESES (technical, economic, social, ecological and strategic) classification of criteria was used. The technical criteria describe the technical obstacles of single technologies, technical parameters of produced electricity and if the technology enables an efficient usage of produced warm, coefficient of hazard in the building-up season etc. Economic criteria describe the project economy on the level of assessment from the investor's point of view that is conclusive for the project realization or rejection. Social criteria describe the social project implications and their benefit for solving of

[1] University of Economics, Prague, Department of Econometrics, W. Churchill Sq. 4, Prague 3, 130 67, Czech Republic, kalcevov@vse.cz.

social and socio-economic problems of the regions and states. Ecological criteria describe the particular benefits of single technology for environment components. And strategic criteria describe the long-term project effect for power industry in the Czech Republic and for the situation in further spheres of the national economy – for example agriculture.

2 DEA point of view

Each alternative mentioned above is evaluated from the multicriteria point of view by each investor. Evaluation of investments into RES is a multiple criteria decision making problem with the evaluated production units as alternatives (energy sources) and with the multiple inputs as criteria to be minimized (investments, recovery, etc.) and multiple outputs as criteria to be maximized (expected output, profit, etc.) in the typical case. This problem can be solved by several modeling techniques. Among them are standard multiple criteria decision making (MCDM) methods used in [6], [7] and [1], data envelopment analysis (DEA) models used in [5] and others (for example using methods of mixed integer programming at optimization of the production used in [9]).

Each renewable energy source is production unit and DEA is suitable tool for evaluation of efficiency of production units. The term efficiency is defined as the ratio of multiple effects (outputs) produced by the evaluated units on the one hand and multiple resources (inputs) that are spent during the transformation process of inputs into the outputs on the other hand. Typically, the number of inputs and outputs is higher than one. That is why the question is how to compare multiple outputs and multiple inputs.

First DEA models were formulated in 1978 but their main expansion starts in the last ten years of the last century. But we do not find using of DEA models for evaluation of efficiency of different renewable energy sources. Only one paper in Web of Science database is focused on using DEA models in this field and it is only with respect to biomass energy (see [13]).

Conclusions given by the evaluation of efficiency of investments can offer important issues for allocation of financial resources. This approach seems to be suitable but it is necessary to analyze and verify this theory, especially with respect to Czech conditions. That is why we would like to apply DEA to evaluation of energy investments.

2.1 Decision units

For this paper the list of decision units is given by RES that can be built in the Czech Republic. The used data are from the real projects and they were gained from [10].

- **Wind power station** – a wind farm with four wind power stations, output 2300kW, lifetime 20 years, costs CZK about 367 millions.
- **Photovoltaic power station** – output 373kW, lifetime 15 years, costs about CZK 30 millions.
- **Geothermal power station** – output 300kW, lifetime 30 years, costs about CZK 23 millions.
- **Hydro-electric power stations** – output 103kW, lifetime 20 years, costs about CZK 10 millions.
- **Biomass energy** – output 1500kW, lifetime 20 years, costs about CZK 170 millions.

Note that DEA models are suitable for evaluation of a large number of production units with respect to a relative small number of inputs and outputs. But this paper is focused on methodology of evaluation and the case study is only illustrative, so we used only five production units. We proposed sixteen suitable criteria for this problem.

2.2 Criteria – inputs and outputs

As was mentioned in the Introduction section the criteria can be divided into five main groups for analysis. Each group contains several sub-criteria.

- **Technical group of criteria includes four sub-criteria**
  - *T1:* annual utilization of installed output (maximal) – ratio of real energy production per year to maximal theoretical energy production,
  - *T2:* expected dissipation of energy (minimal) – dissipation of energy given by unstable weather, device errors, repairs and services,
  - *T3:* expected lifetime of power station (maximal) – how long it will be able to produce energy,
  - *T4:* investment costs of project (minimal) – capital expenditures related with building, complexity of realization, time of realization and technical complexity (evaluated by CZK or points from 0 to 100).
- **Economic group consists of five sub-criteria**
  - *F1:* the net present value – NPV (maximal),

- *F2:* internal rate of return – IRR (maximal),
- *F3:* time of recovery (minimal),
- *F4:* recovery of investment – ROI (maximal),
- *F5:* net profit (maximal).

- **Social group is based on only two sub-criteria**
  - *S1:* number of created employee positions – new jobs (maximal), evaluated by absolute value of by points from 0 to 100,
  - *S2:* user's comfort (maximal) – costingness of services, quality and complexity, evaluated for example by points from 0 to 100.

- **Ecological group includes three sub-criteria**
  - *E1:* decreasing of carbon dioxide (maximal) – evaluated by amount of saved fugitive emissions in comparison with pitcoal,
  - *E2:* scenery derogation (minimal) – evaluated for example by points from 0 to 100,
  - *E3:* other environmental impacts (minimal) – such as audible noise, dustiness, effluvium, appropriation of land, etc., also evaluated for example by points from 0 to 100.

- **Strategic group includes two sub-criteria**
  - *G1:* accessibility of suitable areas (maximal) – accessibility and suitability of area, adequacy of natural environment,
  - *G2:* volume of sources diversification (maximal) – evaluation of increase in number of energy mix sources.

For DEA models we divided criteria into two groups – inputs and outputs. List of inputs contains minimizing criteria and it can involve for example *T2:* expected dissipation of energy, *T4:* investment costs of project, *F3:* time of recovery, *E2:* scenery derogation and *E3:* other environmental impacts. List of outputs contains maximizing criteria and it can involve for example *T1:* annual utilization of installed output, *T3:* expected lifetime of power station, *F1:* the net present value, *F4:* recovery of investment, *F5:* net profit, *S1:* number of created employee positions, *E1:* decreasing of carbon dioxide, *G2:* volume of sources diversification.

For this paper we use two illustrative case studies with only two inputs (*T4:* investment costs of project, *F3:* time of recovery, for both cases) and two outputs (for the first case: *F2:* internal rate of return, *F4:* recovery of investments, and for the second case: *T1:* annual utilization of installed output, *F5:* net profit) – the most discussed ones in energetic and investment practice. Complete data for illustrative case studies are included in Table 1.

	Criterion	Unit	Type	Wind power	Photovoltaic power	Geothermal power	Hydroelectric power	Biomass power
Inputs	T4	point	min	75	40	95	80	60
	F3	year	min	8,5	8,5	2,5	6,5	3,5
Outputs1	F2	%	max	0,115	0,080	0,407	0,157	0,307
	F4	%	max	2,034	1,529	11,09	2,158	0,671
Outputs2	T1	%	max	1,009	0,093	0,799	0,593	0,856
	F5	mil. CZK	max	386,5	28,4	103,7	11,408	86,8

Table 1 Data for DEA case studies

2.3 Metodology of DEA

DEA is a set of non-parametric techniques based on solving of linear programming problems for evaluation of efficiency of the set of homogenous units. The efficient unit consumes minimal amount of inputs and produces maximal amount of outputs. The basic idea of DEA models, developed by Charnes et al. [2] and [3], consists in estimation of so-called efficient frontier, and projects all DMUs onto this frontier. If a DMU lies on the frontier, it is referred to as an efficient unit, otherwise inefficient. All efficient DMUs are considered equally “good” [4]. DEA also provides efficiency scores and reference units for inefficient DMUs. Reference units are hypothetical units on the efficient frontier, which can be regarded as target units for inefficient units. A virtual reference unit is traditionally found in DEA by projecting the inefficient DMU radially onto the efficient frontier.

Note also that DEA models are linear problems those can be solved classically for example by simplex method. As is written below the alternative possibility for solution is to solve dual problem [12]. The review and detailed information about DEA models can be found in [8].

Mathematical model

The mathematical model of DEA considers r decision making units (DMUs) $U_1, U_2, \ldots, U_r$ with m inputs ($i = 1, 2, \ldots, m$) and n outputs ($j = 1, 2, ..., n$). The vector of input values of DMU k ($k = 1, 2, ..., r$) is denoted as $\mathbf{x}_k = (x_{1k}, x_{2k}, ..., x_{mk})^{\mathrm{T}}$ and matrix of all input values for all DMUs is denoted as $\mathbf{X} = \{x_{ik}, i = 1, 2, \ldots, m, k = 1, 2, ..., r\}$. Similarly, the vector of output values of DMU k is denoted as $\mathbf{y}_k = (y_{1k}, y_{2k}, ..., y_{nk})^{\mathrm{T}}$ and matrix of all output values for all DMUs is denoted as $\mathbf{Y} = \{y_{jk}, j = 1, 2, \ldots, n, k = 1, 2, ..., r\}$. The relative technical efficiency of given DMU q can be generally expressed as ratio of weighted sum of outputs and weighted sum of inputs

$$TE_q = \frac{\sum_j u_j y_{jq}}{\sum_i v_i x_{iq}}, \tag{1}$$

where v_i, $i = 1, 2, ..., m$ is a weight for i-th input and u_j, $j = 1, 2, ..., n$ is a weight for j-th output.

DEA models can be oriented to inputs or outputs. In this study we assume fixed level of inputs and so we work with output oriented models. In such case we would like to maximize outputs with respect to given inputs. Such model has the following form

$$\begin{aligned} \text{maximize} \quad & z = \mathbf{u}^{\mathrm{T}}\mathbf{y}_q, \\ \text{subject to} \quad & \mathbf{v}^{\mathrm{T}}\mathbf{x}_q = 1, \\ & \mathbf{u}^{\mathrm{T}}\mathbf{Y} - \mathbf{v}^{\mathrm{T}}\mathbf{X} \leq 0, \\ & \mathbf{u} \geq \varepsilon, \\ & \mathbf{v} \geq \varepsilon. \end{aligned} \tag{2}$$

Model (2) is called primary CCR (Charnes, Cooper, Rhodes) output oriented model (CCR-O) and it assumes constant return to scale (CRS).

For practical solution we can use dual CCR-O model. It has form

$$\begin{aligned} \text{maximize} \quad & f = \theta + \varepsilon(\mathbf{e}^{\mathrm{T}}\mathbf{s}^{+} + \mathbf{e}^{\mathrm{T}}\mathbf{s}^{-}), \\ \text{subject to} \quad & \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \theta\, \mathbf{y}_q, \\ & \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \mathbf{x}_q, \\ & \boldsymbol{\lambda}, \mathbf{s}^{+}, \mathbf{s}^{-} \geq \mathbf{0}, \end{aligned} \tag{3}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_r)^{\mathrm{T}}$ are weights for DMUs, $\mathbf{s}^{+}$ and $\mathbf{s}^{-}$ are vectors of slack variables for inputs and outputs, $\mathbf{e}^{\mathrm{T}} = (1, 1, \ldots, 1)$ a $\varepsilon > 0$ is a number smaller than any positive real number.

The evaluated q-th DMU is efficient if and only if optimal value of $\theta = 1$ and all slack variables ($\mathbf{s}^{+}$ and $\mathbf{s}^{-}$) in the optimal solution are zero.

The models (2) and (3) are used if we assume constant return to scale (CRS). In the case of variable return to scale (VRS) we work with BCC (Banker, Charnes, Cooper) model and the dual model (3) must be extended about constraint $\mathbf{e}^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{1}$. The input oriented models have similar mathematical model as (3), with conditions $\mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^{+} = \mathbf{y}_q$, $\mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^{-} = \theta\,\mathbf{x}_q$.

3 Results

We tried to analyze two illustrative case studies of RES evaluation. For computation we use optimization software Lingo 10.0. In Table 2 we can see values of theta parameters for all input and output oriented CCR and BCC models for the first case study.

	Wind power	Photovoltaic power	Geothermal power	Hydroelectric power	Biomass power
CCR-I	0,3357	0,4422	1,0000	0,4179	1,0000
BCC-I	0,7041	1,0000	1,0000	0,7332	1,0000
CCR-O	2,9790	2,2620	1,0000	2,3930	1,0000
BCC-O	2,8840	1,0000	1,0000	2,3190	1,0000

Table 2 DEA results for outputs *F2* and *F4*

From this table we can see that geothermal and biomass energies are efficient by using all four DEA models. With respect to super-efficiency we can concluded that geothermal power is the best alternative of investments and biomass power is the second one. The third place has photovoltaic power, the forth is hydroelectric and the worst investment is into wind power stations.

Note that this problem is very small, it is only illustrative and the results are dependent on selection of criteria. For correct results it suitable to extend the data set (number of alternatives and numbers of inputs and outputs). If we only change outputs 1 (*F2* and *F4*) to outputs 2 (*T1* and *F5*) the order of RES will change. The best alternative will be wind power (it was the worst in the previous case), the second place is for geothermal and the third for biomass power stations.

4 Conclusion

As was mentioned above the alternatives were evaluated with respect to all sixteen criteria in [7] and the obtained results are in Table 3 (note that higher number in table is better).

Alternative	Wind power	Photovolta-ic power	Geothermal power	Hydroelec-tric power	Biomass power
AHP	0,206	0,124	0,330	0,094	0,246
WSA	0,428	0,221	0,691	0,253	0,551
TOPSIS	0,447	0,223	0,552	0,200	0,546
ELECTRE I.	Ineff.	Ineff.	Eff.	Ineff.	Eff.
PROMETHEE II.	-0,049	-0,244	0,363	-0,245	0,175

Table 3 Evaluation by MCDM methods

From this table we can see that the best RES in the conditions of the Czech Republic is the geothermal power station with respect to all used methods. It is one of two efficient alternatives marked by ELECTRE method (this result is equivalent to DEA). The second one is the biomass power station that is placed as the second by all used methods. The differences between geothermal and biomass power stations are relatively remarkable. The third place belongs to wind power station by all five methods (it differs from DEA). The fourth and fifth place is dependent on used method. By AHP, TOPSIS and PROMETHEE II. the photovoltaic power station is better than hydro-electric one. By contrast WSA evaluates hydro-electric power station better than photovoltaic one. This fact is given by using of linear metric and consequently linear utility function in the case of WSA. For TOPSIS and PROMETHEE methods are used generally non-linear utilities and the results differ.

It is clear that the best alternative in our study is geothermal power station. By higher reflecting investor's point of view the next best alternative is biomass power station and the three worse are the remaining ones, i.e. wind, photovoltaic and hydroelectric power stations. It is quite surprising because the real situation in the Czech Republic is that especially photovoltaic power stations are very popular among investors. The reason consists probably in a special governmental support of this kind of energy in the last years which is not included in the available data set.

The order of RES differs with respect to used method. The question is which results are correct. The advantage of DEA models is a fact that efficiency evaluation is based on the data available without taking into account the decision-maker's preferences. Therefore DEA seems to be more suitable than MCDM methods. Unfortunately, DEA methods evaluate a large set of units with respect to a small set of criteria. In this case we work with only five alternatives and therefore the set of criteria have to be very small. Thus the results are biased and it is necessary to solve the problem of extension of data set.

Renewable energy sources integration may be the key element of new energy policy not only in the Czech Republic because it improves the stability and reliability of the energy system, minimizes environmental impact and significantly saves sources of limited and not ecological fossil fuels. The first place of the geothermal power station is not surprising. It has maximal expected lifetime and also minimal time of recovery. Also other economic criteria as recovery of investment and internal rate of return are the best comparing to other sources. Expected dissipation of energy is minimal. Both social criteria are very good evaluated. The environmental impact of geothermal power stations is small because they are placed under ground. This power station also increases the number of energy mix sources. The problem with installation of geothermal power station is with accessibility of suitable areas for this kind of energy (G1 criterion). In the Czech Republic there are very limited accessible and suitable areas for this kind of renewable energy sources and that is why probably this kind of power stations will not belong to the most significant sources of energy in the Czech Republic in the future.

Acknowledgements

The research is supported by GACR, project no. 402/09/0231.

References

[1] Benesova, A., Skocdopolova, V. and Kuncova, M.: MS Excel ADD-IN Applications in MCDM Education. In: *Efficiency and Responsibility in Education.* Prague: Authors, 2010, 33–49. ISBN 978-80-213-2084-0.

[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] Charnes, A., Cooper, W.W., and Rhodes, E.: Short communication: measuring efficiency of decision making units. *European Journal of Operational Research* **3** (1979), 339.

[4] Hosseinzadeh Lotfi, F., Noora, A.A., Jahanshahloo, G.R., Jablonsky, J., Mozaffari, M.R., and Gerami, J. An MOLP based procedure for finding efficient units in DEA models, *CEJOR* **17**, 2009, 1–11.

[5] Jablonsky, J., and Dlouhy, M.: Solving DEA and MCDM Decision Problems in Spreadsheets. In: *Efficiency and Responsibility in Education.* Prague, 2010, 126–136.

[6] Jablonsky, J., and Kalcevova, J.: Multicriteria Evaluation of Investments into Renewable Energy Sources. In: *Quantitative Methods in Economics – Multiple Criteria Decision Making XV.* Bratislava : University of Economics, 2010, 74–84. ISBN 978-80-8078-364-8.

[7] Jablonsky, J., and Kalcevova, J.: Multicriteria Evaluation of Investments into Renewable Energy Sources. In: *Sustainable Development: Industrial Practice, Education & Research.* Bari : DIMEG Edizioni, 2010, 20–25. ISBN 978-88-904625-1-1.

[8] Pedraja-Chaparro, F., Salinas-Jimenez, J., and Smith, P.: On the Role of weight restrictions in data envelopment analysis. *Journal of Productivity Analysis* **8**, 1997, 215–230.

[9] Pelikan, J.: An Optimisation of the Production System – Case Study. In: *Proceedings of the 21st International Conference Mathematical Methods in Economics*, Prague, 2003, 203–208.

[10] Petríková, T.: Analýza investic do energetických zdrojů (Investment Analysis in Energy Sources). *Diplomová práce (Diploma Theses)*, Vysoká škola ekonomická v Praze (University of Economics Prague), 2010.

[11] Pilavachi, P.A., Chatzipanagi, A.I., and Spyropoulou, A.I.: Evaluation of hydrogen production methods using the Analytic Hierarchy Process. *International Journal of Hydrogen Energy*, no. 34 (2009), 5294–5303.

[12] Thompson, R.G, Singleton F., Thrall R., and Smith, B.: Comparative site evaluations for locating a highenergy physics lab in Texas. *Interfaces* **16**, 1986, 35–49.

Yang, J.A.: Evaluation Study on Developing Biomass Energy Source in State-owned Forestry Region Based on DEA Mode In: *Proceedings of International Forum of Knowledge as a Service*, 2010, 148–152.

Dependent Data in Economic and Financial Problems

Vlasta Kaňková[1]

Abstract. Optimization problems depending on probability measures correspond to many economic and financial applications. The paper deals with the case when an empirical measure substitutes the theoretical one. Especially, the paper deals with a convergence rate of the corresponding estimates. "Classical" results for independent samples are recalled, situations in which the case of dependent sample can be (from the mathematical point of view) reduced to independent case are mentioned. A great attention is paid to weak dependent samples fulfilling the Φ–mixing condition.

Keywords: Stochastic programming, Wasserstein metric, $\mathcal{L}_1$ norm, Empirical estimates, One–stage problems, Multistage problems, Independent samples, m–dependent sequences, Markov dependence, Φ– mixing random sample

JEL classification: C44
AMS classification: 90C15

1 Introduction

Economic activities are usually simultaneously influenced by a random factor and a decision parameter. Constructing their mathematical models we often obtain optimization problems depending on a probability measure. These models can be static or dynamic. Multistage stochastic problems belong to a dynamic types. Employing a recursive definition we obtain a system of one–stage problems. Consequently, the results obtained for one–stage problems can be often employed to study multistage cases.

1.1 One–Stage Model

Let $\xi\,(:= \xi(\omega) = [\xi_1(\omega),\, \ldots,\, \xi_s(\omega)])$ be s–dimensional random vector defined on a probability space $(\Omega,\, \mathcal{S},\, P)$; $F(:= F(z),\, z \in R^s)$ the distribution function of ξ; $F_i,\, i = 1,\, \ldots,\, s$ one–dimensional marginal distribution functions corresponding to F; $P_F,\, Z_F$ the probability measure and support corresponding to F. Let, moreover, $g_0(:= g_0(x,\, z))$ be a real–valued function defined on $R^n \times R^s$; $X \subset R^n$ be a nonempty set. If the symbol E_F denotes the operator of mathematical expectation corresponding to F, then a rather general "classical" one–stage stochastic programming problem can be introduced in the form:

Find
$$\varphi(F) = \inf\{\mathsf{E}_F g_0(x,\, \xi) | x \in X\}. \tag{1}$$

In applications very often the "underlying" probability measure P_F has to be replaced by an empirical one. Evidently, then the solution is sought w.r.t. the "empirical problem":

Find
$$\varphi(F^N) = \inf\{\mathsf{E}_{F^N} g_0(x,\, \xi) | x \in X\}, \tag{2}$$

where F^N denotes an empirical distribution function determined by a random sample $\{\xi^i\}_{i=1}^N$ corresponding to the distribution function F. If we denote the optimal solutions sets of (1) and (2) by $\mathcal{X}(F),\, \mathcal{X}(F^N)$, then $\varphi(F^N),\, \mathcal{X}(F^N)$ are stochastic estimates of $\varphi(F),\, \mathcal{X}(F)$.

The investigation of these estimates started in 1974 by R. Wets (see [27]). In the same time consistency has been investigated under ergodic assumption in [10]. These papers have been followed many times (see e.g. by [5], [19], [20], [22]). The investigation of the convergence rate started in [11], and followed e.g. by [2], [8], [21], [25], [26]. The investigation for weakly dependent samples started in eighties (see [12]) and followed e.g. by [2], [8], [13], [14].

[1] Academy of Sciences of the Czech Republic, Institute of Information Theory and Automation, Pod Vodárenskou věží 4, 182 08 Praha 8 , e-mail: kankova@utia.cas.cz

1.2 Multistage Model

A general multistage stochastic programming problem can be in a rather general form introduced recursively (see e.g. [6], [15]) as the problem:

Find

$$\varphi_{\mathcal{F}}(M) = \inf \{\mathsf{E}_{F^{\zeta^0}} g^0_{\mathcal{F}}(x^0, \zeta^0) |\ x^0 \in \mathcal{K}^0\}, \tag{3}$$

where $g^0_{\mathcal{F}}(x^0, z^0)$ is given recursively

$$\begin{aligned} g^k_{\mathcal{F}}(\bar{x}^k, \bar{z}^k) \quad &= \quad \inf\{\mathsf{E}_{F^{\zeta^{k+1}|\bar{\zeta}^k = \bar{z}^k}}\ g^{k+1}_{\mathcal{F}}(\bar{x}^{k+1}, \bar{\zeta}^{k+1}) \,|\, x^{k+1} \in \mathcal{K}^{k+1}(\bar{x}^k, \bar{z}^k)\}, \quad k = 0, 1, \ldots, M-1, \\ g^M_{\mathcal{F}}(\bar{x}^M, \bar{z}^M) \quad &:= \quad g^M_0(\bar{x}^M, \bar{z}^M). \end{aligned} \tag{4}$$

$\zeta^j = \zeta^j(\omega)$, $j = 0, 1, \ldots, M$ are s–dimensional random vectors defined on a probability space (Ω, S, P); $\bar{\zeta}^k = \bar{\zeta}^k(\omega) = [\zeta^0, \ldots, \zeta^k]$, $\bar{z}^k = [z^0, \ldots, z^k]$, $z^j \in R^s$, $x^j \in R^n$, $\bar{x}^k = [x^0, \ldots, x^k]$, $j = 0, 1, \ldots, k$, $k = 0, 1, \ldots, M$; $F^{\zeta^j}(z^j)$, $F^{\bar{\zeta}^j}(\bar{z}^j)$, $j = 0, \ldots, M$ the distribution functions of the ζ^j and $\bar{\zeta}^j$; $F^{\zeta^k|\bar{\zeta}^{k-1}}(z^k| \bar{z}^{k-1})$, $k = 1, \ldots, M$ denotes the conditional distribution function (ζ^k conditioned by $\bar{\zeta}^{k-1}$). $g^M_0(\bar{x}^M, \bar{z}^M)$ is a function defined on $R^{n(M+1)} \times R^{s(M+1)}$; $\mathcal{K}^{k+1}(\bar{x}^k, \bar{z}^k)$, $k = 0, 1, \ldots, M-1$, are multifunction mappings $R^{n\,(k+1)} \times R^{s\,(k+1)}$ into the space of (mostly compact) subsets of $\mathcal{X}$; $\mathcal{X}$, $\mathcal{K}^0 \subset R^n$ are nonempty sets; $\mathcal{K}^0 \subset \mathcal{X}$. $Z^j_{\mathcal{F}} \subset R^s$, $j = 0, 1, \ldots, M$, denote the supports corresponding to $F^{\xi^j}(\cdot)$; $\bar{Z}^k_{\mathcal{F}} = Z^0_{\mathcal{F}} \times \ldots \times Z^k_{\mathcal{F}}$, $\bar{\mathcal{X}}^k = \mathcal{X} \times \ldots \times \mathcal{X}$, $k = 0, 1, \ldots, M$.

Evidently, the problem given by (3) and (4) is depending on the system:

$$\mathcal{F} = \{F^{\zeta^0}(z^0), \quad F^{\zeta^k|\bar{\zeta}^{k-1}}(z^k|\bar{z}^{k-1}), k = 1, \ldots, M\}. \tag{5}$$

2 Historical Survey

First we recall "classical" results on consistency and convergence rate.

Theorem 1. [10]. If

1. X is a compact set, $g_0(x, z)$ is a uniformly continuous bounded function on $R^s \times X$,
2. $\{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ is a random sample corresponding to ergodic sequence,

then

$$P\{\omega : |\varphi(F^N) - \varphi(F)| \xrightarrow[N \to \infty]{} 0\} = 1.$$

(The ergodicity has been considered in the sense of [1].)

Theorem 2. Let X be a nonempty compact set. If

1. in every $x \in X$ the function $g_0(x, \xi)$ is a continuous function of x for almost every $\xi \in Z_F$,
2. $g_0(x, \xi), x \in X$ is dominated by an integrable function,
3. $\{\xi^i\}_{i=1}^N, N = 1, 2, \ldots$ is independent random sample,

then

$$P\{\omega : |\varphi(F^N) - \varphi(F)| \xrightarrow[N \to \infty]{} 0\} = 1.$$

Proof. The assertion of Theorem 2 follows from Proposition 5.2 and Theorem 7.48 proven in [25].

Theorem 3. [11] Let $t > 0$, X be a nonempty compact, convex set. If

1. $g_0(x, z)$ is a uniformly continuous function on $X \times Z_F$ bounded by $M > 0$ (i.e., $|g_0(x, z)| \le M$),
2. $g_0(x, z)$ is a Lipschitz function on X with the Lipschitz constant L',
3. $\{\xi^i\}_{i=1}^N$, $N = 1, 2, \ldots$ is an independent random sample,

then there exist $K(t, X, L') > 0$ and $k_1(M) > 0$ such that

$$P\{\omega : |\varphi(F) - \varphi(F^N)| > t\} \le K(t, X, L') \exp\{-N k_1(M) t^2\}.$$

Remark. The assertion of Theorem 3 is valid independently of the distribution function F; consequently also for the distribution functions with heavy tails. On the other hand g_0 must be a bounded function. Moreover under the assumptions of Theorem 3 it has been proven in [13] that

$$P\{\omega : N^{\beta}|\varphi(F) - \varphi(F^N)| > t\} \xrightarrow[N \longrightarrow \infty]{} 0 \quad \text{for} \quad \beta \in (0, \tfrac{1}{2}).$$

If the moment generating function $M_{g_0}(t)$, corresponding to $g_0(x,\xi)$, is defined by the relation $M_{g_0}(t) := \mathsf{E}_F\{e^{t[g_0(x,\xi)-\mathsf{E}_F g_0(x,\xi)]}\}$, then we can recall the following assertion.

Theorem 4. [24] Let $X \subset R^n$ be a nonempty closed set, $\|\cdot\| = \|\cdot\|_n^2$ denotes the Euclidean norm in R^n. If

1. for $x \in X$ the moment generating function $M_{g_0}(t)$ is finite valued for all t in a neighbourhood of zero,
2. there exists a measurable function $\kappa\colon Z_F \to R_+$, and a constant $\gamma > 0$ such that
$$|g_0(x',\xi) - g_0(x,\xi)| \le \kappa(\xi)\|x'-x\|^{\gamma} \quad \text{for all} \quad \xi \in Z_F,\ x, x' \in X,$$
3. the moment generating function $M_\kappa(t)$ of $\kappa(\xi)$ is finite valued for all t in a neighbourhood of zero,

then for any $\varepsilon > 0$ there exist positive constants $C = C(\varepsilon)$ and $\beta = \beta(\varepsilon)$, independent of N, such that
$$P\{\sup_{x\in X} |\mathsf{E}_{F^N} g_0(x,\xi) - \mathsf{E}_F g_0(x,\xi)| \ge \varepsilon\} \le C(\varepsilon)e^{-N\beta(\varepsilon)}.$$

3 Wasserstein Metric via Empirical Estimates

Let $\mathcal{P}(R^s)$ denote the set of Borel probability measures on $R^s, s \ge 1$, $\|\cdot\|_s^1$denote $\mathcal{L}_1$ norm in R^s and $\mathcal{M}_1(R^s) = \{P \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\|_s^1 P(dz) < \infty\}$. We introduce the following system of the assumptions:

A.1 • $g_0(x,z)$ is a uniformly continuous function on $X \times R^s$,
- $g_0(x,z)$ is for $x \in X$ a Lipschitz function of $z \in R^s$ with the Lipschitz constant L (corresponding to the $\mathcal{L}_1$ norm) not depending on x,

A.2 • $\{\xi^i\}_{i=1}^{\infty}$ is independent random sequence corresponding to F,
- F^N is an empirical distribution function determined by $\{\xi^i\}_{i=1}^N, N = 1,2,\dots,$

A.3 $P_{F_i}, i = 1,\dots,s$ are absolutely continuous w. r. t. the Lebesgue measure on R^1; F_i, f_i, P_{F_i}, $i = 1,2,\dots,s$ denote one-dimensional marginal distribution function, probability density and the probability measure corresponding to F.

Proposition 1. [16] Let $P_F, P_G \in \mathcal{M}_1(R^s)$, and X be a compact set. If A.1 is fulfilled, then
$$|\varphi(F) - \varphi(G)| \le L \sum_{i=1}^{s} \int\limits_{-\infty}^{+\infty} |F_i(z_i) - G_i(z_i)| dz_i.$$

Replacing G by F^N in Proposition 1 and supposing $s = 1$ we can obtain for the random value $\int\limits_{-\infty}^{+\infty} |F(z) - F^N(z)|dz$ the following result.

Proposition 2. [19] Let $s = 1, t > 0$ and A.2, A.3 be fulfilled. Let, moreover, $\mathcal{N}$ denote the set of natural numbers. If there exists $\beta > 0$, $R := R(N) > 0$ defined on $\mathcal{N}$, $R(N) \xrightarrow[N\to\infty]{} \infty$ and, moreover,
$$\begin{aligned}
&N^\beta \int\limits_{-\infty}^{-R(N)} F(z)dz \xrightarrow[N\to\infty]{} 0, \qquad N^\beta \int\limits_{R(N)}^{\infty} [1-F(z)]dz \xrightarrow[N\to\infty]{} 0,\\
&2NF(-R(N)) \xrightarrow[N\to\infty]{} 0, \qquad 2N[1-F(R(N))] \xrightarrow[N\to\infty]{} 0,\\
&\left(\tfrac{12N^\beta R(N)}{t} + 1\right)\exp\{-2N\left(\tfrac{t}{12R(N)N^\beta}\right)^2\} \xrightarrow[N\longrightarrow\infty]{} 0,
\end{aligned} \tag{6}$$
then
$$P\{\omega:\ N^\beta \int\limits_{-\infty}^{\infty} |F(z) - F^N(z)|dz > t\} \xrightarrow[N\to\infty]{} 0. \tag{7}$$

4 Convergence Rate

4.1 One–Stage Problem: Independent Sample

Theorem 4. [19] Let $t > 0$, $\beta \in (0, \frac{1}{2})$, A.1, A.2 and A.3 be fulfilled. Let, moreover, $\{\xi^i\}_{i=1}^{\infty}$ be independent random sequence corresponding to F. If there exists constants C_1, C_2 and $T > 0$ such that
$$f_i(z_i) \le C_1 \exp\{-C_2|z_i|\} \quad \text{for } z_i \in (-\infty, -T) \cup (T, \infty) \quad \text{and} \quad i = 1,\, 2,\, \dots,\, s,$$

then

$$P\{\omega : N^{\beta}|\varphi(F^N) - \varphi(F)| > t\} \xrightarrow[N\to\infty]{} 0.$$

4.2 One–Stage Problem: m–Dependent Random Sample

To recall a definition of m– sequences, let $\{\xi^i\}_{i=-\infty}^{\infty}$ be a strictly stationary s–dimensional random vectors. We denote by the symbol $\mathcal{F}_c^d$ the σ–algebra generated by ξ^i, $c \le i \le d$.

Definition 1. [4] $\{\xi^i\}_{i=-\infty}^{\infty}$ is said to be m–dependent sequence ($m \ge 2$) if $\mathcal{F}_{-\infty}^a$ and $\mathcal{F}_b^{\infty}$ are independent for $b - a > m$.

Theorem 5. [17] Let $t > 0$, $\{\xi^i\}_{i=1}^{\infty}$ be a strictly stationary m sequence of s–dimensional random vectors corresponding to distribution function F. If A.1, A.3 are fulfilled and if there exist C_1, C_2, $T > 0$ such that

$$f_i(z_i) \le C_1 \exp\{-C_2|z_i|\} \quad \text{for} \quad z_i \in (-\infty, -T) \bigcup (T, \infty), \quad i = 1, \ldots, s,$$

then

$$P\{\omega :\, N^{\beta}|\varphi(F^N) - \varphi(F)| > t\} \longrightarrow_{(N\longrightarrow\infty)} 0 \quad \text{for } \beta \in (0, \frac{1}{2}).$$

4.3 Multistage Problem –Markov Dependence

To investigate problems (3) and (4) we restrict to the case when $\{\zeta^j\}_{j=-\infty}^{\infty}$ fulfils the Markov type of dependence and recall that $\{\zeta^j\}_{j=0}^{\infty}$ corresponds to a homogenous Markov chains iff ζ^j, $j = 0, \ldots,$ can be represented by a recurrence equation $\zeta^j = \bar{H}(\zeta^{j-1}, \varepsilon^j)$, where $\bar{H}$ is a measurable function and ε^j, $j > 0$ is i.i.d. sequence independent of ζ^0 (for more details see [3] or [9]). A (rather general) Markov type dependence has been considered in [15]. We consider only a special case. To this end we assume:

A.4 $\{\zeta^k\}_{k=-\infty}^{\infty}$ follows a (generally) nonlinear autoregressive sequence $\zeta^k = H(\zeta^{k-1}) + \varepsilon^k$, where ζ^0, ε^k, $k = 1, 2, \ldots$ are stochastically independent; ε^k, $k = 1, 2, \ldots$identically distributed. $H := (H_1, \ldots, H_s)$ is a Lipschitz vector function defined on R^s. We denote the distribution function of $\varepsilon^1 = (\varepsilon_1^1, \ldots, \varepsilon_s^1)$ by the symbol F^{ε} and suppose the realization ζ^0 to be known.

If A.4 is fulfilled, then (3), (4) is a s system of one–stage stochastic (mostly parametric) programming problems. Moreover, the system $\mathcal{F}$ is (under A.4) determined by the $P_{F^{\varepsilon}}$ Consequently, empirical estimates $\mathcal{F}^N$ of the $\mathcal{F}$ are determined by i.i.d. $\{\varepsilon^j\}_{j=1}^{N}$, $N = 1, \ldots$ Evidently, the problems (3), (4) is (from the mathematical point of view) transformed to one–stage case.

Theorem 6. [18] If

- i.1 $g_{\mathcal{F}}^k(\bar{x}^k, \bar{z}^k)$, $k = 1, \ldots, M$ are for $\bar{x}^k$, $\bar{z}^{k-1}$ a Lipschitz function of z^k with the Lipsch. const. L_k,
- i.2 $g_{\mathcal{F}}^0(\bar{x}^0, \bar{z}^0)$ is for $x^0 \in \mathcal{K}_0$ a Lipschitz function of z^0 with the Lipschitz const. L_0,
- i.3 $\mathcal{K}^0$, $\mathcal{K}^{k+1}(\bar{x}^k, \bar{z}^k)$, $k = 0, 1, \ldots, k-1$ are compact sets,
- i.3 $P_{F_i^{\varepsilon}}$, $i = 1, \ldots, s$ are absolutely continuous with respect to Lebesgue measure on R^1 (we denote by f_i^{ε}, $i = 1, \ldots, s$ the corresponding probability densities,
- i.4 there exist constants C_1, $C_2 > 0$, $T > 0$ such that
  $f_i^{\varepsilon}(z_i) \le C_1 \exp\{-C_2|z_i|\}$ for $z_i \in (-\infty, -T) \bigcup (T, \infty)$, $i = 1, \ldots, s$.

then

$$P\{\omega : N^{\beta}|\varphi_{\mathcal{F}}(M) - \varphi_{\mathcal{F}^N}(M)| > t\} \longrightarrow_{N\longrightarrow\infty} 0 \quad \text{for} \quad T > 0,\, \beta \in (0, 1/2).$$

(The conditions under which the assumptions i.1, i.2, i.3 are valid can be found in [15].)

4.4 One–Stage Problem: Φ–Mixing Random Samples

Definition 2. The sequence $\{\xi^i\}_{i=-\infty}^{\infty}$ is called Φ–mixing (uniformly mixing) whenever there exists Φ_N, $\Phi_N \to 0$ as $N \to \infty$ fulfilling the relation

$$|P(A \cap B) - P(A)P(B)| \le \Phi_N P(A), \quad A \in \mathcal{F}_{-\infty}^k, \quad B \in \mathcal{F}_{k+N}^{\infty}, \quad -\infty < k < \infty.$$

Remark. [28] (see also [4]) If $\{\xi^i\}_{-i=\infty}^{\infty}$ is fulfilling the conditions of Φ–mixing and simultaneously it is strictly stationary Gaussian sequence, then there exists m such that ξ^i, $i = \ldots, -1, 0, 1 \ldots$ is a m sequence. (The proof of this assertion belongs to Ibragimov ([7]).

Lemma 1. If ξ^i is a uniformly mixing sequence of centered random variables with $|\xi^i| \leq 1$ such that $\sum\limits_{N=1}^{\infty} \Phi_N < \infty$; $b^{-1} = (1 + 4 \sum\limits_{N=1}^{\infty} \Phi_N)$, $a = 2 \exp 3\sqrt{e}$ and if N and $t > 0$ satisfy

$$N\sigma^2 \geq 1, \quad 0 \leq t \leq \frac{\sigma\sqrt{N}}{8bk_N} \text{for} \quad k_N = \inf\{k : \frac{\Phi_k}{k} \leq \frac{1}{N}\}, \quad \sigma^2 = \sup_N \mathsf{E}_F|\xi^N|^2, \tag{8}$$

then

$$P\{\omega : |\sum_{i=1}^{N} \zeta^i| \geq t\sqrt{N}\sigma\} \leq a \exp\{-bt^2\}. \tag{9}$$

Furthermore, let us assume that $\{\xi^i\}_{i=-\infty}^{\infty}$ is one–dimensional strongly stationary random sequence fulfilling Φ–mixing conditions with coefficients Φ_N. If F is the corresponding one dimensional distribution function, then for every $z \in (-\infty, \infty)$, $i = \ldots, -1, 0, 1 \ldots$

$$\mathsf{E}_F I_{(-\infty, z]}(\xi^i) = F(z), \quad \mathsf{E}_F[I_{(-\infty, z]}(\xi^i) - \mathsf{E}_F I_{(-\infty, z]}(\xi^i)]^2 = F(z)(1 - F(z)), \quad \frac{1}{N}\sum_{i=1}^{N} I_{(-\infty, z]}(\xi^i) = F^N(z).$$

Evidently, it is easy to see that for every $z \in R^1$ it is possible to apply the assertion of Lemma 1 to random variable $I_{(-\infty, z]}(\xi^i)$ with $\sigma^2 := \sigma^2(\xi^i(z)) = F(z)(1 - F(z))$. However, admitting unbounded support and setting $R \in R^1, R > 0$ then (according to the properties of the distribution functions) it is easy to see that the relation (8) can be simultaneously for $z \in (R, R)$ fulfilled only for "large" N. Moreover, generally then N fulfilling the relation (8) (simultaneously for $z \in (-R, R)$ converges (generally) to ∞ if R converges to ∞. Analyzing the proof of Proposition 2 it is easy to see that employing the approach of Proposition 1 we have to restrict our consideration to the case of bounded support. The following assertion can be proven.

Theorem 7. Let $t > 0$, $\{\xi^i\}_{i=-\infty}^{\infty}$ be a uniformly mixing sequence of s–dimensional random vectors with a common distribution function F and a mixing coefficient Φ_N such that $\sum\limits_{N=1}^{\infty} \Phi_N < \infty$. Let, moreover, X be a nonempty, compact set, A.1, A.3 be fulfilled. If

1. there exist $U_i > 0$, $i = 1, \ldots, s$ such that the support of P_{F_i} is the interval $\langle -U_i, U_i \rangle$,
2. there exist ϑ_i, ϑ^i, $i = 1, \ldots, s$ such that $\vartheta_i \leq f_i(z_i) \leq \vartheta^i$, $\quad z_i \in \langle -U_i, U_i \rangle$,
3. the mixing coefficients Φ_N fulfil the relations (8) for $z = \max\limits_i \sigma_i(z_i)$, $z = (z_1, \ldots, z_s$ and $R := R(z)$,

then there exists $\beta_0 \in (0, 1/2)$ such that

$$P\{\omega : N^\beta |\varphi(F) - \varphi(F^N)| > t\} \longrightarrow_{N \longrightarrow \infty} 0 \quad \text{for} \quad \beta \in (0, \beta_0).$$

Remark. It is easy to see that the value of β_0 depends on the value of Φ mixing coefficients.

5 Conclusion

The paper deals with empirical estimates in the case of stochastic programming problems. First, a consistency results (including ergodic case) has been recalled. However, the aim of the paper has been to compare the results (concerning convergence rate) obtained under the assumption of independent data and some types of dependent samples. Consequently, again the "classical" results (Theorem 3 and Theorem 4) have been recalled. These results have been followed by Theorem 5 and Theorem 6 in which the the assertions have been proven on the basis of a "transformation" of dependent case to the independent one. At the end a result for Φ–mixing sequences has been introduced. To present a detail proof of this last assertion is over the possibility of this contribution. Summarizing we can constant that the former results (considering the problem (1) and depending samples) published former (e.g. in [14]) has been extended.

Acknowledgment This work was supported by the Grant Agency of the Czech Republic under grants P402/10/0956, P402/11/3024 and P402/10/1610.

References

[1] P. Billingsley: Ergodic Theory and Information. Wiley, New York, 1965.

[2] L. Dai, C. Chen, and J.: Birge:. (2000). Convergence Properties of Two–Stage Stochastic Programming. Journal of Optimization Theory Applications *106* (2000), 489–509.

[3] J. Dedecker, P. Doukham, G. Lang, J.,R. León R., S. Louhich and C. Prieur: Weak Dependence (With Examples and Applications): Springer, Berlin 2007.

[4] P.Doukham: Mixing: Properties and Examples. Lectures Notes and Statistics 85. Springer, Berlin, 1994.

[5] J. Dupačová and R. J.-B. Wets: Asymptotic Behaviour of statistical Estimates and Optimal Solutions of Stochastic Optimization Problems. Annals of Statistics *16* (1984), 1517–1549.

[6] J. Dupačová: Multistage Stochastic Programs: the State–of–the–Art and Selected Bibliography. Kybernetika *31* (1995), 2, 151–174.

[7] I. A. Ibragimov and Y. V. Linnik: Independent and Stationary Sequences of Random Variables. Walter Nordhoof Pub., Gronongen 1971.

[8] Homen de Mello, T.: On Rates of Convergence for Stochastic Optimization Problems under Non–i.i.d. Sampling. Speps 2006.

[9] O. Kallenberg: Foundations of Modern Probability. Springer, Berlin 1997.

[10] V. Kaňková: Optimum Solution of a Stochastic Optimization Problem with Unknown Parameters. In: Trans. 7th Prague Conf. 1974. Academia, Prague 1977, 239–244.

[11] V. Kaňková: An Approximative Solution of Stochastic Optimization Problem. In: Trans. 8th Prague Conf. 1977. Academia, Prague 1978, 349–353.

[12] V. Kaňková: On the Convergence Rate of Empirical Estimates in Chance Constrained Stochastic Programming. Kybernetika *26* (1990), 6, 449–461.

[13] V. Kaňková and P. Lachout: Convergence Rate of Empirical Estimates in Stochastic Programming. Informatica *3* (1992), 497–523.

[14] V. Kaňková: A Note on Estimates in Stochastic Programming. J Comput. Math. *56* (1994), 97–112.

[15] V. Kaňková: A Remark on the Analysis of Multistage Stochastic Programs: Markov Dependence. Z. angew. Math. Mech. *82* (2002), 11–12, 781–793.

[16] V. Kaňková and M. Houda: Empirical Estimates in Stochastic Programming. In: Proceedings of Prague Stochastics 2006. (M. Hušková and M. Janžura, eds.), MATFYZPRESS, Prague 2006, 426–436.

[17] V. Kaňková. Empirical Estimates via Stability in Stochastic Programming. Research Report No. 2192, Institute of Information Theory and Automation Prague 2007.

[18] V. Kaňková: Multistage Stochastic Programs via Autoregressive Sequences and Individual Probability Constraints. Kybernetika *44* (2008), 2, 151–170.

[19] V. Kaňková: Empirical Estimates in Stochastic Optimization via Distribution Tails. Kybernetika 2010 *46*, 3, 459–471.

[20] G. Ch. Pflug: Scenario Tree Generation for Multiperiod Financial Optimization by Optimal Discretization. Mathematical
Programming Ser. B. *89* (2001), 251–271.

[21] G. Ch. Pflug: Stochastic Optimization and Statistical Inference. In: Stochastic Programming (A. Ruszczynski& A. A. Shapiro, eds.). Handbooks in Operations Research and Managemennt Science, Vol 10, Elsevier, Amsterodam 2003, 427–480.

[22] W. Römisch and R. Schulz. Stability of Solutions for Stochastic Programs with Complete Recourse. Math. of Oper. Res. *18* (1993), 590–609.

[23] A. Shapiro: Quantitative Stability in Stochastic Programming. Mathematical Programming *67* (1994), 99–108.

[24] A.Shapiro and H. Xu, H.: Stochastic Mathematical Programs with Equilibrium Cnstraints, Modelling and Sample Average Approximation. Optimization *57* (2008), 395–418.

[25] A.Shapiro, D. Dentcheva, D. and A. Ruszczynski: Lectures on Stochastic Programing ((Modeling and Theory). Published by Society for Industrial and Applied Mathematics and Mathematical programming Society, Philadelfia 2009.

[26] R. Schulz. Rates of Convergence in Stochastic Programs with Complete Integer Recourse. SIAM J. Optimization *6, 4* (1996), 1138–1152.

[27] R. J. B. Wets: A statistical Approach to the Solution of Stochastic Programs with (Convex) Simple Recourse. Research Report, University Kentucky, USA 1974.

[28] Yoshihara, K.: Wekly Dependent Stochastic Sequences and Their Applications, vol I Summation Theory for Weakly Dependent Sequences. Sanseido, Tokyo 1992.

On limit theorems for weakly dependent sequences

Jana Klicnarová[1]

Abstract. Classical limit theorems – a central limit theorem, an invariance principle and others – are established for independent random variables. But, in many economics applications, the condition of independence is often violated and limit theorems for weakly dependent random variables must be used.

Therefore, the aim of the paper is to introduce some useful limit theorems which can be used for weakly dependent sequences, to study their conditions and to illustrate their usage.

Keywords: central limit theorem, invariance principle, weak dependent sequences

JEL classification: C49
AMS classification: 60F05

1 Introduction

In many practical economics and econometrics problems, we deal with time series or other sequences of data which come as observations of the sequence of random variables and we use statistical technics for achieving some results. A lot of statistical technics is based on the assumption of the independence of random variables in the time series or in our sequences. But in practical time series or data sequences, the condition of independence is often violated. If we want to use statistical technics in such cases, we need rigorous limit theorems for non independent random variables.

2 Notation

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and $T : \Omega \to \Omega$ be a bijective bimeasurable and measure preserving transformation. Let X_0 be a zero mean random variable with finite second moment. We define the stationary sequence $(X_k)_{k \in \mathbf{Z}}$ by $X_k = X_0 \circ T^k$. (It is well known fact that every stationary sequence can be represented in this way.) By $(\mathcal{F}_k)$ we will denote an increasing filtration where $T^{-1}(\mathcal{F}_k) = \mathcal{F}_{k+1}$, by H_k we will denote the space $L^2(\mathcal{F}_k)$, P_k will denote the orthogonal projection onto $H_k \ominus H_{k-1}$. If X_0 is $\mathcal{F}_0$-measurable, we say that the process (X_k) is adapted (to the filtration $(\mathcal{F}_k)$).

Let us put

$$S_n := \sum_{k=1}^{n} X_k.$$

Then, by σ_n^2 we denote $\mathrm{E}(S_n^2)$.

We say that the sequence satisfies a Central Limit Theorem (CLT) (for some sequence of positive numbers $(s_n)_{n \in \mathbb{N}}$) if

$$\left(\frac{S_n}{s_n}, n \in \mathbb{N} \right)$$

converges in distribution to a mixture of normal distribution. If

$$\left(\frac{S_{\lfloor nt \rfloor}}{s_n}, t \in [0, 1] \right)$$

[1]University of South Bohemia in České Budějovice, Faculty of Economics, Studentská 13, 370 05 České Budějovice, e-mail: janaklic@ef.jcu.cz

converges in distribution to a mixture of Wiener measures in the space of cadlag functions – $D([0,1])$ then we say that the the sequence (S_n) satisfies Invariance Principle (IP) for some sequence of positive numbers $(s_n)_{n\in\mathbb{N}}$.

3 The types of Dependence

In case of a sequence of independent random variables, we can use classical limit theorems and classical inequalities, as we mentioned above. But what happen if our sequence is not the sequence of independent random variables? (Unfortunately, it is very frequent problem in a practical usage.) We know, that the sequence contains any dependence, but how to specify the dependence, how to measure the dependence in the sequence?

At first let us recall what it is independence. If we have two real random variables X, Y on $(\Omega, \mathcal{A}, P)$ then we say that X and Y are **independent** if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B), \quad \text{for all } A \in \mathcal{B}_X,\ B \in \mathcal{B}_Y.$$

(Where $\mathcal{B}_X, \mathcal{B}_Y$ are σ-fields generated by random variables X and Y. The second way how to define the independence is based on use of bounded real functions. Real random variables X and Y are **independent** if

$$\operatorname{cov}(f(X), g(Y)) = 0, \quad \text{for all } f, g \text{ bounded real functions (right measurable).}$$

And how to handle with data, if the previous inequalities are not satisfied? There are three main ways of approach to weak dependent data. First of them, and probably the most commonly used one, is using of mixing coefficients or also mixing conditions. Mixing conditions were introduced by Rosenblatt (see [14]) and are based on σ-fields generated by a random sequence. For instance, for two sub-σ-fields $\mathcal{B}$ and $\mathcal{C}$ subsets of $\mathcal{A}$, the α- and ϕ-mixing coefficients are defined as follows. (The mixing coefficient for random variables correspond to mixing coefficient of their σ-fields.)

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{B\in\mathcal{B}, C\in\mathcal{C}} |P(B \cap C) - P(B) \cdot P(C)|,$$

$$\phi(\mathcal{B}, \mathcal{C}) = \sup_{B\in\mathcal{B}, C\in\mathcal{C}} \left| \frac{P(B \cap C)}{P(B)} - P(C) \right|.$$

Their usage is quite technical. For more details see [2].

The second approach to the weak dependence follows the work of Gordin ([7]) and uses so called martingale approximations or mixingales. Let us recall one of the basic definition of martingale approximation. (By $||\cdot||_2$ we denote the Euclidian norm.)

Definition 1. We say that X_0 (process $(X_0 \circ T^i)$) has **a martingale approximation** if there exists a martingale difference sequence $(m \circ T^i)_{i\in\mathbf{Z}}$ such that $||S_n(X_0 - m)||_2^2 = o(n)$.

Because the CLT holds for stationary sequences of L^2 martingale differences, this condition guarantees a CLT for $(X_0 \circ T^i)$ (if μ is ergodic, the limit law is normal and for μ non ergodic it can be mixture of normal laws). Since 1969, the date of publication of Gordin's paper [7], martingale approximations have been an important tool in the research of limit theorems for stationary processes. Conditions for a martingale approximation can be found e.g. in [8], [13], [5]. We will focus on this approach.

The last most commonly used way of handling with weak dependent data is the idea coming from Bickel and Bü hlmann ([3]) and Doukhan and Louhichi ([6]). Their conception of weak dependence is based on an asymptotic independence between "past" and "future", respectively that

$$\operatorname{cov}(f(\text{"}past\text{"}), g(\text{"}future\text{"}))$$

is for all function f, g small enough if the distance between the "past" and "future" is sufficiently large.

Obviously, we can also speak about strong dependent data, but it is not a subject of this paper, for more details see for example [1].

4 Limit Theorems

There exists a lot of papers on limit theorems for weakly dependent sequences. If we focus on the limit theorems under martingale type's conditions we can mention, as above, for example [8], [13], [5].

One of the most useful seems to be following theorem which is by Dedecker, Merlevède and Volný (see [4]). Recall that $\mathrm{E}(X|\mathcal{F})$ denotes the orthogonal projection of X on to the linear subspace of $\mathcal{F}$-measurable random variables, see for example [9].

Theorem 1. *Let $m \in H_0 \ominus H_{-1}$. The following conditions are equivalent:*

1. $\lim_{n\to\infty} \left|\left|\frac{S_n}{s_n} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m \circ T^i\right|\right|_2 = 0,$
2. (a) $||\mathrm{E}(S_n|\mathcal{M}_0)||_2 = o(s_n)$ *and* $||S_n - \mathrm{E}(S_n|\mathcal{M}_n)||_2 = o(s_n)$,

   (b) $\lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n} \left|\left|\frac{\sqrt{n}}{s_n}\sum_{i=1-l}^{n-l} P_0(X_i) - m\right|\right|_2^2 = 0.$

If one of these conditions holds then S_n/s_n converges in distribution to $\sqrt{\mathrm{E}(m^2|\mathcal{I})}N$, where N is a standard Gaussian random variable independent of $\mathcal{I}$.

Recall, that we say that the a sequence X_n is $o(n)$ if $\frac{X_n}{n}$ tends to zero as n tends to infinity.

Remark 1. The first part of the second condition in the theorem is in fact the Wu-Woodroofe condition (if we take $s_n = \sigma_n$) (see [16]) which is equivalent to existence of martingale array approximation. But this condition is not sufficient for CLT nor Wu-Woodroofe condition and linear growth of variances are not sufficient conditions for CLT, see [11].

Remark 2. In fact, in practical usage, the sequences are adapted, so the projections $P_0(X_i) = 0$ for $i < 0$ and $||S_n - \mathrm{E}(S_n|\mathcal{M}_n)||_2 = 0$, too.

In the same paper, they also presented an invariance principle. (By $\mathbb{I}$ we denote a characteristic function.)

Theorem 2. *Assume that $s_{\lfloor nt \rfloor}/s_n$ is bounded for any $t \in [0,1]$. If*

- $$\lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n} \left|\left|\frac{\sqrt{n}}{s_n}\sum_{i=1-l}^{n-l} P_0(X_i) - m\right|\right|_2^2 = 0,$$
- $$\left|\left|\sup_{1\le k\le n} |\mathrm{E}(S_k|\mathcal{F}_0)\right|\right|_2 = o(s_n), \quad \left|\left|\sup_{1\le k\le n} |S_k - \mathrm{E}(S_k|\mathcal{F}_n)\right|\right|_2 = o(s_n),$$
- *for some positive sequence $(u_i)_{i\in\mathbf{Z}}$ such that $\frac{\sqrt{n}}{s_n}\sum_{i=-n}^{n} u_i$ is bounded,*
$$\lim_{A\to\infty} \limsup_{n\to\infty} \frac{\sqrt{n}}{s_n} \sum_{i=-n}^{n} \mathrm{E}\left(\frac{P_0^2(X_i)}{u_i} \mathbb{I}_{P_0^2(X_i) > Au_i^2}\right) = 0,$$

then $(S_{\lfloor nt \rfloor}/s_n\ ,\ t \in [0,1]$ converges in distribution in $(D([0,1]), d)$ to $\mathrm{E}(m^2|\mathcal{I})W$, where W is a standard Brownian motion independent of $\mathcal{I}$.

Similar results (to the previous given by Dedecker, Merlevede and Volný) are presented in the paper by Zhao and Woodroofe ([17]).

They introduced the plus-norm (cf. [17])

$$||f||_+^2 = \limsup_{n\to+\infty} \frac{1}{n}||S_n(f)||_2^2$$

and a condition which we call Zhao-Woodroofe's condition (ZW):

Definition 2. Let the process $(X_k)_{k\in\mathbf{N}}$ be adapted. We say that it satisfies **Zhao-Woodroofe's condition (ZW)** if

$$\lim_{m\to+\infty}\frac{1}{m}\sum_{k=1}^{m}||\mathrm{E}(X_0|\mathcal{F}_{-k})||_+^2=0.$$

Zhao and Woodroofe proved the following theorem (see [17]).

Theorem 3. *Let $X_0\in L^2$, then an adapted process $(X_k)_{k\in\mathbf{N}}$ has a martingale approximation if and only if it satisfies ZW condition.*

Remark 3. Zhao and Woodroofe stated their Theorem only for adapted sequences. But in [12] it is shown, how to arrange the Zhao-Woodroofe's condition for non-adapted cases and then the theorem is true under the non-adapted ZW condition, too.

Remark 4. The second main difference of Zhao-Woodroofe's Theorem from Dedecker-Merlevède-Volný's Theorem is that the Zhao-Woodroofe's one requires linear growth of variances against Dedecker-Merlevede-Volný's one. But Zhao and Woodroofe proved the existence of martingale approximation in L^2, what is a stronger result then the satisfaction of CLT. And recall that in DMV's theorem we need to know the limit martingale m, but in ZW's one we have conditions only on our sequence.

As we mentioned above, since we suppose the finite second moment of our variables, the martingale approximation ensures the CLT (even the coditional CLT (see [13])). But the Invariance principle is not guaranteed. For Invariance Principle we need some more conditions. We can mention the Invariance Principle given by Wu and Woodroofe in [16]. It is again only for adapted sequences, but in [10] it is possible to find the non-adapted version of this Theorem.

Theorem 4. *Let P be ergodic and let $X_0\in L^p$ for some $p>2$ be $\mathcal{F}_0$-measurable, and*

$$\|E(S_n|\mathcal{F}_0)\|_2=o\Big(\frac{\sqrt{n}}{\log^q n}\Big) \tag{1}$$

for a $q\geq 2$. Then the process of $S_n(t)$ converges in distribution to a Brownian motion in the space $D[0,1]$, conditionally with respect to $\mathcal{F}_0$.

5 Applications

In a variety of applied fields, the linear process $(X_k)_{k\in\mathbf{Z}}$ is widely used. It is defined for any squared summable sequence $(a_i)_{i\in\mathbf{Z}}$ as follows:

$$X_k=\sum_{i=-\infty}^{\infty}a_i\varepsilon_{k-i},$$

where $\varepsilon_0\in H_0\ominus H_{-1}$ with zero mean and finite second moment, $\varepsilon_k=\varepsilon_0\circ T^k$ and $(a_i)_{i\in\mathbf{Z}}$ is a sequence of real constants. It is easy to see that the satisfying of conditions for CLT or IP by the linear process (X_k) depends on the sequence of $(a_k)_{k\in\mathbf{Z}}$. So, the question is, under which conditions for the sequence (a_i) the previous theorems or some others are satisfied.

One of the useful theorems is Corollary 4 in [4].

Corollary 1. If the linear process is defined as above, let us put $s_n=\sqrt{n}|a_{-n}+\dots a_n|$. If

- $$\limsup_{n\to\infty}\frac{\sum_{i=-n}^{n}|a_i|}{|\sum_{i=-n}^{n}a_i|}<\infty,$$
- either $\sum_{k=1}^{n}\sqrt{\sum_{i:|i|\geq k}a_i^2}=o(s_i)$ or $\sum_{i\in\mathbf{Z}}|a_i|<\infty$,

then the conclusion of Theorem 2 is true.

In [15] we can find following statement:

Proposition 5. *For the linear process defined above, if* $\mathrm{E}|\varepsilon_0|^{2+\delta} < \infty$ *for some positive constant* δ*, and*

- $\sum_{i=0}^{n-1}(\sum_{k=0}^{i} a_k^2) \to \infty$
- $\sum_{i=0}^{\infty}(\sum_{k=i+1}^{n+i} a_k)^2 = o\left(\sum_{i=0}^{n-1}(\sum_{k=0}^{i} a_k)^2\right)$,

then the sequence satisfies the Invariance Principle.

Let us put $b_n = \sum_{i=0}^{n} a_i$. Then the Wu-Woodroofe's condition ($||\mathrm{E}(S_n|\mathcal{F}_0)||_2 = o(\sigma_n)$) could be for linear process rewritten as

$$\sum_{j=0}^{\infty}(b_{j+n} - b_j)^2 = o\left(\sum_{k=1}^{n-1} b_k^2\right).$$

And Zhao-Woodroofe's condition for linear process is

$$\lim_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} \limsup_{n\to\infty} \frac{1}{n}\left(\sum_{l=-\infty}^{-k}(b_{n-l} - b_{-l})^2 + \sum_{l=-k+1}^{-k+n}(b_{n-l} - b_{-k})^2\right) = 0.$$

When $a_i = i^{-\beta}$, where $\beta : 1/2 < \beta < 1$, then Wu-Woodroofe's condition (and Zhao-Woodroofe's, too) is failed. Wu-Woodroofe's condition is satisfied for example for $a_0 = 0$, $a_1 = 1/\log 2$ and $a_k = 1/\log(k+1) - 1/\log(k)$ for $k > 1$. If $a_0 = 1$ and $a_i = -1/i$, then the Wu-Woodroofe's condition is satisfied but the Zhao-Woodroofe's one is not.

Remark 5. The IP by Wu-Woodroofe is satisfied for linear processes with coefficients $a_i = i^{-\beta}(-1)^i$ for $i \geq 1$ and $\beta :\ 1/2 < \beta < 1$.

Acknowledgements

This paper is supported by a grant GAČR P201/11/P164.

References

[1] Beran, J.: *Statistics for long memory processes.* Chapman& Hall, 1994.

[2] Bradley, R., C.: *Introduction to strong mixing conditions.* Kendrick Press, Heber City, UT, 2007.

[3] Bickel, P., Bühlmann, P.: A new mixing notion and functional central limit theorems for a sieve bootstrap in time series, *Bernoulli*, **5-3** (1999), 413-446.

[4] Dedecker, J., Merlevède, F., Volný, D.: On the weak invariance principle for non-adapted sequences under projective criteria, *J. Theor. Probab.*, **20** (2007), 971-1004.

[5] Dedecker, J., Rio, E.: On the functional central limit theorem for stationary processes, *Ann. Inst. Henri Poincaré Prob. Statist.*, **36** (2000), 1-34.

[6] Doukhan, P., Louchichi, S.: A new weak dependence condition and applications to moment inequalitites, *Stoch. Proc. Appl.* **84** (1999), 313-342

[7] Gordin, M. I.: A central limit theorem for stationary processes, *Soviet Math. Dokl.*, **10** (1969), 1174-1176.

[8] Hall, P., Heyde, C. C.: *Martingale limit theory and its application.* Academic Press, New York, 1980.

[9] Kallenberg, O.: *Foundations of Modern Probability.* Springer, New York, 1997.

[10] Klicnarová, J., Volný, D.: An Invariance Principle for non Adapted Processes. *C.R. Acad. Sci. Paris*,**345** (2007), 283-287.

[11] Klicnarová, J., Volný, D.: On the exactness of the Wu-Woodroofe approximation, *Stoch. Procc. Appl.*, **119** (2009), 2158-2165.

[12] Klicnarová, J., Volný, D.: On Zhao-Woodroofe's condition for martingale approximation, *submitted to El. Journal of Prob.*, (2011).

[13] Maxwell, M., Woodroofe, M.: Central limit theorems for additive functionals of Markov chains, *The Annals of Probability* **28** (2000), 713-724.

[14] Rosenblatt, M.: A central limit theorem and a strong mixing condition, *Proc. Nat. Ac. SC.* **42** (1956), 43-47.

[15] Wu, W.B., Min, W.: On linear processes with dependent innovations for sums of stationary processes, *Stoch. Proc. Appl.* **115** (2005), 939-958.

[16] Wu, W.B., Woodroofe, M.: Martingale Approximations for Sums of Stationary Processes, *The Annals of Probability* **32** (2004), 1674-1690.

[17] Zhao, O., Woodroofe, M.: On Martingale Approximations, *The Annals of Applied Probability* **18** (2008), 1831-1847.

Testing the New Keynesian Model on Czech data

Jan Kodera[1], Tran Van Quang[2]

Abstract. With inflation targeting regime, New Keynesian model provides an essential theoretical framework to central bankers. In its reduced version, the New Keynesian model consists of two fundamental equations: the dynamic IS curve which captures relationship among outputs, inflation and interest rates and new Phillips curve which describes the relationship of inflation as a nominal variable with a real economic variable. Though there have been numerous works both on monetary policy and on New Keynesian model, there still is a lack of knowledge how New Keynesian model fits well with real data. For this reason, in our work, we will test the validity of New Keynesian model on the real Czech economic data by using a VAR model.

Keywords: New Keynesian Model, Dynamic IS Curve,New Keynesian Phillips Curve,VAR Models

JEL classification: C44
AMS classification: 90C15

1 Introduction

New Keynesian model is the principal theoretical framework for central banks to conduct monetary policy in the regime of inflation targeting. The basic New Keynesian model with the dynamic IS curve and new Phillips curve relates output as a real economic variable, with inflation and interest rates as nominal variables as well as it connects the real world with the world of expectation. Though there have been numerous works both on monetary policy and on New Keynesian model, there still is a lack of knowledge how New Keynesian model fits well with real data. Further, as the Czech National Bank has conducted the monetary policy through inflation targeting for some time it is worth thoroughly examining how the theory fits well with reality. For this objective, we use the real quarterly data from the Czech Republic from 1st quarter of 2000 to 4th quarter of 2010 to verify the validity of the New Keynesian model.

2 Basic Principles of New Keynesian model

In this introductory section we try to give a short overview of a simplified New Keynesian model. The behaviour of aggregate household will be shown at first, then the individual firm will be presented with its price setting policy. Our explanation of New Keynesian model will be concluded by equilibrium conditions. Necessary conditions for of New Keynesian model are used for the derivation of dynamic IS curve and New Keynesian Phillips curve. The system of these derived curves is very often called as reduced form of New Keynesian model. What is typical for New Keynesian models is using Calvo price mechanism, which is based on the assumption, that some firms keep prices of the last period and the others reset the prices so that they maximise their expected discounted profit in infinite period. More complex New Keynesian models include other components such as government, central bank eventually others.
The amount of parameters in real business cycle models is not too high and are included in utility, production functions and in composite indices of consumption and price levels. There are in fact two methods setting up of model parameters. The first one is calibration which should be adapted in such a way that the trajectories of model variable correspond to the actual process. The second approach is the econometric one which verifies the correctness of the models by using its tests. In our paper we are

[1]University of Economics, Prague/Centre of Dynamic Economics and econometrics, Nam W. Churchilla 4, Praha 3, e-mail: kodera@vse.cz
[2]University of Economics, Prague/Centre of Dynamic Economics and econometrics, Nam W. Churchilla 4, Praha 3, e-mail: tran@vse.cz

going to use second approach and test the model by comparison to VAR model with the same variables estimated on actual data.

2.1 Household

In a New Keynesian model the utility function is related to a composite index of consumption C_t and to labour supply N_t. The set of commodities is continuum, so for indexation we use all numbers from interval $[0,1]$. The composite index of consumption is given by

$$C_t = \left[\int_0^1 C_t^{1-\frac{1}{\epsilon}}(i)\mathrm{d}i\right]^{\frac{\epsilon}{\epsilon-1}},$$

where ϵ is elasticity of substitution between 2 goods. A representative infinitely living household maximizes its utility function

$$\sum_{t=1}^{\infty} \beta^t U(C_t, N_t), \text{ s.t. budget restriction } \int_0^1 P_t(i)C_t(i)\mathrm{d}i + Q_t B_t \leq B_{t-1} + W_t N_t + Z_t, \tag{1}$$

where β is the household's discount factor. Let us recall that $P_t(i)$, B_t, W_t, Z_t denote the price of ith commodity, discount bond, wage rate, and household subsidies respectively. The variable Q_t is defined $Q_t = 1/(1+i_t)$, where i_t denotes market interest rate. Let us introduce price level P_t and demand for commodity i function:

$$P_t = \left[\int_0^1 P_t^{1-\epsilon}(i)\mathrm{d}i\right]^{\frac{1}{1-\epsilon}}, \qquad C_t(i) = \left[\frac{P_t(i)}{P_t}\right]^{-\epsilon} C_t.$$

Demand function shows how the composite index of consumption is distributed into different commodities with regard to their relative prices. Using demand function and the definition of price index we can rearrange the budget restriction from (1)

$$P_t C_t + Q_t B_t \leq B_{t-1} + W_t N_t + Z_t. \tag{2}$$

Let us assume $U(C_t, N_t) = \dfrac{C_t^{1-\sigma}}{1-\sigma} - \dfrac{N_t^{1+\varphi}}{1+\varphi}$, where σ and φ are the inverses of the intertemporal elasticities of substitution for consumption and labour supply. For derivation of necessary conditions of household equilibrium we use Lagrange method applied for infinite dimensional space (the time horizon is infinite). After using it, we get the following log-linear version of necessary conditions:

$$\sigma c_t + \varphi n_t = w_t - p_t, \qquad c_t = \mathrm{E}_t c_{t+1} - \frac{1}{\sigma}(i_t - \mathrm{E}_t \pi_{t+1} - \rho), \tag{3}$$

where $\beta = 1/(1+\rho)$ and ρ is the household's interest rate.

2.2 Firm

The technology of firm is described by one-factor production function

$$Y_t(i) = A_t L_t^{1-\alpha}(i),$$

where α is the parameter of the simplified production function. Let us recall that the firm produces only one product i amount of it is $Y_t(i)$ using one factor labour which quantum is $L_t(i)$. The firms in the environment of New Keynesian model are assumed to be monopolist sui generis. In New Keynesian models the Calvo's price mechanism of staggered prices is used. It means that a Neo-keynesian firm keeps the prices constant with probability θ or change them with probability $1-\theta$ in order to maximize its discounted expected profit in infinite horizon

$$\max_{P_t^*} \sum_{k=t}^{\infty} \theta^k \mathrm{E}_t \left\{ Q_{t,t+k}[P_{t+k}^* Y_{t+k|t} - \Psi(Y_{t+k|t})]\right\}., \text{ subject to } Y_{t+k|t} = \left[\frac{P^*}{P_{t+k}}\right]^{-\epsilon} C_{t+k}.$$

Variable P_t^* is a common price of firms resetting their price in period t, $Q_{t,t+k} = 1/[(1+i_t)\dots(1+i_{t+k-1}]$ is market discount factor discounting profit unit reached in the time of $t+k$ to a period t and $Y_{t+k,t}$ is production of firms resetting their price in time t.

2.3 Market equilibrium and the derivation of IS and Phillips curves

There is a continuum of commodity markets for each commodity and one labour market, because the household supply homogeneous labour for continuum of firms which demand for labour is $L_t(i)$. Commodity markets and labour market are assumed to be in equilibrium state

$$C_t(i) = Y_t(i), \quad \int_0^1 L_t(i)\mathrm{d}i = N_t.$$

Aggregating the equations of commodity equilibrium, we get the equation of aggregated equilibrium in commodity market $C_t = Y_t$. Taking logarithm we get $c_t = y_t$ and substituting it into (3), we get:

$$y_t = \mathrm{E}_t y_{t+1} - \frac{1}{\sigma}(i_t - \mathrm{E}_t\pi_{t+1} - \rho). \tag{4}$$

Let us begin with the New Keynesian Phillips curve. The relation between the real marginal costs mc_t and the real wage $w_t - p_t$ and the marginal product of labour mpn_t can be expressed as (see Gali [2]):

$$\begin{aligned} mc_t =& w_t - p_t - mpn_t = (\sigma y_t + \varphi n_t) - (y_t - n_t) - \log(1-\alpha) = \\ =& \left(\sigma + \frac{\varphi + \alpha}{1-\alpha}\right) y_t - \frac{1+\varphi}{1-\alpha} a_t - \log(1-\alpha) \end{aligned} \tag{5}$$

For the model with flexible prices, i.e. firms reset their prices in each instant of time $mc = \mu$, where μ is a constant. The level of production in the economy with flexible prices is called natural level of production denoted by y_t^n.
The relation between real marginal cost and real wage and marginal product of labour in the environment with flexible prices is expressed as:

$$mc = \left(\sigma + \frac{\varphi + \alpha}{1-\alpha}\right) y_t^n - \frac{1+\varphi}{1-\alpha} a_t - \log(1-\alpha). \tag{6}$$

Subtracting (6) from (5) obtains

$$\widetilde{mc}_t = \left(\sigma + \frac{\varphi + \alpha}{1-\alpha}\right)(y_t - y_t^n). \tag{7}$$

Now we use a very important equation derived from the theory of staggered prices and applied in New Keynesians models for the price setting process (see [2]):

$$\pi_t = \beta \mathrm{E}_t \pi_{t+1} + \lambda \widetilde{mc}_t, \quad \text{where } \lambda \text{ is a positive parameter.} \tag{8}$$

Replacing from (7) to (8) we get the New Keynesian Phillips curve:

$$\pi_t = \beta \mathrm{E}_t \pi_{t+1} + \kappa \hat{y}_t, \tag{9}$$

$$\text{where} \quad \mathrm{\hat{y}_t = y_t - y_t^n} \quad \text{and} \quad \kappa = \lambda\left(\sigma + \frac{\varphi + \alpha}{1-\alpha}\right).$$

To introduce the Dynamic IS curve, we begin with reformulating the equation (4), we obtain:

$$y_t^n = \mathrm{E}_t y_{t+1}^n - \frac{1}{\sigma}(i_t^n - \mathrm{E}_t \pi_{t+1}^n - \rho). \tag{10}$$

Subtracting (10) from (4)gives:

$$\hat{y}_t = \mathrm{E}_t \hat{y}_{t+1} - \frac{1}{\sigma}(i_t - \mathrm{E}_t \pi_{t+1} - r_t^n) \tag{11}$$

Equations (9) and (11) form the reduced model of economy. The equation (9) generate price dynamics provided production gap is given and the equation (11) gives production dynamics provided inflation is given. Both equations form system of difference equations which could be adapted including behaviour of central bank into system. This behaviour is described by reaction equation of central bank

$$i_t = \rho + \phi_\pi \pi_t + \phi_y \hat{y}_t + v_t.$$

The variable v_t represents shocks in monetary policy and is assumed to be generated by autoregressive process

$$v_t = \rho_v v_{t-1} + \epsilon_t^v.$$

From the system of equations (9), (11) the following system of difference equations is derived:

$$\begin{bmatrix} \hat{y}_t \\ \pi_t \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathrm{E}_t \hat{y}_{t+1} \\ \mathbf{E}_t \pi_{t+1} \end{bmatrix} + \mathbf{B}(\tilde{r}_t^n - v_t), \quad \text{where} \quad \tilde{\mathrm{r}}_\mathrm{t}^\mathrm{n} = \mathrm{r}_\mathrm{t}^\mathrm{n} - \rho.$$

3 Verification of New Keynesian model on Czech quarterly data

The two fundamental equations thoroughly derived above combined with a Taylor type policy rule equation form the following system:

$$\hat{y}_t = \beta_1 \mathrm{E}_t \hat{y}_{t+1} + \beta_2 (i_t - \mathrm{E}_t \pi_{t+1}) + \varepsilon_{1t} \tag{12}$$

$$\pi_t = \beta_3 \mathrm{E}_t \pi_{t+1} + \beta_4 \hat{y}_t + \varepsilon_{2t} \tag{13}$$

$$i_t = \beta_5 \pi_t + \beta_6 \hat{y}_t + \varepsilon_{3t}. \tag{14}$$

There are four endogenous variables in the system: y_t, $\hat{y}_t$, π_t and i_t. Provided that the agents in the economy are rational and have rational expectation, if the system is valid, then all these variables must be integrated of order 1, then they must also be cointegrated and finally there must be three cointegrating vectors among them. Therefore, in order to verify the validity of New Keynesian model, we need to test whether our endogenous variables in the model y_t, $\hat{y}_t$, π_t are integrated of order 1 and if there are actually 3 cointegrating vectors among them. For this purpose we use series of seasonally unadjusted quarterly data on GDP of the Czech Republic from 1. quarter of 2000 to 4. quarter of 2010. This series is deflated to get a series of real GDP of the Czech Republic for the period we are interested in and after that it was transformed into a log series titled lhdp. We use CPI index as the measure for inflation from the same period and in the paper the series is named as inflace. Both two series are publicly available from the datasource of the Czech Statistical Office. We choose the series of interbank deposits interest rate Pribor 3 months as the short term interest rate. The reason for this choice is there is no better alternative in the Czech Republic and this interest rate is directly affected by the monetary policy interest rate "repo sazba". This seires is denoted as p3m. In order to obtain series of output gap, we use Prescott-Hodrick filter to generate it from the series of quarterly real GDP from the period 1.2000 to 4.2010. The principle of Prescott-Hodrick filter is very simple as following: suppose that real GDP y_t is consisted of two components: the trend component T_t and the cyclical one C_t. For a given value of parameter λ, we must find the trend component such that it minimizes the weighted sum of all squared cyclical components and the sum of second differences of trend components:

$$\min \left(\sum_{t=1}^{T} (y_t - T_t)^2 + \lambda \sum_{t=2}^{T-1} \left[(T_{t+1} - T_t) - (T_t - T_{t-1}) \right]^2 \right). \tag{15}$$

We use the widely recommended value of $\lambda = 1600$ for quarterly data and make use of the econometric package Eviews to get the series of log real output gap which we denote as lhdp-hp. These series then are tested for unit root by using augmented Dickey - Fuller test by using Eviews again. The test result is shown in table 1. The result of the testing clearly shows that we cannot reject the null hypothesis that the

Table 1: The results of testing for unit root

series	test stats	p-value	H0: has unit root
lhpd	-2.398854	0.3744	Not rejected
Lhdp-hp	7.833797	1.0000	Not rejected
inflace	-3.369826	0.0703	Not rejected
P3m	-2.375409	0.3863	Not rejected

examined series has a unit root in any case at the 5% significant level, which implies that all series in the model are integrated of order one. This result also justifies our proceeding to test for cointegration among our endogenous variables in the model. For doing so, we use the Johansen procedure. The essence of the

Table 2: The results of testing for cointegration:

λ_{trace} test				
H0:the number of CV	Eigenvalue	Trace statistic	Critical Value at 5	p-value
None	0.964426	211.0587	54.07904	0.0000
At most 1	0.589926	70.94040	35.19275	0.0000
At most 2	0.487594	33.50085	20.26184	0.0004
At most 3	0.121027	5.418044	9.164546	0.2407
λ_{max} test				
None	0.964426	140.1183	28.58808	0.0000
At most 1	0.589926	37.43955	22.29962	0.0002
At most 2	0.487594	28.08280	15.89210	0.0004
At most 3	0.121027	5.418044	9.164546	0.2407

Johansen procedure is following (see Enders [1]) . Suppose that $x_t = (lhdp, lhdp - hp, inflace, p3m)$' is a VAR(1) process (for simplicity), therefore we can write:

$$x_t = Ax_{t-1} + \epsilon_t. \tag{16}$$

We can rewrite it as:

$$\Delta x_t = (A - I)x_{t-1} + \epsilon_t = \pi x_{t-1} + \epsilon_t. \tag{17}$$

From this it is clear that the rank of matrix π is the number of independent relationship among Δx_{it} and x_{t-1} which are stationary. And since the rank of π equals the number of its characteristic roots which are different from 0, the Johansen test examines how many characteristic roots of π are statistically different from zero. The Johansen procedure gives us two test statistics: λ_{max} and λ_{trace}. λ_{trace} tests a specific null hypothesis against a general alternative (for example the number of nonzero characteristic root $r = 1$ as H0 against $r > 1$ as Ha) while λ_{max} tests a specific null hypothesis against a specific alternative (that is for example the number of nonzero characteristic root $r = 1$ as H0 against $r = 2$ as the alternative). The result of testing for cointegration in our four endogenous variables by the Johansen procedure is shown in table 2. Let's interpret the results of testing for cointegration step by step. According to the λ_{trace} test, if we test the null hypothesis $r = 0$, e.g. the number of nonzero eigenvalue of matrix π equals 0 or the number or cointegrating vectors is zero against the alternative hypothesis that $r > 1$, we have to reject H0 at 5% significance level for the alternative hypothesis. We get the same results for H0 when $r = 1$ and $r = 2$ against $r > 1$ and $r > 2$. When we test the null hypothesis for $r = 3$ against the alternative $r > 3$, according to test statistic and the critical value for 5% significance level, as well as the p-value, we cannot reject the null hypothesis. This means that there are three statistically nonzero eigenvalues of matrix π, and therefore there also three coitegrating vectors among our endogenous variables: the real GDP, the GDP gap, the inflation rate and the interbank deposits interest rate. These results are confirmed by the results of testing for cointegration by λ_{max} test. In this case, we also reject the null hypothesis when it is $r = 0$, $r = 1$ and $r = 2$, against the alternative $r = 1$, $r = 2$ and $r = 3$ respectively. If we test for the null hypothesis $r = 3$ against the alternative $r = 4$, we cannot reject the null hypothesis by λ_{max} test. The result of testing for cointegration by the Johansen procedure indicates that the system of three equations of the New Keynesian model is valid when being verified on the Czech quarterly data. Having found that the variables included in the New Keynesian model are cointegrated and together they are related in threefold cointegrating relationship corresponding to the New Keynesian model, we take another step to look into their interaction by examining the exogeneity or endogeneity of each of them. In order to do so, first we regress $lgdp$ on the remaining variables: $lgdp - hp, inflace$ and $p3m$. The results of this regression is shown in table 4 below: The result clearly shows the relevance of the relationship between $lhdp$ and the remaining variables in the model as the whole as well as the statistic significance of each explanatory variable. We use the residuals obtained from this regression to estimate the error correction model. In this case, we generate the first difference of our four variables and then regress them on the residuals of the previous regression and the lagged first difference of all variables (up to lag 4). The results for the error correction coefficients are reported below. These results tell us that while error correction coefficients for real GDP and real GDP gap are statistically different from zero, and therefore, significant, it is not the case for the estimated error correction coefficients of the inflation rate and the interest rate. This means that inflation and interest rate are two weakly exogenous variables

Table 3: The results of regression of $lgdp$ on $lgdp - hp$, $inflace$, $p3m$

Regressor	Regression Coefficient	Standard Error	t-Statistic	p-value
C	-1.488350	0.408885	-3.640021	0.0008
LHDP-HP	1.105828	0.029649	37.29752	0.0000
INFLACE	-0.016172	0.004839	-3.342284	0.0018
P3M	0.022451	0.004344	5.167775	0.0000

Table 4: The results of EC models: error correction coefficients

Regressor	Coefficient of Res(-1)	Standard Error	t-Statistic	p-value
dlhdp	-0.797701	0.296379	-2.691486	0.0133
dlhdp-HP	-3.62E-11	1.19E-11	-3.032005	0.0061
dINFLACE	-12.80629	24.80171	-0.516347	0.6108
dP3M	7.371114	7.227879	1.019817	0.3189

in the New Keynesian model. We can interpret the result of our testing the New Keynesian model on the Czech data that unlike the results of other authors ([3],[4]), all variables present in the model are connected in a long-run relationship described by the two fundamental equations of the New Keynesian model and the corresponding policy rule equation. On the other hand, when being deviated from their long run equilibrium, output and output gap tend to act in such a way so that the equilibrium can be re-established, inflation and interest rate do not react to the deviation that way. It is understandable in the case of interest rate when the central bank's activity affect it directly. In the case of inflation, it might indicate that the role of money in the long run is neutral as well as inflation is a monetary phenomenon.

4 Conclusion

The relevance of the New Keynesian model is well documented and it has become the theoretical backbone for central bankers around the world to conduct monetary policy. In this paper we use an econometric approach to verify the validity of the model. To achieve this objective, we use the publicly available quarterly data from the Czech Statistical Office and the Czech National Bank for the period from 2000 to 2010 when inflation targeting is used to perform monetary policy. Assuming the rational expectation, we use the cointegration test suggested by Johansen to have found a long run relationship among variables present in the New Keynessian model which empirically justifies the model. Nevertheless, using error correction technique, we also find in the real data that inflation is a weakly exogenous variable which indicates that in the long run, the role of money is neutral.

Acknowledgements

The authors are grateful for the financial support of Czech Science Foundation No. 402/09/0273 and of Ministry of Education No. LC06075.

References

[1] Enders, W: *Applied Econometric Time Series, 2nd edition.* Wiley & Sons, Inc., New Jersey, 2004

[2] Gali, J.: *Monetary Policy Inflation and the Business Cycle, An Introduction to the New Keynesian Framework.* Princeton University Press, Princeton and Oxford, 2008.

[3] Juselius, M.: Cointegration implications of linear rational expectation model. Bank of Finland Research Discussion Paper, Bank of Finland, Helsinki, 2008

[4] Kim Hae-min: Empirical Study of New Keynesian Model using Cointegrated VAR: What New Zealand data tell us? Partial fulfilment of the requirement of Master of Science, Massachusetts Institute of Technology, Cambridge (Mass), 2009

Comparison of different approaches to the control of cyclic processes

Michal Kohání[1], Michal Rusek[2]

Abstract. Cyclic process is a process in which the activities are repeatedly carried out in subsequent regular time intervals in the same order. There are several approaches to solving of cyclic processes. In the first one the start and finish time of all activities must be in the same interval. In the second approach some given activities are allowed to start in one interval and finish in the next interval. In the last case there are no restrictions for time of finish of the activity in the subsequent interval. One example of a cyclic process is the road intersection signal plan. An optimal intersection signal plan minimizes the total waiting time of all vehicles in the intersection. Vehicles come to intersection in flows. The start time and end time of green signal is the activity mentioned above. This time can be affected by existence of "red point". It is the moment when all flows in the intersection have red signal. In this paper we focused on the explanation of the three approaches to control of cyclic processes and on the task of designing signal plans we show the impact of these approaches on computational time and quality of the solution.

Keywords: cyclic process, intersection signal plan, red point, linearization

JEL Classification: C44
AMS Classification: 90C10

1 Introduction

Cyclic process is a process in which the activities are repeatedly carried out in subsequent regular time intervals in the same order. There is lot of examples of cyclic processes, for example processes on production lines in factories, road intersection signal plans, etc.

Several approaches to solve cyclic processes can be used. In the first one the start and finish time of all activities must be in the same interval. In the second approach some given activities are allowed to start in one interval and finish in the next interval. The third approach allows to all activities to finish in the subsequent interval.

An example of a cyclic process is the road intersection signal plan. The principle of solving the problem is to find the time of beginning and end of the green signal for each traffic flow respecting the collision times between flows and minimal given technical time for the green signal. We can simplify the model by introducing of phases, which gathers non-collision flows together.

As was mentioned above, there are three possibilities, how to solve this problem. In the first approach the "red point" exists. It is the moment when all flows in the intersection have red signal. In the second approach can the green signal of last phase's flows finish in the next interval. The last approach doesn't use phases and the green signal of all flows can finish in the subsequent interval.

The total waiting time of all vehicles in vehicle-seconds can be the quality criterion of optimization in this problem. This criterion has the disadvantage that it leads to a nonlinear model and this model needs to be rearranged to linear form [2] [3].

In this paper we focus on the impact of mentioned approaches on computational time and quality of the solution. We study how the restrictions of phases can influence the overall waiting time of vehicles in the intersection and the solution time. As the solving method we use linearization of the quadratic objective function based on piecewise linear function. We verify the feasibility of the proposed modifications of the model on 13 intersections and compare them. For the computational study and comparison of the approaches we use optimization tool XPRESS-IVE [6].

[1] University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Network, Univerzitná 1, 01026 Žilina, Slovakia, Michal.Kohani@fri.uniza.sk
[2] VSB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, 70833 Ostrava-Poruba, Czech Republic, michal.rusek@atlas.cz

2 Mathematical Formulation of the Road Intersection Signal Plan

Let the traffic flows entering the intersection constitute a set I. Each traffic flow i from the set I is specified by the intensity f_i. During the red signal vehicles in the flow create the queue and after the beginning of the green signal vehicles leave the intersection with saturated intensity f_i^s. Technical standards set value τ_i for each type of traffic flow and this value is the minimum time for green signal of the flow i. [1]

For every two collision flows i and j we assume the minimum time of delay between the end of the green signal of flow i and the beginning of the green signal of flow j as m_{ij}. The period of intersection signal plan will be considered as the interval $<0, t_{max}>$.

The total waiting time of the vehicles in the flow i during the period $<0, t_{max}>$ where the time t_i^r is the duration of the red signal correspond to the gray area in Figure 1.

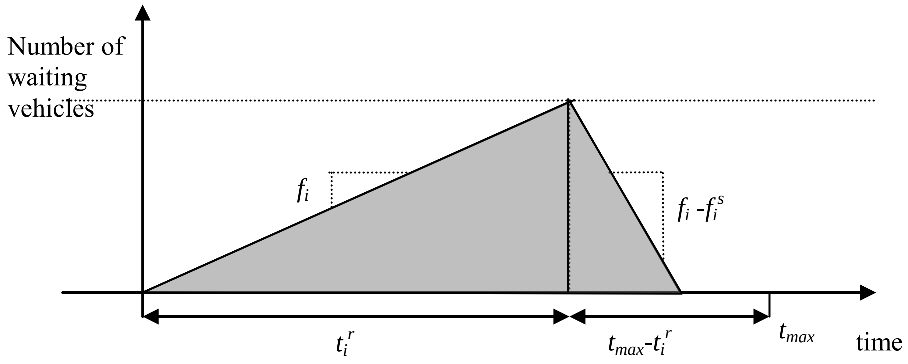


Figure 1 Waiting time of vehicles in flow i: $0.5*(t_i^r)^2 * f_i * f_i^s /(f_i^s - f_i)$

We introduce variables u_i to model the length of the red signal of the flow i and variables x_i and y_i to model the start and end of the green signal of the flow i. We also introduce phases of the set $K = \{F_1, F_2, ..., F_r\}$, where F_k is the set of non-collision traffic flows, ie. traffic flows which vehicles can pass through the intersection simultaneously [1]. This situation represents the first approach where the start and finish time of all activities must be in the same interval. The mathematical model can be written in the form *(RedPoint)*[4][5]:

$$Minimize \sum_{i \in I} 0.5 * \left(\frac{f_i * f_i^s}{f_i^s - f_i} \right) * (u_i)^2 \tag{1}$$

$$Subject\ to \qquad t_{\max} - y_i + x_i = u_i \quad for \ \ i \in I \tag{2}$$

$$y_i - x_i \ge \left(\frac{f_i t_{\max}}{f_i^s} + 1 \right) \quad for \ \ i \in I \tag{3}$$

$$y_i - x_i \ge \tau_i \quad for \ \ i \in I \tag{4}$$

$$x_j - y_i \ge m_{ij} \quad for\ k = 1,\dots, r-1, s = k+1.r,\ i \in F_k,\ j \in F_s,\ i, j - collision \tag{5}$$

$$x_j - y_i \ge m_{ij} - t_{\max} \quad for\ k = s+1,..,r,\ s = 1,..,r-1,\ i \in F_k,\ j \in F_s,\ i, j - collision \tag{6}$$

$$x_i \le y_i \quad for\ i \in I \tag{7}$$

$$x_i \le t_{\max} \quad for\ i \in I \tag{8}$$

$$y_i \le t_{\max} \quad for\ i \in I \tag{9}$$

$$x_i \in Z^+ \quad for\ i \in I \tag{10}$$

$$y_i \in Z^+ \quad for\ i \in I \tag{11}$$

$$u_i \geq 0 \quad for\ i \in I \tag{12}$$

Quadratic objective function *(1)* models the total waiting time of vehicles crossing the intersection during the period. Conditions (2) are the coupling conditions between the beginning and end of the green signal and duration of the red signal of the flow *i*. Conditions (3) ensure that the flow *i* will have a sufficiently long period of green signal that all vehicles in the flow could leave the intersection. Conditions (4) ensure that duration of the green signal will be at least as long as the prescribed standard value τ_i. Conditions (5) ensure adequate minimum delay for flows to avoid in collision. The conditions (6) solve this situation for all flows in the last phase.

In the second approach some given activities are allowed to start in one interval and finish in the next interval. In the case of road intersection signal plan green signal of flows from the last phase can end in the next interval. To describe this situation we introduce binary variables z_i, which obtain value 1 when $x_i > y_i$ and value 0 otherwise. The model after rearrangement can be written in the form *(LastPhase)*[5]:

$$Minimize\ (1)$$

$$Subject\ to \quad (8)-(12)$$

$$t_{\max}(1-z_i) - y_i + x_i = u_i \quad for\ \ i \in I \tag{13}$$

$$y_i + t_{\max} z_i - x_i \geq \left(\frac{f_i t_{\max}}{f_i^s} + 1 \right) \quad for\ \ i \in I \tag{14}$$

$$y_i + t_{\max} z_i - x_i \geq \tau_i \quad for\ \ i \in I \tag{15}$$

$$y_i + (t_{\max} + 1) z_i - x_i \leq t_{\max} \quad for\ i \in I \tag{16}$$

$$x_j - y_i \geq m_{ij} - t_{\max} \quad for\ k = s+1,..,r-1,\ s = 1,..,r-1,\ i \in F_k,\ j \in F_s,\ i,j-collision \tag{17}$$

$$x_j - y_i \geq m_{ij} - t_{\max} z_i \quad for\ s = 1.r-1,\ i \in F_r,\ j \in F_s,\ i,j-collision \tag{18}$$

$$x_j + t_{\max} - y_i \geq m_{ij} - t_{\max}(1-z_i) \quad for\ s = 1.r-1,\ i \in F_r,\ j \in F_s,\ i,j-collision \tag{19}$$

$$z_i \in \{0,1\} \quad for\ i \in I \tag{20}$$

Constraints (13)-(19) replace constraints (2)-(6) from *RedPoint* model and respect the situation when flows from last phase can end in the next interval.

The last approach doesn’t use phases and the green signal of all flows can finish in the subsequent interval. In addition to variables z_{ij} we introduce binary variables w_{ij} for all pairs of collision flows where this variable obtains value 1 when $y_i > x_j$ and value 0 otherwise. The model after rearrangement can be written in the form *(NoPhase)*:

$$Minimize\ (1)$$

$$Subject\ to \quad (8)-(12), (13)-(16), (20)$$

$$x_j + t_{\max} w_{ij} - y_i \geq m_{ij} \quad for\, i \in I,\ j \in I, i, j-collision \tag{21}$$

$$x_j + (t_{\max} + 1) w_{ij} - y_i \leq t_{\max} \quad for\, i \in I,\ j \in I, i, j-collision \tag{22}$$

$$z_i + z_j + w_{ij} + w_{ji} \leq 1 \quad for\, i \in I,\ j \in I, i, j-collision \tag{23}$$

$$w_{ij} \in \{0,1\} \quad for\ i \in I, j \in I, i, j - collision \tag{24}$$

3 Solving method

Due to the nonlinearity of the objective function in all models we will use another optimizing criterion. We will use the linearization, which can be used to replace objective function (1) by piecewise linear function without loss of preciseness [3].

As was described in [3], to replace the non-linear item $(u_i)^2$ by a piecewise linear function in <0, u_i^{max}>, where u_i^{max} is integer, we introduce a set of auxiliary variables v_{ij}, where $0 \le v_{ij} \le 1$ for j=1, …, u_i^{max}. Then, the relation between variables u_i and v_{ij} can be expressed by the equation (25).

$$u_i = \sum_{j=1}^{u_i^{\max}} v_{ij} \tag{25}$$

The non-linear item $(u_i)^2$ can be replaced by right-hand-side of the equation (26).

$$(u_i)^2 = \sum_{j=1}^{u_i^{\max}} (2 * j - 1) v_{ij} \tag{26}$$

This way, the models become linear and linear programming solvers can solve the associated problems.

4 Numerical experiments

The verification of all approaches to road intersection signal plan design was made on series of 13 intersections in the city of Ostrava.

We used the general optimization software environment XPRESS-IVE [6] to perform the experiments. This software includes the branch-and-cut method and it also enables to solve large linear programming problems. The experiments were performed on a personal computer equipped with Intel Core 2 Duo E6850 with parameters 3 GHz and 3,5 GB RAM. In the Table 1 and Table 2 there are results for all approaches mentioned in this study.

Results in the Table 1 represent the total waiting time of all vehicles in intersection in vehicle-seconds. Results in the Table 2 represent the computing time in seconds.

In the column *"Linearised Model - Red Point"* there are results for model in which we use phases and the "red point" policy. In the column named "*Linearised Model - Last Phase*" are results of the model in which we divide trafic flows into phases and the green signal of flows from the last phase can finish in the next interval. In the column "*Linearised Model - No Phase*" there are the results for the model, in which we don't use division into phases. In the column *"No. of Traffic Flows"* the size of the problem is expressed by the number of flows in the intersection.

Intersection Name	No. of Traffic Flows	Linearised Model - Red Point	Linearised Model - Last Phase	Linearised Model - NoPhase
		Waiting Time [v-s]	Waiting Time [v-s]	Waiting Time [v-s]
Marianskohorska x Jirska	17	648	518	492
Marianskohorska x Nadrazni	26	1335	1174	1026
Ceskobratrska x Nadrazni	14	569	569	569
Bohuminska x 28.rijna	8	1629	770	756
Bohuminska x Tesinska	5	347	259	255
Rudna x Lidicka	13	964	895	866
Rudna x Mistecka	11	205	195	195
Opavska x 17. Listopadu	20	1621	1284	1251
Opavska x Nabrezi SPB 1	15	2846	1223	1197
Opavska x Nabrezi SPB 2	15	3641	1484	1466
Opavska x Nabrezi SPB 3	15	4939	2213	2195
Opavska x Porubska	19	765	762	731
Opavska x Sjizdna	21	582	559	522

Table 1 Results of numerical experiments

Intersection Name	No. of Traffic Flows	Linearised Model - Red Point	Linearised Model - Last Phase	Linearised Model - NoPhase
		Computing Time [s]	Computing Time [s]	Computing Time [s]
Marianskohorska x Jirska	17	0,23	0,27	0,7
Marianskohorska x Nadrazni	26	0,28	0,42	1,05
Ceskobratrska x Nadrazni	14	0,25	0,05	0,44
Bohuminska x 28.rijna	8	0,25	0,33	0,33
Bohuminska x Tesinska	5	0,12	0,72	0,94
Rudna x Lidicka	13	0,25	0,28	0,25
Rudna x Mistecka	11	0,17	0,64	0,97
Opavska x 17. listopadu	20	0,28	0,33	30,72
Opavska x Nabrezi SPB 1	15	0,27	0,36	6,45
Opavska x Nabrezi SPB 2	15	0,28	0,38	5,41
Opavska x Nabrezi SPB 3	15	0,27	0,48	12,59
Opavska x Porubska	19	0,25	0,25	4,31
Opavska x Sjizdna	21	0,25	0,33	1,41

Table 2 Computational Time

5 Conclusion

In this paper we presented the impact of three approaches on computational time and quality of the solution of cyclic processes problem. When comparing results in terms of total waiting time, the usage of "*No Phase*" model gives better solutions than "*Last Phase*" model and "*Red Point*" model. In terms of calculation time, solving time of “*Last Phase*” model and “*Red Point*” model was lower than one second in all cases. The computational time of “*No Phase”* model was maximally several of seconds, tens of seconds in some cases.

According to results, in the need of planning the activities in the cyclical processes is better, when all activities are not limited by the time of start and end in the same interval and the end of activity can be in the subsequent interval. The existence of red point seems to be inefficient according to criterion of waiting time of cyclical processes.

Acknowledgements

This article was supported by the grant project of the Faculty of Mechanical Engineering Technical University of Ostrava, No. SP2011/129 “Research on modeling to support management of urban traffic” and by the grant VEGA 1/0361/10 „Optimal design of public service systems”

References

[1] Černý, J. & Kluvánek P.: *Základy matematickej teórie dopravy*, VEDA: Bratislava, 1991. (in Slovak).
[2] Janáček, J., Cenek, P., Jánošíková, L., Buzna, L., Bednár, P.: The design of optimal transportation system structure, Communications–Scientific Letters of the University of Žilina, vol. 5., 2003, pp. 70-72
[3] Janáček, J., Koháni, M.: *Waiting Times in Public Transport and Their Modelling for IP-Solver*. Communications–Scientific Letters of the University of Žilina, vol. 3., 2010
[4] Koháni, M., Rusek, M.: *Využití kaskádového přístupu pro návrh signálního plánu křižovatek,* In: Úlohy diskrétní optimalizace v dopravní praxi : kvantitativní metody optimalizace v dopravních a logistických systémech I - sborník příspěvků, Pardubice 29.-30.4. 2010, Univerzita Pardubice, 2010, ISBN 978-80-7395-297-6, pp. 131-143.
[5] Koháni, M., Rusek, M.: *The impact of "red point" for vehicle waiting times in intersections*, In: Journal of Information, Control and Management Systems., ISSN 1336-1716, Vol. 8, No. 3 (2010), pp. 199-206.
[6] XPRESS-Mosel “*User guide*”. Dash Associates, Blisworth, 2005, UK, p. 99.

Comparison of various approaches to portfolio efficiency

Miloš Kopa[1]

Abstract. This paper deals with portfolio efficiency testing with respect to various criteria. Basically, two approaches can be employed. If expected utility approach is considered one can test portfolio efficiency with respect to stochastic dominance relation. We focus on the second-order stochastic dominance portfolio efficiency that allows for risk averse decision makers. Alternatively, portfolio efficiency with respect to mean-risk criteria is analyzed, when considering the most favorite risk measures (variance, semivariance, value at risk, conditional value at risk). We assume discrete distribution of monthly returns. As the basic assets, we consider ten representative US industry portfolios and a riskfree asset. For all these efficiency approaches we test more than forty thousand portfolios from a regular grid and we identify sets of efficient portfolios. We compare these sets and corresponding efficient frontiers between each other in classical mean-variance space.

Keywords: mean-risk models, risk measures, second-order stochastic dominance, portfolio efficiency.

JEL classification: D81, G11
AMS classification: 91B16, 91B30

1 Introduction

The classical portfolio efficiency analysis is based on well-known mean-variance criteria introduced already in 1952 by Harry Markowitz (see [10]). Since that time, many improvements have been proposed and implemented. The new risk measures, such as semivariance [11], Value at Risk [4] (VaR) or Conditional Value at Risk [13] (CVaR) where proposed and analysed. These alternative measures model the risk of investments in a more sophisticated way, focusing more on the investments losses.

Stochastic dominance (SD) is another appealing approach to analyzing investments and portfolio choice problems. Stochastic dominance relations offer an approach that effectively considers the entire return distribution rather than a finite set of moments. Assuming risk averse investors, we limit our attention to second-order stochastic dominance relation. We say that portfolio $\boldsymbol{\lambda}$ dominates portfolio $\boldsymbol{\tau}$ with respect to second-order stochastic dominance if expected utility of $\boldsymbol{\lambda}$ is not lower than expected utility of $\boldsymbol{\tau}$ for all concave utility functions. Put differently, if portfolio $\boldsymbol{\lambda}$ dominates portfolio $\boldsymbol{\tau}$ with respect to second-order stochastic dominance then no risk averse investor prefers $\boldsymbol{\tau}$ to $\boldsymbol{\lambda}$. The more general notion of first-order portfolio efficiency was discussed in [8] and [7].

In the last decade, several portfolio efficiency tests with respect to second-order stochastic dominance were developed. First test, based on the representative set of utility functions, was developed in 2003 [12]. A year later, another test using majorization theorem for dual stochastic dominance approach were introduced, see [8]. Finally, [6] presents a test formulated in terms of CVaRs.

The aim of this paper is to analyze the efficiency of portfolios with respect to mean-risk criteria and empirically compare it with SSD portfolio efficiency. We consider 2001 - 2010 period of monthly returns of ten representative US industry portfolios and a riskfree asset. For this data set, we construct more than 40 000 portfolios from a regular grid. For each portfolio, we test its efficiency with respect to mean-risk criteria and with respect to SSD criterion. Finally, we construct and compare the empirical efficient frontiers and sets of efficient portfolios.

[1]Institute of Information Theory and Automation of the ASCR, Department of Econometrics, Pod Vodárenskou věži 4, 182 08 Prague, Czech Republic, e-mail: kopa@karlin.mff.cuni.cz

The remainder of this paper is structured as follows. Section 2 introduces basic definitions of second-order stochastic dominance relation and portfolio efficiency with respect to this criterion. Section 3 presents mean-risk efficiency algorithms (tests) in terms of mathematical programming, when variance, semivariance, VaR and CVaR are employed. It is followed by an empirical study where we compare the corresponding efficient sets and frontiers. Section 5 summarizes and concludes the paper.

2 Portfolio efficiency with respect to second-order stochastic dominance relation

We consider a random vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ of returns of N assets with a discrete probability distribution described by T equiprobable scenarios. 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 representing the assets returns along t-th scenario. We assume that the decision maker may also combine the alternatives into a portfolio. We will use $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_N)^{\mathrm{T}}$ for a vector of portfolio weights and $X\boldsymbol{\lambda}$ represents returns of portfolio $\boldsymbol{\lambda}$. The portfolio possibilities are given by a simplex

$$\Lambda = \{\boldsymbol{\lambda} \in R^N | \mathbf{1}'\boldsymbol{\lambda} = 1, \quad \lambda_j \geq 0, \quad j = 1, 2, \ldots, N\},$$

which arises as the relevant case if we exclude short sales and impose a budget restriction. Moreover, the tested portfolio is denoted by $\boldsymbol{\tau}$.

Following [9] and references therein, portfolio $\boldsymbol{\lambda}$ *dominates* portfolio $\boldsymbol{\tau}$ *with respect to second-order stochastic dominance* ($\boldsymbol{\lambda} \succ_{SSD} \boldsymbol{\tau}$) if $Eu(\mathbf{r}\boldsymbol{\lambda}) \geq Eu(\mathbf{r}\boldsymbol{\tau})$ for all non-decreasing and concave utility functions with strict inequality for at least one such utility function. Alternatively, one can consider as a definition of this relation some of its necessary and sufficient conditions summarized in, for example, [5]. In any case, if portfolio $\boldsymbol{\lambda}$ dominates portfolio $\boldsymbol{\tau}$ with respect to second-order stochastic dominance then every risk averse decision maker prefers $\boldsymbol{\lambda}$ to $\boldsymbol{\tau}$ or is indifferent between them.

Following [12], [8] and [6], we define the efficiency of a given portfolio with respect to second-order stochastic dominance relative to all portfolios that can be created from a considered set of assets.

Definition 1. A portfolio $\boldsymbol{\tau}$ is *SSD inefficient* if there exists portfolio $\boldsymbol{\lambda} \in \Lambda$ such that $\boldsymbol{\lambda}$ dominates $\boldsymbol{\tau}$ by SSD. Otherwise, the portfolio $\boldsymbol{\tau}$ is *SSD efficient.*

Since its relation to CVaR (one of the considered risk measure), we choose the portfolio efficiency test developed in [6]. Let $\alpha_k = k/T, \quad k \in K = \{0, 1, \ldots, T-1\}$. Consider the following linear program:

$$\begin{aligned} D^*(\boldsymbol{\tau}) = \max_{D_k, \lambda_n, b_k, w_k^t} & \sum_{k=1}^{T} D_k \\ \text{s.t. } \mathrm{CVaR}_{\frac{k-1}{T}}(-\mathbf{r}'\boldsymbol{\tau}) - b_k - \frac{1}{(1-\frac{k-1}{T})T} \sum_{t=1}^{T} w_k^t &\geq D_k, \qquad k \in K \\ w_k^t + \mathbf{x}^t\boldsymbol{\lambda} &\geq -b_k, \quad t, k \in K \\ w_k^t &\geq 0, \qquad t, k \in K \\ D_k &\geq 0, \qquad k \in K \\ \boldsymbol{\lambda} &\in \Lambda. \end{aligned}$$

The optimal objective value $D^*(\boldsymbol{\tau})$ can be seen as a measure of SSD inefficiency of portfolio $\boldsymbol{\tau}$ and [6] uses it for the portfolio efficiency testing as follows.

Theorem 1. *If* $D^*(\boldsymbol{\tau}) > 0$ *then* $\boldsymbol{\tau}$ *is SSD inefficient and* $\boldsymbol{\lambda}^* \succ_{SSD} \boldsymbol{\tau}$. *Otherwise,* $D^*(\boldsymbol{\tau}) = 0$ *and* $\boldsymbol{\tau}$ *is SSD efficient.*

3 Portfolio efficiency with respect to mean-risk criteria

The classical optimization task which leads to mean-risk efficient portfolios can be written as:

$$\begin{aligned} &\min_{\boldsymbol{\lambda}} \; risk_{\boldsymbol{\lambda}} \\ &\text{s. t. } mean_{\boldsymbol{\lambda}} \geq mean_e \\ &\qquad \boldsymbol{\lambda} \in \Lambda, \end{aligned} \tag{1}$$

where $risk_{\boldsymbol{\lambda}}$ and $mean_{\boldsymbol{\lambda}}$ represent risk and mean return of portfolio $\boldsymbol{\lambda}$, respectively, $mean_e$ is the minimal required expected return.

Definition 2. A portfolio $\boldsymbol{\tau}$ is *mean-risk inefficient* if there exists portfolio $\boldsymbol{\lambda}$ satisfying $risk_{\boldsymbol{\lambda}} \leq risk_{\boldsymbol{\tau}}$ and $mean_{\boldsymbol{\lambda}} \geq mean_{\boldsymbol{\tau}}$ with at least one strict inequality. Otherwise, portfolio $\boldsymbol{\tau}$ is *mean-risk efficient.*

Since we want to test the efficiency of portfolio $\boldsymbol{\tau}$ with respect to mean-risk criterion we reformulate (1) in the following way:

$$\begin{aligned} \xi(\boldsymbol{\tau}) = &\min_{\boldsymbol{\lambda}, s_m, s_r} \; s_m + s_r \\ &\text{s. t. } mean_{\boldsymbol{\lambda}} - mean_{\boldsymbol{\tau}} \geq s_m \\ &\qquad risk_{\boldsymbol{\tau}} - risk_{\boldsymbol{\lambda}} \geq s_r \\ &\qquad \boldsymbol{\lambda} \in \Lambda \\ &\qquad s_m, s_r \geq 0, \end{aligned} \tag{2}$$

where $\xi(\boldsymbol{\tau})$ can be understood as a measure of mean-risk inefficiency that is composed from possible improvements in both mean and risk. These improvements are represented by slack variables s_m and s_r.

Theorem 2. *If $\xi(\boldsymbol{\tau}) > 0$ then $\boldsymbol{\tau}$ is mean-risk inefficient. Otherwise, $\xi(\boldsymbol{\tau}) = 0$ and $\boldsymbol{\tau}$ is mean-risk efficient.*

In all mean-risk efficiency tests, we will use $mean_{\boldsymbol{\lambda}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^t\boldsymbol{\lambda}$ and $mean_{\boldsymbol{\tau}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}^t\boldsymbol{\tau}$. To get the test for mean-variance efficiency, we apply $r_{\boldsymbol{\lambda}} = \boldsymbol{\lambda}' V \boldsymbol{\lambda}$ and $r_{\boldsymbol{\tau}} = \boldsymbol{\tau}' V \boldsymbol{\tau}$ to (2) where V is the covariance matrix of asset returns. In the semivariance case, we use variables z^t corresponding to returns of portfolio $\boldsymbol{\tau}$ that are smaller than $mean_{\boldsymbol{\tau}}$. The general test (2) is modified as follows:

$$\begin{aligned} \xi(\boldsymbol{\tau}) = &\min_{\boldsymbol{\lambda}, s_m, s_r, z^t} \; s_m + s_r \\ &\text{s. t. } mean_{\boldsymbol{\lambda}} - mean_{\boldsymbol{\tau}} \geq s_m \\ &\qquad semivariance_{\boldsymbol{\tau}} - \frac{1}{T}\sum_{t=1}^{T} \left(z^t\right)^2 \geq s_r \\ &\qquad z^t \geq -\mathbf{x}^t\boldsymbol{\lambda} + mean_{\boldsymbol{\lambda}}, \; t = 1,..,T \\ &\qquad z^t \geq 0, \; t = 1,..,T \\ &\qquad \boldsymbol{\lambda} \in \Lambda \\ &\qquad s_m, s_r \geq 0. \end{aligned}$$

In the case of VaR we have to employ integer variables δ^t, what leads to more computationally demanding problem:

$$\begin{aligned} \xi(\boldsymbol{\tau}) = &\min_{\nu, \boldsymbol{\lambda}, \delta^t, s_m, s_r} \; s_m + s_r \\ &\text{s. t. } mean_{\boldsymbol{\lambda}} - mean_{\boldsymbol{\tau}} \geq s_m \\ &\qquad \mathrm{VaR}_{\boldsymbol{\tau}} - \nu \geq s_r \\ &\qquad -\mathbf{x}^t\boldsymbol{\lambda} \leq \nu + K\delta^t, \; t = 1,..,T \\ &\qquad \sum_{t=1}^{T} \delta^t = \lfloor (1-\alpha)\, T \rfloor \\ &\qquad \delta^t \in \{0, 1\}, \; t = 1,..,T \\ &\qquad \boldsymbol{\lambda} \in \Lambda \\ &\qquad s_m, s_r \geq 0, \end{aligned}$$

where $\lfloor x \rfloor = \max\{n \in \mathbb{N}_0, n < x\}$ for $x \in \mathbb{R}^+$, and K is sufficiently large constant, for example, $K \geq \max_{t,j} x_j^t - \min_{t,j} x_j^t$. When CVaR is chosen as the risk measure, we get the following linear program:

$$\begin{aligned}
\xi(\boldsymbol{\tau}) = \min_{\boldsymbol{\lambda}, s_m, s_r, z^t, a} \quad & s_m + s_r \\
\text{s. t.} \quad & mean_{\boldsymbol{\lambda}} - mean_{\boldsymbol{\tau}} \geq s_m \\
& \mathrm{CVaR}_{\boldsymbol{\tau}} - a - \frac{1}{(1-\alpha)T} \sum_{t=1}^{T} z^t \geq s_r \\
& z^t \geq -\mathbf{x}^t \boldsymbol{\lambda} - a, \; t = 1, .., T \\
& z^t \geq 0, \; t = 1, .., T \\
& \boldsymbol{\lambda} \in \Lambda \\
& s_m, s_r \geq 0.
\end{aligned}$$

4 Empirical application

We consider monthly returns of ten US representative industry portfolios and a risk free asset which represent $N = 11$ basic assets. The returns can be found in data library of Kenneth French [14] and we proxy risk free asset by CRSP index. We consider ten years period 2001 - 2010, that is $T = 120$ historical scenarios. The Table 1 shows descriptive statistics of the data set. We consider a regular grid

	mean	st. dev.	min	max	skewness
Non-durables	0.622	3.583	-12.980	9.310	-0.622
Durables	0.689	8.515	-32.790	43.090	0.462
Manufactory	0.778	5.525	-20.840	17.950	-0.664
Energy	1.038	5.991	-17.020	19.160	-0.320
HiTech	0.310	7.879	-26.230	19.400	-0.385
Telecom	0.116	5.956	-15.460	21.380	0.027
Shops	0.531	4.773	-15.060	12.310	-0.390
Health	0.097	3.968	-10.930	9.220	-0.391
Utilities	0.519	4.487	-12.480	10.200	-0.924
Other	0.157	5.594	-19.560	16.280	-0.619
Riskfree	0.180	0	0.180	0.180	0

Table 1: The basic descriptive statistics.

on feasibility set Λ with step size $\frac{1}{8}$ and we create 43 758 portfolios from these assets. Specifically, each element of each created portfolio $\boldsymbol{\lambda}$ is equal to one of the following numbers: $0, \frac{1}{8}, \frac{2}{8}, ..., 1$ such that the elements of each portfolio sum up to 1, in order to satisfy the conditions of set Λ.

Firstly, applying Theorem 1, we test SSD efficiency of all considered portfolios. We find only 0,11% SSD efficient portfolios and Figure 1 (left) presents the set of SSD efficient portfolios in the classical mean-variance space.

Secondly, we identify a set of mean-variance, mean-semivariance, mean-VaR and mean-CVaR efficient portfolios using tests in Section 3. Perhaps surprisingly, the set of mean-semivariance efficient portfolios coincides with that for mean-CVaR criteria. For all risk measures we construct efficient frontiers: mean-variance frontier (solid line), mean-semivariance and mean-CVaR frontier (dashed line), mean-VaR frontier (dashdotted line) and we compare the results in Figure 1 (right). These frontiers are only empirical ones, that is, they are constructed only from portfolios on the grid. For large values of mean (larger than 0.88%), the mean-variance frontier coincides with mean-semivariance and mean-CVaR ones.

Finally, we compare the efficiency sets among each other. We find in our study that mean-semivariance efficiency is equivalent to mean-CVaR efficiency. Moreover, each of these efficient portfolios is also efficient with respect to mean-variance criteria. Finally, all mean-risk efficient portfolios, except of mean-VaR efficient ones, are classified as SSD efficient as well. Figure 2 summarizes this efficiency sets comparison.

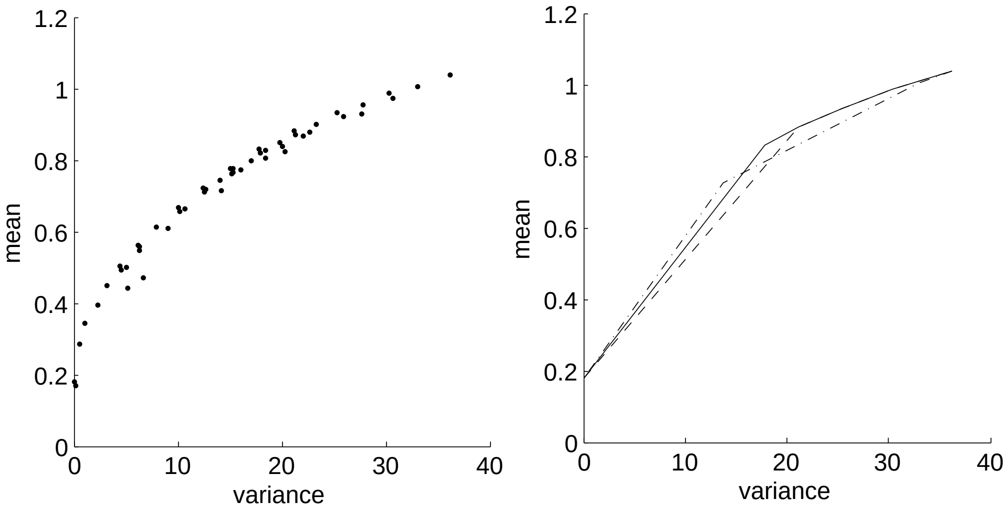


Figure 1: Empirical set of SSD efficient portfolios (left) and mean-risk efficiency frontiers (right).

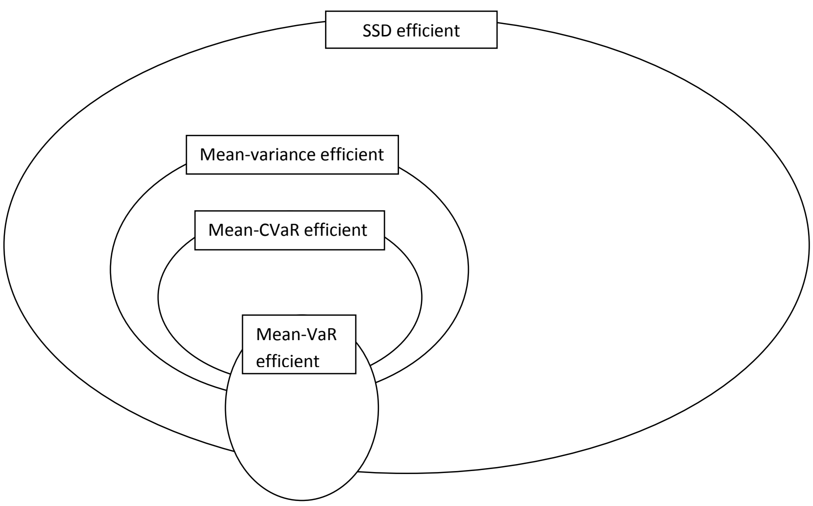


Figure 2: Efficient portfolio sets, mean-CVaR efficient set coincide with mean-semivariance one.

5 Conclusion

This paper compares the several portfolio efficiency sets when using stochastic dominance and mean-risk criteria. We constructed and compared these sets and corresponding efficient frontiers. We found that SSD efficiency set is larger than any mean-risk efficiency set. This is not a surprising result because SSD efficiency test can be seen as T-criteria problem while any mean-risk efficiency test is only two-criteria problem.

For future research, this study can be improved in various ways. For example, longer historical data can be used. In addition, one can consider the portfolio efficiency in a more robust way as it was done in [5] and [2] using, for example, contamination techniques discussed e.g. in [1]. Alternatively, one can apply fuzzy approach recently used in [3]. Unfortunately, all these improvements would lead to more computationally demanding efficiency tests what requires much better hardware equipment than is currently available.

Acknowledgements

The research was partly supported by the Czech Science Foundation (grant P402/10/1610).

References

[1] Branda, M. and Dupačová, J.: Approximation and contamination bounds for probabilistic programs, *Annals of Operations Research*, published on-line November 11, 2010.

[2] Dupačová, J. and Kopa, M.: Robustness in stochastic programs with risk constraints, *Annals of Operations Research*, published on-line January 4, 2011.

[3] Holčapek, M. and Tichý, T.: A probability density function estimation using F-transform, *Kybernetika*, **46** (3), (2010), 447–458.

[4] J.P. Morgan Risk Metrics: *Technical Document, 4th Edition*, Morgan Guaranty Trust Company, New York, 1995.

[5] Kopa, M.: Measuring of second-order stochastic dominance portfolio efficiency, *Kybernetika*, **46** (3), (2010), 488–500.

[6] Kopa, M. and Chovanec, P.: A Second-Order Stochastic Dominance Portfolio Efficiency Measure, *Kybernetika*, **44** (2), (2008), 243–258.

[7] Kopa, M. and Post, T.: A portfolio optimality test based on the first-order stochastic dominance criterion. *Journal of Financial and Quantitative Analysis*, **44** (5), (2009), 1103–1124.

[8] Kuosmanen, T.: Efficient diversification according to stochastic dominance criteria, *Management Science*, **50** (10), (2004), 1390–1406.

[9] Levy, H.: *Stochastic Dominance: Investment Decision Making Under Uncertainty.* Second edition. Springer Science, New York 2006.

[10] Markowitz, H. M.: Portfolio Selection, *The Journal of Finance*, **7** (1), (1952), 77–91.

[11] Markowitz, H. M.: *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York 1959.

[12] Post, T.: Empirical tests for stochastic dominance efficiency, *Journal of Finance*, **58**, (2003), 1905–1932.

[13] Rockafellar, R. T. and Uryasev, S.: Conditional Value-at-Risk for General Loss Distributions, *Journal of Banking and Finance*, **26** (7), (2002), 1443–1471.

[14] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

The time coordination of bus links in both transport directions

Petr Kozel[1]

Abstract. The time coordination is an important tool for solving problems of optimization of time position of bus links of the public mass transport. We can focus on many different problems. One of them is a coordination of links. All links depart from a common bus stop. The problem can be divided into two variants. First of them is a time coordination of one transport direction only. The second problem is a time coordination in both transport directions. This paper is focused on the second problem. Let us consider the section between places A and B. There is bus stop Z. This bus stop is operated by a set of links. The aim of the research is following: Shift arrival time of vehicle at the stop, so that the overall waiting time in the both directions in passenger-minutes is minimal for the passengers. Some of the links operate the bus stop in both directions: in the A – B direction and in the B – A direction. It is necessary to ensure that the departure of these links in one direction will be linked with the departure of links in the opposite direction. The time shift of all of the links will be in allowable interval. This is the basic condition, too.

Keywords: time coordination, cascade approach, max-min criterion.

1 Introduction

The bus link coordination is a complicated matter both in the terms of its mathematical formulation and the terms of methods of solution. The time coordination of bus link can be solved in several modifications. We can solve:

1. A one-way time coordination. The time coordination in one transport direction only. This problem has been solved in the paper [7].
2. A time coordination in both transport directions. This paper deals with this problem.

In both cases we can choose several optimization criterions. First of them is maximize the minimum difference between neighboring bus links. Original approaches were designed in [1]. The problem was solved in the paper [6], [7]. This approach was improved by using the cascade approach in the paper [8]. Another criterion is the total number of waiting time passenger-minutes relative to the unit of time. Problem about using this criterion is that it leads to a non-linear model. This problem was solved in the paper [2].

This paper focuses on problem number 2, the time coordination of bus links in both transport directions. The time coordination of bus links will be solved by using max-min criterion in the first part of this paper. This approach will be improved by using cascade approach in the second part of this paper. The mathematical models of both cases will be presented in this paper, too. These approaches are tested on real problems in the chapter 4. The problem was solved in the part of the traffic net in the town Frýdek-Místek. More specifically, this problem was solved in the section Frýdek-Místek – Dobrá and in the opposite section Dobrá – Frýdek-Místek.

2 Mathematical model of the time coordination bus links in the both transport directions by using max-min criterion

Let us consider the section between the places *A* (Frýdek-Místek) and *B* (Dobrá). There is a bus stop *Z*. This bus stop is operated by a set of links *F* in the transport direction Frýdek-Místek – Dobrá and a set of links *D* in the transport direction Dobrá – Frýdek-Místek. Let ft_i be arrival time of vehicle *i* at the bus stop *Z* in the transport direction F-M – Dobrá and let dt_i be arrival time of vehicle *i* at the bus stop *Z* in the transport direction Dobrá – F -M. The earliest possible arrival time of the vehicle *i* is identified as $fa_i; da_i$, and this time may be postponed until the time $fa_i + fc_i$ or $da_i + dc_i$, where fc_i, dc_i is the maximum possible shift of arrival at the bus stop. It is necessary to find such time positions of the individual arrivals so that the total passengers waiting time is minimal. The arrival times ft_0, dt_0 and ft_n, dt_n are fixed.

[1]Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17.listopadu 15, 70833 Ostrava-Poruba, Czech Republic, kozelp@seznam.cz.

Some of the links operate the bus stop in both directions, in the $A-B$ direction and in the $B-A$ direction. For these links, it is necessary to ensure that the departure of links in one direction will be linked with the departure of links in the opposite direction. The time shift of all of the links will be in allowable interval. This is the basic condition, too.

Mathematical model for the time coordination in both transport direction with using max-min criterion has the form:

$$Maximize \quad y+w \tag{1}$$

$$Subject \quad to$$

$$fx_1 + fa_1 - ft_0 \geq y \tag{2}$$

$$fx_i + fa_i - fx_{i-1} - fa_{i-1} \geq y \quad for \ i = 2,..,F-1 \tag{3}$$

$$ft_n - fx_{n-1} - fa_{n-1} \geq y \tag{4}$$

$$dx_1 + da_1 - dt_0 \geq w \tag{5}$$

$$dx_i + da_i - dx_{i-1} - da_{i-1} \geq w \quad for \ i = 2,..,D-1 \tag{6}$$

$$dt_n - dx_{n-1} - da_{n-1} \geq w \tag{7}$$

$$fx_i \leq fc_i \quad for \ i = 1,..,F-1 \tag{8}$$

$$dx_i \leq dc_i \quad for \ i = 1,..,D-1 \tag{9}$$

$$fx_i \geq 0 \quad for \ i = 1,..,F-1 \tag{10}$$

$$dx_i \geq 0 \quad for \ i = 1,..,D-1 \tag{11}$$

$$y \geq 0 \tag{12}$$

$$w \geq 0 \tag{13}$$

Objective function (1) models the minimal difference between neighboring bus links. The variable y models minimal difference between neighboring bus links in the transport direction F-M – Dobrá and the variable w models the same in the opposite direction. The conditions (2) – (4) and conditions (5) – (7) determine the gap between two consecutive arrivals for both transport directions. Conditions (8), (9) ensure that the arrival time shift due to the coupling potential first arrival was less than the maximum permissible value of shift fc_i or dc_i. Conditions (10) – (13) are conditions for nonnegativity calculated variables of problem.

Some of the bus links operate both transport directions. For these bus links it is necessary to introduce additional conditions. These conditions ensure that some of bus links in both transport directions will be linked. For each bus link that is linked with another one, we must introduce a special condition. Creating these conditions is as following:

For each pair of linked bus links we know the earliest possible arrival time of the vehicle at the appropriate bus stop $fa_i; da_i$. As following we know travel time t_j between default and final bus stop and time for the manipulation t_m. The time for manipulation has a fixed value. This value is 10 minutes. Situation is depicted in Figure 1.

For each pair of linked bus links following must be respected:

$$fa_i + fx_i + t_j + t_m \leq da_i + dx_i \tag{14}$$

For example: let us consider a vehicle V. This vehicle operates the bus links 860 310/5 and 860 310/4. The earliest possible arrival time of the vehicle V on the default bus stop is 7:53 o´clock and the earliest possible departure time of the vehicle V from the final bus stop is 8:22 o´clock. The travel time between default and final bus stops is 17 minutes and time of manipulation is 10 minutes. For pair of these links 860 310/5 and 860 310/4 the following must be respected: $7:53 + fx_2 + 17 + 10 \leq 8:22 + dx_3$; $fx_2 \leq 2 + dx_3$.

Situation is depicted in Figure 2.

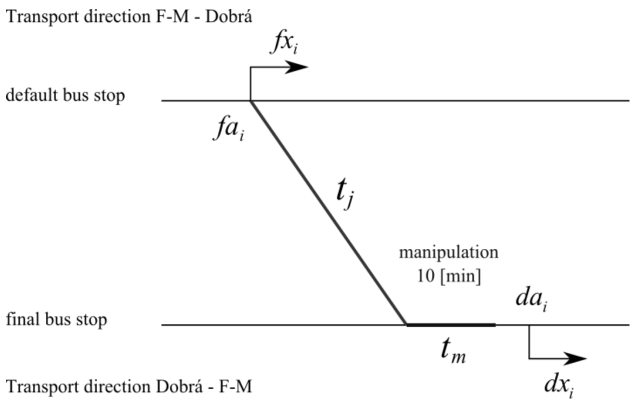


Figure 1 Construction of additional conditions

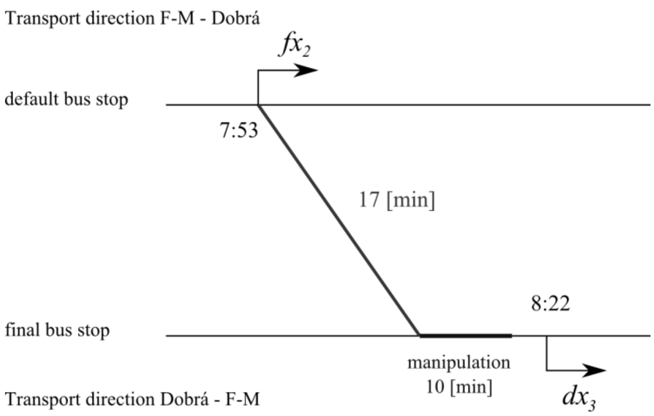


Figure 2 Construction of additional conditions – real example

The mathematical model (1) – (14) was supplemented by these conditions:

$fx_2 \leq 2 + dx_3$; $fx_1 \leq 47 + dx_5$; $fx_3 \leq dx_7 - 3$; $fx_7 \leq 68 + dx_8$; $dx_1 \leq 29 + fx_5$; $fx_4 = dx_6$

3 Mathematical model of the time coordination of bus links in both transport directions by using max-min criterion with cascade approach

The results of mathematical model with the max-min criterion can be improved. We can use the criterion max-min type with cascade approach. This approach was designed in [8]. This part of paper focuses on the use of the cascade approach to the time coordination in both transport directions. At first the mathematical model will be presented. Mathematical model of cascade approach for time coordination in both transport directions has this form:

$$Maximize\ \ 10 \cdot F \cdot y + \sum_{i}^{F} fu_i + 10 \cdot D \cdot y + \sum_{i}^{D} du_i \tag{15}$$

Subject to

$$fx_1 + fa_1 - ft_0 \geq y + fu_1 \tag{16}$$

$$fx_i + fa_i - fx_{i-1} - fa_{i-1} \geq y + fu_i \quad for\ i = 2,..,F-1 \tag{17}$$

$$ft_n - fx_{n-1} - fa_{n-1} \geq y + fu_n \tag{18}$$

$$dx_1 + da_1 - dt_0 \geq w + du_1 \tag{19}$$

$$dx_i + da_i - dx_{i-1} - da_{i-1} \geq w + du_i \quad for\ i = 2,..,D-1 \tag{20}$$

$$dt_n - dx_{n-1} - da_{n-1} \geq w + du_n \tag{21}$$

$$fx_i \leq fc_i \quad for\ i = 1,..,F-1 \tag{22}$$

$$dx_i \leq dc_i \quad for\ i = 1,..,D-1 \tag{23}$$

$$fu_i \leq \varepsilon \quad for\ i = 1,..,F \tag{24}$$

$$du_i \leq \varepsilon \quad for\ i = 1,..,D \tag{25}$$

$$fx_i \geq 0 \quad for\ i = 1,..,F-1 \tag{26}$$

$$dx_i \geq 0 \quad for\ i = 1,..,D-1 \tag{27}$$

$$y \geq 0 \tag{28}$$

$$w \geq 0 \tag{29}$$

The mathematical model has been supplemented by perturbation variables for each transport direction. These perturbation variables fu_i, du_i identify the loose coupling conditions. Furthermore a constant ε has been supplemented to the mathematical model. This constant determines accuracy of perturbation variables. Those are conditions (24) and (25). The conditions (16) – (18) and conditions (19) – (21) determine the gap between two consecutive arrivals for both transport directions. Conditions (22), (23) ensure that the arrival time shifts due to the coupling potential first arrival was less than the maximum permissible value of shift fc_i or dc_i. Conditions (26) – (29) are conditions for nonnegativity calculated variables of problem.

As in the previous case we have to add additional conditions to the mathematical model: $fx_2 \leq 2 + dx_3$; $fx_1 \leq 47 + dx_5$; $fx_3 \leq dx_7 - 3$; $fx_7 \leq 68 + dx_8$; $dx_1 \leq 29 + fx_5$; $fx_4 = dx_6$. These conditions ensure that appropriate bus links will be linked.

4 Numerical experiment

The numerical experiments using the user interface XPRESS-IVE. Numerical experiments in this paper were made for time coordination of communications within suburban public transport in the area of Frýdek-Místek – Dobrá and Dobrá – Frýdek-Místek during the morning carriage seat. There are 11 bus links in the first transport direction (Frýdek-Místek – Dobrá), two of them are outboard bus links that are fixed. There are 10 bus links in the second transport direction (Dobrá – Frýdek-Místek) and two of them are outboard bus links which are fixed, too. The earliest possible arrival at the stop fa_i, da_i and maximum possible shift fc_i, dc_i are known for each bus link $i = 1, ..., F-1$ in the first transport direction and for each bus link $i = 1, ..., D-1$ in second transport direction. These values are converted to minutes and they are shown in Table 1 and Table 2.

i	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
fa_i [min]	458	469	473	545	546	552	555	585	684	701	722
fc_i [min]	0	57	65	7	63	78	71	83	5	44	0

Table 1 The earliest possible arrival of vehicles fa_i and maximum time shift fc_i

i	0	1	2	3	4	5	6	7	8	9
da_i [min]	461	480	501	509	542	567	610	647	687	727
dc_i [min]	0	34	75	148	39	10	63	47	36	0

Table 2 The earliest possible arrival of vehicles da_i and maximum time shift dc_i

The results of both mathematical models are summarized in Table 3.

Optimization criterion		Max-min [min]		Max-min with cascade approach [min]	
Transport direction		F-M – Dobrá	Dobrá – F-M	F-M – Dobrá	Dobrá – F-M
The resulting interval between neighboring bus links [min]	$t_1 - t_0$	19	23.2	27	28.333
	$t_2 - t_1$	19	23.2	27	28.333
	$t_3 - t_2$	49	23.2	33	28.333
	$t_4 - t_3$	19	23.2	27	15.5
	$t_5 - t_4$	19	23.2	27	15.5
	$t_6 - t_5$	19	51	27	59
	$t_7 - t_6$	19	23.2	29	30.333
	$t_8 - t_7$	63	35.8	29	30.333
	$t_9 - t_8$	19	40	19	30.333
	$t_n - t_9$	19	-	19	-
Total waiting time-value of objective function [passenger-minutes]		4629	4356,04	3569	4559,22

Table 3 Comparison of results of time coordination with the various optimization criterion

The total waiting time of passengers in a period $\langle ft_0, ft_n \rangle$ resp. $\langle dt_0, dt_n \rangle$ can be expressed as: $\sum_{i=1}^{n} \frac{1}{2} f\,(dt_i - dt_{i-1})^2$ resp. $\sum_{i=1}^{n} \frac{1}{2} f\,(ft_i - ft_{i-1})^2$. This relationship was used for comparison of results. Suppose that passengers arrive at the stop evenly, with intensity *f*. This relationship was shown in [2] and [8].

5 Conclusions

This paper was focused on the time coordination of bus links in the both transport directions. The solution was made with two optimization criteria. The total waiting time of passengers in an appropriate period was used for comparison of the results. We can see that total waiting time-value of objective function is lower for transport direction Frýdek-Místek – Dobrá with using max-min criterion with cascade approach than with using max-min only. The situation is reversed in the opposite direction. This situation may be caused by linking pairs of bus links. In next research, attention will be focused on solution of this problem. The other variations of optimization criterion in the solution of time coordination in the both transport directions will be tested and the results will be compared as well.

Acknowledgements

This work has been supported by the grant VEGA 1/0361/10 and grant project of the Faculty of Mechanical Engineering Technical University of Ostrava SP2011/129.

References

[1] Černý, J., Kluvánek, P.: *Základy matematickej teórie dopravy*. VEDA: Bratislava, 1991, (in Slovak).

[2] Gábrišová, L., Kozel, P.: *Accuracy of linear approximation used in non-linear problem optimization.* Journal of Information, Control and Management Systems of the University of Žilina Vol 8, No 4, 2010.

[3] Janáček, J.: *Matematické programování.* EDIS, Žilina, 2003, (in Czech).

[4] Janáček, J., Koháni, M.: *Waiting time optimalization with IP-Solver*. Communications–Scientific Letters of the University of Žilina No 3/a, 2010.

[5] Koháni, M., Marton, P.: *Methods and techniques for design of effective and competitive single wagon load transportation*. Communications–Scientific Letters of the University of Žilina No 3, 2009.

[6] Janáček, J., Širc, J.: *The column generation technique for public transport line planning by IP-Solver*. Communications–Scientific Letters of the University of Žilina No 3/a, 2010.

[7] Kozel, P.: *Časová koordinace spojů veřejné hromadné dopravy*. Sborník konference „Dopravní systémy 2009“, Pardubice 3. 12. 2009. Univerzita Pardubice, Dopravní fakulta Jana Pernera, 2009.

[8] Kozel, P.: *Využití kaskádového přístupu při koordinaci spojů*. In Proceedings of the Conference Univerzity of Pardubice, 2010, (in Czech).

Stability of Mean-Risk Models with Log-Normal Distribution of Returns

Václav Kozmík[1], Miloš Kopa[2]

Abstract. This paper deals with portfolio selection problems formulated as mean-risk models with several favorite risk measures (variance, VaR, CVaR, absolute deviation and semivariance). Log-normal and discrete distributions of daily returns are assumed. Analytical approximation based on the assumption that the sum of correlated log-normal random variables follows univariate log-normal distribution is used. An extensive numerical study investigates the stability properties of solutions obtained for sample-based approximations and solutions obtained using the analytical approximation. The sensitivity of the convergence results to the chosen risk measure and parameters of the analytical approximation is analyzed. Optimal approximation parameters for considered financial data are identified. The accuracy and computational burden of the analytical approximation is compared with that of the sample-based approaches.

Keywords: mean-risk models, risk measures, stocks, sample-based approximation.

JEL classification: C63, D81, G11
AMS classification: 49M25, 49M37

1 Introduction

Mean-risk models are well-known instruments in portfolio selection analysis. These two-criteria models both maximize the portfolio mean return as well as minimize the risk which is linked to stock market trading. Basics of portfolio selection theory using variance and semivariance as a measure of risk were published in the article [9] and the book [10] by Harry Markowitz already in 1950s. Since that time, many improvements have been proposed and implemented. New risk measures, such as mean absolute deviation [15] (MAD), Value at Risk [6] (VaR) or Conditional Value at Risk [13] (CVaR), have been proposed. Moreover, several different continuous distributions, for instance log-normal distribution, have been applied in these models. Other popular techniques include scenarios and sampling approaches. Every possible future asset price is called scenario and has some probability attached. Usually, the perfect information about the probability distribution of returns is not available. Its estimation offers a question how the choice of probability distribution affects the optimal portfolio composition. To address the problem, several stability approaches were introduced. General quantitative stability results valid under suitable assumptions such as metric regularity were introduced in [14]. Alternatively, partial knowledge of probability distribution can be included into the model formulation using the ambiguity approach, see [12]. Robustness results with respect to the contamination of the original probability distribution were recently derived and applied to the models with probability constraints in [3] as well as to the mean-risk models in [5]. Another stability approach, applied in the portfolio efficiency testing with respect to the stochastic dominance criteria, was introduced in [8].

In this paper we analyse the stability of mean-risk models when using several favorite risk measures and two types of probability distributions: log-normal and empirical. For sampled scenarios we consider the discrete empirical distribution an approximation of the underlying unknown continuous distribution. We find the optimal solutions for all the considered risk measures while increasing the number of generated scenarios from 100 up to 50,000.

[1]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, e-mail: vkozmik@gmail.com
[2]Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague 8, Czech Republic, e-mail: kopa@karlin.mff.cuni.cz

Unfortunately, in the case of the log-normal distribution the portfolio optimization problem cannot be solved precisely. Therefore, we study an analytical approximation of this optimization problem based on the assumption that the sum of correlated log-normal distributions follows univariate log-normal distribution. The log-scale and shape of the univariate log-normal distribution are determined by matching the moment generating functions. Finally, we compare the obtained results with those of the generated scenarios cases. Convergence properties are studied, especially their dependence on the choice of the risk measure.

The paper is structured as follows: Section 2 presents mean-risk models and solutions based on the analytical approximation for the selected risk measures. Section 3 deals with the same models for generated scenarios and the results are compared in Section 4. The conclusion is reached in Section 5.

2 Mean-Risk Models

Let $R_1, \ldots, R_N$ be the random variables representing our returns, e.g. from holding the individual stocks. Then we get the return of the portfolio with weights $\boldsymbol{w}$: $R(\boldsymbol{w}) = \sum_{i=1}^{N} w_i R_i$. We will assume in the following that the portfolio weights are elements of a given polyhedral set W, e.g.

$$W = \left\{ \boldsymbol{w} : \sum_{i=1}^{N} w_i = 1,\ w_1, .., w_N \in \mathbb{R},\ w_1, .., w_N \geq 0 \right\},$$

which is used in our numerical study. Mean-risk model has two functionals, one representing the returns and second representing the risk. We will always use expectation as the first functional, denoted $u_{\boldsymbol{w}} \equiv u(R) = \mathsf{E}R$. Similarly we denote the risk: $r_{\boldsymbol{w}} \equiv r(R(\boldsymbol{w}))$. With reference to [4], we define:

Definition 1. Portfolio of given N assets with weights $\boldsymbol{w}$ is (mean-risk) efficient, if there are no other weights $w_1^*, .., w_N^*$ such that $\sum_{i=1}^{N} w_i^* = 1$, $u_{\boldsymbol{w}^*} \geq u_{\boldsymbol{w}}$ and $r_{\boldsymbol{w}^*} \leq r_{\boldsymbol{w}}$, where at least one inequality is strict.

The standard optimization task which leads to efficient portfolios can be written as:

$$\begin{aligned} &\min_{\boldsymbol{w}} r_{\boldsymbol{w}} \\ &\text{s. t. } u_{\boldsymbol{w}} \geq u_e \\ &\quad\ \boldsymbol{w} \in W, \end{aligned} \tag{1}$$

where u_e represents the minimal required expected return.

Definition 2. Random variable $\boldsymbol{X}$ follows multivariate log-normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, if it can be written as $\boldsymbol{Y} = \exp\{\boldsymbol{X}\}$, where $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Cumulative distribution function of normal distribution will be denoted by $\Phi(x)$. Considering portfolio weights $\boldsymbol{w}$ we want to evaluate the distribution of random variable $\boldsymbol{w}^T\boldsymbol{Y}$. Let $w_i > 0\ \forall i$:

$$\begin{aligned} &\boldsymbol{w}^T\boldsymbol{X} \sim \mathcal{N}\left(\boldsymbol{w}^T\boldsymbol{\mu}, \boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}\right), \\ &\boldsymbol{w}^T\boldsymbol{Y} = \boldsymbol{w}^T\exp\{\boldsymbol{X}\} = \sum_{i=1}^{N} w_i \exp\{X_i\} = \sum_{i=1}^{N} \exp\{\log w_i + X_i\} = \sum_{i=1}^{N} \exp\{Z_i\}, \\ &\quad Z_i \sim \mathcal{N}(\mu_i + \log w_i, \Sigma_{ii}). \end{aligned}$$

We will denote new log-scale parameter $\boldsymbol{\mu}^* = \boldsymbol{\mu} + \log \boldsymbol{w}$. The random variable representing portfolio return $R = \boldsymbol{w}^T\boldsymbol{Y}$ is in general a sum of correlated log-normal random variables. Unfortunately, analytical formula for this sum is not known, which forces us to use the analytical approximation found in [11]. The sum is approximated by a single univariate log-normal variable, but nonlinear equations leading to nonlinear optimization have to be used to fit the parameters (denoted $\mu^{'}$ a $\sigma^{'2}$). Approximation consists of the evaluation of the moment generating functions for both the univariate log-normal approximation and the sum of multivariate log-normal random variables. Both are approximated by Hermite-integration and put equal in two points s_1, s_2 as we need to fit two parameters. Weights c_j and points x_j come from Hermite-integration, they can be found in [1], section 25.10., J influences approximation accuracy, $s_{ik}^{'}$

denotes an element of the square root of matrix $\boldsymbol{\Sigma}$. Authors of the approximation procedure recommend to use $J = 12$ and $s_1 = 0.001, s_2 = 0.005$, which should lead to a good fit in the tails of the distribution (relevant for risk measures like CVaR or VaR). The following equations are used to determine the parameters of the log-normal approximation:

$$
\begin{aligned}
&K_1 = \sum_{j=1}^{J} \frac{c_j}{\sqrt{\pi}} \exp\left\{-s_1 \exp\left\{x_j \sigma^{'} \sqrt{2} + \mu^{'}\right\}\right\} \\
&K_2 = \sum_{j=1}^{J} \frac{c_j}{\sqrt{\pi}} \exp\left\{-s_2 \exp\left\{x_j \sigma^{'} \sqrt{2} + \mu^{'}\right\}\right\} \\
&K_1 = \sum_{j_1=1}^{J} \cdots \sum_{j_N=1}^{J} \frac{c_{j_1} \cdots c_{j_N}}{\pi^{\frac{N}{2}}} \prod_{i=1}^{N} \exp\left\{-s_1 w_i \exp\left\{\sqrt{2} \sum_{k=1}^{N} s_{ik}^{'} x_{j_k} + \mu_i\right\}\right\} \\
&K_2 = \sum_{j_1=1}^{J} \cdots \sum_{j_N=1}^{J} \frac{c_{j_1} \cdots c_{j_N}}{\pi^{\frac{N}{2}}} \prod_{i=1}^{N} \exp\left\{-s_2 w_i \exp\left\{\sqrt{2} \sum_{k=1}^{N} s_{ik}^{'} x_{j_k} + \mu_i\right\}\right\} \\
&s_1 = 0.001,\ s_2 = 0.005,\ J = 12,\ K_1,\ K_2 \in \mathbb{R}.
\end{aligned}
$$

The final optimization task (1) can be easily solved by using one of the following equations for risk measures as the objective function, see Table 1, and expectation $u_{\boldsymbol{w}} = \exp\left\{\mu^{'} + \frac{1}{2}\sigma^{'2}\right\}$ together with approximation equations.

Variance	$\left(\exp\left\{\sigma^{'2}\right\} - 1\right) \exp\left\{2\mu^{'} + \sigma^{'2}\right\}$
VaR	$-\exp\left\{\mu^{'} + \sigma^{'} q_{1-\alpha}\right\}$
CVaR	$-\frac{1}{1-\alpha} \exp\left\{\mu^{'} + \frac{\sigma^{'2}}{2}\right\} \Phi\left(q_{1-\alpha} - \sigma^{'}\right)$
MAD	$\exp\left\{\mu^{'} + \frac{\sigma^{'2}}{2}\right\} \left(4\Phi\left(\frac{\sigma^{'}}{2}\right) - 2\right)$
Semivariance	$\exp\left\{2\mu^{'} + \sigma^{'2}\right\} \left(\exp\left\{\sigma^{'2}\right\} \Phi\left(-\frac{3\sigma^{'}}{2}\right) - 2 + 3\Phi\left(\frac{\sigma^{'}}{2}\right)\right)$

Table 1: Log-normal distribution: risk measures

3 Discrete Scenarios

Suppose we have M scenarios of possible future returns, that means vector $\boldsymbol{r}^j$ for each scenario $j, j = 1, .., M$. We will sample our scenarios from the selected continuous distribution, allowing for a reasonable assumption that the probabilities of all the scenarios are the same, e.g. $p = p^j = \frac{1}{M},\ j = 1, .., M$.

To get the mean-variance portfolio, we first estimate the mean returns vector $\boldsymbol{r}$ and the variance matrix $\boldsymbol{V}$. Then we solve (1) with $r_{\boldsymbol{w}} = \boldsymbol{w}^T \hat{\boldsymbol{V}} \boldsymbol{w}$ and $u_{\boldsymbol{w}} = \boldsymbol{w}^T \hat{\boldsymbol{r}}$.

In the case of VaR we have to use integer variables, which leads to a slower mixed integer program, see [7]:

$$
\begin{aligned}
\min_{\nu, \boldsymbol{w}, \delta^j}\ & \nu \\
\text{s. t. } & -\boldsymbol{w}^T \boldsymbol{r}^j \leq \nu + K\delta^j,\ j = 1, .., M \\
& \sum_{j=1}^{M} \delta^j = \lfloor (1-\alpha) M \rfloor \\
& \delta^j \in \{0, 1\},\ j = 1, .., M \\
& \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^j \geq u_e \\
& \boldsymbol{w} \in W
\end{aligned}
$$

where $\lfloor x \rfloor = \max \{n \in \mathbb{N}_0, n < x\}$ for $x \in \mathbb{R}^+$, and $K \geq \max_{i,j} r_i^j - \min_{i,j} r_i^j$.

When CVaR is chosen as the risk measure, we get the following linear program which can be found for instance in [17]:

$$\begin{aligned}
\min_{a,\boldsymbol{w},z^j} \ & a + \frac{1}{(1-\alpha)\, M} \sum_{j=1}^{M} z^j \\
\text{s. t. } & z^j \geq -\boldsymbol{w}^T \boldsymbol{r}^j - a,\ j = 1, .., M \\
& z^j \geq 0,\ j = 1, .., M \\
& \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^j \geq u_e \\
& \boldsymbol{w} \in W.
\end{aligned}$$

When deriving the form of the optimization problem for mean absolute deviation, we can use a similar approach as above. From the basic properties of absolute value ($x \leq |x|, -x \leq |x|$), we get a linear program:

$$\begin{aligned}
\min_{\boldsymbol{w},z^j} \ & \frac{1}{M} \sum_{j=1}^{M} z^j \\
\text{s. t. } & \boldsymbol{w}^T \boldsymbol{r}^j - \frac{1}{M} \sum_{k=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^k \leq z^j,\ j = 1, .., M \\
& -\boldsymbol{w}^T \boldsymbol{r}^j + \frac{1}{M} \sum_{k=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^k \leq z^j,\ j = 1, .., M \\
& \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^j \geq u_e \\
& \boldsymbol{w} \in W.
\end{aligned}$$

For semivariance, we use a similar approach with variable z representing our losses, $\max(0, \mathsf{E}R - R)$:

$$\begin{aligned}
\min_{\boldsymbol{w},z^j} \ & \frac{1}{M} \sum_{j=1}^{M} \left(z^j\right)^2 \\
\text{s. t. } & z^j \geq -\boldsymbol{w}^T \boldsymbol{r}^j + \frac{1}{M} \sum_{k=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^k,\ j = 1, .., M \\
& z^j \geq 0,\ j = 1, .., M \\
& \frac{1}{M} \sum_{j=1}^{M} \boldsymbol{w}^T \boldsymbol{r}^j \geq u_e \\
& \boldsymbol{w} \in W.
\end{aligned}$$

4 Results

In this section, we summarize the numerical results obtained by using the presented models for stock market data. We used daily returns of stock indices in Japan, USA (Dow Jones), Great Britain, Czech Republic and Germany from 09-15-2008 to 09-18-2009. The computations were executed using own software written in C++ which cooperates with GAMS to solve the optimization tasks. We used TRNG [16] and its non-linear random number generator to generate random samples from continuous distributions. Each sampling experiment was repeated 50 times, which means we have found 50 optimal portfolio weights and taken their average as the sought approximate optimal solution. Scenarios were always generated from the scratch with random seed based on the current system time. The confidence level was always set to

95 %. Maximal number of scenarios to be processed and solved on a desktop machine was approximately 50,000 (Core 2 Duo 2,4 GHz and 4 GB RAM) for all risk measures except VaR, where the maximum was approximately 1,000 scenarios.

The points recommended in [11] to match the moment generating functions were specified as 0.001 and 0.005 in order to approximate the tails of the distribution, which is suitable for VaR and CVaR. In the case of other risk measures the tails of the distribution do not influence the results that much, and we also have to find good approximation of the expected value, because it is used as a constraint. Alternative recommended points were 0.1 and 0.5, but neither choice led to satisfactory results. We have used several alternatives lying between these recommendations and found out that parameters 0.01 and 0.05 provide the best approximation in our models. The best parameters were found by comparison with sample-based optimal solutions, we assume that with growing number of scenarios the sample-based approximations would be more and more precise. In an ideal case the distance should be shortening as the number of scenarios grows, meaning that our analytical approximation could be more precise than the sample-based solution for a given number of scenarios. This was achieved with the parameters mentioned, as opposed to the default parameters which provided at least twice as bad results measured by the distance of both approximations. Our approach was to study optimal weights and compare both approximations mentioned, but there are also other procedures dealing with the optimal objective function values, see [2].

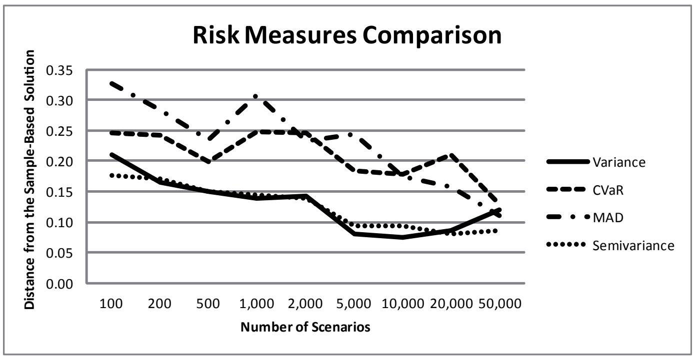


Figure 1: Results with different risk measures

We measure the distance between the optimal solution based on the analytical approximation and the sample-based solution as the number of scenarios grows. This is shown in Figure 1 for all risk measures except VaR. We conclude that variance and semivariance provide the best results for a lower number of scenarios, but for a high number of scenarios all risk measures perform more or less the same. With regard to the decrease of the distance between our approximations for all risk measures as the number of scenarios grows, we expect that this decrease stops for some greater number of scenarios, were we able to calculate such results. Optimal solutions using the analytical approximation should therefore be more precise and we recommend them for the setup with few assets in the portfolio. Unfortunately, when the number of assets grows the approximation complexity grows exponentially, which leads to the conclusion that for setups with a high number of assets we would recommend sample-based approximation instead. We have achieved similar results for the VaR risk measure, but the complexity of the mixed integer program for the sample-based approximation allows us to optimize only up to 1,000 scenarios, so we would always recommend to use analytical approximation when VaR is chosen as the risk measure.

5 Conclusion

This paper studies portfolio optimization under the assumption of the log-normal distribution and various risk measures. We have managed to derive an analytical approximation procedure based on the assumption that the sum of correlated log-normal random variables follows log-normal distribution and compared it with the sample based approximation. Moreover, we have proposed the optimal parameters of the analytical approximation for the mean-risk models with financial data.

The presented results indicate that both approximative procedures provide good results for all risk measures. Analytical approximation should be used for the cases with a low number of assets in the portfolio and the sample-based approximation otherwise, mainly for computational reasons. Comparing the selected risk measures, variance and semivariance provide the best convergence results, but differences tend to vanish as the number of scenarios grows.

Our work could be extended in many ways. We have included only a few basic risk measures involving linear programming, quadratic programming and mixed integer programming, but there are also other ways of measuring the risk. The future work could include dynamic risk measures, other continuous distributions or a study of the case when negative weights are allowed.

Acknowledgements

The research was partly supported by the project 'Methods of modern mathematics and their applications' – MSM 0021620839, the Czech Science Foundation (grants P402/11/0150, P402/10/1610) and the grant SVV 261315/2011.

References

[1] Abramowitz, M. and Stegun I. A.: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York, 1964, ISBN 0-486-61272-4.

[2] Bayraksan, G. and Morton, D.: Assessing Solution Quality in Stochastic Programs via Sampling, *Tutorials in Operations Research* (2009), 102–122, ISBN 978-1-877640-24-7.

[3] Branda, M. and Dupačová, J.: Approximation and Contamination Bounds for Probabilistic Programs, *Annals of Operations Research*, published on-line November 11, 2010.

[4] Dupačová, J., Hurt, J. and Štěpán J.: *Stochastic Modeling in Economics and Finance*, Kluwer Academic Publishers, 2002, ISBN 1-4020-0840-6.

[5] Dupačová, J. and Kopa, M.: Robustness in Stochastic Programs with Risk Constraints, *Annals of Operations Research*, published on-line January 5, 2011.

[6] J.P. Morgan Risk Metrics: *Technical Document, 4^{th} Edition*, Morgan Guaranty Trust Company, New York, 1995.

[7] Kopa, M.: *Utility Functions in Portfolio Optimization*, Doctoral dissertation thesis, Charles University in Prague, Faculty of Mathematics and Physics, 2006.

[8] Kopa, M.: Measuring of Second-Order Stochastic Dominance Portfolio Efficiency, *Kybernetika* **46 (3)** (2010), 488–500.

[9] Markowitz, H. M.: Portfolio Selection, *The Journal of Finance* **7 (1)** (1952), 77–91.

[10] Markowitz, H. M.: *Portfolio Selection: Efficient Diversification of Investments*, John Wiley & Sons, New York, 1959, ISBN 978-0300013726.

[11] Mehta, N., Molisch, A., Wu, J. and Zhang, J.: *Approximating the Sum of Correlated Lognormal or Lognormal-Rice Random Variables*, Mitsubishi Electric Research Laboratories, 2006.

[12] Pflug, G. and Wozabal, D.: Ambiguity in Portfolio Selection, *Quantitative Finance* **7** (2007), 435–442.

[13] Rockafellar R. T. and Uryasev S.: Conditional Value-at-Risk for General Loss Distributions, *Journal of Banking and Finance* **26 (7)** (2002), 1443–1471.

[14] Römisch, W.: Stability of Stochastic Programming Problems. In: *Stochastic Programming, Handbooks in OR & MS*, Ruszczyński, A., Shapiro, A. (Eds.), 2003, 483–554.

[15] Sharpe, W. F.: Mean-Absolute-Deviation Characteristic Lines for Securities and Portfolios, *Management Science* **18 (2)** (1971), B-1–B-13.

[16] *Tina's Random Number Generator Library* http://trng.berlios.de/.

[17] Uryasev, S.: *Probabilistic Constrained Optimization: Methodology and Applications*, Kluwer Academic Publishers, 2002, ISBN 0-7923-6644-1.

Agent-Based Model of an Urban Retail Market

Ondřej Krčál[1]

Abstract. The paper presents an agent-based model of an urban retail market that explains the stylized facts about the location of shops in Brno/CZ. More specifically, it explains why some types of shops are more dominant in the center, some in the inner city, and some in the periphery of the city. In the model, consumers spend their incomes in different types of shops located randomly in a square city. The types of products sold in the shops differ in the transportation costs and in the proportion of the income spent on them. The consumers buy each type of product in the shops with the lowest price plus the per-unit transportation cost. The price of each good is determined by the profit margin assigned randomly to each shop. In each period, new shops of all types enter the market and locate randomly in the city, and some of the shops in loss exit the market. This way the urban retail market evolves toward an optimal structure; i.e. toward the optimal prices, sizes, and locations of shops. The model is calibrated using the data from the Brno retail market in 2009, and for a realistic setting of the model, it generates patterns consistent with the stylized facts.

Keywords: ACE, spatial competition, location, retail market, Brno

JEL classification: L13, R39

1 Introduction

Imagine getting on a tram in the center of a European city and traveling through the inner city to the terminal stop in the outskirts. On the way, you may notice that the structure of shops changes as the distance from the center grows. Doing this experiment in Brno/Czech Republic, you would find that there are three types of shops. The first type, the supermarkets, take up relatively low proportion of the total shop area in the center of the city and relatively high proportion of the area in the periphery. The second type, the shops selling cars, furniture, electronics, building materials or sports equipment, take up relatively more area in the inner city and relatively less area in the center and periphery. And finally, the third type of shops selling clothes, shoes, music, books, cosmetics, stationery or PC equipment is relatively more frequent in the center of the city, and relatively less frequent in the inner city and in the periphery.

This paper presents an agent-based model that explains the stylized facts about the location of shops in Brno. The model follows the tradition of spatial competition models (see [5], [1], [7], [2]). But in order to provide a more realistic picture, consumers and firms are located in a two-dimensional square city. In this respect, the paper is closely related to location models in a two-dimensional space (e.g. see [4], [3], [8]). Contrary to these papers, the model presented in this paper finds an evolutionary rather than a game-theory equilibrium. The intuition behind the explanation provided by the model is as follows. Shops of the third type are most frequent in the center, because their products are relatively easy to transport. Unlike in other sectors, people can purchase these products while being in the center for work or other reasons, and transport them home without much inconvenience. And because most people go to the center regularly, the shops of the third type located in the center are able to attract more consumers than the competition in other parts of the city. Shops of the second type are more common in the inner city, because the optimal number of shops in these sectors is relatively low. That is, they serve customers within a wider range. These shops are less likely to locate near the edge of the city, because part of their range would reach out of the city. And similarly, they are less likely to locate in the center because of tough competition from the inner market. Finally, the supermarkets locate evenly in the city, but they are relatively more common in the outskirts, simply because the other types of shops locate relatively less frequently there.

The paper proceeds as follows: Section 2 introduces the model. Section 3 presents the stylized facts about the Brno retail market. Section 4 specifies the setting of the model. And Section 5 explains the stylized facts and discusses the robustness of the explanation.

[1]Masaryk University/FEA, Economics department, Lipová 41a, 602 00 Brno, e-mail: krcalo@mail.muni.cz

2 Description of the model

Time is discrete. The city has a form of a square with side a. It is inhabited by a population of n consumers with their *homes* located randomly in the city. Each consumer commutes to *work* with the probability $1 - h$ or stays at home with the probability h. The work of the commuting consumers is located on a straight line between home and the center of the square at a random distance from the center $0 \leq d_c \leq wa/2$, where $w \in \langle 0, 1 \rangle$. Each consumer divides her income I between different types of consumer goods (*products*). Each product j differs from other products in three aspects: 1) in the proportion of income g_j spent on product j, 2) in the transportation cost t_j, and 3) in whether consumers are willing to buy and transport the product also from their workplaces.

Each product is sold in a different *sector*. Each consumer spends the corresponding proportion her income Ig_j in the shop i in sector j that minimizes the price plus the unit transportation cost

$$p_i + d_i t_j \frac{p_i}{Ig_j}, \tag{1}$$

where p_i is the price in shop i, and d_i is the distance from home to shop i. If consumers are willing to purchase the product also on the way from the work, each commuting consumer considers two distances to each shop i in sector j: the distance from home and from the workplace. The price of shop i is given by $p_i = c + m_i$, where c is the constant marginal cost (the same for all the shops), and m_i is a profit margin assigned randomly to shop i from the interval $m_i \in \langle 0, m_{max} \rangle$. The profit of shop i is then given by

$$\pi_i = m_i * R_i - F, \tag{2}$$

where R_i is the revenue determined by the consumer choice (1) and F is the fixed cost (the same for all the shops).

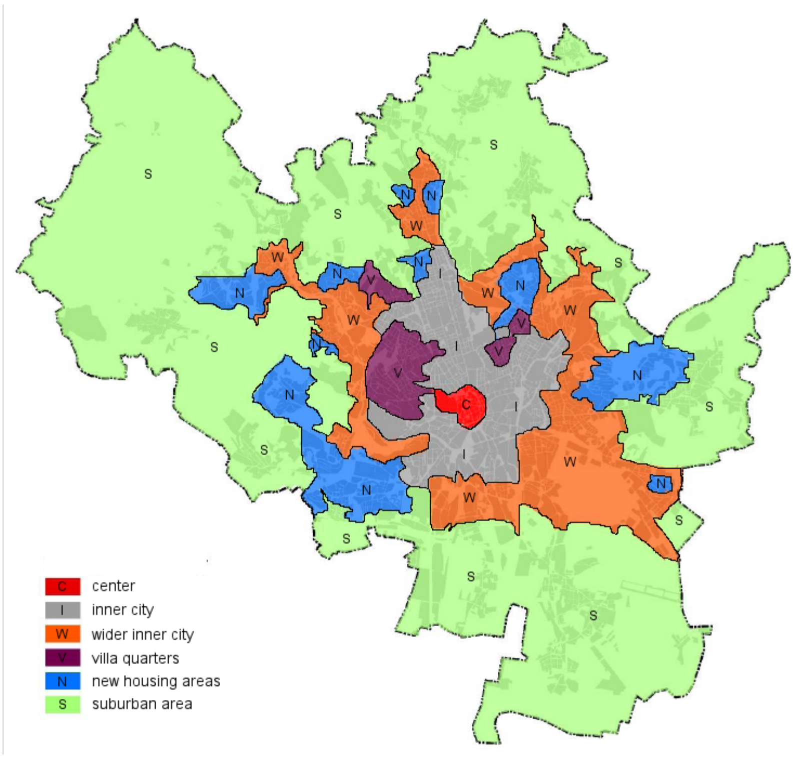
The number of shops in different sectors, their locations, revenues and profit margins evolve in the course of the simulation. At the beginning of the simulation, the number of shops in each sector j is sg_j, where s is the initial number of shops. The shops have random profit margins and are located randomly in the city. At the beginning of each period, u new shops of each type again with random profit margins locate randomly in the area. And at the end of each period, each shop exits the market with the probability $-\pi_i/F$. That is, each shop in loss exits the market with a probability that is increasing in the loss, and each shop with the loss at least as high as F exits the market with certainty. This way, the market evolves towards the optimal structure, because only the shops with correct locations and profit margins survive in the long run.

3 Brno retail market

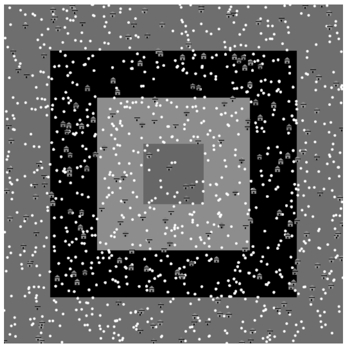
The data about the structure of Brno retail market are from a 2009 study by Mulíček & Osman [6]. This study divides Brno into six morphogenetic zones: center, inner city, wider inner city, villa quarters, new housing areas and suburban area (see Panel 1a). This paper does two adjustments to the division: 1) It merges the inner city and villa quarters into one zone (called *inner city* in this paper), because villa quarters is a relatively small zone, and judging by the distance from the center, it forms a natural part of inner city. And 2) it excludes suburban area from the stylized facts, because the structure of shops there might be biased by the large shopping centers located near D1, D2 and R43 motorways that serve a wider market than Brno. Hence, the paper compares the structure of shops in four zones: center, inner city, wider inner city, and new housing areas.

Furthermore, Mulíček & Osman [6] divide the retail market into 15 sectors. Table 1 shows the percentage of total area taken up by the sectors in different zones of the city. For instance, grocery stores use 7.8% of the total retail area (m^2) in the center of the city, and 35.5% of the retail area in the new housing areas. The sectors can be divided into four categories:

- Type 1 sector (G): the percentage share is increasing in the distance from the center with an exception of wider inner city;
- Type 2 sectors (A, B, E, F, Sp): the percentage share is first increasing and then decreasing in the distance from the center (inverted-U shape);
- Type 3 sectors (St, C, S, Ch, J, M, P): the percentage share is first decreasing and then slightly increasing in the distance from the center (U shape);
- Small sectors with no clear pattern (H, Pe).



(a) Morphogenetic zones in Brno, *Source: [6], p. 21, own translation*



(b) Model city with four zones

Figure 1: Zones of the city

sector	center	inner city	wider inner city	new housing areas	type
grocery (G)	7.8	17.4	10.5	35.5	1
auto (A)	0.1	18.3	34.1	30.5	2
building (B)	0.3	16.1	22	0.3	2
electronics (E)	1.5	3.5	3.2	2.5	2
furniture (F)	1.8	7	7.8	6.8	2
sports (Sp)	2.5	3.9	0.9	0.4	2
stationery (St)	5.1	1.5	0.2	1.8	3
clothes (C)	40.7	12.3	2.5	12.5	3
shoes (S)	6.4	2.1	0.1	2.6	3
chemist's (Ch)	4.7	2.1	0.9	1.9	3
jewelers (J)	4.3	0.8	0.1	0.5	3
music (M)	8.2	0.7	0	0.1	3
PC (P)	1.5	11	0.2	0.3	3
houseware (H)	7.3	2.9	15.6	1.9	–
pet shop (Pe)	0.6	0.9	15.6	1.9	–
other	7.1	9.4	1.4	1.7	
total m^2	42,413	236,379	129,779	79,598	

Source: [6], own calculations

Table 1: Retail market in Brno
The table shows percentage of the total shop area in the four zones of Brno covered by different types of shops and the total shop area in the four zones. The sectors sell the following products: G – food; A – new and used cars, car accessories, etc.; B – building materials; E – consumer electronics; F – furniture; Sp – sports equipment; St – stationery; C – clothes, textile; S – shoes; Ch – cleaning agents, cosmetics, fragrances, etc.; J – jewelry and watches; M – music and books; P – computers and communication technology; H – houseware; Pe – pet supplies.

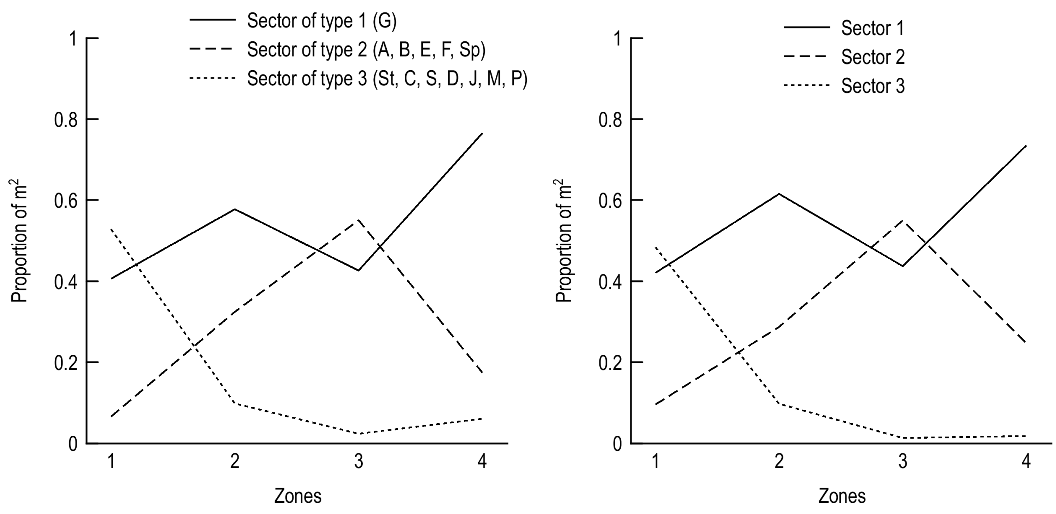
4 Setting of the model

This section specifies the setting of the model using the data from Brno retail market presented in the previous section. The model is implemented in Netlogo 4.0.4. The side length of the model city is $a = 50$ patches (patch is a measure of distance used in Netlogo). The model city is divided into four zones. The center is a square with a side of 8 patches and the other three zones have a form of a square with a square hole in the middle with the width of 7 patches on all sides (see Panel 1b for the picture of the zones in the model city).

The number of sectors in the model is limited to three ($j = 1, 2, 3$), because the sectors of type 1, 2, and 3 sell similar products and this assumption simplifies greatly the presentation of the results in the following section. The parameters of products 1, 2, and 3 are as follows:

- Product 1 (food):
  1) $g_1 = 0.55$: Consumers spend 55% of their income on food;
  2) $t_1 = 3$: Transportation cost is the cost per unit of distance of going to the shop and back during a period of time (cost of time, gasoline, tickets for public transport, etc.). The size of transportation cost is increasing in the frequency of shopping;
  3) Only from home: The food is heavy and bulky. Consumers are not willing to transport it from work in the center of the city.
- Product 2 (cars, car parts, building materials, furniture, etc.):
  1) $g_2 = 0.4$: Consumers spend 40% of their income on a typical product 2;
  2) $t_2 = 0.1$: The frequency of shopping for product 2 is assumed to be 30 times lower than for product 1;
  3) Only from home: A typical product 2 is heavy and bulky.
- Product 3 (clothes, shoes, stationery, books, music, etc.):
  1) $g_3 = 0.05$: Consumers spend 5% of their income on a typical product 3;
  2) $t_3 = 1$: The frequency of shopping for product 3 is assumed to be 10 times higher than for product 2;
  3) From home and work: A typical product 3 is easy to transport.

The remaining parameters are as follows: number of consumers $n = 1000$, income of consumers $I = 250$, initial number of shops $s = 200$, number of shops in each sector entering the market in each period $u = 10$, fixed cost $F = 50$, marginal cost $c = 1$, maximal profit margin $m_{max} = 0.2$, the parameter of location of work $w = 0.4$, and the proportion of consumers staying at home $h = 0.4$.



(a) Stylized facts

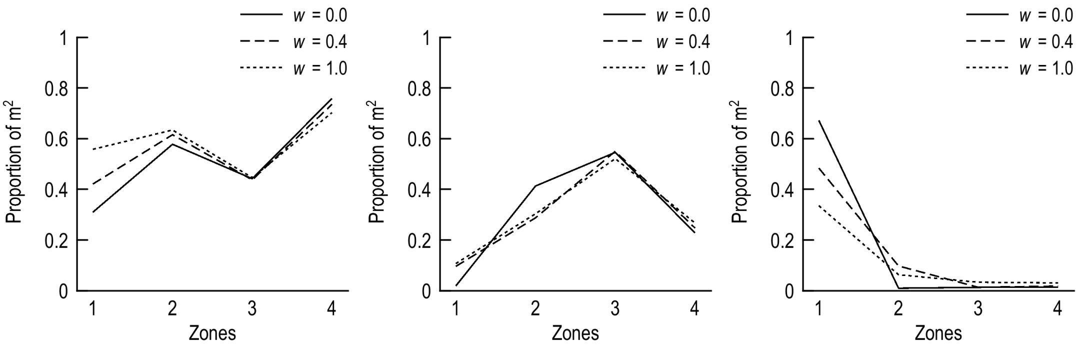
(b) Prediction of the model

Figure 2: Stylized facts vs. prediction of the model
The figure shows the proportion of the total shop area taken up by an average sector of type 1, 2, and 3 in zones: 1 – center of the city, 2 – inner city, 3 –wider inner city, 4 – new housing areas.

5 Results of simulations

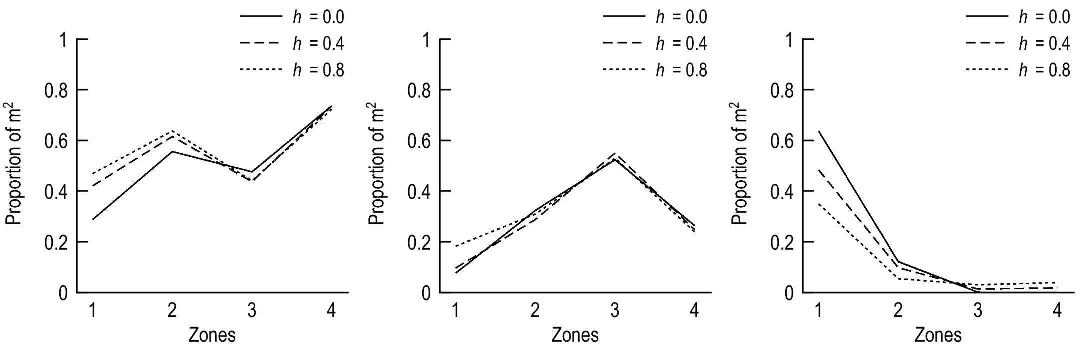
This section presents an explanation of the stylized facts and discusses the robustness of the explanation. Panel 2a shows the stylized facts for three average sectors of type $j = 1, 2, 3$. For each zone of the city, the proportion of the total shop area taken up by each sector is calculated as a sum of the area of all sectors of type j divided by total area of all 13 sectors of types 1, 2, and 3 (see Table 1). Panel 2b and the following figures present the predictions of the model. They use revenue of shops at the end of the period (after some of the shops exit the market) as a proxy for the shop area. The proportion of the total shop area taken up by each sector is calculated as the revenue of the sector divided by total revenue of all shops in a given zone of the city. Each point in the figures is an average of 1,000 data generated in 10 runs, where each run provides 100 numbers generated in the periods from 101 to 200.

The prediction of the model in Panel 2b corresponds to the empirical findings in Panel 2a. Two effects explain the predicted pattern: 1) Because product 3 can be bought both also from work, the shops in the center that offer product 3 for competitive prices serve a relatively high number of consumers. Therefore, the proportion of sector 3 in the center tends to be high, and the proportion of sectors 1 and 2 in the center tends to be low. The first effect is called the *center effect*. 2) Because product 2 has low transportation cost, only a handful of shops survive in the market. The shops located near the edge of the city have a competitive disadvantage, because part of their market reaches out of the city where no consumers live. Therefore, they are likely to lose in direct competition with shops in zone 3. And at the same time, shops located in the center have a difficult position because they are surrounded by large shops in zones 2 and 3. And because the inverted-U relationship in Panel 2b peaks in zone 3, the proportion of sector 1 and 3 in wider inner city is lower than in inner city or new housing areas. The second effect is called the *inner-city effect*.



(a) Sector 1 (b) Sector 2 (c) Sector 3

Figure 3: The effect of w on the predicted pattern.



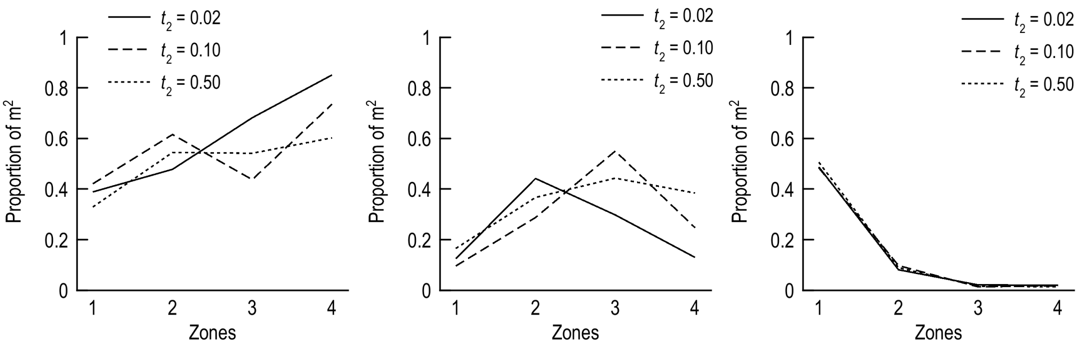
(a) Sector 1 (b) Sector 2 (c) Sector 3

Figure 4: The effect of h on the predicted pattern.

The center and inner-city effects are reasonably robust to a variation in parameters of the model. Because of space limitation, the rest of the paper discusses only the effects of changes in three parameters: w, h, and t_2. The parameter w determines the size of the working area. If $w = 1$, the working area is a circle with the radius of $a/2$.

And if $w = 0$, all the commuting consumers work in the center of the square area. More consumers working in the center means higher potential demand for the shops in the center. Therefore, a reduction in w increases the center effect (see Panel 3c). (In the following figures, all other parameters are at the levels specified in Section 4.) Changes in other sectors (Panel 3a and 3b) are directly related to the changes in sector 3. Figure 3 also shows that a large variation in w has a relatively small effect on the predicted pattern.

The parameter h indicates the proportion of consumers staying at home. If $h = 0$, all consumers go to work. If the workplace is located near the center of the city, the commuting consumers increase the demand in the central part of the city. Hence, a reduction in h increases the center effect (see Panel 4c, note that $w = 0.4$ in this figure). Moreover, Figure 4 shows that the center effect is robust to a variation in h. Even if 80% of consumers don't go to work ($h = 0.8$), the center effect is large enough to generate a pattern that is similar to the empirical findings.



(a) Sector 1 (b) Sector 2 (c) Sector 3

Figure 5: The effect of t_2 on the predicted pattern.

The transportation cost t_2 influences directly the pattern in sector 2. An increase in t_2 increases the optimal number of firms in sector 2. These firms are more likely to survive also in the new housing areas and center of the city. Hence, an increase in t_2 makes the inverted-U relationship flatter (see the dotted line in Panel 5b). On the other hand, a reduction in t_2 reduces the optimal number of shops, which leads to location of shops closer to the center. The inverted-U relationship for very low t_2 is likely to peak in zone 2 (see the solid line in Panel 5b). Figure 5 shows that the inner-city effect (the inverted-U line in sector 2) is robust even to a fivefold change in t_2, i.e. to a fivefold increase or reduction in the frequency of shopping for product 2.

References

[1] d'Aspremont, C., Gabszewicz, J. J. and Thisse, J.-F.: On Hotelling's "Stability in Competition", *Econometrica* **47** (1979), 1145–1150.

[2] Economides, N.: Minimal and Maximal Product Differentiation in Hotellings Duopoly, *Economic Letters* **21** (1986), 67–71.

[3] Economides, N.: Nash Equilibrium in Duopoly with Products Defined by Two Characteristics *The RAND Journal of Economics* **17** (1986), 431–439.

[4] Eaton, B. C., Lipsey, R. G.: The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition, *The Review of Economic Studies* **42** (1975), 27–49.

[5] Hotelling, H.: Stability in Competition, *The Economic Journal* **39** (1929), 41–57.

[6] Mulíček, R., Osman, O: Průzkum maloobchodní sítě na území města Brna [online], (2009) available at http://www.brno.cz/sprava-mesta/magistrat-mesta-brna/usek-rozvoje-mesta/odbor-uzemniho-planovani-a-rozvoje/dokumenty/upp/pruzkum-maloobchodni-site-na-uzemi-mesta-brna.

[7] Salop, S. C.: Monopolistic Competition with Outside Goods, *The Bell Journal of Economics* **10** (1979), 141–156.

[8] Veendorp, E. C. H., Majeed, A.: Differentiation in a Two-Dimensional Market, *Regional Science and Urban Economics* **25** (1995), 75–83.

Modelling disaggregated aging chain of Czech population

Igor Krejčí[1], Roman Kvasnička[2]

Abstract. Paper deals with the design and construction of the disaggregated aging chain for the purposes of dynamic population model of the Czech Republic. The methodology of system dynamics is used for model construction.
The emphasis is laid primarily on specifics of Czech population structure and modelling consequences of these specifics. The necessity of high age cohorts' disaggregation is presented. The model also respects some demographic changes in children bearing.
The population aging chain is necessary for different kinds of models that focus on various socioeconomic problems. Prepared modeling structures used for common problems (Molecules or System ZOO) belong to the good practice of system dynamics. The goal of this aging chain is to perform a similar role for the problems that need to model Czech population. Therefore aging chain is constructed and provided. The goal also affects the model's structure; the wide utility of the aging chain is conditioned by high details of modeled population structure.

Keywords: system dynamics, aging chain, population, death probability, fertility rate, life expectancy.

JEL Classification: C63, C44, J11
AMS Classification: 91D20, 93C15

1 Introduction

Prepared modeling structures used for common problems (Molecules or System Zoo) belong to the good practice of system dynamics. These structures make modeling process easier and faster. Aging chain is a common structure that expresses movement and maturation between different cohorts [13]. As general structure of stock and flow diagram and system of differential equations, the aging chain is a part of so called system Zoo [1].

Meadows et al. [11] used aging chain with four cohorts in the population model from the Limits to Growth; the same aggregation was used for model update [12]. The whole model from the Limits to Growth [11] is also a part of system Zoo [2]. Wang and Steman [15] used highly disaggregated (sixty-five cohorts) aging chain for the system dynamic population model of China.

Furthermore, aging chain structure is also useful for non-human population modeling. Wang and Ma [14] used the same structure for capital aging modeling in a system dynamics urban study in China. Eskinasi et al [8] used two aging chains for social and market housing stocks in a system dynamics model of urban transformation in Haaglanden region in the west of the Netherlands. Kunc [10] used promotion chain to simulate staff qualification advancement.

The paper deals with the design and construction of the disaggregated aging chain for the purposes of system dynamics population model of the Czech Republic. Second section of the paper deals with demographic data sources used for parameters quantification. This section also contains the explanation of basic system dynamics components used for aging chain construction. The third section contains whole stock and flow diagram of the aging chain and focuses on the comparison of results of the aging chain model and surveyed demographical data. The last section summarises the previous two sections and explains the versatility of the model, the approach how to implement the aging chain into more complex model is suggested as well.

The emphasis is laid primarily on specifics of Czech population structure and modelling consequences of these specifics. The necessity of high disaggregation of age cohorts' is presented. The model also contains demographic changes in child bearing. The forecasting is not an objective of this paper; however it is possible to use this model for these purposes.

[1] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, krejcii@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, kvasnicka@pef.czu.cz.

2 Materials and methods

We used data from Czech statistical office to quantify the parameters: Czech demographic hand book [3], Demographic yearbook of the Czech Republic [4] and life tables [5] and [7]. We used simulation software Vensim DSS 5.10e to design the stock and flow diagram and for simulation run.

The model disaggregates the population by sex, thus two aging chains were used. Concerning the model as a molecule for more complex system dynamics models with different objectives, the aging chain is highly disaggregated. Focusing on the retirement reform policy testing requires different age groups in comparison to model that focuses on school reform or health care policy. Thus, the aging chains are disaggregated to 72 age cohorts. One stock variable is for each age from 0 to 70 and one stock variable for age 71 and higher.

Each stock variable has two inflows and two outflows. The first inflow is a net immigration, that is concerned as exogenous variable and for years 2011 to 2050 it is represented as a constant equal to the average from the years 2000 to 2010. The second inflow is aging from previous cohort or births for cohort of age 0. Outflows involve aging to higher cohort and deaths. Deaths are function of the life expectancy at birth. Births are function of total fertility rate, average mother's age, and population in female cohorts. Figure 1 presents a simplified structure of aging chain, index s represents sex identification (M, F) and i is age, the stock variables are in boxes, the flow variables are represented by pipes with arrow and faucet, simple arrows are causal connections.

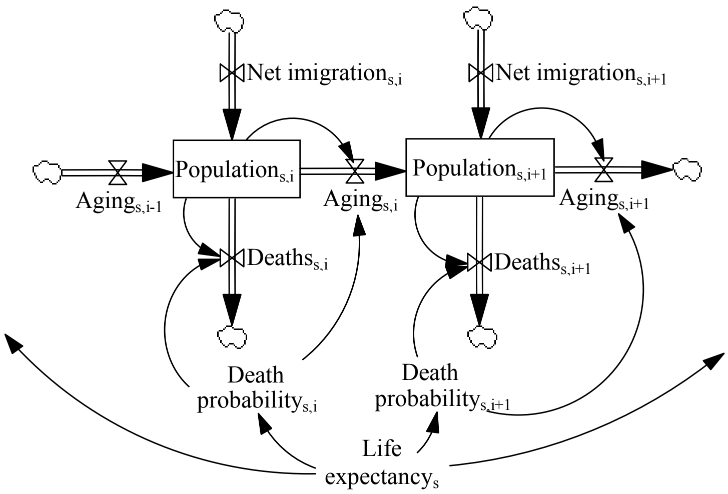


Figure 1 Stock and flow diagram for population aging chain

Population is calculated on the basis of (1). $A_{s,i}$ represents aging, $D_{s,i}$, is deaths' outflow from population $P_{s,i}$, $I_{s,i}$ is net immigration, T is current time, T_0 is initial time and t is any time between T and T_0.

$$P_{s,i}(T) = \int_{T_0}^{T} \left(A_{s,i-1} - A_{s,i} + I_{s,i} - D_{s,i}\right) dt + P_{s,i}(T_0) \quad (1)$$

Equations (2), (3) and (4) are used for outflows calculations of the flow variables. $DP_{s,i}$ is death probability LE_s is life expectancy, $O_{s,i}$ is total outflow from population cohort $P_{s,i}$.

$$D_{s,i} = O_{s,i} \cdot DP_{s,i} \quad (2)$$

$$A_{s,i} = O_{s,i} \cdot \left(1 - DP_{s,i}\right) \quad (3)$$

$$O_{s,i} = P_{s,i} \quad (4)$$

The right-hand side of the equation (4) is usually divided by average number of years per cohort [13]. In our case this denominator is equal to one. Data for surveyed death probability $SDP_{s,i}$ are obtained from life tables [7]. Equation (5) is used for its calculation, $m_{s,i}$ is age specific mortality rate, i.e. number of deaths in cohort to number of people in the cohort [5]. Equation (5) is appropriate to continuous simulation [13].

$$SDP_{s,i} = 1 - e^{-m_{s,i}} \quad (5)$$

Life expectancy is taken as an exogenous variable. Death probability was transformed into Vensim lookup graphical function with life expectancy as an independent variable. The ordinary least square method was used. For each age cohort under 15 years, the determination coefficient R^2 was more than 0.75, for cohorts with age higher than 30, R^2 was more than 0.9.

Fertility is calculated only for female cohorts of age 15 – 49. The births B are a sum of children born to mothers in each specific age cohort (6), FI_i is cohort fertility index and TF is total fertility rate. Contrary to the references in the introduction it was necessary to implement the average mothers age into calculations and dynamise the fertility distribution.

$$B = \sum_{i=15}^{49} P_{F,i} \cdot TF \cdot FI_i \qquad (6)$$

Figure 2 shows the distribution shift in years 2000 – 2009, this shift is reason for the cohort fertility index implementation. FI_i is function of average mother age (R^2 is higher than 0.94 for each age cohort; for these purposes the marginal cohorts 15-17 and 42-49 were aggregated). Total fertility rate and cohort indexes are taken as exogenous variables. 51.4 % of the births are the inflow to male aging chain; the rest is the inflow to female aging chain. The used ratio is 20-year average.

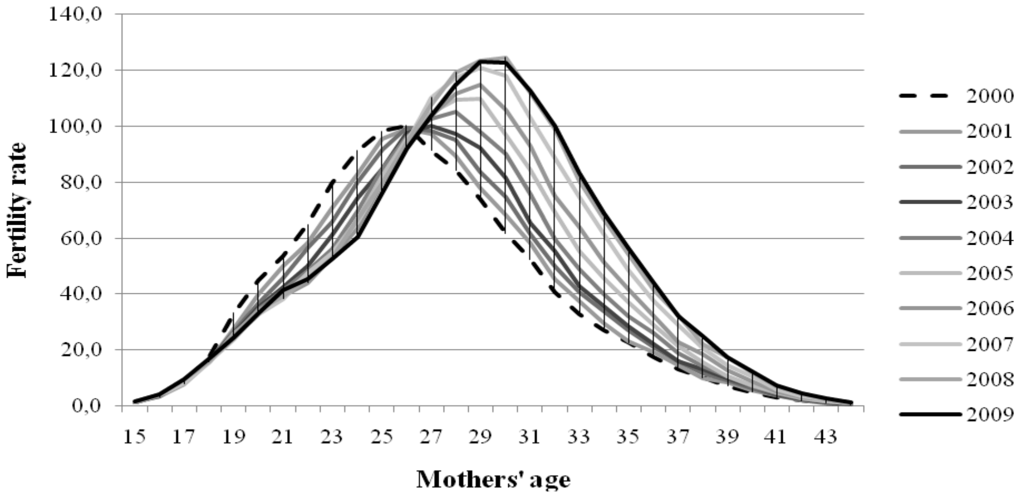


Figure 2 Fertility rate distribution

For simulation, fourth order Runge-Kutta integration with time step 0.03125 was used. The T_0 is 2003.5, i.e. 1st July 2003.

3 Results and discussion

Final aging chain molecule PopulCZ72 contains more than five hundred variables and similar number of equations. The stock and flow diagram (Figure 3) was simplified by vector definition of the variables.

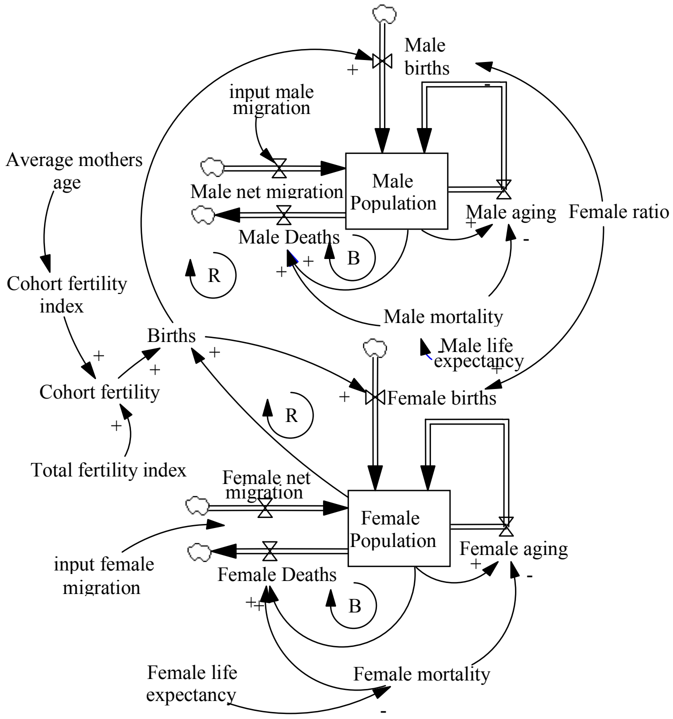


Figure 3 Stock and flow diagram for the disaggregated aging chain

The simulation run was tested against the surveyed data from 2003 to 2010. The mean absolute percentage error MAPE for total population is less than 0.08%, for male population the MAPE is 0.09%, for females its value is 0.06%. The MAPE for single age cohorts is always lower than 2.1%.

For comparison, the figure 4 shows extrapolations from Eurostat [9] and three extrapolations from the Czech statistical office [6]. The tested model does not focus on extrapolations; the exogenous variables are very simply estimated for years 2011 – 2050 by ordinary least square method and in such case it reaches the peak within this interval; the Δ for the next years was calculated by exponential decay with 10% rate of change to avoid the change of the slope sign of the estimated function.

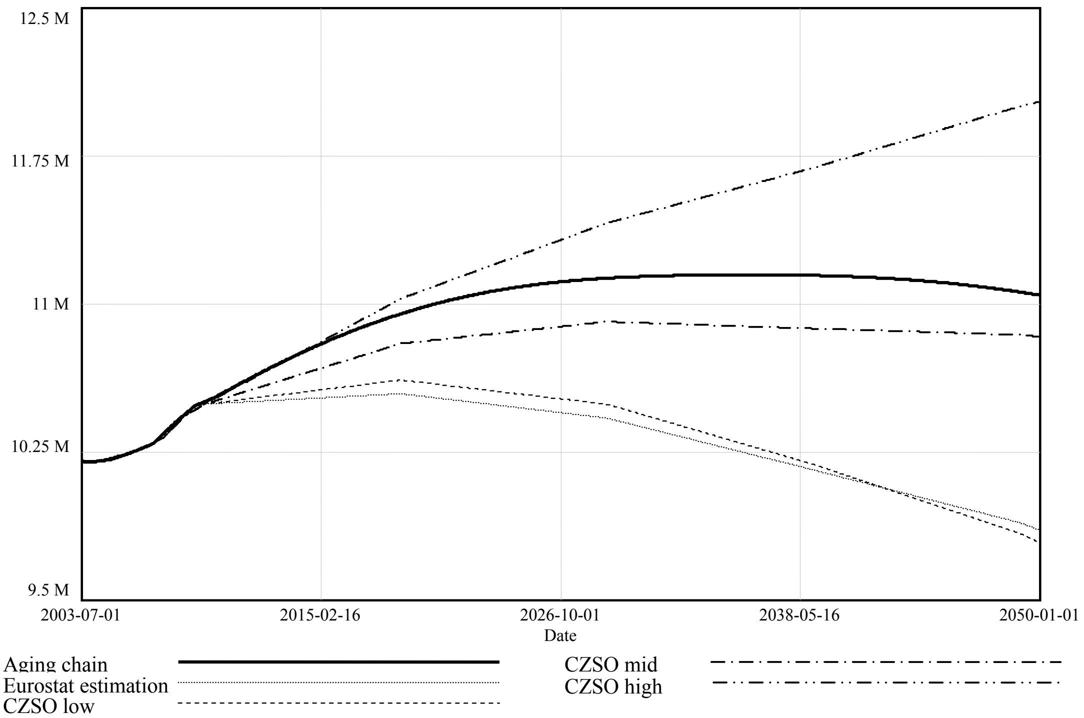


Figure 4 Estimations of Czech population in millions of people

To implement the molecule into more complex models, the exogenous variables work as an interface of such aging chain. Thus, the molecule should be connected by male and female life expectancy, average mother's age and total fertility. For some purposes the migration could be also included as an endogenous variable.

4 Conclusions

The presented aging chain PopulCZ72 should not be interpreted as a final model. Even though with more sophisticated extrapolation of the exogenous variable it is possible to use it for forecasting, the main objective was to prepare and also to provide the molecule for complex system dynamics models.

The model is highly disaggregated. This allows higher precision of calculations and enables to avoid high simplification and aggregation in any model whose part is the aging chain of Czech population. Disaggregation also provides versatility of the aging chain, which was the main demand for that molecule. The model contains mothers' age shift, which is specific for Czech population.

In the implementation into complex system dynamics models the current exogenous variables act as the aging chain interface. In full system dynamics models these variables should be transformed into endogenous parts of feedback loops to obtain a causally closed model. Such transformation is recommended to be done in the multiplication form of equations. Summation equations should be avoided for relationships modeling to prevent model behavior that violate requirement for model robustness [13].

References

[1] Bossel, H.: *System ZOO 1 Simulation models – Elementary systems, Physics, Engineering*. Books on Demand, Norderstedt, 2007.

[2] Bossel, H.: *System ZOO 3 Simulation models – Economy, Society, Development*. Books on Demand, Norderstedt; 2007.

[3] Czech Statistical Office: *Czech demographic handbook 2009*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4032-10 [25 Apr 2011]

[4] Czech Statistical Office: *Demographic Yearbook of the Czech Republic 2009*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4019-10 [25 Apr 2011]

[5] Czech Statistical Office: *Life Tables for the Czech Republic, Areas and Regions*, [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2010edicniplan.nsf/engp/4002-10 [26 Apr 2011]

[6] Czech Statistical Office: *Projekce obyvatelstva české republiky do roku 2065* [Czech population projection up to 2065], [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/2009edicniplan.nsf/p/4020-09 [15 May 2011]

[7] Czech Statistical Office: *Úmrtnostní tabulky* [Life Tables], [online] Czech Statistical Office, Prague, 2010, Available: http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky [23 Apr 2011]

[8] Eskinasi, M., Rouwette, E., Vennix, J.: Simulating urban transformation in Haaglanden, the Netherlands. *System Dynamics Review*, vol. 25, no. 3, 2009, 182 – 206.

[9] Giannakouris, K.: *Ageing characterises the demographic perspectives of the European societies*, [online] Eurostat, Luxembourg; 2008, Available: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-08-072/EN/KS-SF-08-072-EN.PDF [cit 10. May 2011].

[10] Kunc, M.: Achieving a balanced organizational structure in professional services firms: some lessons from a modelling project. *System Dynamics Review*, vol. 24 no. 2, 2008, 119 – 143.

[11] Meadows, D. et al.: *The Limits to Growth*, Universe Books, New York. 1972.

[12] Meadows, D. H., Randers, J., Meadows, D.: *The Limits to Growth: The 30-year update*. Chelsea Green Publishing Company, Vermont, 2004.

[13] Sterman, J. D.: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill, Boston, 2000.

[14] Wang, Q., Ma, S.: Some of the limits to growth of big city in China. In: *Proceedings of The 5th International Conference of the System Dynamics Society*. System Dynamics Society, Shanghai, 1987, 615–628.

[15] Wang, Q., Sterman, J. (1985): A disaggregate population model of China. *Simulation*, vol. 45 no. 1, 1985, 7 – 14.

Gross Fixed Capital Dynamics

Igor Krejčí[1], Jaroslav Švasta[2]

Abstract. The consumption and the gross and net stocks of fixed capital are important indicators in the system of national accounts. According to the current international standards these indicators are estimated by Perpetual Inventory Method (PIM). Processes of retirement, decay and obsolescence of fixed assets in national economy are modeled. The required indicators are annually computed on the basis of surveyed and estimated gross fixed capital formation (GFCF), price indexes and also determined by the main model parameters (the average service life of assets, the type of mortality function, depreciation pattern).

The paper deals with the modelling of gross capital stock as a part system dynamics model of fixed capital. This means change from discrete time model to continuous alternative. The principles of system dynamics modelling and national accounts based on former econometric research are compared. The advantages of both models are presented.

This comparison is necessary for full model testing. The system dynamics model allows extensions of the model that lead to more precise estimations. But the implementation of extensions biases comparison of dynamic model with non-improved original data. That's why the continuous alternative of PIM is essential.

Keywords: fixed capital, national accounts, system dynamics, simulation, retirement pattern

JEL Classification: C44, C63, E22
AMS Classification: 91B82, 93C15

1 Introduction

The consumption and the gross and net stocks of fixed capital are important indicators in the system of national accounts. The System of National Accounts 1993 [13] provides the systemic approach to stocks and flow indicators in national economy [8]. According to the current international standards [9] the consumption and stocks of fixed capital are recommended to be calculated by Perpetual Inventory Method (PIM).

Such calculation must be done because the business accounting does not provide useful information about value of fixed capital stock [3]. The indicators from business accounting are in historical prices (i.e. business accounts sums prices from different periods of time) and also the capital is depreciated on basis of convention and law and usually does not correspond with real economic service lives of the assets. Furthermore, it is nearly impossible to evaluate existing assets in national economy otherwise than by model calculation.

PIM is based on accumulation of previous investments (Gross Fixed Capital Formation, GFCF) and modeling of retirement, decay, moral obsolescence etc. Thus, the required indicators are annually computed on the basis of surveyed and estimated GFCF, price indexes and also determined by the main model parameters (the average service life of assets, the type of mortality function, depreciation pattern) [9], [11].

The paper deals with the modelling of gross capital stock as a part of system dynamics model of fixed capital. Its aim lies in the model structure design. From the beginning of system dynamics, different kinds of fixed assets are simulated in system dynamics models. E.g. the aging and retirement of dwellings (one of the specified types of fixed assets [13]) and urban development planning is quite common [4], [5] or [14]. Also the system dynamics models of whole investment process use some kind of retirement pattern [6], the retirement is modelled by distributed lag function or in terms of system dynamics they are modelled by different kinds of delays [6].

The transformation of the PIM model into system dynamics model is necessary for full model testing. The system dynamics model allows extensions of the model that lead to more precise estimations. But the implementation of extensions biases comparison (i.e. quality check) of dynamic model with non-improved

[1] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, krejcii@pef.czu.cz.
[2] Czech University of Life Sciences Prague, Department of Systems of Engineering, Kamýcká 126, Prague 6, svasta@pef.czu.cz.

original data. That's why the system dynamics alternative of PIM is essential for further development of the system dynamics model of fixed capital with investment function as structure of endogenous variables.

2 Materials and methods

The original PIM used at Czech statistical office uses the lognormal function for retirement pattern [1]. The indicators of fixed capital are calculated in four-dimensional classification (institutional sector, industry, kind of asset and time). Approach used at Czech statistical office is fully described in [11] and can be completely converted into Markov chain model [7].

For this paper the real data from Czech statistical office are used. The example is computed for transport equipment in agriculture industry in institutional sub-sector of public nonfinancial institutions. For this equipment the average service life is 16 with standard deviation equal to 6. For practical purposes the maximal service life is set on 99.5 quantile, which means 38 years for that kind of assets.

The transformation was made in few steps. At first, the original model was transformed into system dynamics model that provides similar results. Then, the quarter year indexes for GFCF were applied on yearly GFCF. It is impossible to use quarterly GFCF data because these do not contain lot of necessary adjustments [1], therefore quarterly data are used only as indexes. Next, the continuous retirement and continuous service life decrease were tested. The model was also compared to common system dynamics retirement modeling [12].

For the aging and retirement, the aging chain modelling structure was used [12]. In this structure, each capital of age i is represented by individual stock variable s_i and the aging a_i is flow variable between these stocks. Another flow is the retirement of the assets r_i. Part of assets is retired p_i, remainder ($1-p_i$) ages to higher stock variable after one year (fixed delay function based on FIFO). This retirement fraction is function of age of the stock i and kind of retirement pattern. Figure 1 and equations (1) – (4) represent general model structure. Simple arrows stand for causal links, tube arrows are flows, and stock variables are denoted by boxes.

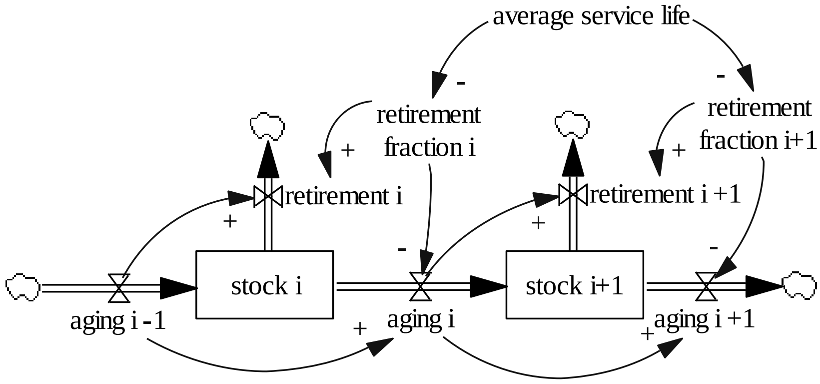


Figure 1 General structure of fixed capital model

$$s_i(T) = \int_{T_0}^{T} [a_{i-1}(t) - a_i(t) - r_i(t)] dt \tag{1}$$

$$a_i(t) = (1 - p_i) a_{i-1}(t-1) \tag{2}$$

$$r_i(t) = (p_i) a_{i-1}(t-1) \tag{3}$$

$$p_i = f(i) \tag{4}$$

Where T is current time, T_0 is initial time and t is any time between T and T_0.

Vensim DSS 5.10e was used for the design of stock and flow diagram and calculations. Since this software does not contain tabled values of the log-normal distribution function and model contain the dynamical change of the average service life, retirement function is computed on basis of approximation of error function proposed in [15]. Such approximation (equation (5)) gives less than 2% error.

$$\text{erf}(x) = \frac{x}{|x|}\sqrt{1-\exp\left(x^2\,\frac{\frac{4}{\pi}+ax^2}{1+ax^2}\right)}, \quad a = \frac{8(\pi-3)}{3\pi(4-\pi)} \tag{5}$$

The used time step was 0.03125 of year and 4th order Runge-Kutta method was used for numerical solving of differential equations.

3 Results and discussion

The testing model starts in the beginning of year 1952 and ends at the end of the 2008. For the comparison only published data from 1995 to 2008 were used, long time series of GFCF are necessary for correct stock accumulation. Stock and flow diagram of the model is in figure 2. This diagram contains only one stock variable because the chained variables were defined as vectors. Whole model contain more than 200 variables and similar number of equations.

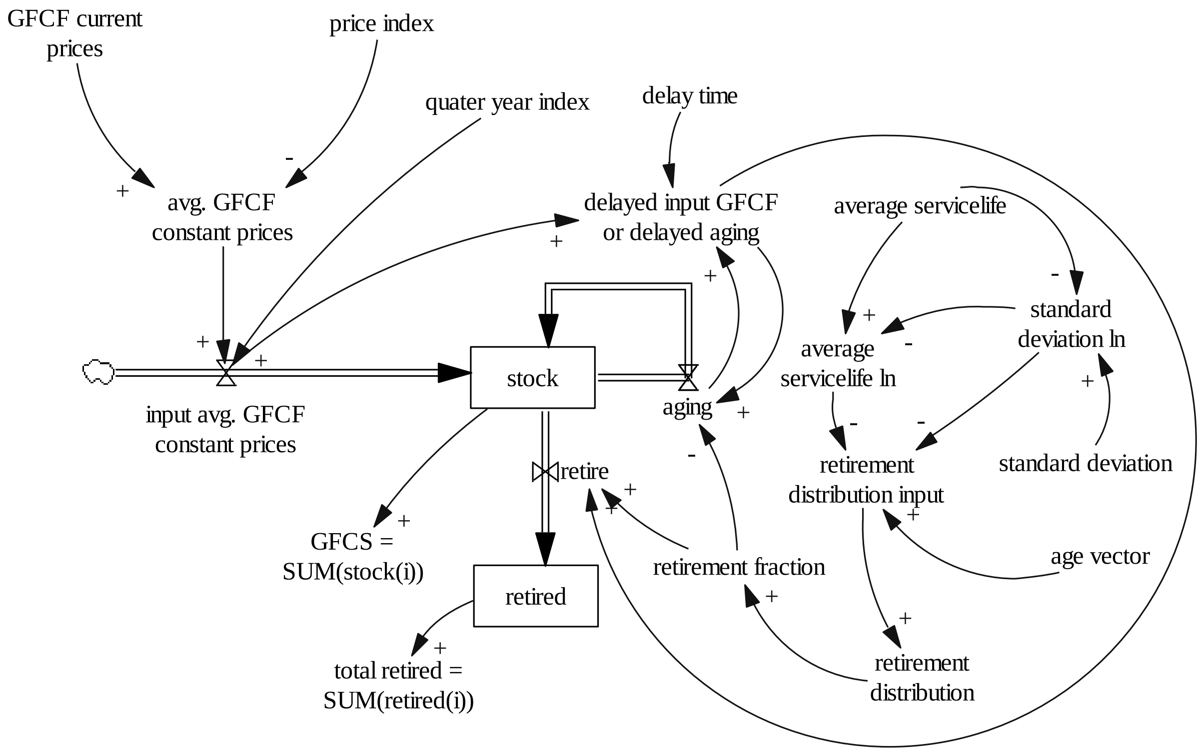


Figure 2 Simplified stock and flow diagram of gross fixed capital in constant prices

The original data were compared to the model with other extensions. At first the quarter-year indexes were added. The implementation of the indexes causes small differences in total GFCF but the relative difference is smaller than 0.003%. Furthermore, the decreasing average service life was tested. Such continuous decreasing is applied in some countries for chosen assets with long service lives [9]. In another run the retirement was changed from fixed delay to first order delay that generates exponential decay [2].

Furthermore, the model was compared to common system dynamics approach using the *n*-order delay that generates distributed lag with Erlang distribution function; this can be compared with German PIM version where Gamma distribution function is used as retirement pattern [10]. The *n* order delay is constructed from *n* chained first order delays; the equation (6) from [2] was used for order calculation. Result of the equation (6) is 7.11 for the data on transport equipment, thus the order that must be integer was defined as 7 and 8.

$$n = \frac{\bar{x}^2}{s^2} \tag{6}$$

Figure number 3 shows dynamics of gross fixed capital stock during the period 1995 – 2008. It is necessary to realize the date 200X represents closing stock of 200(X-1) and opening stock of 200X. 1% rate of change for decreasing average service life is only as illustrative example for general model structure. For transport equipment such rate was not observed but still the periodic actualization of service lives is recommended [9]. Addition of quarterly indexes for GFCF has only small impact (maximum difference in studied period was 0.05% in 2001). Common system dynamics approach gives 3.14% lower result for 7-order delay and 4.15% lower for 8-order result in comparison with original PIM based on surveyed data for service life and retirement pattern. The biggest difference to original PIM is for 8-order delay with decreasing service life, the gross fixed capital stock at the end of year 2008 was 9.58% smaller than original PIM.

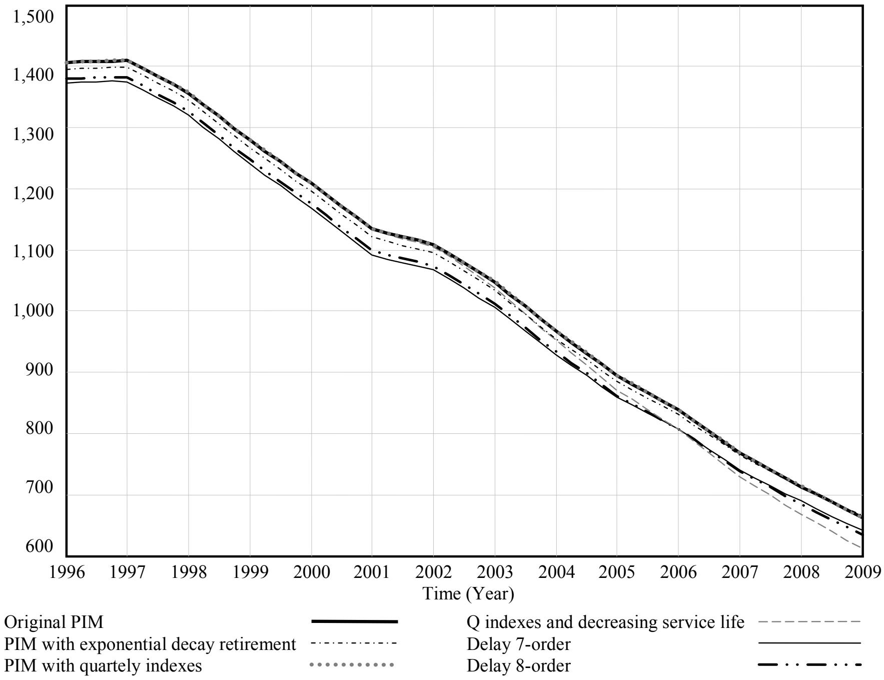


Figure 3 Alternative calculation of Gross Fixed Capital Stock in millions of crowns

4 Conclusions

The model uses real surveyed and estimated data from Czech statistical office. The proposed structure is tested on data on transport equipment in non-profit public organisation in agriculture industry, but for different industries, sectors or types of assets the structure remains, only the parameters changes.

Use of system dynamics allows dynamization of parameters and simple construction of model structure by copying its basic molecules. Addition of other characteristics like quarterly indexes or service life decreasing causes relatively small changes, but the presented calculations are basis for calculation of consumption of fixed capital, which is part of the GDP of nonmarket producers (government, non-profit organisations, households) [1]. Thus, one percent change means billions of crowns. Obviously decreasing capital stock is connected with increasing capital consumption and increasing GDP.

In comparison with basic system dynamics modelling structures the proposed structure has advantage that is the access to the age structure. The structure is useful for evaluation of so called other changes like e.g. disasters [13] that correspond with the modelling purposes. Furthermore not only gross stock is necessary to calculate, for net stock and consumption of fixed capital the age structure is necessary.

Use of combination of fixed delay for aging and exponential decay for retirement does not fit to the proposed model. The model is focused on precision. The exponential decay provides fully continuous behaviour but some assets remain in one year category for more than one year, in first year more than 60% are retired and it takes three years to retire 95% of the one year stock. According to PIM purposes where stocks are used for balance construction to specified date [7], [11] it is useful to disaggregate the one year stock to quarter year stock but with the fixed delay retirement.

The design of the whole PIM model is the prerequisite of wider system dynamics model concerning fixed capital. Fixed capital formation (investment function) will be used as endogenous variable and policies testing will be possible. The first step – PIM transformation is necessary for model quality check and it is its main objective. But it is also possible to use transformed model with exogenous variables for more precise calculations of the balances of fixed capital.

Acknowledgements

The paper is supported by the grant project IGA FEM CULS 201011160003 "Dynamics of fixed capital".

References

[1] Czech statistical offiece: *Gross national income inventory*. [online] 2002 Available: http://apl.czso.cz/nufile/GNI_CZ_en.pdf [11 May 2011].
[2] Hamilton, M. S. Estimating Lengths and Orders of Delays in System Dynamics Models, In: *Proceedings of The 1976 International Conference on System Dynamics*, The System Dynamics Society, Geilo, 1976, 593-628.
[3] Hulten, Ch., R., Wykoff, F., C.: Issues in the measurement of economic depreciation introductory remarks, *Economic Inquiry*, vol. 34, 1995, 10-23.
[4] Eskinasi, M., Rouwette, E., Vennix, J.: Simulating urban transformation in Haaglanden, the Netherlands. *System Dynamics Review*, vol. 25, no. 3, 2010, 182 – 206.
[5] Forrester, J.: W.: *Urban dynamics*, Pegasus Communications, Waltham, 1969.
[6] Forrester, N. B.: The role of econometric techniques in dynamic modeling: systematic bias in the estimation of stock adjustment models, *System Dynamics Review*, vol. 3, no. 1, 1987, 45–67.
[7] Krejčí, I: Computation of Stocks and Consumption of Fixed Capital with use of Markov Chain. In: *Proceedings of the 28th International Conference on Mathematical methods in Economics,* University of south Bohemia in České Budějovice, České Budějovice, 2010, 365–370.
[8] Moulton, B., R.: The system of national accounts for the new economy: what should change?, *Review of Income and Wealth*, vol. 50, no. 2, 2004, 261–278.
[9] OECD: *Measuring Capital - OECD Manual 2009*: *Second edition*, OECD Publishing, 2009.
[10] Schmalwasser, O., Schidlowski, M.: Kapitalstockrechnung in Deutschland. *Wirtschaft und Statistik* vol. 11, Statistisches Bundesamt. 2006, 1107–1123.
[11] Sixta, J.: Odhady spotřeby fixního kapitálu. *Statistika*, vol. 87, no. 2, 2007, 156 -163.
[12] Sterman, J. D.: *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill, Boston, 2000.
[13] Eurostat, United Nations, International Monetary Fund, OECD, World Bank: *System of National Accounts 1993*, New York. 1993.
[14] Wang, Q., Ma, S.: Some of the limits to growth of big city in China. In: *Proceedings of The 5th International Conference of the System Dynamics Society*. System Dynamics Society, Shanghai, 1987, 615–628.
[15] Winitzki, S.: Uniform approximations for transcendental functions, In: *Proceedings International Conference on Computational Science and Its Applications 2003*, Lecture Notes in Computer Science 2667,Montreal, 2003, 780–789.

Economical Aspects of the Vehicle Scheduling Optimization

Michal Krempl[1]

Abstract. The paper deals with the vehicle scheduling problem related to regional public transport. Linear programming methods are used to solve the problem. A mathematical model is created including the constraints and the objective function minimizing costs and the number of vehicles. A minimum costs and a number of vehicles is forced at the same time by special economical input data analysis and an allocation of costs. Determining of the costs coefficients is done by 3 methods, which differs primarily by how much of the total costs they take into account. The decomposition of the set of lines into disjoint subsets can be used instead of the "direct" optimization. The decomposition has proven to be a suitable alternative in solving large optimization problems. The problem was applied to optimize vehicle scheduling in the region, which is situated in the north-east of the Czech Republic. There is used Xpress - IVE software, which solve the problem by simplex algorithm and branch and bound method.

Keywords: transport, optimization, vehicle scheduling, linear mathematical modeling.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Grant demands of the public transport have increased substantially in the Czech and Slovak Republic in the last 20 years. It's an interest of the state to reduce subsidy for the public transport. The main goal for carriers to the future should be preparation and intensive searching for the savings in their management. Significant savings have to be achieved simply by changing an organization of work. It is important to use linear programming methods (LP) for solving the optimization problem of the vehicle scheduling. What objectives can be achieved using these methods? Which opportunities are opening for carriers by solving the problem? See the SWOT analysis on fig. 1.

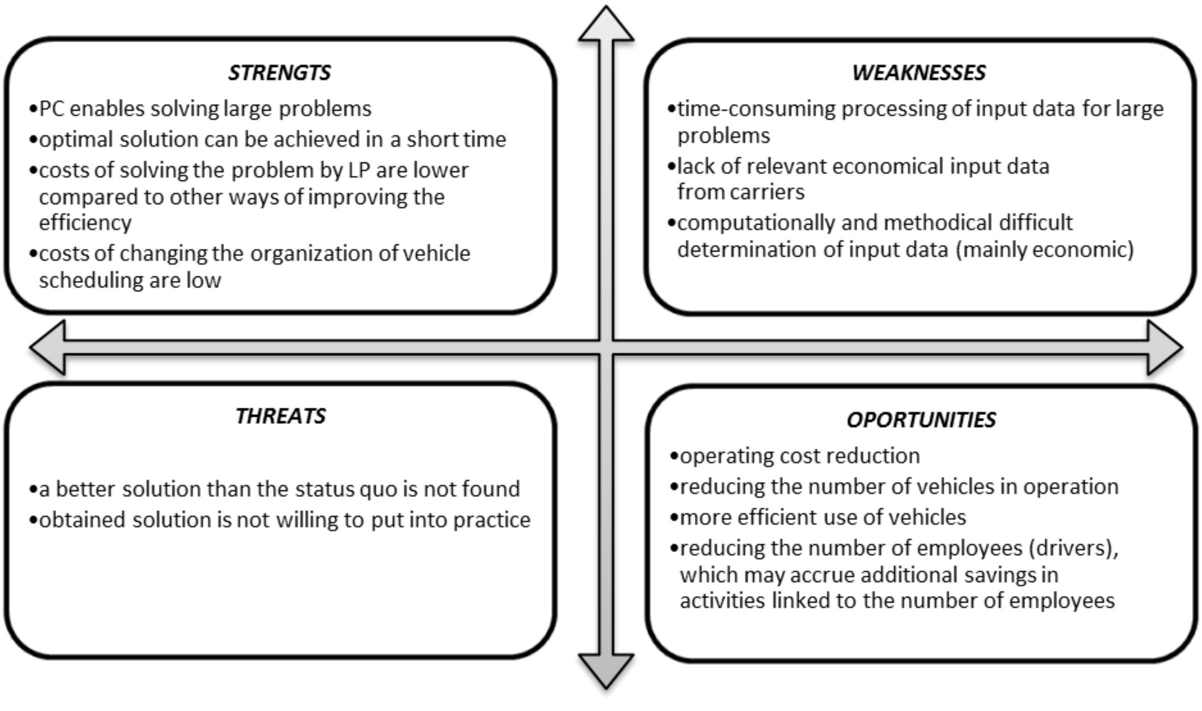


Fig. 1: The SWOT analysis of solving the vehicle scheduling problem by linear programming methods.

[1] VŠB - Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, Ostrava-Poruba, 708 33, Czech Republic, michal.krempl.st@vsb.cz

2 Basic reasoning

The main goal in the task of the optimizing the vehicle scheduling is proposing such a sequence of every single transport link (i.e. train, airplane, bus connection etc.) operations for each vehicle, in order to achieve the minimum or the maximum value of an optimization criteria. It seems most appropriate to choose as an optimization criterion one of two: the total costs of operation and the number of vehicles. However, the minimum would be necessary to search all at once. Using a multi-criterion optimization function has many disadvantages. That's why it is appropriate to use the following solution. We choose the transport link operation costs as an optimization criterion. The depreciation (i.e. the cost of the vehicle) for one planning period includes costs of the vehicle crossing from the garage to the first stop of the first circuit which operates the vehicle that day. If the vehicle throughout the planning period does not serve any link, it is not necessary to put the vehicle into the fleet carrier. Then the objective function is not burdened by a depreciation expense. It is ensured that the given set of transport links will be operated at the lowest costs. It is also expected that the number of deployed vehicles will be minimal.

3 The Mathematical Model

First, it is necessary to define what is about to decide in the vehicle scheduling problem:

- z_{ijk} … bivalent variable modeling the decision of whether the vehicle $i \in I$ may pass between the end stop of the transport links $j \in J$ a $k \in K$.

The variable z_{ijk} is introduced only in those cases for which the introduction is necessary:

- $z_{ijk} \in \{0, 1\}$ for $i \in I$, $j \in J$ a $k \in K$.

The definition of the sets can be elaborate as follows:

- I … the set of vehicles, which can be deployed on the operation of the transport link,
- J … the set of the transport links j after which, you can change the default stop of transport link k to the time of departure of the link k,
- K … the set of the transport links k to which it is possible to drive after finishing of operating the link j in the time of departure of the link k.
- S ... the set of the garages.

The objective function minimizing the costs takes the form (1):

$$\min f(z) = \sum_{i \in I} \sum_{j \in J \cup S} \sum_{k \in K \cup S} a_{ijk} \cdot z_{ijk} \qquad (1)$$

- a_{ijk} … total costs of the vehicle operation of the transport link j plus the costs of the crossing between the links j and k plus the costs of vehicle's waiting between the operation of the link j and k.

It is necessary to specify all the requirements on the mathematical model:

- deploying the required type of the vehicle on the link (1a),
- avoid the "return of the vehicle at the time" (1b),
- the correct sequence of the technological activities (ensured by the constraint 2),
- the operating each transport link (3),
- every vehicle exits the garages only once during the planning period (4),
- limiting the length of the shifts and the hours of operation of the vehicle (5).

The simplicity of the model is ensured by the introduction of the variable z_{ijk} only in cases when the declaration is necessary. This declaration also meets some of the above requirements: (1a), (1b). The "return at the time" means a situation where the vehicle after finishing operating link in a given time shall be deployed on operation of the other link with the start time earlier.

The correct sequence of technological operations (vehicle crossing from the garage to stop of the link, operating link, vehicle crossing between the final stop of links, crossing between the final stop of the vehicle and garages) can ensure (2):

$$\sum_{k \in K \cup S} z_{ijk} = \sum_{k \in J \cup S} z_{ikj} \quad for\ \forall i \in I \tag{2}$$

The operating of each link guarantees (3):

$$\sum_{i \in I} \sum_{k \in K} z_{ijk} = 1 \quad for\ \forall j \in J \tag{3}$$

The constraint (4) ensure that each vehicle exits from the garages at most once:

$$\sum_{k \in K} z_{ijk} \leq 1 \quad for\ \forall i \in I,\ j \in S \tag{4}$$

It is necessary to define a variable t_{ijk}. It represents the duration time of the transport link j plus the transit time from the end of the link j to the starting point of link k plus waiting to start operating link k.

The constraint (5) ensure that the time of the operation of the vehicle i will not exceed T_i. If we assume that the driver is over the whole operation "assigned" to the same vehicle, the constraint analogous limits the length of the driver’s shifts to T_i.

$$\sum_{j \in J} \sum_{k \in K} t_{ijk} \cdot z_{ijk} \leq T_i \quad for\ \forall i \in I \tag{5}$$

The constraints (5), (6) define the domain of variables:

$$z_{ijk} \in \{0,1\} \quad for\ \forall i, j, k \tag{6}$$

$$t_{ijk} \in Z_0^+ \quad for\ \forall i, j, k \tag{7}$$

4 Decomposition of the Set of Transport Links

The number of the transport links is a major factor affecting the computational time and the time of processing the input data. The decomposition set of links to n disjoint subsets can be also used to achieve optimum solution. The components of the set of links are given by the time interval. The link is inserted into the subset of links based on time of departure from the default stop. The problem is solved separately for each set of components. Separated links sequences operated by the vehicle are suitably connected to achieve the minimum total costs. This procedure does not guarantee an optimal solution, but it can be assumed that the obtained solution is not far from the optimum.

5 The Economical Input Data Analysis

It is necessary to accurately determine the value of all inputs appearing in the mathematical model in order to achieve valid results with good predictive value. It is difficult to obtain relevant information about the management of private carriers. That is the reason why it is appropriate to calculate the costs of driving and parking of vehicles in several ways, based on different and independent data. The results are compared in tab. no. 4. Among other may be used following procedures:

I. The determination of the relative costs coefficients.
II. The determination of the costs coefficients of the costing formula.
III. The determination of the costs coefficients according to the Ostrava Transport Company (DPO).

We need to find the minimum costs and the minimum number of vehicles at the same time. It is done the following way. The depreciation (i.e. the cost of the vehicle) for one planning period is included into costs of the vehicle crossing from the garage to the first stop of the first circuit which operates the vehicle that day.

Procedure no. I. allocates the costs into several groups as you can see in the fig. 2 and tab. 3.

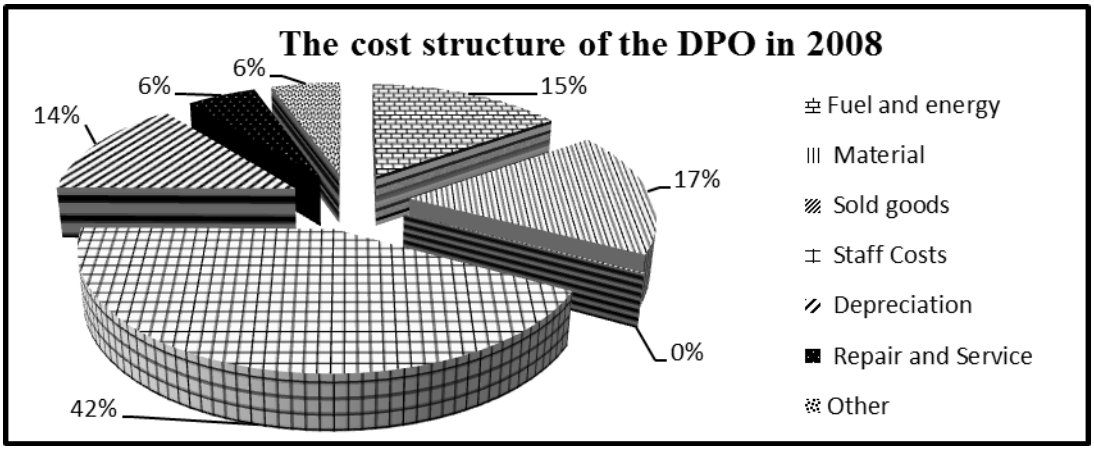


Fig. 2: The cost structure of the DPO in 2008.

The type of driving the vehicle	Percentage of the total costs
Driving and crossing the vehicle by driver	80,61%
Parking of the vehicle with driver	47,91%
Parking of the vehicle without driver	5,64%
Vehicle ownership (depreciation + repair, service)	19,39%

Tab. 3: The share of the total costs for each type of vehicle driving and parking.

Procedure no. II. is based on the calculation formula for the road transport. This formula divides the cost items in dependent and independent. To the first group belong:

- direct materials (fuel and oil, tires and other direct material)
- direct wages (drivers wage, health and social insurance).

To the second group belong:

- depreciation,
- the other direct costs (repairs and maintenance, insurance, road tax, fees, other direct costs). Results of the calculations are listed in the tab. 4.

Procedure no. III. is based on the information, which carriers are willing to tell. It is costs in $CZK \cdot km^{-1}$ and average speed of vehicles in $km \cdot h^{-1}$. According to the DPO the costs are:

- bus (standard construction) 1184 $CZK \cdot h^{-1}$,
- “long” bus 1408 $CZK \cdot h^{-1}$.

Results done by three mentioned methods are listed in the tab. 4. The methods of calculation differ primarily by how much of the total costs they take into account (see comments).

The computational method	The costs of the driving vehicle	The costs of the parking vehicle	Comments
The determination of the relative costs coefficients	1,0000 $[-]$	0,6755 $[-]$	The calculation takes into account the 100% of the costs of the carrier.
The determination of the costs coefficients of the costing formula	614,75 $[CZK \cdot h^{-1}]$ *(8 years depreciation)*	251,59 $[CZK \cdot h^{-1}]$ *(8 years depr.)*	The calculation takes into account of the costs referred in the costing formula.
	666,12 $[CZK \cdot h^{-1}]$ *(5 years depreciation)*	302,96 $[CZK \cdot h^{-1}]$ *(5 years depr.)*	
The determination of the costs coefficients according to the DPO	1184/1408 $[CZK \cdot h^{-1}]$ *(standard/long bus)*	-	The calculation takes into account all of the costs arising when the vehicle is in motion.

Tab. 4. The summary of the costs associated with the operate links according to different methods of the calculation.

It is necessary to use only the values referred in the costing formula with a depreciation of vehicles 5 years, because the optimization of the vehicle scheduling can only affects the operating costs. The costs are therefore considered 666 $CZK \cdot h^{-1}$ for a vehicle in motion and at the 303 $CZK \cdot h^{-1}$ for the parking vehicle with a driver.

6 Application into the Practice

The proposed attitude to the problem was applied to optimize vehicle scheduling in the region, which is defined in the south by the towns of Ostrava, Hlučín, Dolní Benešov, Kravaře and in the north by the border between the Czech Republic and Poland. This is the link, no 70, 72, 281, 282 and 283. All of them are finished at the stop called Přívoz, Muglinovská in the urban district of Ostrava-Přívoz. There are used Xpress - IVE software, which solve the problem by simplex algorithm and branch and bound method. Results are listed in the tab. 5.

	The current state	The optimum solution	The index of the original/ optimal solution	The solution obtained by the decomposition	The index of the original/ decomposition
The total calculated costs [CZK]	65 526	62 937	0,960	63 268	0,966
The costs (without the operating links and the depreciation) [CZK]	15 739	13 150	0,836	13 481	0,857
The total time of the vehicles being outside the garage [hours]	131	113	0,863	114	0,870
The number of vehicles [-]	8	8	1,000	8	1,000
The number of drivers [-]	14	13	0,929	13	0,929
The number of crossings between the links [-]	0	5	-	6	-

Tab. 5. The comparison of the current state with decomposition and with optimal solution.

The optimization was performed in a relatively small group of links. The implementation of the new vehicle scheduling would bring significant costs reductions in amount of 16,4% for the optimal solution and in amount of 14,3% for the decomposition solution. The number of drivers has decreased from 14 to 13 because the total time of the vehicles being outside the garage has also decreased by about 13%.

If the minor timetable changes are accepted, further substantial savings could be achieved by reducing the number of vehicles and the number of drivers.

7 Conclusion

The decomposition of the set of links has proven to be the suitable alternative in solving large optimizing problems. It affects the value of the objective function by only 0.6%. I believe that using linear programming techniques with powerful computers is the way to the rational organization of public passenger transport, which will leads to its improvement and subsequent revival.

References

[1] Výroční zpráva 2008: *Dopravní podnik Ostrava, a.s., Ostrava*. In . Ostrava : DPO, a.s., 2009 [2009-12-14]. From WWW: <http://dpo.cz/spolecnost/v_zpravy/2008.pdf>.

[2] SUROVEC, P. *Provoz a ekonomika silniční dopravy I.* Ostrava: VŠB-TU Ostrava, 2000. 119 p. ISBN 80-7078-735-X.

[3] *Xpress-Mosel Reference manual*: *Release 2.4*. *Dash Optimization*. 2008, 1, p. 1-382.

[4] *Xpress-Mosel User guide: Release 2.4*. *Dash Optimization*. 2008, 1, p. 1-162.

Mathematical Methods in Economics 2011

29th International Conference

Mathematical Methods in Economics 2011

Proceedings – part II

We celebrate 20th Anniversary of the Czech and Slovak Operational Research Societies

September 6 – 9, 2011

Janská Dolina, Slovakia

The conference is organised by:

Faculty of Informatics and Statistics, University of Economics, Prague

Faculty of Economic Informatics, University of Economics in Bratislava

Czech Society of Operational Research

Slovak Society of Operational Research

Czech Econometric Society

Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011 – part II
Publisher: Professional Publishing
Mikulova 1572/13, 149 00 Praha 4, Czech Republic
Editors: Martin Dlouhý, Veronika Skočdopolová
Pages: 791

ISBN 978-80-7431-059-1

The publication has not passed the language correction.

Programme Committee

Helena Brožová
Faculty of Economics and Management, University of Life Science, Prague

Michal Fendek
Faculty of Economic Informatics, University of Economics in Bratislava

Petr Fiala
Faculty of Informatics and Statistics, University of Economics, Prague

Jana Hančlová
Faculty of Economics, Technical University of Ostrava

Zlatica Ivaničová
Faculty of Economic Informatics, University of Economics in Bratislava

Josef Jablonský
Faculty of Informatics and Statistics, University of Economics, Prague

Ladislav Lukáš
Faculty of Economics, University of West Bohemia, Pilsen

Vladimír Mlynarovič
Faculty of Social and Economic Sciences, Comenius University in Bratislava

Jan Pelikán
Faculty of Informatics and Statistics, University of Economics, Prague

Karel Sladký
Institute of Information Theory and Automation of the Academy of Sciences of CR, Prague

Tomáš Šubrt
Faculty of Economics and Management, University of Life Science, Prague

Milorad Vidović
Faculty of Transport and Traffic Engineering, Uiversity of Belgrade, Serbia

Karel Zimmermann
Faculty of Mathematics and Physics Charles University, Prague

Organization Committee

Martin Dlouhý, Jan Fábry Jana Kalčevová, Martina Kuncová

Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic

Ivan Brezina, Juraj Pekár, Kvetoslava Surmanová, Marian Reiff

Faculty of Economic Informatics, University of Economics in Bratislava, Slovakia

Contents – Part II

International equity risk modeling by NIG model

Aleš Kresta[1], Tomáš Tichý[2]

Abstract. Financial risk modeling and its subsequent management is a very important and no less challenging task of quantitative units of financial institutions. Due to the nature of complex portfolios and based on recent evolution at financial markets, contemporary research is focused either on tails modeling or dependency modeling or both. The main task of this paper is to examine a potential contribution of Lévy based subordinated NIG model as a tool to estimate the risk of positions in foreign equities. In order to model the joint evolution of equities and FX rates, Student and Gaussian copula functions are assumed, ie. marginal distributions in terms of NIG model are joined together. We examine several horizons to estimate the model parameters and evaluate the efficiency of risk measure by backtesting procedure. We also study the effect of particular positions, as well as the impact of single positions on the quality of risk estimation.

Keywords: Market risk, equity, FX rate, subordinated model, backtesting

1 Introduction

Financial institutions play an inherent role within the economic system, since their existence allow an efficient transfer of funds, liquidity, maturity, and also risks. The efficiency is given by the fact that the markets (or the world, in general) are not ideal or perfect – doing business is more simple with huge amount of funds, information are not available for free, financial instruments are difficult to understand. However, financial institutions have at their disposal a huge amount of funds and high-skilled staff. Thus, the operations can be done quickly and effectively.

Although the well-managed financial institutions help to improve the conditions of the economic system, their faults can obviously result into the contrary. To prevent the failures of financial institutions and to increase the confidence in the financial system, some sort of regulation and subsequent supervision is necessary. One of its most important parts within the banking industry is to specify which model is qualified to measure the risk exposure. In this paper we concentrate ourselves at market risk models and their ability to estimate the risk of a portfolio exposure soundly. (For further discussion on risk management of financial institutions see eg. Hull (2010) or Resti and Sironi (2007).)

The market risk is commonly linked to a prespecified quantile of a probability distribution of a financial institution's portfolio. The most often used measure, VaR (Value at Risk), informs us what is the lowest return with a given confidence $(1-\alpha)$ or, alternatively, that with a given significance α the loss will be equal to or higher than VaR. This measure of risk was introduced in JP Morgan as a simplifying, but a unique proxy to a collection of many diverse risk measures of complex portfolios. Since many criticism to the way the VaR was defined arose shortly after its introduction to the public (see eg. Artzner et al., 1999), it has been suggested to measure also the conditional mean for a given probability, ie. what is the average loss when things go wrong. Such measures are denoted by various authors as cVaR (conditional Value at Risk), ES (expected shortfall) or tail VaR.

Recently, there have been published several papers dealing with the analysis of risk models via backtesting (preferably according to Kupiec's test). While eg. Alexander and Sheedy (2008) assumed Gaussian/Student/GARCH/Empirical models for a simple position and Rank (2007) analyzed similar multi-position models joined by several copula functions, in Tichý (2010a,b) the performance of ordinary

[1]Technical University Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, e-mail: ales.kresta@vsb.cz.

[2]Technical University Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, e-mail: tomas.tichy@vsb.cz.

elliptical copula Lévy driven models for FX rate sensitive portfolio (ie. allowing skewness and kurtosis in the marginal distribution and potentially non-linearity, but still symmetry in the dependency) with/without respect to various time span for parameter estimation was analyzed. Similar models were also applied by Kresta et al. (2010) for risk estimation of international portfolio and tested later in Kresta (2010).

In this paper we extend such studies by analyzing the contribution of single positions and zero-value positions within an international equity portfolio assuming again the Lévy type NIG model for marginal distributions. We proceed as follows. In the next section, the most important findings on VaR calculation via NIG model and/or its backtesting are briefly reviewed. Next in Section 3 we describe the data and then the backtesting results are provided (Section 4).

2 Risk measure estimation and model evaluation

Assuming a random variable X following a Gaussian distribution,[1] VaR over a time length Δt with a significance α (i.e. with confidence $1-\alpha$) can be obtained as follows:

$$VaR_X(\Delta t, \alpha) = -F_X^{-1}(\alpha) = -\mu_X(\Delta t) - \sigma_X(\Delta t)F_{\mathcal{N}}^{-1}(\alpha). \tag{1}$$

Here, $-F_X^{-1}(\alpha)$ denotes the inverse function to the distribution function[2] of random variable X for α (similarly, $F_{\mathcal{N}}$ is used for standard normal (Gaussian) distribution). It follows, that:

$$Pr\left(X < -VaR_X(\Delta t, \alpha)\right) = \alpha. \tag{2}$$

However, it is a rare case that random variable, eg. a return of financial asset, follows Gaussian distribution. Usually, we have to select a distribution with some additional parameters, so that we can control even higher moments of the distribution. In that case, it can be inevitable to run a Monte Carlo simulation procedure to obtain VaR as an estimate to $-F_X^{-1}(\alpha)$. Several useful models belongs to the family of Lévy processes or subordinated Lévy processes in particular (see eg. Cont and Tankov (2004).

Let us define a stochastic process $\mathcal{Z}(t;\mu,\sigma)$, which is a Wiener process. As long as $\mu = 0$ and $\sigma = 1$ its increment within infinitesimal time length dt can be expressed as:

$$dZ = \varepsilon\sqrt{dt}, \quad \varepsilon \in \mathcal{N}[0,1], \tag{3}$$

where $\mathcal{N}[0,1]$ denotes Gaussian distribution with zero mean and unit variance. Then, a subordinated Lévy model can be defined as a Brownian motion[3] driven by another Lévy process $\ell(t)$ with unit mean and positive variance κ. The only restriction for such a driving process is that it is non-decreasing on a given interval and has bounded variation. Hence, we replace standard time t in

$$X(t;\mu,\sigma) = \mu dt + \sigma\mathcal{Z}(t) \tag{4}$$

by its function $\ell(t)$:

$$X(\ell(t);\theta,\vartheta) = \theta\ell(t) + \vartheta Z(\ell(t)) = \theta\ell(t) + \vartheta\varepsilon\sqrt{\ell(t)}. \tag{5}$$

Due to its simplicity (tempered stable subordinators with known density function in the closed form), the most suitable models seem to be either the variance gamma model – the overall process is driven by gamma process from gamma distribution with shape a and scale b depending solely on variance κ, $G[a,b]$, or normal inverse Gaussian model – the subordinator is defined by inverse Gaussian model based on inverse Gaussian distribution, $IG[a,b]$. In this paper, we will apply the latter. Therefore, the risk of a single position can be estimated by evaluating the log-return model:

$$X(I(t;\kappa);\theta,\vartheta) = \mu t + \theta(I(t)-t) + \vartheta Z(I(t)) = \mu t + \theta(I(t)-t) + \vartheta\varepsilon\sqrt{I(t)}, \tag{6}$$

where μ is average return (long-term drift), which is fitted by deducing θt from the original model, and I is an inverse Gaussian process independent of ε.

[1] Due to its simplifying nature and overwhelming usage, the Gaussian distribution is also denoted as *normal* distribution.
[2] To simplify the notation, $F_X^{-1}(\alpha)$ is the inverse to the distribution function, if it exists, or at least its generalization $F_X^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}$.
[3] For our purposes a Brownian motion is a Wiener process without any premise on μ and σ.

Since VaR is often calculated for portfolios, the models for marginal distribution of asset returns as the one in (6) must be joined. A very useful way is to utilize copula functions. Assuming for simplicity two potentially dependent random variables with marginal distribution functions F_X, F_Y and joint distribution function $F_{X,Y}$, we get, according to Sklar's theorem (see Nelsen, 2006 for more details):

$$F_{X,Y}(x, y) = \mathcal{C}(F_X(x), F_Y(y)). \tag{7}$$

Thus, being equipped by copula function and marginal distribution function, the joined evolution can be modeled easily.

The ability of (market) risk models to estimate the risk exposure soundly is commonly assessed by the so called backtesting procedure. Within the backtesting procedure on a given time series $\{1, 2, \ldots, T\}$, two situations can arise – the loss L is higher than its estimation or lower (from the stochastic point of view, the equality shouldn't arise). While the former case is denoted by 1 as an exception, the latter one is denoted by zero:

$$I_X(t+1, \alpha) = \begin{cases} 1 & \text{if} \quad L_X(t, t+1) > VaR_X(t, t+1; \alpha) \\ 0 & \text{if} \quad L_X(t, t+1) \leq VaR_X(t, t+1; \alpha). \end{cases} \tag{8}$$

On the sequence $\{I_X(t+1, \alpha)\}_{t=1+m}^{T-1}$, where m is a number of data (days) needed for the initial estimation, it can be tested whether the number of *ones* (exceptions) corresponds with the assumption, ie. $\alpha \times n$ (where $n = T - 1 - m$), whether the estimation is valid either unconditionally or conditionally, whether bunching is present, etc. Generally the most simple way is to compare the true number of exceptions to the assumption about them (Kupiec, 1995). A one step further is to evaluate the one-step dependency of exceptions in line with Christofferson (1998). The review of some further techniques can be found eg. in Berkowitz et al. (2010).

3 Data description

The data set we consider in this study comprises of daily closing prices of four well established equity indices – Down Jones Industrial Average (DJI) from the US market, FTSE 100 (FTSE) from London (UK), Nikkei 225 (N225) from Tokyo (Japan) and Swiss Market Index (SMI) from Switzerland – over preceding 20 years (January 1, 1990 to December 31, 2009). Since the trading days at particular markets are not always harmonized, we had to interpolate missing data. In this way we obtained four time series of 4,939 log-returns. Basic descriptive statistics of daily log-returns are apparent from Table 1.

Table 1: Basic descriptive statistics of equity indices (daily log-returns)

Index	Min	Max	Mean	Median	St.dev.	Skewness	Kurtosis
DJI	−8.201%	10.508%	0.029%	0.045%	1.100%	−0.060	12.042
FTSE	−9.265%	9.384%	0.019%	0.042%	1.136%	−0.103	9.983
N225	−12.111%	10.086%	−0.017%	0.004%	1.479%	−0.248	7.894
SMI	−8.383%	10.788%	0.032%	0.082%	1.177%	−0.127	9.420

However, the indices are denominated in four distinct currencies, in particular the US dollar (USD), British pound (GBP), Japan yen (JPY) and Swiss franc (CHF). This fact extends our data set to 8 distinct time series. As a reference currency Czech koruna (CZK) is chosen – it implies that there is no riskless investment opportunity. However, it can be assumed that investments into European equities will imply lower FX rate risk than investment to equities in Japan or USA due to tighter links of economies (see Table 2).

Table 2: Basic descriptive statistics of FX rates (daily log-returns)

Index	Min	Max	Mean	Median	St.dev.	Skewness	Kurtosis
USD	−4.818%	8.208%	−0.008%	−0.023%	0.694%	0.449	10.160
GBP	−4.937%	7.922%	−0.012%	−0.005%	0.602%	0.249	14.535
JPY	−5.253%	7.995%	−0.001%	−0.034%	0.819%	0.464	8.552
CHF	−3.063%	7.313%	−0.004%	−0.013%	0.519%	0.790	14.791

4 Results

In order to examine the power of the suggested model to estimate the risk, marginals defined in terms of NIG model coupled together by ordinary symmetric copula functions, we will assume (i) Brownian motion and/or NIG model for marginal distribution of single position in equity indices and/or FX rates from the point of view of CZK; (ii) zero value position of two assets, ie. all combinations of long and short positions, $(1, -1)$ and $(-1, 1)$, under Gaussian or Student copula function.[4]

The power of the model to estimate the risk will be assessed by backtesting procedure applied to VaR on daily basis calculated for several significance levels: $\alpha = 0.0003, 0.005, 0.01, 0.05, 0.15, 0.5$. While the former can be connected to internal capital management and target rating (say, AA), the others are implied by Basel II / Solvency II. Finally, $\alpha = 0.5$ allows us to check the estimation for the median.

In order to estimate the parameters of the model, various time spans, τ, will be assumed with $\tau_{max} = 2,000$ days, ie. approximately 8 years of daily data. Since the data were collected over 20 years, we can apply the backtesting procedure on a rolling basis for almost 12 years, $N = 2,939$ log-returns in particular. The power of the model will be assessed by comparing the assumed number of exceptions and related observations (recall, that exception is given by $I_{L>VaR(\alpha)}$). For particular α's we therefore get the following numbers of assumed exceptions: $0.89, 14.7, 29.39, 146.95, 440.85, 1469.5$.

4.1 Single position backtesting

We will first evaluate particular models for single positions, four FX rates from the Czech point of view (in CZK), four indices in local currencies and also the indices after recalculating into CZK; however, the dependency of equity indices and FX rates will be ignored. Therefore, we are interested only in the marginal distribution.

The standard way to risk estimation is to suppose that the (log)returns of financial quantities follow Gaussian distribution, a symmetric probability distribution with two parameters. In this case, both approaches to the parameter estimation, the method of moments and maximal likelihood method, provide equivalent result. A common approach is to use a one year window to parameter estimation. However, quite high kurtosis of daily log-returns suggests that Gaussian distribution is not a good candidate to model the marginals. Indeed, setting α to 0.005 or 0.001 the observed number of exception is about two or three times higher than the assumption and assuming $\alpha = 0.0003$, it is even 10times higher.

Obviously, there are important differences in results among particular assets. One can assume, that Gaussian distribution might work better for indices than currencies, since higher kurtosis was documented for the latter. However, we should take into account also the skewness – the risk is measured for the left tail, but the FX rate returns are significantly right-skewed. Therefore, the best results are obtained for JPY with quite high (and positive) skewness and low kurtosis. Among the equity indices, the best results of Gaussian distribution are obtained for DJI, probably through insignificant empirical skewness. The Gaussian distribution might be accepted as a valid model for risk estimation of single positions only for $\alpha = 0.05$. Next, also the median, $\alpha = 0.5$, is fitted well – the error is 1% or 2%, ie. about 20 observations. By contrast, VaR at the significance level of 15% is overestimated – the observed number of exceptions is lower by 15% – 20%. We have also tried to increase or decrease the time span, but we have not observed any significant impact. It is therefore clear, that Gaussian distribution should not be used to model the risk of marginal positions.

Considering the NIG model, the situation is more challenging since there exist many ways how to define and estimate the model parameters. In theory, the maximal likelihood method (NIG_{mlm}) should be preferred over the methods of moments (NIG_{mm}). Indeed, when we compared the results for a given τ, NIG_{mlm} works better. Although the results are better comparing with Gaussian distribution, there are still apparent significant deficiencies, especially for lower τ. The reason is that the parameters of the distribution are of different memory. While the second moment, the standard deviation, describes rather a short term variability, the fourth moment is related to the heaviness of the tails, ie. the observations of rare events – and rare events are by definition rare, ie. can be observed only in a long horizon. We have therefore decided to separate the time span to estimate the mean and variance from the time span to estimate the kurtosis, while leaving the skewness unsolved. Unfortunately, this approach significantly complicates the application of NIG_{mlm}. This is the reason why we continue further only with NIG_{mm}.

[4]Due to the lack of space, only the most important results will be reported in tables; the rest is only briefly commented.

In Table 3 we review the results obtained by application of NIG model via method of moments assuming the time span in days for (mean, variance, skewness, kurtosis) as follows: $(60, 60, 60, 2000)$, $(250, 250, 250, 2000)$, and $(250, 250, 2000, 2000)$. We can observe that both time spans for skewness, $\tau = 250$ and $\tau = 2000$, give us similar results, although the former is slightly better. Time span of 60 days works well only for the median; however, the differences among all approaches are not very significant.

Table 3: Number of exceptions over 1994-2009 (NIG_{mm})

Signific.level	0.0003	0.005	0.01	0.05	0.15	0.50
Assumption	0.89	14.70	29.39	146.95	440.85	1469.50
DJI	1/1/2	14/11/17	29/29/32	197/188/187	564/503/502	1481/1458/1476
FTSE	0/0/0	23/22/21	42/37/37	185/162/166	521/479/478	1475/1450/1446
N225	5/3/3	24/20/23	46/33/36	178/147/157	509/461/467	1486/1483/1458
SMI	3/1/1	16/18/22	37/35/35	193/157/158	558/498/500	1491/1445/1445
USD	0/3/0	12/14/9	24/33/34	190/189/200	595/572/566	1463/1490/1437
GBP	6/1/1	14/8/8	28/23/26	217/182/201	605/590/583	1476/1485/1452
JPY	0/1/0	14/14/10	27/25/22	180/167/178	556/525/518	1473/1468/1452
CHF	2/1/1	9/13/14	23/23/27	206/187/197	655/617/601	1454/1473/1439
DJI&USD	1/1/1	18/19/20	35/35/34	172/164/168	527/475/477	1495/1473/1466
FTSE&GBP	0/1/1	21/12/17	32/32/32	192/172/173	499/470/472	1468/1439/1443
N225&JPY	4/1/1	19/14/19	30/28/32	142/132/141	451/414/416	1469/1486/1456
SMI&CHF	2/1/1	11/10/10	22/25/27	156/139/137	489/452/456	1469/1446/1444

When comparing the observed number of exceptions with the assumption, we can regard the NIG_{mm} model as a very good for all tail VaR levels. Similarly to the Gaussian assumption, also the NIG model provides better results for FX rates in the tails. By contrast, the risk is underestimated for $\alpha = 15\%$. Since we have reported no positive dependency among the FX rates and indices, it is also natural that the model works well for equity positions in CZK even when no dependency is captured by the model. More particularly, since in reality there is some negative dependency, the number of exceptions is lower.

4.2 Zero value position backtesting

After evaluating the single position models, we can proceed to simple dependency modeling – ie. only two positions are assumed, one long, the other short. This time, however, we use only one time span, $\tau = (250, 250, 250, 2000)$. Instead, we compare NIG model under Gaussian (Ga) and Student (St) copula with Brownian motion (BM) under the same copulas. Although, we have made clear that Gaussian distribution (ie. Brownian motion) is not a suitable candidate for risk modeling of single positions, it might be that using Student copula for dependency among them brings better results. Again, different time spans for the estimation of the kurtosis allow us to apply only the method of moments. By contrast, the parameters of the copula function are estimated on the basis of canonical maximum likelihood for $\tau = 250$, as well – we assume a short memory for the dependency again. The results in Table 4 clearly document that for modeling the risk of left tails the NIG-St model is the best one, while with $\alpha = 0.15$ the Gaussian copula should be preferred. In general, the difference between the assumed and observed number of exceptions is very small. There is also no significance effect of changing the positions, ie. replacing $(1; -1)$ by $(-1; 1)$, except replacing (DJI;SMI) by (SMI;DJI), which is probably implied by specific pattern of tail dependency.

5 Conclusion

The presence of jumps and unexpected decreases (increases) in price provide very challenging task on any risk model. A common approach to evaluate the ability of the model to estimate the risk soundly, is known as backtesting. In this paper we deepen the analysis of subordinated Lévy models joined by ordinary elliptical copulas by considering various time span for single positions and zero-value positions within a complex problem of international equity portfolio modeling.

Table 4: Number of exceptions for zero value portfolio (long&short) over 1994-2009 (in CZK)

Signific.level	0.0003	0.005	0.01	0.05	0.15	0.50
Assumption	0.89	14.70	29.39	146.95	440.85	1469.50
DJI&FTSE	7/4/3/2	40/24/17/12	53/46/38/33	139/143/139/155	382/410/416/459	1491/1491/1492/1491
DJI&N225	5/3/0/0	31/26/19/18	45/41/39/33	138/140/155/158	389/404/429/460	1493/1495/1480/1480
DJI&SMI	12/3/2/1	33/23/19/16	43/37/30/29	127/127/131/146	393/419/437/476	1478/1479/1475/1476
FTSE&DJI	6/4/0/0	26/19/17/16	39/36/33/30	141/143/150/153	424/430/450/465	1479/1478/1466/1467
N225&DJI	10/3/0/0	30/21/15/14	44/36/25/23	160/163/155/168	401/421/432/470	1460/1464/1461/1462
SMI&DJI	16/8/4/1	47/42/31/27	60/52/47/46	148/152/142/155	368/400/405/447	1446/1448/1447/1447
FTSE&N225	13/6/3/2	38/29/22/18	54/50/40/36	148/149/148/153	379/398/430/453	1446/1444/1459/1460
FTSE&SMI	8/5/4/2	34/27/18/15	48/43/33/28	148/148/143/150	394/409/435/459	1461/1461/1473/1472
N225&FTSE	9/5/2/2	31/29/21/18	44/42/33/30	155/158/152/163	387/405/435/451	1446/1445/1461/1459
SMI&FTSE	12/3/0/0	37/26/21/17	47/43/38/33	142/145/145/155	387/410/429/472	1461/1460/1464/1462
N225&SMI	8/2/1/0	26/15/11/10	37/30/24/23	140/141/141/147	376/404/413/459	1477/1475/1477/1476
SMI&N225	5/3/0/0	27/19/14/12	44/41/34/30	135/136/141/148	397/414/440/467	1493/1496/1478/1480

Results are provided as follows: BM-Ga/BM-St/NIG-Ga/NIG-St

Acknowledgement

The support provided by VSB-TU Ostrava under the SGS project SP2011/7 is kindly announced.

References

[1] Alexander, C., Sheedy, E.: Developing a stress testing framework based on market risk models. *Journal of Banking & Finance* **32**, 2220–2236, 2008.

[2] Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. *Mathematical Finance* **9**: 203-228, 1999.

[3] Berkowitz, J., Christoffersen, P.F., Pelletier, D.: Evaluating Value-at-Risk Models with Desk-Level Data, *Management Science*, 2010.

[4] Bertoin, J.: *Lévy Processes.* Cambridge: Cambridge University Press, 1998.

[5] Christoffersen, P.F.: Evaluating Interval Forecasts, *International Economic Rev.* **39**, 841–862, 1998.

[6] Cont, R., Tankov, P.: *Financial Modelling with Jump Processes.* Chapman & Hall/CRC press, 2004.

[7] Hull, J.C.: *Risk Management and Financial Institutions.* 2e. Upper Saddle Rv.: Prentice Hall, 2010.

[8] Kresta, A.: Portfolio Value at Risk estimation utilizing Lévy models coupled together by copula functions. In *Proceedings of 47th EWGFM meeting.* Ostrava: VSB-TU Ostrava, 2010.

[9] Kresta, A., Petrová, I., Tichý, T.: Innovations at financial markets: How to model higher moments of portfolio distribution. *Actual Problems of Economics* **2** (12): 59-71, 2010.

[10] Kupiec, P.: Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3**, 73–84, 1995.

[11] Nelsen, R. B.: *An Introduction to Copulas.* 2nd ed. Springer, 2006.

[12] Rank, J.: *Copulas. From theory to application in finance.* Risk books, 2007.

[13] Resti, A., Sironi, A.: *Risk Management and Shareholders' Value in Banking: From Risk Measurement Models to Capital Allocation Policies.* Chichester: Wiley, 2007.

[14] Tichý, T.: Posouzení vybraných možností zefektivnění simulace Monte Carlo při opčním oceňování. *Politická ekonomie* **56** (6), 772–794, 2008.

[15] Tichý, T.: Posouzení odhadu měnového rizika portfolia pomocí Lévyho modelu. *Politická ekonomie* **58** (4): 504–521, 2010.

[16] Tichý, T.: Market risk backtesting via Lévy models and parameter estimation. *In Proceedings of Mathematical Methods in Economics*: 634639. Ceske Budejovice: JCU, 2010.

Multifractal height cross-correlation analysis

Ladislav Krištoufek [1]

Abstract. We introduce a new method for detection of long-range cross-correlations and cross-multifractality – multifractal height cross-correlation analysis (MF-HXA). MF-HXA is a multivariate generalization of the height-height correlation analysis. We show that long-range cross-correlations can be caused by a mixture of the following – long-range dependence of separate processes and additional scaling of covariances between the processes. Similar separation applies for cross-multifractality – standard separation between distributional properties and correlations is enriched by division of correlations between auto-correlations and cross-correlations. We further apply the method on returns and volatility of NASDAQ and S&P500 indices as well as of Crude and Heating Oil futures and uncover some interesting results.

Keywords: multifractality, long-range dependence, cross-correlations

JEL classification: C19, C49
AMS classification: 91Gxx, 91C15

1 Introduction

The research of long-range dependence and multifractality in various time series has grown significantly during the last years, e.g. [8, 22, 7, 13]. An efficient detection of long-range dependence and estimation of Hurst exponent is crucial for financial analysts as its presence has important implications for a portfolio selection, an option pricing and a risk management. There are several methods for the long-range dependence detection, among the most popular are the rescaled range analysis [15], the modified rescaled range analysis [20], the rescaled variance analysis [12], the detrended fluctuation analysis [24] and the detrending moving average [1]. For the detection of multifractality, there are three popular methods – the multifractal detrended fluctuation analysis (MF-DFA) of [17], the generalized Hurst exponent approach (GHE) of [2, 9], which is based on the height-height correlation analysis of [3], and the wavelet transform modulus maxima (WTMM) of [23]. The precision of various methods has been discussed as well [6, 14, 27, 4, 18].

Recently, the examination of long-range cross-correlations has become of interest as it provides more information about the examined process. Podobnik et al. [25] generalized the detrended fluctuation analysis for two time series and introduced the detrended cross-correlation analysis (DCCA). Zhou [28] further generalized the method and introduced the multifractal detrended cross-correlation analysis (MF-DXA). In this paper, we introduce two new methods, which are a generalization of the height-height correlation analysis of [3] – the multifractal height cross-correlation analysis (MF-HXA) and its special case of the height cross-correlation analysis (HXA).

The paper is structured as follows. In Section 2, we briefly discuss the basic notions of long-range correlations and multifractality. Section 3 introduces the method of MF-HXA. In Section 4, we apply MF-HXA on daily returns and volatility of NASDAQ and S&P500 indices as well as of the Crude and Heating Oil. We show that these two pairs of processes posses very different properties. The long-rage correlations and cross-multifractality of the stock indices cannot be distinguished from two pairwise independent processes. On the other hand, the commodity pair shows very complex dynamics with long-range cross-correlations and cross-multifractality deviating from the pairwise independent behavior. Section 5 concludes.

[1] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod Vodárenskou věži 4, CZ-182 08, Prague, Czech Republic; Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, Opletalova 26, CZ-110 00, Prague, Czech Republic, e-mail: kristouf@utia.cas.cz

2 Long-range correlations and multifractality

In this section, we present basic notions of multifractality, long-range correlations and long-range cross-correlations. As the subject is widely discussed in the recent literature, we present only a brief description. For more detailed reviews, see [5, 16, 11].

A stationary process is long-range dependent if an autocorrelation function $\rho(k)$ of the said process decays as $\rho(k) \sim Ck^{2H-2}$ for lag $k \to \infty$ where parameter $0 < H < 1$ is Hurst exponent [15, 21].

A critical value of Hurst exponent is 0.5 and suggests two possible processes – either an independent process [5] or a short-term dependent process [19]. If $H > 0.5$, auto-covariances decay hyperbolically and are positive at all lags, the process is then called long-range dependent with positive correlations [11] or persistent [21]. On the other hand, if $H < 0.5$, auto-covariances again decay hyperbolically and are negative at all lags and the process is said to be long-range dependent with negative correlations [11] or anti-persistent [21]. The persistent process implies that a positive movement is statistically more likely to be followed by another positive movement or vice versa. On the other hand, the anti-persistent process implies that a positive movement is more statistically probable to be followed by a negative movement and vice versa [26].

If the process can be described by a single Hurst exponent H, it is called monofractal (or unifractal). If different Hurst exponents are needed for various scales, the process exhibits crossovers and is called a multiscaling process. Further, there can be different Hurst exponents for parts of the series, which is solved by a use of the time-dependent (or local) Hurst exponent [13]. The most complicated is the case when there is a whole spectrum of Hurst exponents which is needed for a full description of the process made up of many complex fractal processes [17].

Both long-range dependence and multifractality can be present in the relation between two separate series. The series may be long-range dependent but can also have a long memory of a different process so that it is pairwise long-range dependent with Hurst exponent H_{xy}. Cross-correlation function $\rho_{xy}(k)$ of processes x_t and y_t then decays hyperbolically as $\rho_{xy}(k) \sim Ck^{2H_{xy}-2}$. Similarly to the standard case, if the whole spectrum of cross-correlation Hurst exponents H_{xy} is needed for the description of the cross-correlations between two processes, the relation is cross-multifractal. Further features of the long-range cross-correlations and cross-multifractality are discussed in the following sections.

3 Multifractal height cross-correlation analysis

The detection of long-range dependence and the estimation of the generalized Hurst exponent $H(q)$ of [3] is based on the q-th order height-height correlation function of time series $X(t)$ with time resolution ν and $t = \nu, 2\nu, ..., \nu[\frac{T}{\nu}]$ (where [] is a lower integer operator). For better legibility, let us define a τ-order difference of function $X(t)$ as $\Delta_\tau X(t) = X(t+\tau) - X(t)$. Height-height correlation function of $X(t)$ is then defined as

$$K_q(\tau) = \frac{\nu}{T} \sum_{t=1}^{T/\nu} |\Delta_\tau X(t)|^q \tag{1}$$

where time interval τ generally ranges between $\nu = \tau_{min}, \dots, \tau_{max}$. $K_q(\tau)$ then scales as

$$K_q(\tau) \propto \tau^{qH(q)}. \tag{2}$$

We generalize the method presented above and introduce the multifractal height cross-correlation analysis (MF-HXA) which can be used for the detection of long-range correlations and multifractality between two separate time series.

Note that it makes sense to analyze the scaling according to Equation 2 only for a detrended series $X(t)$ and only for $q > 0$ [8]. A type of detrending can generally take various forms – polynomial, moving averages and other filtering methods[1]. Such a detrending must take place for each time resolution ν separately. For $q = 2$, Equation 1 then describes scaling of variance with changing time scale τ.

[1]Simple linear trend is most frequently used detrending method, see e.g. [9, 10, 8]. Nevertheless, many times the detrending is not taken into consideration at all, see e.g. [2]

We generalize the method presented above and introduce the multifractal height cross-correlation analysis (MF-HXA) which can be used for the detection of long-range correlations and multifractality between two separate time series. Let us consider two series $X(t)$ and $Y(t)$ of length T with time resolution ν and $t = \nu, 2\nu, ..., \nu[\frac{T}{\nu}]$. Generalizing Equation 1 for two time series, we obtain

$$K_{xy,q}(\tau) = \frac{\nu}{T}\sum_{t=1}^{T/\nu} |\Delta_\tau X(t)\Delta_\tau Y(t)|^{\frac{q}{2}} \tag{3}$$

where again time interval τ generally ranges between $\nu = \tau_{min}, \ldots, \tau_{max}$. For $q = 1$, the generalized height correlation function represents a scaling of the absolute deviations of the covariates; and for $q = 2$, it corresponds to the cross-covatiance function. We propose the multifractal height cross-correlation analysis (MF-HXA) based on the generalization of Equation 2. Scaling relationship between $K_{xy,q}(\tau)$ and the generalized cross-correlation Hurst exponent $H_{xy}(q)$ is obtained as

$$K_{xy,q}(\tau) \propto \tau^{qH_{xy}(q)}. \tag{4}$$

We again consider $q > 0$ and detrended series $X(t)$ and $Y(t)$ for all ν used. For $q = 2$, the method can be used for the detection of long-range cross-correlations solely and we call it the height cross-correlation analysis (HXA). Obviously, for $X(t) = Y(t)$ for $t = 1, 2, \ldots, T$, MF-HXA turns into the generalized Hurst exponent approach of [2], which is equivalent to the height-height correlation analysis of [3].

4 Application

To show the usefulness of MF-HXA, we apply the method on two different types of financial assets – the US stock indices (NASDAQ and S&P500) and the commodity prices (Crude and Heating Oil). We choose such pairs as we expect strong correlations and therefore potential long-range cross-dependence and multifractality. Indeed, NASDAQ and S&P500 are highly correlated (the correlation coefficient for logarithmic returns is $\rho_{xy}(0) = 0.8652$) as well as the pair of the oil commodities (with $\rho_{xy}(0) = 0.6476$). For the stock index prices, we analyze a period between 1.1.2000 and 31.12.2009 (2531 observations). For the commodity prices, we use a nearest-future basis prices based on the Commodity Research Bureau for a period between 1.11.1993 and 16.2.2010 (4115 observations). Note that all the series contain some extreme events as well as strong trends and reversals. For the stock indices, the examined period contains a long-term decreasing trend in the beginning of the 2000s as well as the current financial crisis. For the oil prices, the series contains several strong speculative trends, to name the most evident one – the bubble that started in 2007, peaked in July 2008 (when the price of both oil commodities almost tripled in 18 months) and reversed into a strong bullish market (returning below the levels of the beginning of 2007).

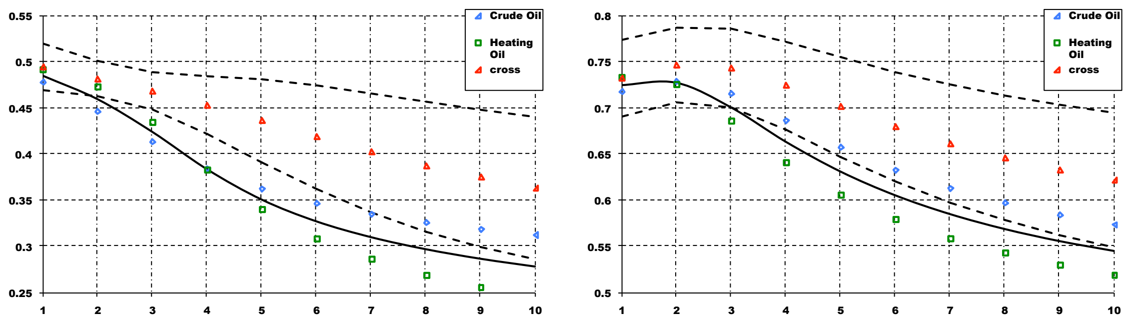


Figure 1: Estimates of $H_{Crude}(q)$, $H_{Heat}(q)$ and $H_{xy}(q)$ (y-axis) for returns *(left)* and absolute returns *(right)* of Crude and Heating Oil for $q = 1, 2, \ldots, 10$ (x-axis).

We research on the potential long-range dependence and cross-correlations in returns and volatility. As a measure of volatility, we take the absolute returns, which is standard in the financial literature and also intuitive as returns can be taken as a product of a sign and a magnitude (absolute return). All the examined assets share similar properties – average return close to zero, negative skewness and excess kurtosis. Such a deviation from normality is supported by both Jarque-Bera and Shapiro-Wilk tests, which reject normality at all meaningful significance levels. Stationarity is supported by ADF and KPSS test – strong rejection of a unit root and inability to reject stationarity. With a use of a standard Q-statistic, we reject that both returns and volatility show no significant autocorrelations. Importantly, the autocorrelations are much stronger for volatility than for returns. Such results indicate potential long-range dependence in volatility of the examined series. All the previous results hold for all the examined series.

To examine possible long-range (cross-)dependence and (cross-)multifractality in both returns and volatility of the series, we apply MF-HXA with $\tau_{min} = 1$ and $\tau_{max} = 20$ for $q = 1, 2, \dots, 10$. Such values are standardly used as 20 days correspond to a trading month [8].

Figures 1 - 2 show the estimates of the generalized Hurst exponents for returns and volatility of NASDAQ, S&P500, Crude Oil, Heating Oil and corresponding joint processes. Moreover, we present the 99% confidence intervals for $H_{xy}(q)$.

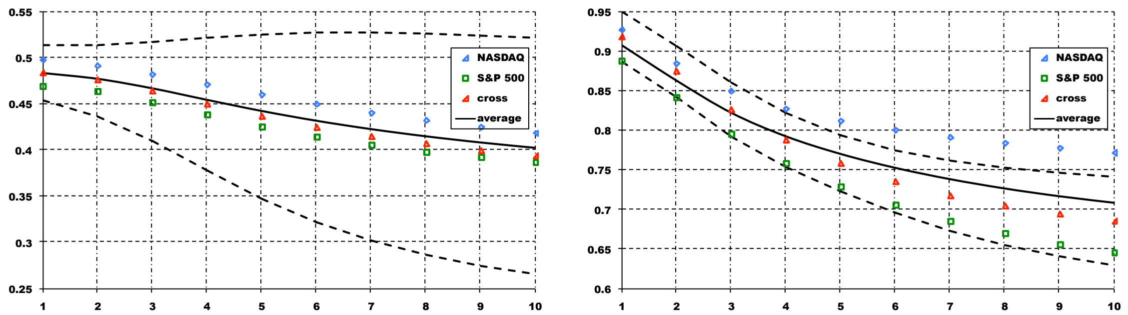


Figure 2: Estimates of $H_{NASD}(q)$, $H_{S\&P}(q)$ and $H_{xy}(q)$ (y-axis) for returns *(left)* and absolute returns *(right)* of NASDAQ and S&P500 for $q = 0.1, 0.2, \dots, 10$.

The generalized Hurst exponents for the returns vary for NASDAQ from $H_{NASD}(1) = 0.498$ to $H_{NASD}(10) = 0.418$ and for S&P500 from $H_{S\&P}(1) = 0.469$ to $H_{S\&P}(10) = 0.387$. The cross-correlated Hurst exponents vary from $H_{xy}(1) = 0.484$ to $H_{xy}(10) = 0.394$. Long-range dependence Hurst exponents $H(2)$ are estimated as $H_{NASD}(2) = 0.491$, $H_{S\&P}(2) = 0.463$ and $H_{xy}(2) = 0.476$. Therefore, there are no strong signs of persistence in returns for any of the series, rather very weak signs of anti-persistence . Moreover, there are no signs of additional scaling in covariances of the processes, i.e. there is no statistically significant deviation of cross-correlation Hurst exponent from the averages of $H_{xx}(q)$ and $H_{yy}(q)$ for all qs. Figure 2 illustrates this finding with 99% confidence intervals based on standard errors of $H_{xy}(q)$. Note that the confidence intervals get very wide for higher qs indicating that scaling law becomes less stable with increasing q.

For the commodity futures, the generalized Hurst exponents for the returns vary for Crude Oil from $H_{Crude}(1) = 0.478$ to $H_{Crude}(10) = 0.312$ and for Heating Oil from $H_{Heat}(1) = 0.492$ to $H_{Heat}(10) = 0.245$. The cross-correlated Hurst exponents vary from $H_{xy}(1) = 0.495$ to $H_{xy}(10) = 0.364$. Further, long-range dependence Hurst exponents are estimated as $H_{Crude}(2) = 0.446$, $H_{Heating}(2) = 0.473$ and $H_{xy}(2) = 0.482$. Such results show very weak signs of anti-persistence in both Heating Oil and Crude Oil series while the joint process shows only very weak signs of negative long-range dependence. Nevertheless, the cross-correlation Hurst exponent $H_{xy}(2)$ deviates from the average Hurst exponents of the separate processes indicating additional scaling in the covariances of the joint process. The deviation is even statistically significant on 99% significance level for $q = 3, 4, \dots, 7$. However, the deviation is very weak for $q = 2$ indicating no statistically significant long-range cross-correlations. Nevertheless, for the higher moments, there is strong additional scaling of covariances between the series.

We now turn to the analysis of the volatilities. For NASDAQ and S&P500, the results are quite similar – the generalized Hurst exponents vary with q with very high long-range dependence Hurst exponents $H(2)$ around 0.9. More specifically, the generalized Hurst exponents vary from $H_{NASD}(1) = 0.927$ to $H_{NASD}(10) = 0.771$ for NASDAQ and $H_{S\&P}(1) = 0.888$ to $H_{S\&P}(10) = 0.646$ for S&P500. The cross-correlation Hurst exponent $H_{xy}(q)$ does not deviate from the average of $H_{xx}(q)$ and $H_{yy}(q)$ for all qs.

The results are different for the volatilities of Crude and Heating Oil. Both commodities show strong long-range dependence with $H(2)$ around 0.75. The generalized Hurst exponents for the returns vary for Crude Oil from $H_{Crude}(1) = 0.718$ to $H_{Crude}(10) = 0.574$ and for Heating Oil from $H_{Heat}(1) = 0.733$ to $H_{Heat}(10) = 0.519$. For the cross-correlation Hurst exponent, the results are very similar to the ones of commodities' returns – $H_{xy}(q)$ deviates from $(H_{xx}(q) + H_{yy})/2$ for $q > 2$ at 99% confidence level. Thus, there is again an additional scaling in covariances for moments higher than 2. This also implies that for standard long-range correlations ($q = 2$), the relationship between the two series cannot be distinguished from uncorrelated series.

5 Conclusions

In the paper, we introduced a new method for the detection of long-range cross-correlations and cross-multifractality – the multifractal height cross-correlation analysis (MF-HXA). We showed that long-range cross-correlations can be caused by long-range dependence of separate processes and/or by additional dependence between the two series caused by scaling of the covariances. Similarly for cross-multifractality, the causes can be separated into three groups – multifractality due to the joint-distributional properties and due to correlations, which can be further divided into the auto-correlations and the cross-correlations.

To show the usefulness of the method, we applied MF-HXA on returns and volatility of NASDAQ, S&P500, Crude Oil and Heating Oil. We showed that the two pairs of series are characterized by very different behavior. Whereas the long-range cross-correlations and cross-multifractality between the stock indices cannot be distinguished from a pairwise independent processes, the relationship is more complex for the oil commodities where we find strong scaling in covariances of the processes.

Acknowledgements

The support from the Czech Science Foundation under Grants 402/09/H045 and 402/09/0965, the Grant Agency of Charles University (GAUK) under project 118310 and Ministry of Education MSMT 0021620841 are gratefully acknowledged.

References

[1] E. Alessio, A. Carbone, G. Castelli, and V. Frappietro. Second-order moving average and scaling of stochastic time series. *European Physical Journal B*, 27 (197), 2002.

[2] J. Alvarez-Ramirez, M. Cisneros, C. Ibarra-Valdez, and A. Soriano. Multifractal hurst analysis of crude oil prices. *Physica A*, 313:651–670, 2002.

[3] A.-L. Barabasi, P. Szepfalusy, and T. Vicsek. Multifractal spectre of multi-affine functions. *Physica A*, 178:17–28, 1991.

[4] J. Barunik and L. Kristoufek. On hurst exponent estimation under heavy-tailed distributions. *Physica A*, 389:3844–3855, 2010.

[5] J. Beran. *Statistics for Long-Memory Processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994.

[6] M. Couillard and M. Davison. A comment on measuring the hurst exponent of financial time series. *Physica A*, 348:404–418, 2005.

[7] L. Czarnecki, D. Grech, and G.: Pamula. Comparison study of global and local approaches describing critical phenomena on the polish stock exchange market. *Physica A*, 387, 2008.

[8] T. Di Matteo. Multi-scaling in finance. *Quantitative Finance*, 7(1):21–36, 2007.

[9] T. Di Matteo, T. Aste, and M. Dacorogna. Scaling behaviors in differently developed markets. *Physica A*, 324, 2003.

[10] T. Di Matteo, T. Aste, and M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, (29):827–851, 2005.

[11] P. Embrechts and M. Maejima. *Selfsimilar Processes*. Princeton University Press, 2002.

[12] L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssiere. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112, 2003.

[13] D. Grech and Z. Mazur. Can one make any crash prediction in finance using the local hurst exponent idea? *Physica A*, 336, 2004.

[14] D. Grech and Z. Mazur. Statistical properties of old and new techniques in detrended analysis of time series. *Acta Physica Polonica B*, 36(8), 2005.

[15] H.E. Hurst. Long term storage capacity of reservoirs. *Transactions of the American Society of Engineers*, 116, 1951.

[16] J. Kantelhardt. *Encyclopedia of Complexity and Systems Science*, chapter Fractal and Multifractal Time Series. Springer, 2009.

[17] J. Kantelhardt, S. Zschiegner, E. Koscielny-Bunde, A. Bunde, S. Havlin, and E. Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, 316(1-4), 2002.

[18] L. Kristoufek. Classical and modified rescaled range analysis: Sampling properties under heavy tails. *IES Working Papers*, 26, 2009.

[19] F. Lillo and J. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics and Econometrics*, 8(3), 2004.

[20] A. Lo. Long-term memory in stock market prices. *Econometrica*, 59(5), 1991.

[21] B. Mandelbrot and J. van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10(422), 1968.

[22] J. Matos, S. Gama, H. Ruskin, A. Sharkasi, and M. Crane. Time and scale hurst exponent analysis for financial markets. *Physica A*, 387(15), 2008.

[23] J. Muzy, E. Bacry, and A. Arneodo. Wavelets and multifractal formalism for singular signals : application to turbulence data. *Physical Review Letters*, 67:3515, 1991.

[24] C. Peng, S. Buldyrev, S. Havlin, M. Simons, H. Stanley, and A Goldberger. Mosaic organization of dna nucleotides. *Physical Review E*, 49(2), 1994.

[25] B. Podobnik and H.E. Stanley. Detrended cross-correlation analysis: A new method for analyzing two nonstationary time series. *Physical Review Letters*, 100:084102, 2008.

[26] N. Vandewalle, M. Ausloos, and Ph. Boveroux. Detrended fluctuation analysis of the foreign exchange market. In *Econophysic Workshop*, Budapest, Hungary, 1997.

[27] R. Weron. Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, 312, 2002.

[28] W.-X. Zhou. Multifractal detrended cross-correlation analysis for two nonstationary signals. *Physical Review E*, 77:066211, 2008.

Solution of the Time Limited Vehicle Routing Problem by Different Approximation Methods Depending on the Number of Necessary Vehicles

Petr Kučera[1]

Abstract. The time limited vehicle routing problem (TLVRP) stems from the vehicle routing problem. The main difference is that the routes are paths (not cycles), i.e. vehicles do not return (or we do not mind how they return) to the central city. Costs are given for the straight routes between each pair of the cities and represent the time necessary for going through. Each path must not exceed a given time limit. The sum of time for all routes is to be minimized. For the exact definition see [8].
This problem is NP-hard. One of the possibilities how to solve the TLVRP is to use heuristics (approximation methods), and thus to obtain a sufficiently good solution. In this paper we have chosen three of these approximation methods, test them on some different instances and asses the performance of single heuristics depending on the number of vehicles necessary for currying out the desired transportation and number of cities on single routes.

Keywords: time limited vehicle routing problem; vehicle routing problem; traveling salesman problem; heuristics (approximation method).

JEL Classification: C61
AMS Classification: 90B06

1 Introduction

The time limited vehicle routing problem (TLVRP) is defined as follows: One central city and other n (ordinary) cities are given and for each pair of the cities a cost is given, representing time necessary for going through the straight route between them. The cost matrix is supposed to be symmetric. The goal is to find a set of paths so that each of them has one of its endpoints in the central city, its length does not exceed a given time limit L and each city lies on exactly one of the paths except for the central city.

Let us introduce the following notation. The central city will be indexed by 0 and the other cities by numbers from 1 to n. The cost matrix will be denoted by **C** (and so single costs c_{ij}, $i, j = 0, \dots, n$). The decision variables x_{ij} are bivalent and indicate whether the straight route from the city i to city j is in the solution, the decision variables u_i mean the length of the route from city 0 to city i. M is a sufficiently large constant (in comparison to L and c_{ij}). The mathematical model of the TLVRP according to [8] is:

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min \qquad (1)$$

$$\sum_{j=1}^{n} x_{ij} = 1,\ i = 0,1,2,\dots,n \qquad (2)$$

$$\sum_{i=1}^{n} x_{ij} = 1,\ j = 0,1,2,\dots,n \qquad (3)$$

$$u_i + c_{ij} - M(1 - x_{ij}) \le u_j,\ i = 0,1,2,\dots,n,\ j = 1,2,\dots,n,\ i \ne j \qquad (4)$$

$$u_i \le L,\ i = 0,1,2,\dots,n \qquad (5)$$

[1] Czech University of Life Sciences, Department of Systems Engineering, Kamýcká 129, 165 21 Prague, Czech Republic, kucera@pef.czu.cz.

The objective function (1) represents the total length (time) of all the routes. Equations (2), resp. (3) assure that just one edge goes from, resp. to each city. Constrains (4) define the sense of variables u_i as mentioned above and constrains (5) assure that no route exceeds the time limit L.

This problem has many practical instances, e.g. the transportation of newspapers from a publishing house to shops, grocery products (dumplings etc.) from a manufactory to restaurants, daily reports from affiliated branches to headquarters (this example is mentioned in [8]) etc. Each vehicle is required to visit all the cities on its route until a given time, but we do not mind how it gets from the end back to the start of its route to realize it next time.

Nevertheless, the TLVRP has been studied relatively little. It belongs among the NP-hard problems, for which there is no efficient algorithm for finding their theoretical optimum. It stems from the vehicle routing problem (VRP), where routes are cycles instead of paths, and so from the traveling salesman problem (TSP), i.e. the task to construct one cyclic route containing all the cities, too. Thus, heuristics (approximation methods) for the TLVRP can be derived from the methods for the VRP and the TSP.

Generally, we can differentiate two basic types of heuristics: the former one constructs a solution (from the beginning) while the latter one improves an initial solution (which is either randomly given or obtained using another heuristics). For deriving methods for the TLVRP (or some of other related problems) from methods for the TSP, the former type is more suitable because the latter one, improving the solution, utilizes special properties of TSP solutions hardly modifiable for the solutions of different tasks. Perhaps the most famous heuristics improving the solution is by Lin and Kernighan [7].

Let us mention the following examples of heuristics for the TSP constructing a solution: The nearest neighbour method, tree approaches, nearest merger method and insertion methods were investigated by Rosenkrantz, Stearns, and Lewis [9]. Another one is the savings method by Clarke and Wright [2]. The most accurate method for the TSP is the Christofides method [1] based on the combination of the tree approach and the construction of a minimum matching which always find a solution at most 1.5-times longer than the optimum. Let us mention also the patching method by Karp [4] and the loss method by Webb [11] and Van der Cruyssen and Rijckaert [10]

In this paper we compare the performance of three heuristics on different types of cases. The first one is a modification of the nearest neighbour method as one of the simplest possible methods. Then follows an approach using so called Habr frequencies which are, on the contrary, one of the most sophisticated tools how to asses single routes between pairs of cities in different types of transportation tasks. The last one is a parallel version of the savings method which is based in a similar procedure to the previous one, but uses the savings which are not so thorough tool for assessing the single routes but their substance is closer to the VRPs. All these three methods are described in details below.

2 Nearest Neighbour Method

The algorithm of the nearest neighbour method (NNM) for the TSP looks as follows:

1. Choose an arbitrary city and join the closest city to it.
2. While there exist some cities not joined to the route, join the closest one of these to the last joined one.
3. Finally, join the last joined city to the one chosen at the start to close the cyclic route.

It is recommended, starting the algorithm, to try to choose all the cities and then select the shortest one of the obtained routes.

Such simple methods as the NNM often do not give good results. As regards this NNM version for the TSP, Rosenkrantz, Stearns, and Lewis [9] have shown that if we require any accuracy ratio for the solution, there exists an instance for which the NNM cannot achieve it.

The following modification for the TLVRP which we will use and test is taken from [8].

1. To start the construction of a new path, join the closest city from the ones, which have not yet been put on any route, to the central one.
2. Repeat the following procedure: join the closest city to the last joined one to the path, while upon joining this city the total length of the path does not exceed the time limit L. The algorithm terminates as soon as all the cities lie on some of the paths. Otherwise, go to step 1 to construct another path..

3 Habr Frequencies

Habr (author of e.g. [3], however, in Czech only) introduced so called frequencies, which compare an edge (a straight route between a pair of cities) with all others, even non-adjacent edges. He applied them in approximation methods for different transportation problems. He designed even several heuristics for the TSP using them.

Habr frequency for the edge is the value

$$F_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n}(c_{ij} + c_{kl} - c_{il} - c_{kj}) \tag{6}$$

This form obviously shows its sense. There exists another form, called the modified frequency, more suitable for computations: $F`_{ij} = c_{ij} - r_i - s_j$, where $r_i$ and $s_j$ are the arithmetic means of the costs of $i$-th row and $j$-th column of **C**, respectively. $F`_{ij}$ can be derived from F_{ij} by linear transformation

. Habr frequencies consider all the edges with the same importance. But in the case of TLVRP the edges from/to the central city are more important (more frequently and often used) than the others. Now let us show how big this difference is: Let us suppose that we will use p vehicles (routes) for the transportation (p can be estimated e.g. as $\lceil w/v \rceil$, where w is the sum of the capacities of all the cities and v is the vehicle capacity). Let us consider a randomly chosen (with an uniform probability distribution, without respect to the costs) solution with p vehicles used. The probability of the choice of an edge non-incident to the central city is $\frac{2(n-p)}{n(n-1)}$ while for the edges incident to the central city this probability is equal to p/n. So the edges incident to the central city are $\frac{p(n-1)}{2(n-p)}$-times more important than the others (they occur in a solution with $\frac{p(n-1)}{2(n-p)}$-times greater probability). Thus the frequencies for the TLVRP will be computed by the formula

$$F_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n}(c_{ij} + c_{kl} - c_{il} - c_{kj}) + \frac{p(n-1)}{2(n-p)}\sum_{m=1}^{n}(2c_{ij} + c_{m0} - c_{i0} - c_{mj} + c_{0m} - c_{im} - c_{0i}) \tag{7}$$

or the modified frequencies $F`_{ij}$ can be computed by a formula derived from (7) by an analogous linear transformation as in the general case from (6) above, which we will not specify here more precisely.

Now we can propose a a method based on the Habr frequencies. We will call it the Habr frequencies approach (HFA):

1. For all pairs of non-central cities (i, j) compute the frequencies using (7).
2. - Process the edges according to the ascending order of the frequencies using the following rules:
   - Upon adding the edge, if all the edges so far added to the solution form the set of vertex disjoint paths and for each path its length after joining the city 0 to the closer end of the path does not exceed the time limit L, then add it to the solution.
   - Repeat the procedure until each city lies on some of the paths and joining arbitrary two paths the allowed time limit is exceeded.
3. In the end add the city 0 to the closer end of all the paths.

4 Savings Method

The savings method (SM) is based on comparing lengths of a straight route between any two cities and a route via another selected city. We will use for testing its parallel version which is denoted as a non-limited savings method in [5] and performed best from all the SM modifications tested there:

1. For all pairs of non-central cities (i, j) compute the savings $s_{ij} = c_{i0} + c_{0j} - c_{ij}$.
2. - Process the edges according to the descending order of the savings s_{ij} using the following rules:
   - Upon adding the edge, if all the edges so far added to the solution form the set of vertex disjoint paths and for each path its length after joining the city 0 to the closer end of the path does not exceed the time limit L, then add it to the solution.
   - Repeat the procedure until each city lies on some of the paths and joining arbitrary two paths the allowed time limit is exceeded.
3. In the end add the city 0 to the closer end of all the paths.

5 Test computations and their results

First let us recall some results published in recent time. In [5] we tested different heuristics on randomly generated cases with 12 non-central cities where the cities were located in a circle with 100 time unit diameter and the time limit *L* set to 250 time units. In accordance with our expectation, the HFA was doing better than both the NNM and the SM, and, although we tested two different types of cases which mutually significantly differed in the location of the central city, there was in general each time the same difference between the results by HFA and both the remaining methods. The results are summarized in Table 1 in a percentage form, where 100 p.c. is the result of the HFA. Cases 1 to 10 are of the former type while cases 11 to 15 are of the latter type. Let us note that all there methods always found solutions with 3 vehicles except one case by NNM and one case by HFA (different at each method) where the number of the vehicles was 4. The number of cities on a vehicle route was maximally 6 except one case by NNM where the number of cities on one of the routes was 7.

	NNM	SM	HFA
Case 1	112,9%	100,0%	100,0%
Case 2	122,4%	100,0%	100,0%
Case 3	105,5%	102,3%	100,0%
Case 4	97,4%	111,7%	100,0%
Case 5	99,8%	105,9%	100,0%
Case 6	100,0%	109,8%	100,0%
Case 7	107,3%	100,0%	100,0%
Case 8	107,2%	100,0%	100,0%
Case 9	104,0%	103,7%	100,0%
Case 10	107,1%	102,5%	100,0%
Case 11	97,6%	106,5%	100,0%
Case 12	115,1%	100,1%	100,0%
Case 13	115,5%	91,4%	100,0%
Case 14	109,5%	106,1%	100,0%
Case 15	84,9%	90,9%	100,0%

Table 1 Test case results – part 1

Another type of randomly generated cases was tested in [6]. It differed from the previous one in the number of cities. There were 24 of them. Unfortunately, we have not results by the SM, however, the performance of both the NNM and HFA was again similar as above, as shown in Table 2. All the solutions by the NNM consisted of 4 vehicle routes while the HFA found in 7 cases solution with 4 routes and in 3 cases with 3 routes.

	NNM	HFA
Case 1	112,2%	100,0%
Case 2	108,2%	100,0%
Case 3	131,4%	100,0%
Case 4	115,0%	100,0%
Case 5	102,3%	100,0%
Case 6	119,0%	100,0%
Case 7	120,8%	100,0%
Case 8	96,0%	100,0%
Case 9	120,7%	100,0%
Case 10	87,3%	100,0%

Table 2 Test case results – part 2

In contrast, the experience from the solution of practical cases of various types of VRPs, although not just the TLVRP (e.g. from bachelor and diploma theses), is rather different. Thus we constructed three more test cases using cost matrices from such real cases. The transportations took place in a region with relatively dense road network, so we can assume that these cases were of a similar type as the previous ones. The number of non-central cities was between 16 and 22. The cost matrix was given in minutes, the central city was located in the middle of the region, and the remotest cities were about 60 minutes from the central one. For each cost matrix we tried to compute two TLVRP tasks which differed in the time limit *L*: the former one with 80 minutes and the latter one with 120 minutes. In our opinion, such time limits well corresponded with the size of the served region and might occur in real tasks. The results are summarized in table 3 where 100 p.c. is the result of the NNM for the task with 120 minutes time limit.

Above all, it is worth mentioning that for all six (i.e. three cost matrices times two time limits) tasks both the NNM and the SM found better solutions than the HFA. Beside this, in the case of 120 minutes tasks was the difference of these two results bigger than in the case of 80 minutes tasks. In accordance with general expectation, for a given cost matrix the NNM always found for the 120 minutes task a shorter solution than for the 80 minutes task. However, the HFA always found a shorter solution for the 80 minutes task than for the 120 minutes one (let us remind that an 80 minutes task solution is properly also a feasible solution of the 120 minutes task with the same cost matrix, thus the performance of the HFA for tasks with high time limit L showed out to be very terrible) and the SM has similar troubles in the case with the biggest number of cities, although not so considerable. As one can easily see from Table 3, the differences between the results obtained by single methods and for single time limits increased with increasing number of cities. As the number of vehicle routes is concerned, the solutions by the NNM for the 80 minutes tasks consisted of two routes, and for the 120 minutes tasks either consisted of three routes or the last at most two cities left for the fourth vehicle. The average number of cities on a route was about 10 in case of 120 minutes tasks and about 6 in case of 80 minutes tasks. The solutions by the HFA contained usually one vehicle more in comparison with the solution by NNM and SM for the same task.

		80 Minutes Task			120 Minutes Task		
	# Cities	NNM	SM	HFA	NNM	SM	HFA
Case 1	16	107,4%	94,0%	115,3%	100,0%	90,7%	118,1%
Case 2	20	103,5%	99,1%	125,7%	100,0%	90,9%	143,5%
Case 3	22	113,7%	102,2%	130,6%	100,0%	106,0%	142,1%

Table 3 Test case results – part 3

6 Conclusion

When solving tasks of the TLVRP with possibility to use long routes containing many cities for single vehicles, the NNM shows out to be more suitable than the HFA. The procedure of the NNM, accentuating the route arrangement of short segments between two consecutive cities, is more successful than the procedure of the HFA, joining pairs of cities with longer but according to the Habr frequencies overall more advantageous straight joins. Namely, the problem of the HFA is that the distance between such two cities is often unnecessarily long, and thus makes single routes longer or exacts using more routes (vehicles). On the other hand, in case of tasks with too short transportation time limit for creating such long routes, these advantageous but long edges cannot be so often of use for the HFA during its performance, and thus it provides better solutions than the NNM.

The SM acts similarly as the NNM, when we compare its performance with the HFA. However, in accordance with general expectance, the SM mainly achieves slightly better results than the NNM.

If a user is deciding between the NNM, SM and HFA, and do not want much to speculate what of these two methods is more suitable, we can recommend the following plan. First he/she should use the NNM. If the solution consists of routes containing many cities, then case we propose to try also the SM and in the opposite case the HFA. In both these cases, the remaining not used method should not prevailingly provide a better solution than the better one from the already obtained ones.

References

[1] Christofides, N.: *Worst-case Analysis of a New Heuristic for the Traveling Salesman Problem*, Technical Report CS-93-13, Carnegie Mellon University, 1976

[2] Clarke, G., Wright, J.W.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research* **12** (1964), 568-581.

[3] Habr, J.: *Jednoduché optimalizační metody pro ekonomickou praxi*. SNTL, Prague, 1964

[4] Karp, R.M.: A Patching Algorithm for the Non-symmetric Traveling Salesman Problem, SIAM Journal of. Computers **8** (1979), 561-573

[5] Kučera, P: Different Versions of the Savings Method for the Time Limited Vehicle Routing Problem. In: Proceedings of the 28th International Conference on Mathematical Methods in Economics 2010. Czech Society for Operations Research, Prague, 2010, 381-385

[6] Kučera, P: Quality of the Solutions of the Time Limited Vehicle Routing Problem Obtained by Selected Methods. In: Proceedings of the 27th International Conference on Mathematical Methods in Economics 2009. Czech Society for Operations Research, Prague, 2009, 200-203

[7] Lin, S., Kernighan, B.V.: An Effective Heuristic Algorithm for the Traveling Salesman Problem, *Operations Research*, **21** (1973), 972-989
[8] Pelikán, J.: A time limited vehicle routing problem. *Komunikacie* **3** (2006), 9-11.
[9] Rosenkrantz, D.J., Stearns, R.E. Lewis, P.M.: An Analysis of Several Heuristics for the Traveling Salesman Problem. *SIAM Journal of Computation* **6** (1977), 563-581
[10] Van der Cruyssen, P., Rijckaert, M.: Heuristic for the Asymmetric Traveling Salesman Problem, *Journal of Operations Research Society* **30** (1978), 697-701
[11] Webb, J.: An effective heuristic algorithm for the traveling salesman problem, *Operations Research* **21** (1971), 498-516

Comparison of the Cluster Analysis and the Methods of the Multi-criteria Evaluation of Alternatives Used to Create a Groups of Countries Similar in the Take up of the Internet Services

Martina Kuncová[1], Petr Doucek[2]

Abstract. Cluster analysis is one of the statistical methods that can be used to find the clusters, the groups made by objects with similar characteristics that are different from another cluster. On the other hand some methods of multi-criteria evaluation of alternatives (TOPSIS, WSA, PRIAM) usually try to order the alternatives according to the given criteria. The aim of this article is to compare the results when using these different principles in case of finding the groups of EU (27 members) and some more countries (Island, Norway, Croatia) that are similar in the take up of the internet services. This take up is characterized by 9 criteria such as ordering goods via internet, internet banking, seeking information, reading online newspapers and others. We also describe the change among the groups (clusters) of the countries during the years 2008 and 2009 and we observe the position of the Czech Republic.

Keywords: multi-criteria evaluation of alternatives, cluster analysis, countries comparison, internet services

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Information and communication technologies belong to the sectors that are rapidly developing during last couple of years all over the world. One of the main factors that caused this development is internet. Every year we can see new possibilities how to use World Wide Web or internet services to buy, sell, communicate, share information etc. A lot of people around us use internet to find information, to buy things, to communicate through emails or social nets – so nowadays we usually cannot imagine a life without internet. The globalization leads to the fact that the trends in one country became the trends to other countries, so they spread nearly all over the world. But each country has its own possibilities, own habits, own infrastructure and that is why the situation is not similar in each country no matter how developed it is. That is the reason why we would like to compare the countries according to the internet services usage and find out the position of the Czech Republic.

The analysis and comparison can be made from various points of view. A lot of web pages [3] show the order of countries from the number of the internet users point of view. European Union is on the top, and among their countries Germany, United Kingdom and France lead. Adam et al. [1] have tried to compare business use of the internet and World Wide Web across Australia, New Zealand and the UK. In [5],[6],[7] the comparisons of two selected countries from the ICT point of view is made using some of the method of multi-criteria evaluation of alternatives and also [13] used these methods to analyze the broadband network development in Vysocina Region in the Czech Republic. In [4] they tried to analyze the digital gap among the countries of 15 European Union members by factor and cluster analysis. So for the comparison of the countries some methods of multi-criteria evaluation of alternatives are usually used but sometimes if groups of similar countries or regions are needed, cluster analysis is suitable. In this article we try to combine these principles and describe the difference among them. On the other hand our analysis is not only about the comparison of methods but also about the comparison of EU and some other countries in connection with the internet services usage. Another aspect is to compare the situation relating to the internet service usage in the Czech Republic with the other European countries and detect where the conditions are similar, so which countries can lay in one cluster.

[1] University of Economics Prague, Department of Econometrics, W.Churchill Sq. 4, 13067 Prague 3, Czech Republic, kuncovam@vse.cz
[2] University of Economics Prague, Department of System Analysis, W.Churchill Sq. 4, 13067 Prague 3, Czech Republic, doucek@vse.cz

2 Material and methods

Before we start the analysis we have to select the criteria, the alternatives and the methods for the comparison. The data we use come from the Digital Competitiveness Report made by the European Commission in 2010 [8]. As alternatives we take 27 European Union countries plus Iceland, Norway and Croatia – we compare two years 2008 and 2009. The data from the year 2009 are in Table 1.

Country / criterion	info about goods	uploading self-created content	reading online	internet banking	playing or downloading music, games	seeking health info	looking for a job	online course	seeking info for learning
Austria	54	17	41	35	21	36	10	1	24
Belgium	59	18	34	46	33	33	13	4	18
Bulgaria	17	8	21	2	24	10	9	1	12
Cyprus	39	17	27	15	25	16	5	1	10
Czech Re-public	50	5	43	18	23	20	8	1	11
Denmark	74	33	64	66	34	46	27	4	37
Estonia	54	30	63	62	35	32	23	6	24
Finland	73	18	64	72	38	56	24	13	31
France	60	20	24	42	26	37	16	7	24
Germany	69	23	27	41	26	48	18	3	28
Greece	33	9	21	5	19	15	6	2	12
Hungary	48	29	36	16	29	36	18	2	19
Ireland	54	13	19	30	19	24	14	5	27
Italy	33	17	23	16	17	21	9	3	19
Latvia	50	34	46	42	38	29	25	7	23
Lithuania	44	25	49	32	35	29	15	8	22
Luxembourg	75	38	55	54	33	54	13	6	38
Malta	48	9	32	32	28	30	14	4	26
Netherlands	79	26	46	73	49	50	17	5	28
Poland	29	11	18	21	20	22	9	1	12
Portugal	40	12	28	17	20	28	10	2	27
Romania	12	14	21	2	21	16	5	3	14
Slovakia	50	7	35	26	31	30	16	1	11
Slovenia	49	23	34	24	27	32	12	3	21
Spain	47	19	38	24	30	32	16	7	32
Sweden	77	21	50	71	35	36	22	4	28
United King-dom	64	33	43	45	36	34	25	7	30
Iceland	80	43	72	72	42	37	17	10	41
Norway	83	24	76	77	39	40	22	5	33
Croatia	33	21	36	16	22	26	14	2	20

Table 1 Alternatives and criteria – percent of users of given internet service in countries in 2009 [8]

The criteria (all of them are of maximization type) that describe the internet service usage (Table 1) are:

- Looking for information about goods and services
- Uploading self-created content
- Reading online newspapers/magazines
- Internet banking
- Playing or downloading games, images, films or music

- Seeking health information on injury, disease or nutrition
- Looking for a job or sending a job application
- Doing an online course
- Looking for information about education, training or courses offers

To compare the countries we use three selected methods of multi-criteria evaluation of alternatives but to find the groups of countries with similar conditions it is usually better the cluster analysis.

Multi-criteria evaluation of alternatives belongs to the category of discrete multi-criteria decision making models where all the alternatives and criteria are known. To solve this kind of model it is necessary to know the preferences of the decision maker. These preferences can be described by aspiration levels (or requirements), criteria order or by the weight of the criteria [9], [10], [11].

The model of multi-criteria evaluation of alternatives contains a list of alternatives $A = \{a_1, a_2, \ldots, a_p\}$, a list of criteria $F = \{f_1, f_2, \ldots, f_k\}$ and an evaluation of the alternatives by each criterion in the criteria matrix:

$$Y = \begin{array}{c c} & \begin{array}{cccc} f_1 & f_2 & \cdots & f_k \end{array} \\ \begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_p \end{array} & \begin{bmatrix} y_{11} & y_{21} & \ldots & y_{k1} \\ y_{12} & y_{22} & \ldots & y_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1p} & y_{2p} & \ldots & y_{kp} \end{bmatrix} \end{array},$$

where y_{ij}, $i = 1, 2, ..., p$, $j = 1, 2, ..., k$ represent information about the evaluation of each alternative by each criterion.

The theory of multi-criteria evaluation of alternatives is very good established and there are available many different methods for this kind of problems. For the analysis we have used these methods: WSA, TOPSIS and PRIAM (for more information see [10] or [11]) implemented in full version of IZAR [2] and Sanna [12].

WSA (Weighted Sum Approach): This method sorts the alternatives based on the values of their utility functions which in this case are assumed to be linear. It requires the information about the weights of the criteria [9]. Higher value of utility means better alternative.

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution): The basic concept of this method is that the best alternative should have the shortest distance from the ideal alternative and the farthest from the basal alternative. The method is also able to rank the alternatives using the relative index of distance of the alternatives from the basal alternative. Higher relative index of distance means better alternative. The user must supply only the information about the weights of criteria [11].

PRIAM (Programme utilisatnt l´Intelligence Artificiele en Multicritere): This method belongs to the group of methods that need information about aspiration levels for each criterion that should be matched. The aim is to find out one alternative that meets the requirements (aspiration levels). If there is no alternative like this it is possible to change the aspiration levels or calculate the distance of each alternative from the vector of aspiration levels [10]. In our study we suppose the aspiration level to be the maximum of each criterion. Then we calculate the distance for each criterion according to formula:

$$d_i = \sum_{j=1}^{n} \frac{a_j - y_{ij}}{a_j} \quad ,$$

where d_i is the relative index for the distance of alternative i, a_j is the aspiration level (maximum) for the criterion j and y_{ij} is the real value for the alternative i and criterion j (i=1,2,…,m, j=1,2,…, n). As the best alternative has the lowest index, for better comparison with the previous methods we use the final value for each alternative as (1-d_i). To be able to compare the results of the methods we suppose that all the criteria have the same weight.

Cluster analysis is different principle that uses a number of different algorithms and methods for grouping objects of similar kind into respective categories [14]. We have chosen Ward's method that uses an analysis of variance approach to evaluate the distances between clusters. In short, this method attempts to minimize the Sum of Squares of any two (hypothetical) clusters that can be formed at each step. In general, this method is regarded as very efficient, however, it tends to create clusters of small size [14]. The results of the method are usually described by dendrogram (a kind of a chart that shows the clusters). We used the software STATISTICA [14] to find the results.

3 Results

As we are looking for the position of the Czech Republic we started with the multi-criteria methods. The order of the countries according to the three methods is not equal but it is very similar to each other both for 2008 and 2009. The difference between the years is not so big, it is clear that nearly all the countries increased the internet service usage from 2008 to 2009. Czech Republic unfortunately holds 25-26th position among 30 studied countries – it means that the level of the internet service usage in our country is poorer than in most European Union countries. Only 1% of Czech people use internet for doing an online course (in 2009), 5% for uploading self-created content or 8% are looking for a job or sending a job application. For example in Iceland it is 10%, 43% and 17% users of these services. As the methods for multiple-criteria evaluation of alternatives uses different principles, the results are not the same but they are very close. Figure 1 shows that the PRIAM method gives a little bit higher coefficients than other methods but all the correlation coefficients between the results of this three methods are higher than 0.991 (for 2008) and 0.993 (for 2009).

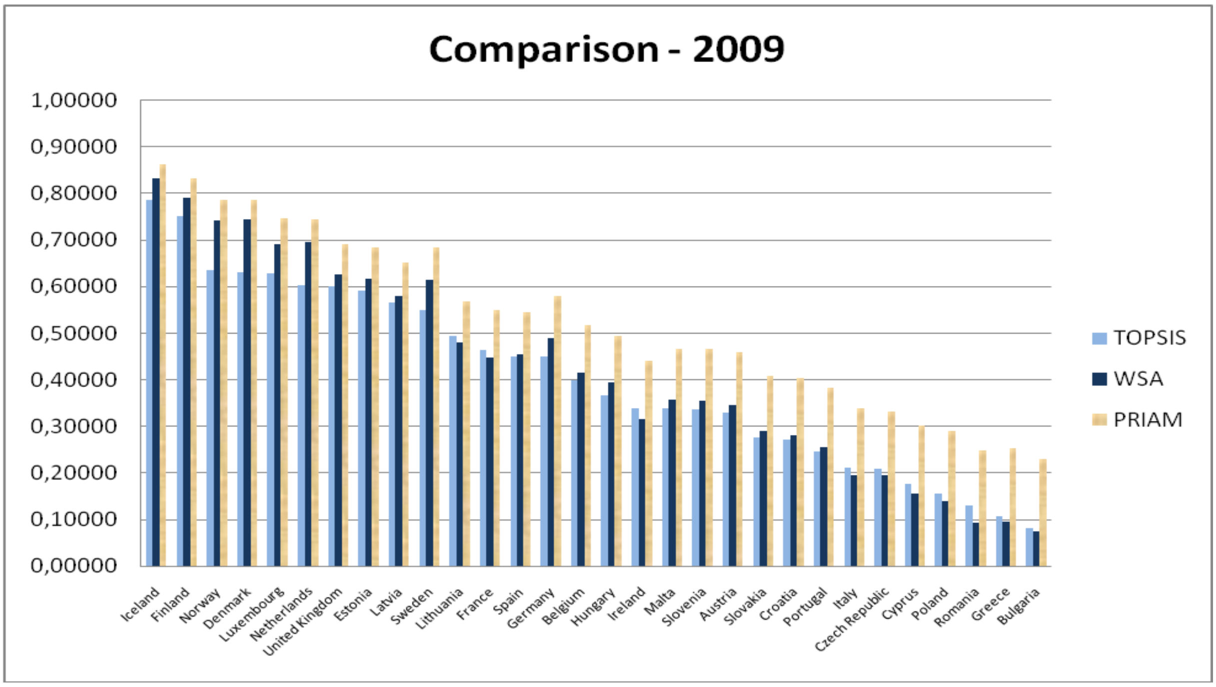


Figure 1 Comparison of the results of multi-criteria evaluation of alternatives methods

To be able to compare the results with the cluster analysis we divided the countries into 9 groups according to the final relative index counted in each method. The differences of the relative indexes of the countries in one group should be lower than 0.05. In the Tables 2,3 and 4 you can see the results, the groups (clusters) and the average index for each cluster. On the top there is always Iceland and the North-European countries, on the bottom Bulgaria and Romania, in 2009 also Greece.

The results from the Ward's method can be displayed only in dendrogram (Figure 2). The problem is that this kind of analysis gives us the results how to distribute the alternatives among clusters (and we can choose how much clusters we want) but it is not possible to find out the order of the clusters from the chart. So according to the previous results we have tried to order the clusters (Table 5). As there is a lot of clusters it is clear that there might be some differences (and they are there) but if we think about 3-4 clusters, the composition of the groups would be probably the same.

Just to compare the positions of the countries the Spearman's rank correlation coefficient has been used. For 2008 (we suppose n=9, for n=30 all the coefficient are higher than 0.9), the coefficient is higher than 0.725, the best is (as can be expected according to the used principles in these methods) for the rankings of TOPSIS and PRIAM (0.958), Ward's method results are closest to TOPSIS (0.9). In the year 2009 it is a little bit different - the coefficient are higher than 0.766, rankings of TOPSIS and PRIAM are again the closest ones but the Ward's results are now closer to WSA ranking (0.892).

2008	**TOPSIS**	**Avg.**	**WSA**	**Avg.**	**PRIAM**	**Avg.**	**WARD**
1.goup	ICE, FIN	0,7663	ICE, NOR, FIN	0,8000	ICE, NOR, FIN	0,8283	ICE, NOR
2.group	NOR	0,6745	NET, DEN	0,6753	NET, DEN, LUX, SWE	0,6923	FIN, NET
3.group	DEN, LUX, NET, SWE	0,5448	LUX, SWE, FRA, UK, EST, LAT	0,6225	FRA, EST, UK	0,6094	LUX, SWE, DEN
4.group	EST, FRA, UK, LAT	0,4919	GER	0,5280	LAT, GER	0,5639	EST, LAT, UK
5.group	GER	0,4202	HUN, SLO, SPA, LIT	0,3714	HUN, SLO, SPA, LIT	0,4521	FRA, GER
6.group	HUN, SPA, SLO, LIT	0,3406	SVK, BEL, AUS, MAL	0,2983	BEL, AUS, SVK, MAL	0,3906	AUS, BEL, IRE
7.group	BEL, AUS, SVK, MAL, IRE, POR, POL, CRO, ITA, **CR**, GRE	0,2644	IRE, POR, POL, CRO	0,2253	IRE, POR, POL, CRO	0,3330	HUN, LIT, SVK, SLO, SPA, MAL
8.group	CYP	0,1719	**CR** ITA, CYP, GRE	0,1441	ITA, **CR**, CYP, GRE	0,2613	CYP, GRE, ITA, CRO, POL, POR, **CR**
9.group	BUL, ROM	0,0696	BUL, ROM	0,0545	BUL, ROM	0,1823	BUL, ROM

Table 2 Groups (clusters) from the results of TOPSIS, WSA, PRIAM and Ward's methods (2008)

2009	TOPSIS	Avg.	WSA	Avg.	PRIAM	Avg.	WARD
1.goup	ICE, FIN	0,7686	ICE, FIN	0,8117	ICE, FIN	0,8477	ICE, NOR
2.group	NOR, DEN, LUX, NET, UK, EST	0,6150	DEN, NOR, NET, LUX	0,7184	DEN, NOR, LUX, NET	0,7660	DEN, FIN, LUX
3.group	LAT, SWE	0,5578	UK, EST, SWE	0,6198	UK, EST, SWE, LAT	0,6782	NET, SWE
4.group	LIT, FRA, SPA, GER	0,4642	LAT	0,5791	GER, LIT, FRA, SPA	0,5613	UK, LAT, EST
5.group	BEL, HUN	0,3830	GER, LIT, SPA, FRA	0,4675	BEL, HUN	0,5066	FRA, GER, BEL
6.group	IRE, MAL, SLO, AUS	0,3353	BEL, HUN, MAL, SLO, AUS	0,3732	SLO, MAL, AUS, IRE	0,4583	IRE, SLO, HUN, LIT, SPA, MAL, AUS
7.group	SVK, CRO, POR, ITA, CR	0,2433	IRE, SVK, CRO, POR	0,2857	SVK, CRO, POR	0,3982	CR, SVK
8.group	CYP, POL, ROM	0,1550	ITA, CR, CYP, POL	0,1715	ITA, CR, CYP, POL	0,3160	CYP, GRE, ITA, POL, POR, CRO
9.group	GRE, BUL	0,0949	GRE, ROM, BUL	0,0875	GRE, ROM, BUL	0,2437	BUL, ROM

Table 3 Groups (clusters) from the results of TOPSIS, WSA, PRIAM and Ward's methods (2009)

4 Conclusions

The purpose of this paper was to compare the results of the selected method of multi-criteria evaluation of alternative and cluster analysis. As we choose the data connected with internet service usage in European countries in the year 2008 and 2009, we also would like to know if the situation in the Czech Republic is similar to other countries. For this problem the cluster analysis should be used but the problem is that this analysis usually tells us the information about the clusters and how far they are from each other but we also would like to know the ranking or ordering of the clusters and countries. Any of the used methods of multi-criteria evaluation of alternatives give us this information. The results from the methods are similar (according to the correlation coefficients and Spearman's rank correlation coefficients), so we may say that it is possible to use selected multi-criteria methods not for ordering but also for clustering. The position of the Czech Republic in internet services usage is poor in comparison with other European countries – usually similar to Poland, Portugal, Italy or Slovakia. The reasons of this situation are behind the scope of this article but it can be a part of another research.

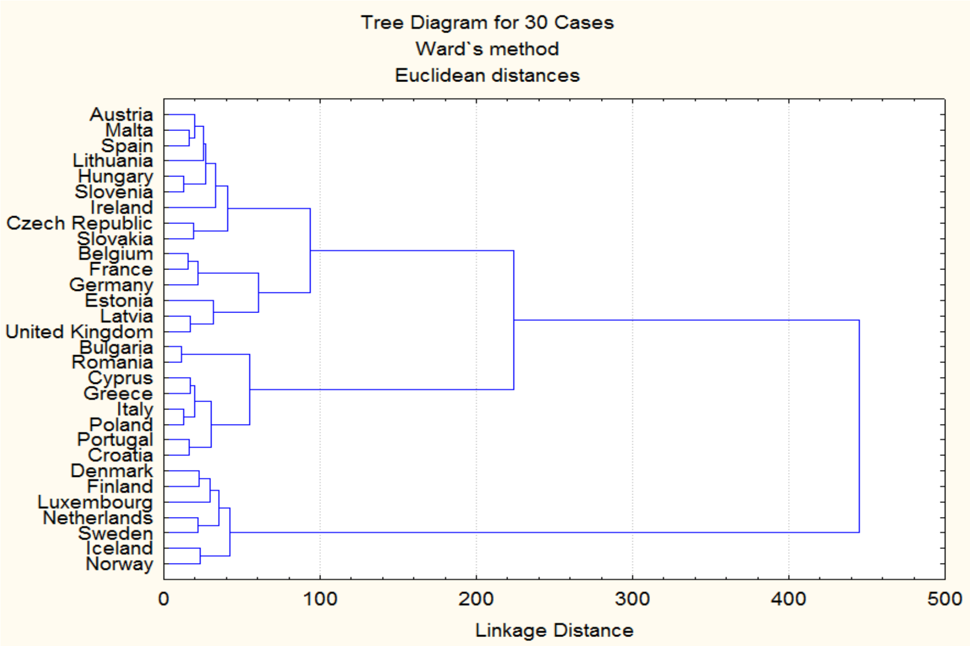


Figure 2 Dendrogram – results from the Ward's method (2009) – STATISTICA

Acknowledgements

This paper was processed with contribution of GAČR by handling task GAČR 402/09/0385 "Human Capital in IS/ICT Operations and Development: Competitivness of Czech Tertiary Education Graduates".

References

[1] Adam, S., Mulye, R., Deans, K.R., Palihawadana, D.: E-marketing in perspective: a three country comparison of business use of the Internet. *Marketing Intelligence & Planning*, vol. 20, issue: 4, 2002, 243 - 251

[2] Boksteflova, B., Kalcevova, J.: IZAR – the multiattribute evaluation decision support for Linux. In: *7th International Conference Efficiency and Responsibility in Education*, Czech University of Life Sciences Prague, 2010, 50-58

[3] Central Intelligence Agency – the World Factbook: Country comparison – Internet users [online], 2011, [cit.2011-5-11], available at: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2153rank.html

[4] Cuervo, V., Rosalia, M.: A multivariate framework for the analysis of the digital divide: Evidence for European Union-15. *Information & Management*, vol.43, issue 6, 2006, 756-766.

[5] Doucek, P., Nedomová, L.: How does financial crises influence ICT? In: *Človek in organizacija* (CD-ROM), Maribor, Univerza v Mariboru, 2010, 349-355.

[6] Doucek, P., Nedomová, L., Novotný, O.: How ICT Affect the Czech Economy? *ECON*, 2011, vol. 19, No. 1, 106–116.

[7] Doucek, P., Nedomová, L.: ICT and Economy – Slovenia and Czech Republic Experience. In: *Organizacija prihodnosti – Future Organization* [CD-ROM]. Maribor : Univerzita v Mariboru, 2011, s. 268–278.

[8] European Commission, 2010, Europe's Digital Competitivness Report 2010, Luxembourg Publications Office of the European Union, 2010

[9] Evans, G.W.: An Overwiev of Techniques for Solving Multiobjective Mathematical Programs. *Management Science*. 1984, vol. 30, No. 11, 1268-1282.

[10] Fiala, P.: *Modely a metody rozhodování*. Praha: Oeconomica 2006.

[11] Figueira, J., Greco, S., Ehrgott M.: *Multiple Criteria Decision Analysis – State of the Art Surveys*. New York : Springer Science + Business Media Inc., 2005.

[12] Jablonský, J. SANNA – A Spreadsheet Based System for Multicriteria Decision. In: *Book of Abstracts. Chania*, 2006, p. 148.

[13] Kuncová, M., Kalčevová, J., Novotný, J., Vokáčková, H.: Multicriteria Decision in the Analysis of the Market Environment for the Notification Process of the Broadband Network Development. In: *Mathematical Methods in Economics* [CD-ROM]. České Budějovice : University of South Bohemia, 2010, 386–391

[14] StatSoft: Cluster analysis. [online], 2011, [cit.2011-5-16], available at: http://www.statsoft.com/textbook/cluster-analysis/

Multi-Agent Simulation of Tiebout Model

Michal Kvasnička[1]

Abstract. In this paper, the Tiebout model (JPE, 1956) is explored by means of an agent-based computational simulation. The simulation suggests that the Tiebout's conjectures holds true only when the consumers are homogeneous in all respects other than their taste for public goods. If they are heterogeneous in other aspects (like productivity) too, some consumers try to parasitize on others and the others try to escape from the parasites. A system with small relocation cost does not reach an equilibrium. If the relocation cost is high enough, the system finds an equilibrium, but the equilibrium does not separate the consumers according to their taste for public goods. The aggregate quantity of public goods is lower than the optimal one and public goods and the tax burden are distributed among the inhabitants in an inefficient way. The system is preferred by a majority sufficient to establish it only under a rather restricting conditions.

Keywords: fiscal federalism, Tiebout's model, agent-based simulation

JEL classification: H210, H710, C690
AMS classification: 37M05, 68T42, 68U20, 91B12, 91B18

1 Introduction

In his seminal paper [4], Tiebout tried to solve the problem of the optimal provision of public goods in a world with heterogeneous consumers and imperfect information. The task was to make the consumers reveal their true "willingness to pay" for public goods. Tiebout's solution was ingenious: he proposed a decentralized system where public goods were provided by local governments. Each government would provide a different package of public goods and services and a different scheme to finance them. Each consumer would then patronize the region with the combination of the amounts of public goods and the tax structure she prefers most. Tiebout conjectured that this way the optimal level of public goods would have been provided (if there was a sufficient diversity among the regions) because each consumer would have had an incentive to reveal her true preferences. The essential condition for the system to work is that public goods are not truly non-rival, so the average cost of providing public goods is not ever-decreasing but is minimized by a certain quantity.

The problem with the Tiebout "model" is that it is only verbal. Its premises are not stated precisely, and its conclusions are not established properly. Moreover, there might be a serious flaw in the "model". In particular, it is not clear what Tiebout had in mind when he claimed that "[t]he consumer-voter may be viewed as picking that community which best satisfies his preference pattern for public goods" [4, p. 418]. A consumer may patronize a region for three reasons: 1) she likes the combination of public goods and the tax rate as such, i.e. she would patronize the region if it was populated by inhabitants with "tastes" for public goods and other characteristics identical to hers; or 2) she may patronize the region to "parasitize" on the contribution of the others, e.g. because the other inhabitants of the region are more productive than her, and thus pay higher taxes; she would not patronize the region if it was populated with inhabitants with the characteristics identical to hers; or 3) she may patronize the region to escape from "parasites" settled in a region which she would otherwise like better. It seems that Tiebout abstracted from similar interactions among the consumers. However, it has been shown in a different context (see e.g. [1]), that such interactions might change the relocation process and its outcomes considerably. The goal of this paper is to asses the validity of the Tiebout model for a world where consumers are heterogeneous both in their "tastes" for public goods and their productivity and where all above mentioned interactions among the consumers may occur. Since the problem involves heterogeneous, locally interacting agents, my approach is an agent-based computational simulation, see [3].

[1]Masaryk University, Economic Faculty, Lipová 41a, Brno, 602 00, the Czech Republic, e-mail: michal.kvasnicka@econ.muni.cz

2 Description of the Model

The model world consists of regions and inhabitants. Each inhabitant lives in one region, produces a homogeneous product in her own business, and pays part of it to the local government as taxes. She consumes what is left over from her production after the taxes are paid and public goods provided by the local government. There is no trade or credit among the inhabitants. The tax rate in each region is constant. The local government uses the the funds collected to provide a single kind of public goods produced with constant returns to scale. The inhabitants can relocate if they believe they would be better off in another region (taking into account a cost of relocation).

More formally, there are K regions and N inhabitants. There are $N_{t,j}$ inhabitants in region j at time t. Let us denote $A_{t,j}$ the set of indices of all inhabitants in region j in time t. All inhabitants of region j pay a flat-rate tax. The proceedings are used to provide public good in quantity $p_{t,j}$. The region j's budget constraint is then

$$N_{t,j} \cdot p_{t,j} = \sum_{i \in A_{t,j}} \tau_j \cdot y_{t,i}, \tag{1}$$

where τ_j is the tax rate in region j and $y_{t,i}$ is the product of the region j's inhabitant i at time t. Region j's tax system thus can be described as a couple $(\tau_j, p_{t,j})$. Budget deficits or surpluses are not allowed.

Every inhabitant i produces a homogeneous product $y_{t,i}$ at time t. Her production function f is

$$y_{t,i} = f(l_{t,i}) = a_i \cdot l_{t,i}, \tag{2}$$

where $l_{t,i}$ is her labor effort at time t (labor is the only factor of production) and a_i is her time-invariant productivity.

Each inhabitant i maximizes her utility function

$$u_i(c_{t,i}, p_{t,\nu_{t,i}}, l_{t,i}) = \sqrt{c_{t,i}} + \alpha_i \cdot \sqrt{p_{t,\nu_{t,i}}} + \sqrt{1 - l_{t,i}} \tag{3}$$

subject to constraints

$$c_{t,i} = (1 - \tau_{\nu_{t,i}}) \cdot y_{t,i} = (1 - \tau_{\nu_{t,i}}) \cdot a_i \cdot l_{t,i}, \quad l_{t,i} \in [0, 1], \tag{4}$$

where $\alpha_i \in (0, 1]$ is her time-invariant parameter of "taste" for public goods, $(1 - l_{t,i})$ is her leisure time at time t, $c_{t,i}$ is her consumption at time t, and $\nu_{t,i}$ is an index function that returns the index of the region where she is presently living. The inhabitants differ in two respects: in their productivity a_i and in their "taste" for public goods α_i (relative to their "taste" for private goods and leisure time).

The inhabitant i's optimal labor effort $l^*_{t,i}(j)$ in the region j is

$$l^*_{t,i}(j) = \frac{a_i(1 - \tau_j)}{1 + a_i(1 - \tau_j)}. \tag{5}$$

The optimal labor effort $l^*_{t,i}(j)$ (and hence optimal level of production $y^*_{t,i}(j)$) decreases in the tax rate τ_j. The amount of public goods $p_{t,j}$ provided in the region affects neither the optimal labor effort $l^*_{t,i}(j)$, nor the optimal level of production $y^*_{t,i}(j)$ because each inhabitant considers himself to be too small to affect the community's level of public goods.

Let us denote $u^*_i(\tau_j, p_{t,j}, \gamma)$ the maximal level of utility that the inhabitant i can have given the tax rate τ_j, the amount of public goods $p_{t,j}$ and a cost γ

$$u^*_i(\tau_j, p_{t,j}, \gamma) = u_i(c^*_{t,i}(j) - \gamma, p_{t,j}, l^*_{t,i}(j)), \tag{6}$$

where $l^*_{t,i}(j)$ is the inhabitant i's optimal labor effort and $c^*_{t,i}(j)$ is her optimal level of consumption given the tax rate τ_j, i.e. $l^*_{t,i}(j)$ and $c^*_{t,i}(j)$ are the levels of labor effort and consumption maximizing (3) subject to (4). The parameter γ will be explained below. (However, if $c^*_{t,i}(j) < \gamma$, then $u^*_i(\cdot) = -\infty$.)

In specified moments, the inhabitants can relocate from their home region to another one. Each inhabitant i wants to move to a region where she would obtain the highest level of utility. There is a relocation cost γ paid in terms of consumption. Each inhabitant i acts like this: let us denote j her present home region. Further, let us denote k the region where her utility $u^*_i(\tau_k, p_{t,k}, 0)$ is maximal. Inhabitant i then relocates from region j to region k if $u^*_i(\tau_k, p_{t,k}, \gamma) > u^*_k(\tau_j, p_{t,j}, 0)$; otherwise she stays in region j. (There is an exception when an inhabitant is considering a relocation into an empty region, where $A_{t,k} = \emptyset$, and hence $p_{t,k} = 0$. To allow the re-occupation of an empty region, I set its $p_{t,k}$ to the level it would have after the region k was patronized by the inhabitant i.) When inhabitant i relocates from region j to region k, the index i is removed from set $A_{t+1,j}$ and added to set $A_{t+1,k}$, the counter $N_{t+1,j}$ is decreased by 1, the $N_{t+1,k}$ is increased by 1, and the index function $\nu_{t+1,i}$ returns k instead of j. After each relocation step, all local governments recalculate the level of public goods provided in their regions so that the regions' budget constraint (1) hold.

The inhabitants are *boundedly rational* in two respects: 1) When they relocate they assume they would change the amount of public goods provided neither in their former home region, nor in their new one. Thus they ignore the possibility that their own relocation changes the amounts of public goods provided in the regions, which could provoke other inhabitants to relocate. They also ignore the possibility that other inhabitants can relocate at the same time too—there is no coordination among them. 2) When considering relocation, the inhabitants expect to stay in the target region for a fixed amount of time T (possibly forever). Intuitively, they have to pay a relocation cost Γ once and for all when they relocate. To compare this cost with their consumption (repeated every period), they have to set a unit cost γ such that Γ is the present value of annuity γ, i.e. $\Gamma = \sum_{l=1}^{T} \beta^l \gamma$ where $\beta \in [0, 1]$ is a discount factor. Fully rational inhabitants should establish the proper value of the unit cost γ from the model; but γ is a parameter of the model here.

The model is an agent-based computational model (see [3]), hence it is simulated, not solved. I describe the simulation only for one particular set of parameters because the limited length of this paper does not allow a more general treatment. The simulation proceeds like this: In the beginning of each simulation, the model is initialized. $K = 6$ regions are created, and assigned the tax rate τ_j equal to 0, 6, 12, 18, 24 and 30 % respectively. Then $N = 1000$ inhabitants are created. Each inhabitant i is assigned a random productivity a_i and a random "taste" for public goods α_i (a_i is drawn from the continuous uniform distribution over the interval $[0, 1]$ in one treatment and $[0.3, 1]$ in other treatment; α_i is drawn from the continuous uniform distribution over the interval $[0, 1]$). Each inhabitant i is positioned randomly in one of the regions (with the same probability in each one). When all inhabitants are created and located, each region j calculates the amount of public goods $p_{0,j}$ to be provided. Each simulation consists of 500 steps. In each step, each inhabitant is allowed to relocate (if she wishes to). When all inhabitants relocated, the amount of public goods provided in each region is recalculated. There are two treatment variables: the minimal productivity $a_{\min} = 0$ or 0.3 and the relocation cost $\gamma = 0, 0.002, 0.004, \ldots, 0.15$. The model was run 30 times for each value of the treatments with random seeds 1, 2, ..., 30. The model was implemented and simulated in NetLogo 4.1.2 (see [2]); data were processed and analyzed in R. The interactive web interface of the model is available at http://www.econ.muni.cz/~qasar/models.html.

3 Method of comparison to the Tiebout model

The comparison of the Tiebout model to the simulated model is not easy because the Tiebout model is stated only verbally, and hence imprecisely. For this reason, only qualitative comparison is possible. We can compare the models based on the Tiebout's four conjectures: 1) the strategic interactions among the inhabitants (parasitism and attempts to escape it) do considerably affect neither the relocation process, nor its outcomes, 2) the inhabitants patronize the regions which best satisfies their "tastes" for public goods, 3) public goods are provided in the optimal amount, and 4) the inhabitants prefer the Tiebout's world to the system with a uniform provision of public goods financed with a uniform tax rate set by a democratic vote. (The second and third claims are explicit, the first and fourth ones are implicit in the Tiebout's paper [4].) The first conjecture can be evaluated on two grounds: 1) the evolution of the simulated system can be observed directly, or 2) it can be claimed that the conjecture holds true if the other three conjectures hold true and vice versa.

The second conjecture apparently always holds true because the inhabitants always patronize their most preferred region. However, in the simulation model, each inhabitant takes into account the strategic considerations and the relocation cost when she chooses a region—the two aspects ignored by Tiebout. Thus to compare each inhabitant i's real location to her theoretic optimal one, I define inhabitant i's optimal region $\kappa(i)$ to be the region she would prefer most if all regions were populated with inhabitants identical to her (i.e. inhabitants with the same parameters a_i and α_i). Thus there would be no parasites and each inhabitant i would contribute precisely her share to public goods. That is, the most preferred region $\kappa(i)$ maximizes $u_i^*(\tau_j, p_{t,j}, 0)$ subject to constraint $p_{t,j} = \tau_j a_i l_{t,i}^*(j)$.

To test the Tiebout's third conjecture, the optimal level of public goods must be calculated. Tiebout did not define what amount of public goods is optimal when public goods are not completely non-rival. For the sake of comparison, I define the optimal aggregate level of public goods as the sum of public goods that would be provided to all inhabitants in all regions if each inhabitant i patronized her optimal region $\kappa(i)$. This quantity is compared to the real amount provided (the sum of public goods provided to all inhabitants in all regions where they are), and to the amount that would be provided to all inhabitants if there was a uniform tax rate and a uniform provision of public goods set democratically. The uniform democratic tax rate is calculated with a simple numeric algorithm.

In the outset, the uniform democratic tax rate is set to 100 %. If the number of inhabitant who prefer the tax rate to be lowered by 0.01 is at least one half of all inhabitants, the tax rate is lowered by 0.01. The procedure is repeated until less than one half of all inhabitants prefer the lower tax rate.

I also calculate how many inhabitants prefer the uniform democratic system to the Tiebout's hypothetic one, how many inhabitants prefer the uniform democratic system to the outcomes of the simulation model, and how many inhabitants prefer the Tiebout's hypothetic system to the outcomes of the simulation model. (Notice that these three preference relations need not to be transitive due to the Condorcet paradox.)

4 Results of the Selected Simulations

The results of the simulations are strikingly at odds with the Tiebout's conjectures. First, there is a phase transition in the model, see Figure 1. The realized phase depends on the relocation cost γ and the minimal productivity $a_{\min}$, though the phases cannot be classified neatly. The phases are three: 1) the model does not reach an equilibrium if the relocation cost is low—the inhabitants relocate forever; 2) no inhabitants relocate if the relocation cost is high; and 3) the equilibrium is reached after some inhabitants relocated if the relocation cost is intermediate. This fact shows that the strategic interactions among the inhabitants are important. If they were not, each inhabitant i would relocate to her optimal region $\kappa(i)$ and would stay there forever. But in fact, the less productive inhabitants pursue the more productive ones and the more productive inhabitants try to run away from them.

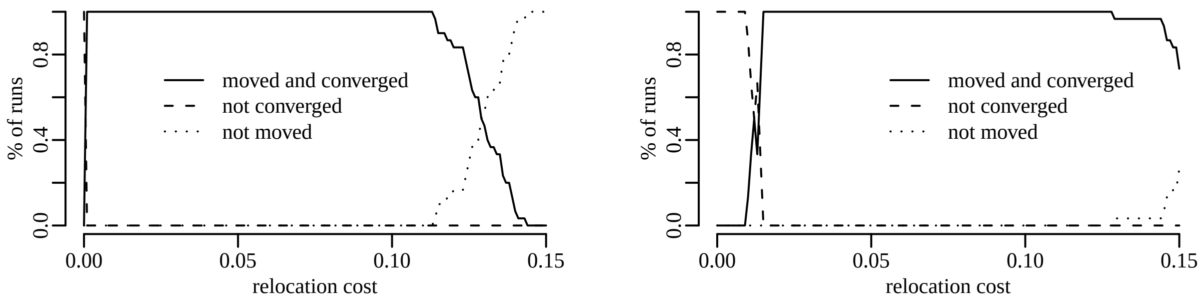


Figure 1: The phase transition in the model. The "not converged" denotes the share of runs where no equilibrium was reached. The "not moved" denotes the share of runs where no inhabitant relocated. The "moved and converged" denotes the share of runs where some inhabitants relocated and yet the equilibrium was reached. The left panel is for $a_{\min} = 0$, the right panel for $a_{\min} = 0.3$.

It is the relocation cost that stops the "hunting game" (if it is sufficiently high to do it at all). However, the inhabitants' equilibrium location is inefficient. For $a_{\min} = 0$ and low γ, most inhabitants relocate to low tax regions, usually to the region with the tax rate 6 %, see the left panel of Figure 2. It is because there are some extremely unproductive inhabitants in all regions which cannot relocate at all (their $c^*_{t,i}(\cdot) < \gamma$). The more productive inhabitants leave these regions to avoid parasitism, and can never invade these regions back because they are occupied by the unproductive inhabitants. The situation is similar but not identical for $a_{\min} = 0.3$. Most inhabitants end up in two or three regions in this case. One of them is the region with the tax rate 0 %, the other regions are picked randomly. That is why the mean occupation rate of all regions looks less diverse in the right panel of Figure 2 than in its left panel. When the relocation cost rises, fewer inhabitants relocate and more inhabitants stay in their original random position. Thus all occupation rates converge to its mean value $1/6$. The same can be seen in Figure 3. When the relocation cost is low, most inhabitants are driven out to regions with lower than their most preferred tax rate (this effect is more pronounced when $a_{\min} = 0$). If the relocation cost is high, fewer inhabitants relocate and more stay in their original random region, but even in this case, more inhabitants end up in regions with lower then the most preferred tax rate than in regions with higher than the most preferred tax rate. It is because the more productive inhabitants can relocate more easily than the less productive ones, and they are biased in favor of the regions with lower taxes to avoid parasitism.

The amount of public goods that would by provided by a uniform democratic system is always higher than the hypothetic optimal amount, see Figure 4. It is because the inhabitants whose "taste" for public goods is low can "opt-out" of the system and because there is no parasitism in the Tiebout's hypothetic system too (contrary to the democratic system, see [1]). However, the really provided aggregate amount of public goods is even lower that the optimal one. It is especially low for $a_{\min} = 0$ and low γ because most inhabitants ends up in the regions with inefficiently low taxes in this case. The mean amount provided with $a_{\min} = 0.3$ and low γ is higher because the

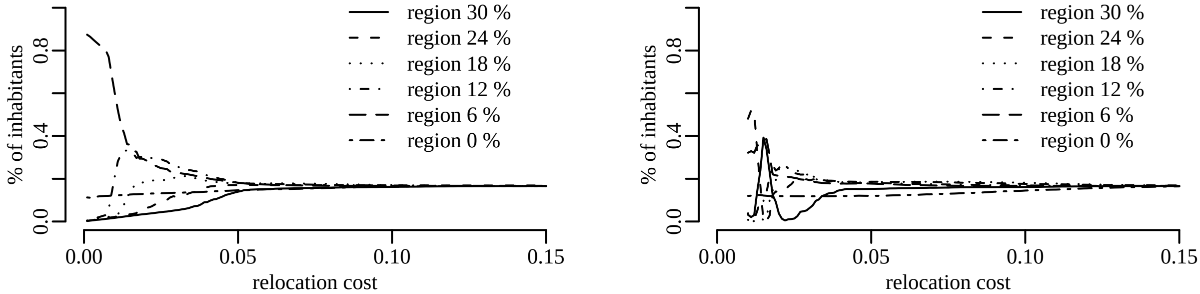


Figure 2: The equilibrium occupation rate of the regions in the runs where an equilibrium is reached. The left panel is for $a_{\text{min}} = 0$, the right panel for $a_{\text{min}} = 0.3$.

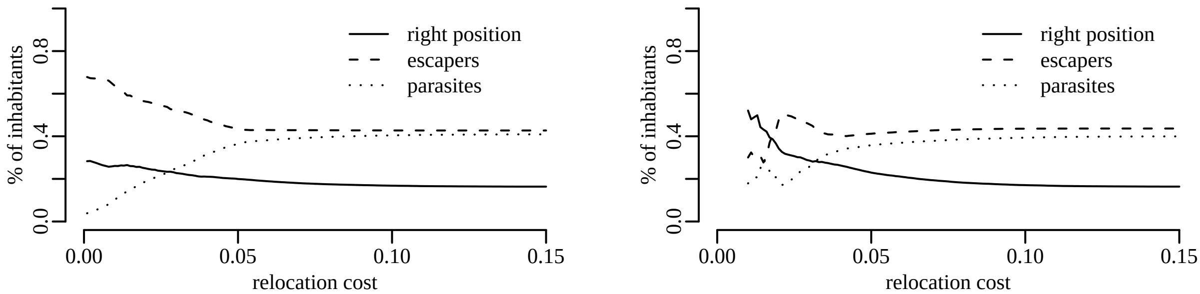


Figure 3: The mean frequency with which the inhabitants end up in their optimal region (denoted "right position"), in the region with a lower tax rate than in their optimal region (denoted "escapers"), or with a higher tax rate than in their optimal region (denoted "parasites"). Only the runs where an equilibrium is reached are considered. The left panel is for $a_{\text{min}} = 0$, the right panel for $a_{\text{min}} = 0.3$.

majority of the inhabitants ends up in random regions in this case, i.e. the tax rate paid by an average inhabitant is higher. As γ rises, the real aggregate amount of public goods converges to the optimal one. However, it is because most inhabitants stay in their original random location. Thus the aggregate amount of public goods is almost efficient but its distribution is random: inhabitants whose taste for public goods is high might pay a low tax rate and consume low quantity of public goods, and vice versa. In any case, the real aggregate amount of public goods is always lower than the optimal amount because the more productive inhabitants can relocate more easily than the less productive ones—and they tend to relocate to the regions with suboptimally low tax rates to avoid parasitism.

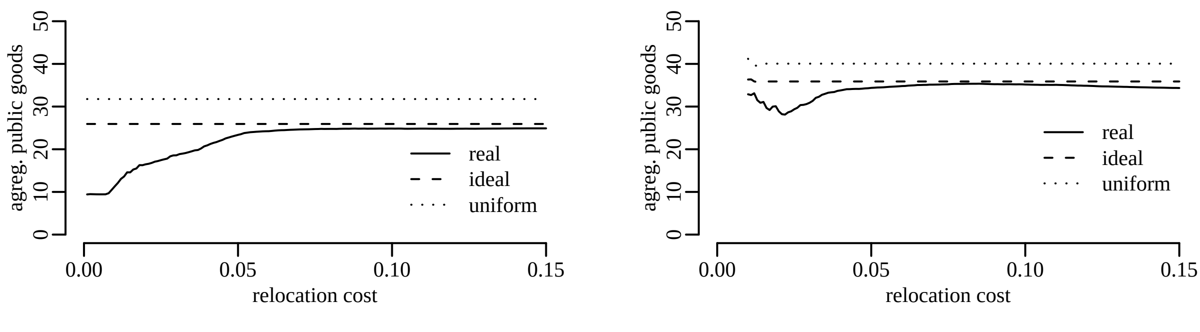


Figure 4: The mean aggregate amount of public goods provided. The "real" denotes the amount really provided in the model, the "ideal" denotes the Tiebout's optimal level of public goods, and the "uniform" denotes the amount that would be provided if all inhabitants paid the same tax rate established by a democratic vote. Only the runs where an equilibrium is reached are considered. The left panel is for $a_{\text{min}} = 0$, the right panel for $a_{\text{min}} = 0.3$.

The majority of inhabitants prefers the uniform democratic system to the Tiebout's hypothetic one and the hypothetic system to its real realization in the majority of runs, see Figure 5. For $a_{\text{min}} = 0$, the majority also prefers the uniform democratic system to the real one in the majority of runs, hence the preferences are transitive. For $a_{\text{min}} = 0.3$, the majority of inhabitants prefers the real system to the uniform in most runs when the relocation cost is low and yet the system reaches an equilibrium. When γ is higher, the majority prefers the uniform democratic system in most runs. However, the mean preference for the uniform democratic system to the Tiebout's ideal is mild, see Figure 6: roughly one half inhabitants prefer the uniform democratic system and the other half the Tiebout's hypothetic optimal system.

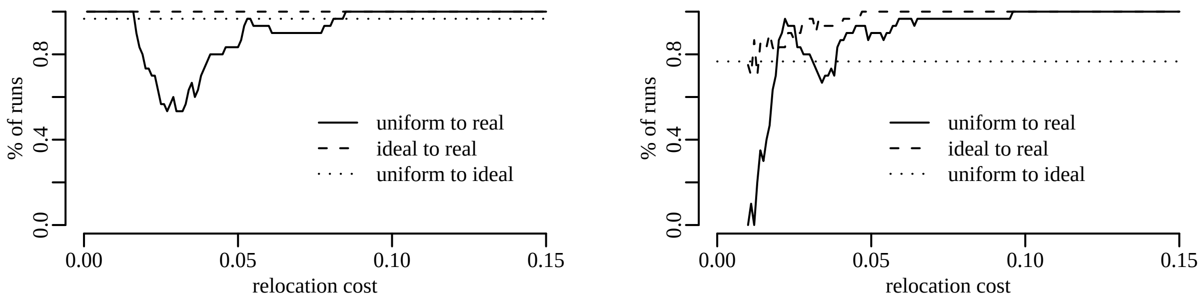


Figure 5: The share of runs where the majority of inhabitants preferred one system to another. Only the runs where an equilibrium is reached are considered. The left panel is for $a_{\min} = 0$, the right panel for $a_{\min} = 0.3$.

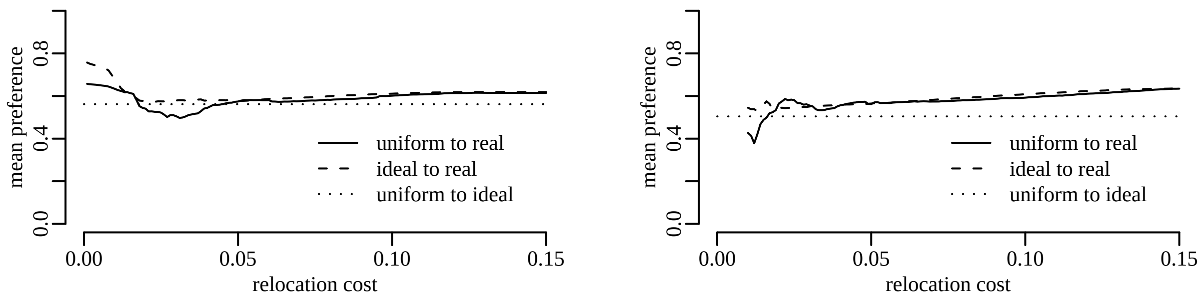


Figure 6: The mean share of inhabitants who preferred one system to another. Only the runs where an equilibrium is reached are considered. The left panel is for $a_{\min} = 0$, the right panel for $a_{\min} = 0.3$.

5 Conclusion

The simulation model shows that Tiebout [4] was overly optimistic. His conjectures hold true only in a world with inhabitants that are identical in all respects other than their "taste" for public goods. If the inhabitants differ in their productivities, the strategic interactions among them considerably modify their choice of a region. There need not to be an equilibrium if the relocation cost is too small. But even if there is an equilibrium, it does not possess the desirable features expected by Tiebout. First, the inhabitants are not located in their optimal regions. Second, the aggregate quantity of public goods is lower than optimal, and public goods and the tax burden are distributed among the inhabitants in an inefficient way. As a result, the real version of the Tiebout's system is preferred by a majority sufficient to establish it only under a rather restrictive conditions.

References

[1] Kvasnička, M.: Agent-based Computational Model of Democratic Choice of Redistribution under Almost Perfect Human Mobility, *Mathematical Methods in Economics 2010* **28** (2010), part II, 392–398.

[2] NetLogo Homepage, http://ccl.northwestern.edu/netlogo/.

[3] Tesfatsion, L.: *Agent-Based Computational Economics: Growing Economies from the Bottom Up*, http://www.econ.iastate.edu/tesfatsi/ace.htm.

[4] Tiebout, Ch. M.: A Pure Theory of Local Expenditures, *Journal of Political Economy* **64** (1956), 416–424.

Monopoly Supply Chain Management via Rubinstein Bargaining[1]

Michal Kvasnička [2], Rostislav Staněk [3], Ondřej Krčál [4]

Abstract. This paper argues that the Spengler's (JPE, 1950) double-markup story does not capture the strategic situation faced by a monopoly supply chain. We show that 1) the Spengler's double-markup model is equivalent to an extensive game in which the upstream monopoly has all power to set the intermediate price, and the downstream monopsony has no bargaining power. This assumption seems unrealistic. 2) We model this situation using the Rubinstein bargaining. The equilibrium of our model does not yield a double markup. Hence, the joint output, the consumers price, and the total profit is as high as those of a vertically integrated firm. Moreover, the profit is split roughly equally between the two firms. 3) Our model has different implications for the competition policy.

Keywords: vertical integration, double markup, Rubinstein bargaining, industrial organization, competition policy

JEL classification: L110, L120, L190, L420
AMS classification: 91A05, 91A10, 91A18, 91A80, 91B26

1 Introduction

The established conception of the supply chain behavior was derived from the Spengler's model (it is the basic model of supply chain management in most books on industrial organization and competition policy, see e.g. [2, 5, 7, 8]; it is also the main model tested experimentally, see e.g. [1]). In his seminal paper [6], Spengler showed that the firms' attempt to maximize their profit independently leads to the "double markup" if the firms constitute a supply chain whose stages operate on non-competitive markets. Such a chain produces lower quantity, charges higher price, and gains lower profit than a vertically integrated firm. The Spengler's model was refined later for the case of a monopoly supply chain. In this case, all Spengler's conclusions hold true. Moreover, the total profit of the chain is split highly unevenly among the firms: the upstream firms always gain more than the downstream firms. The fact that supply chain might be even less efficient than one vertically integrated monopoly has important, but ambiguous implications for the competition policy. In some market structures, vertical mergers of firms in a monopoly supply chain should be allowed, for they increase the efficiency, see [2]. In other market structures (e.g. when there are two upstream firms in Bertrand competition and two downstream firms in Cournot competition), mergers should not be allowed, see [5].

In this paper, we deal with the monopoly refinement of the model only. We argue 1) that the Spengler's double markup story does not capture the strategic situation faced by the bilateral monopoly he envisioned. We show that the double markup model is equivalent to an extensive game in which the upstream monopoly is first on the move: it has all power to set the intermediate price, and the downstream monopsony has no bargaining power and acts as a price taker. This is the reason why the upstream firm has always higher profit than the downstream firm. This assumption seems unrealistic—the downstream firm could try to bargain for a better deal. Therefore, it would be more realistic to model the strategic situation as a bargaining. 2) Our model based on the Rubinstein bargaining has strikingly different results from the Spengler's one. First, there is no double markup: the output, final price and the total chain's profit are the same as those set by a vertically integrated monopoly. Second, the chain's total profit is split between the firms based on which one is first on the move, not based on their position in the supply chain, and is split roughly evenly under fairly realistic conditions. 3) Since the monopoly supply chain is precisely as (in)efficient as the vertically integrated monopoly, the established implications for the competition policy are no

[1] The work on this paper has been supported by the project of specific research MUNI/A/0053/2011 of the Masaryk University.
[2] Masaryk University/Economic Faculty, Lipová 41a, Brno, 602 00, the Czech Republic, e-mail: michal.kvasnicka@econ.muni.cz
[3] Masaryk University/Economic Faculty, Lipová 41a, Brno, 602 00, the Czech Republic, e-mail: rostanek@mail.muni.cz
[4] Masaryk University/Economic Faculty, Lipová 41a, Brno, 602 00, the Czech Republic, e-mail: krcalo@mail.muni.cz

longer valid.

2 Review and Critique of Spengler's Model

In this section, we discuss the simplest version of the double markup model; a more general treatment can be found e.g. in [9, 10]. We assume an industry consisting of two monopoly firms. The upstream firm produces an intermediate good and sells it to the downstream firm for an intermediate price p. The downstream firm produces a consumer good and sells it to the ultimate consumers for a final price $\hat{p}$. The technology is such that the quantity produced by the downstream firm is determined by the quantity produced by the upstream firm, i.e. both firms produce the same quantity q. We assume that the upstream firm is the only supplier of the intermediate good used by the downstream firm and the downstream firm is the only consumer of the intermediate good produced by the upstream firm, hence the firms face a bilateral monopoly. The profits of the upstream and the downstream firms can be represented by the profit function $\Pi_u(q, p)$ and $\Pi_d(q, p)$ respectively

$$\Pi_u(q, p) = p \cdot q - C_u(q), \tag{1}$$
$$\Pi_d(q, p) = R(q) - p \cdot q - C_d(q), \tag{2}$$

where p is the price of the intermediate good produced by the upstream firm, q is the quantity produced by each firm, $R(q) = \hat{p} \cdot q$ is the total revenue of the downstream firm on the consumer market, and $C_u(q)$ and $C_d(q)$ are the total costs of the upstream and the downstream firms respectively.

For simplicity's sake, we assume in this section that the total cost of the upstream firm is $C_u(q) = c_u \cdot q$, the total cost of the downstream firm is $C_d(q) = c_d \cdot q$, and that the downstream firm's total revenue $R(q) = \hat{p} \cdot q$ is given by the inverse demand for its production $\hat{p} = a - b \cdot q$, where $a > c_u + c_d$ and $b > 0$. The profits of the upstream and the downstream firms respectively are then

$$\pi_u = p \cdot q - c_u \cdot q, \tag{3}$$
$$\pi_d = \hat{p} \cdot q - p \cdot q - c_d \cdot q = (a - b \cdot q) \cdot q - (p + c_d) \cdot q. \tag{4}$$

The conventional solution of the Spengler's model proceeds in three steps: 1) we derive the demand for the upstream firm's production, 2) we calculate the intermediate price p set by the upstream firm, and 3) we calculate the price and quantity set by the downstream firm. In other words, first we calculate the output that maximizes the downstream firm's profit (4) taking the intermediate price p as given. The downstream firm's optimal output q° is then

$$q^\circ = (a - p - c_d)/2b. \tag{5}$$

Since both firms produce the same quantity q°, the equation (5) is the demand for the upstream firm's production. Second, we substitute (5) into (3) and calculate the intermediate price p° that maximizes the upstream firm's profit (3). Third, substituting the optimal upstream firm's price p° into (5) we calculate the total chain's output q°, the downstream firm's optimal price $\hat{p}^\circ$, and the profits of both firms. It holds that

$$q^\circ = \frac{a - c_u - c_d}{4b}, \quad p^\circ = \frac{a + c_u - c_d}{2}, \quad \hat{p} = \frac{3a + c_u + c_d}{4}, \quad \pi_u = \frac{(a - c_u - c_d)^2}{8b}, \quad \text{and} \quad \pi_d = \frac{(a - c_u - c_d)^2}{16b}.$$

Notice that the profit of the upstream firm π_u is always higher than the profit of the downstream firm π_d; $\pi_u = 2\pi_d$ in the linear case.

We can compare the outcomes of the monopoly supply chain behavior to those of the vertically integrated monopoly, i.e. a firm created by a merger of the upstream and the downstream firms. We assume there is no economy of scale or scope and the total cost $C_m(q)$ of the vertically integrated monopoly is $C_m(q) = C_u(q) + C_d(q)$. The profit $\pi_m = \Pi_m(q)$ of the vertically integrated monopoly is

$$\pi_m = \Pi(q) = R(q) - C_m(q) = R(q) - C_u(q) - C_d(q). \tag{6}$$

The vertically integrated monopoly with the linear demand $p = a - b \cdot q$ and the total cost $C_m(q) = (c_u + c_d) \cdot q$ would produce the quantity $q^m = (a - c_u - c_d)/(2b)$, sell it for the price $p^m = (a + c_u + c_d)/2$, and reach the profit $\pi_m = (a - c_u - c_d)^2/(4b)$. It is easy to see that the monopoly output q^m is higher than the chain's output q° ($q^m = 2q^\circ$ in the linear case), the monopoly final price p^m is lower than the chain's final price $\hat{p}^\circ$, and the monopoly profit π_m is higher than the chain's total profit $\pi_u + \pi_d$. Thus in terms of welfare the chain's output is inferior to the output of the vertically integrated monopoly since both the consumers' and producers' surpluses are higher in case of the vertically integrated monopoly.

The conventional solution may seem straightforward but it is tricky. It is equivalent to the backward induction used to solve extensive games. In fact, the double markup model is a two stage extensive game. There are two players in this game: the upstream and the downstream firm. The strategy of the upstream firm is to set the intermediate price $p \geq 0$, the strategy of the downstream firm is to set the chain's output $q \geq 0$. The upstream firm is first on the move: it sets the intermediate price p. The downstream firm is on the move in the second stage of the game, where it takes the intermediate price p as given, and sets the chain's quantity q. The payoffs of the game are given by the profits π_u and π_d, i.e. equations (3) and (4), respectively. The game is solved by the backward induction. First, we solve the second stage of the game—the subgame where the downstream firm is on the move. We find its best response to a given intermediate price p. This best response is the demand for the production of the upstream firm (5). Second, we calculate the optimal behavior of the upstream firm given the downstream firm's best response function. Third, we solve for the remaining variables. It is evident that the particular steps of the conventional solution correspond precisely to the steps of the backward induction, and hence the models are equivalent. The fact that the double markup model is equivalent to this extensive game also explains why the upstream firm's profit is always higher then the downstream firm's profit—it is advantageous to be first on the move in this game because $\hat{p}^\circ = (a + c_d + p^\circ)/2$. Thus if the upstream firm rises its price p by 1, the downstream firm rises its price $\hat{p}$ only by $1/2$, i.e. the downstream firm accommodates a part of the price increase. The upstream firm then has a strong incentive to increase its price p as long as its markup rises faster that the quantity demanded decrease.

The extensive game presented above does not grasp the strategic situation faced by the two firms in the chain properly. In the double markup model, the upstream monopoly has a full monopoly power to set the intermediate price p, while the downstream firm (which is really a monopsony) acts as a price taker. There is no reason why a monopsony should accept such a passive role, especially when the resulting split of the total chain's profit is unfavorable to it. It seems more likely that the downstream firm would try to bargain for a better deal. Moreover, it is not obvious why the upstream firm should be first on the move and why any firm would like to play the double-markup game at all. There are many other extensive games the firms could play. For instance, any firm might try to play "take it or leave it" game. If the upstream firm was first on the move, it could ask the downstream firm to sell the monopoly quantity q^m for the monopoly price p^m while charging $p = p^m - c_d$. Since the downstream firm's payoff is zero regardless of whether it accepts and rejects the offer, it would accept it, and the upstream firm would gain full monopoly profit π_m. On the other hand, if the downstream firm was first on the move, it could ask the upstream firm to sell it the monopoly quantity q^m for the upstream firm's marginal cost c_u, and hence gain the whole monopoly profit π_m itself. There are many other extensive games the firms could try to play. However, all of them are susceptible to the same criticism: the other firm has no reason to accept the game and would most likely try to bargain for a better deal. This is why we believe that the strategic situation faced by the bilateral monopoly of the firms in the monopoly supply chain is best modeled by a model of bargaining.

3 Supply Chain Management via Rubinstein Bargaining

In this section we provide a model of a bargaining process between the upstream and the downstream firm. The model is based on the Rubinstein bargaining game [4] where players take turns to make offers to each other until agreement is reached. In the Rubinstein game players bargain over the division of fixed size pie. In our model, bargaining determines not only the division of the pie but also the size of the pie. Consider an industry described in section 2 where profits of the upstream firm and the downstream firm are given by equations (1) and (2) respectively. We assume that there is a unique quantity $q^m > 0$ that maximizes the profit of vertically integrated monopoly $R(q) - C_u(q) - C_d(q)$, but we do not impose any other restrictions on the revenue function or cost functions.

We define the bargaining game with alternating offers between the upstream firm and the downstream firm as follows: Firms bargain over the intermediate price p and production level q sold by the downstream firm to final consumers. (Alternatively, the firms can also bargain over the intermediate price p and the price $\hat{p}$ at which the downstream firm sells the product to the final consumers.) Time is discrete. At time 0 one of the firms, for example the upstream firm, offers to the downstream firm a price p_u and production quantity q_u. If the downstream firm accepts the offer, then the agreement is reached. On the other hand, if the downstream firm rejects the offer, then it makes an counteroffer (q_d, p_d) in the next period. The upstream firm can again accept or reject the offer. This process of making offers and counteroffers continues until the agreement is reached. Preferences of the upstream firm and the downstream firm are given by their discounted profits $\delta_u^T \Pi_u(q, p)$ and $\delta_d^T \Pi_d(q, p)$, where $\delta_u \in (0, 1)$ denotes the discount factor of the upstream firm, $\delta_d \in (0, 1)$ denotes the discount factor of the downstream firm and T is the period when the agreement is reached.

Note that since the game has infinite horizon, we cannot use backward induction method. We proceed as follows. First, we find strategy profile that satisfies two properties stated below and constitutes a reasonable candidate for the subgame perfect equilibrium (SPE). Next, we prove that one-deviation property holds for this strategy profile and thus this strategy profile constitutes a SPE of the game. Finally, we prove that this SPE is unique. We assume that players reach an agreement in the SPE. Further because of the stationary structure of the game, we assume that the SPE strategies are stationary. Although the stationary structure of the game does not necessarily imply that the game has equilibrium in stationary strategies, the stationarity of strategies provides a reasonable starting point. These two assumptions imply that whenever a player makes an equilibrium offer, the offer is accepted.

Consider a subgame in which the upstream firm is making an offer. It follows from the two above stated assumptions that the downstream firm's profit from rejecting the offer is $\delta_d \Pi_d(q_d^*, p_d^*)$ where (q_d^*, p_d^*) is the equilibrium offer made by the downstream firm. The SPE requires that the downstream firm accepts every offer (q_u, p_u) such that $\Pi_d(q_u, p_u) \geqq \delta_d \Pi_d(q_d^*, p_d^*)$. This condition will be binding, because if $\Pi_d(q_u, p_u) > \delta_d \Pi_d(q_d^*, p_d^*)$, then the upstream firm can increase its profit by offering a slightly higher price. The upstream firm is thus solving the following problem

$$\max_{p_u, q_u} p_u q_u - C_u(q_u) \quad \text{s. t.} \quad R(q_u) - p_u q_u - C_d(q_u) = \delta_d(R(q_d) - p_d q_d - C_d(q_d)).$$

The solution of this problem is given by the following equations

$$R'(q_u) = C_u'(q_u) + C_d'(q_u), \tag{7}$$
$$R(q_u) - p_u q_u - C_d(q_u) = \delta_d(R(q_d) - p_d q_d - C_d(q_d)). \tag{8}$$

By a similar argument we can find that the downstream firm is solving the following problem when making an offer

$$\max_{p_d, q_d} R(q_d) - p_d q_d - C_d(q_d) \quad \text{s. t.} \quad p_d q_d - C_u(q_d) = \delta_u(p_u q_u - C_u(q_u)).$$

The solution of this problem is again given by two equations

$$R'(q_d) = C_u'(q_d) + C_d'(q_d), \tag{9}$$
$$p_d q_d - C_u(q_d) = \delta_u(p_u q_u - C_u(q_u)). \tag{10}$$

As we can see both firms offer the same quantity of production q^* which is the unique solution to equation (7) and (9). The level of production is such that marginal revenue equals to the sum of marginal costs. So, it is the same production level that produces vertically integrated monopoly, i.e. $q^* = q^m$. Consequently, total profit of both firms is also the same as the profit of vertically integrated monopoly Π. It is given by q^* and two following equations hold

$$\Pi_u(q^*, p_u) + \Pi_d(q^*, p_u) = R(q^*) - C_u(q^*) - C_d(q^*) = \Pi, \tag{11}$$
$$\Pi_u(q^*, p_d) + \Pi_u(q^*, p_d) = R(q^*) - C_u(q^*) - C_d(q^*) = \Pi. \tag{12}$$

The equilibrium price is then determined by equations (8), (10), (11) and (12). These equations have a unique solution

$$p_u^* = \frac{1 - \delta_d}{1 - \delta_u \delta_d} \frac{R(q^*) - C_d(q^*)}{q^*} + \frac{\delta_d(1 - \delta_u)}{1 - \delta_u \delta_d} \frac{C_u(q^*)}{q^*}, \tag{13}$$
$$p_d^* = \frac{\delta_u(1 - \delta_d)}{1 - \delta_u \delta_d} \frac{R(q^*) - C_d(q^*)}{q^*} + \frac{1 - \delta_u}{1 - \delta_u \delta_d} \frac{C_u(q^*)}{q^*}. \tag{14}$$

By substituting the equilibrium quantity and the equilibrium prices into the profit functions we can easily calculate the profit of the downstream and the upstream firm.

$$\Pi_u(q^*, p_u^*) = \frac{1 - \delta_d}{1 - \delta_d \delta_u} \Pi, \quad \Pi_d(q^*, p_u^*) = \frac{\delta_d(1 - \delta_u)}{1 - \delta_d \delta_u} \Pi, \tag{15}$$

$$\Pi_u(q^*, p_d^*) = \frac{\delta_u(1 - \delta_d)}{1 - \delta_d \delta_u} \Pi, \quad \Pi_d(q^*, p_d^*) = \frac{1 - \delta_u}{1 - \delta_d \delta_u} \Pi. \tag{16}$$

Now we have a reasonable guess of a equilibrium strategy profile. In the proof of following proposition we verify that this strategy profile constitutes a SPE. Moreover, the proposition 2 rules out the possibility of existence of another equilibria.

Proposition 1. *The following pair of strategies constitutes a SPE of the bargaining game with alternating offers between the upstream firm and the downstream firm*

- *the downstream firm always offers q^* and p_d^* and accepts an offer (q_u, p_u) if and only if $\Pi_d(q_u, p_u) \geqq \Pi_d(q^*, p_d^*)$,*
- *the upstream firm always offers q^* and p_u^* and accepts an offer (q_d, p_d) if and only if $\Pi_u(q_d, p_d) \geqq \Pi_u(q^*, p_u^*)$,*

where p_d^*, p_u^* are given by (13) and (14) and q^* is a unique solution to (7) or (9).

Proof. We have to show that one-deviation property holds, i.e. no player can profitably deviate from equilibrium strategy at the start of any subgame. The game has four types of subgames. We show that the upstream firm's strategy is optimal in every point in the game given the downstream firm's strategy. By a symmetric argument, it follows that the downstream firm's strategy is optimal in every point in the game given the upstream firm's strategy.

1. Consider first the subgame in which the first move is the upstream firm's offer. Suppose that the upstream firm makes an offer $(q_u, p_u) \neq (q^*, p_u^*)$. Obviously it is not profitable to offer any (q_u, p_u) such that $\Pi_u(q_u, p_u) \leqq \Pi_u(q^*, p_u^*)$. So suppose that $\Pi_u(q_u, p_u) > \Pi_u(q^*, p_u^*)$. In this case the downstream firm rejects the offer and the upstream firm's profit in the next period is $\delta_u \Pi_u(q^*, p_d^*)$. But from equations (15) and (16) we can see that $\delta_u \Pi_u(q^*, p_d^*) < \Pi_u(q^*, p_u^*)$ and thus the deviation is not profitable.
2. Now consider the subgame in which the first move is the upstream firm's response to the downstream firm's offer (q_d, p_d). First, assume that $\Pi_u(q_d, p_d) \geqq \Pi_u(q^*, p_d^*)$ and the upstream firm is supposed to accept the offer. If the upstream firm rejects the offer it obtains next period payoff $\delta_u \Pi_u(q^*, p_u^*)$. But from the equations (15) and (16) we can see that $\delta_u \Pi_u(q^*, p_u^*) = \Pi_u(q^*, p_d^*)$ and thus the deviation is not profitable. Second, assume that $\Pi_u(q_d, p_d) < \Pi_u(q^*, p_d^*)$ and the downstream firm is supposed to reject the offer. The rejection is clearly optimal. If the upstream firm rejects the offer it obtains next period payoff $\delta_u \Pi_u(q^*, p_u^*)$ such that $\delta_u \Pi_u(q^*, p_u^*) = \Pi_u(q^*, p_d^*)$ and consequently $\delta_u \Pi_u(q^*, p_u^*) > \Pi_u(q_d, p_d)$. Hence, the deviation is not profitable in this subgame.

□

Proposition 2. *The pair of strategies described in proposition 1 is the unique SPE of the bargaining game with alternating offers between the upstream firm and the downstream firm.*

Proof. The proof is based on the original Rubinstein's proof [4]. Rubinstein assumes that the size of the pie is fixed. To convert our problem into Rubinstein bargaining, we have to show that optimal production level q^* is offered at the beginning of every subgame, and hence the total profit of both firms $\Pi(q_i, p_i) = \Pi_u(q_i, p_i) + \Pi_d(q_i, p_i) = \Pi$ is fixed. Consider a subgame starting with the upstream firm's offer. Assume by contradiction that there is a SPE in which upstream firm offers p_u and $q_u \neq q^*$. There are three possible endings of the game. First, suppose that the downstream firm accept the offer immediately. The downstream firm obtains payoff $\Pi_d(q_u, p_u)$. In this case the upstream firm can offer q^* and p'_u such that $\Pi_d(q^*, p'_u) = \Pi_d(q_u, p_u)$. Because the downstream firm's strategy is subgame perfect, the downstream firm accepts this offer. Moreover, we know that $\Pi(q^*, p'_u) > \Pi(q_u, p_u)$ which implies that $\Pi_u(q^*, p'_u) > \Pi_u(q_u, p_u)$. Hence, we have a contradiction with the assumption of SPE. Second, suppose that the downstream firm rejects the offer and the agreement is made in future period T. The downstream firm obtains payoff $\delta_u^{T-1} \Pi_u(q'', p'')$ and the upstream firm obtains payoff $\delta_d^{T-1} \Pi_d(q'', p'')$. It does not matter who makes the offer (q'', p''). In this case the upstream firm can offer q^* and p'_u such that $\delta_d^{T-1} \Pi_d(q^*, p'_u) = \Pi_d(q''_u, p''_u)$. The downstream firm accepts this offer, because its strategy is subgame perfect. From the fact that $\Pi(q^*, p'_u) \geqq \Pi(q''_u, p''_u)$ and $\delta_u < 1$ follows that $\Pi_u(q^*, p'_u) > \delta_u^{T-1} \Pi_u(q''_u, p''_u)$ which is a contradiction with the assumption of SPE. Third, consider that the agreement is never reached. But this is clearly contradiction with subgame perfection. Similar argument holds when the downstream firm offers. Hence, we know that in every SPE each firm offers q^*. From now we can suppose that the firms bargain only over the price p. The price p divides the fixed size pie given by the production level q^*. The problem faced by the upstream firm and the downstream firm is now the same as in the Rubinstein's paper. So, the rest of the proof is the same as the Rubinstein's proof of conclusion 2, see [4, p. 110]. □

4 Conclusion

In conclusion, we would like to emphasize some important features of our bargaining model. The bargaining process is efficient in two ways. First, the agreement is reached immediately, no resources are wasted in delay. Second, the firms always supply the production level that maximizes their total profit. So, they behave just like a vertically integrated monopoly and double markup does not occur. As we can see from equations (13) and (14), the equilibrium price is a convex combination of average revenue and average cost. It is unrelated to the marginal costs or marginal revenues and it only divides the profit between the upstream and the downstream firm. In the contrast with the double markup model, the division of profits is not determined by a doubtful assumption that the upstream firm moves first and sets the intermediate price. The division of profits in our model depends on the patience of firms and their starting positions. Given the patience of one firm, the other firm's profit increases as it becomes more patient. Equations (15) and (16) imply that for any given value of δ_u, the equilibrium profit of the downstream firm increases as δ_d increases. Symmetrically, fixing δ_d the equilibrium profit of the upstream firm increases as δ_u increases. The firm that starts bargaining obtains greater share of profit. If $\delta_u = \delta_d = \delta$ then the equilibrium profit of a first-mover is given by $\frac{1}{1+\delta}\Pi$ whereas the profit of a second-mover is $\frac{\delta}{1+\delta}\Pi$. But the first-mover advantage disappears if δ of both players goes to 1. In this case each firm gets one half of the profit of vertically integrated monopoly.

References

[1] Holt, Ch. A.: *Markets, Games, & Strategic Behavior*, Pearson/Addison Wesley, 2007.

[2] Motta, M.: *Competition Policy: Theory and Practice*, Cambridge, 2004.

[3] Osborne, M. J.–Rubinstein, A.: *Bargaining and Markets*, Academic Press, 1990.

[4] Rubinstein, A.: "Perfect Equilibrium in a Bargaining Model", *Econometrica*, Vol. 50, No. 1 (Jan., 1982), pp. 97–109.

[5] Shy, O.: *Industrial Organization: Theory and Application*, MIT Press, 1985.

[6] Spengler, J. J.: "Vertical Integration and Antitrust Policy", *Journal of Political Economy*, Vol. 58, No. 4 (Aug., 1950), pp. 347–352.

[7] Tirole, J.: *The Theory of Industrial Organization*, MIT Press, 1988.

[8] Varian, H. R.: *Intermediate Microeconomics: A Modern Approach*, 7th ed., W. W. Norton & comp., 2006.

[9] Zouhar, J.: "Efficiency of Decentralized Supply Chains", *Mathematical Methods in Economics 2007: 25th International Conference*, 2007.

[10] Zouhar, J.: "Multiple Marginalization in Serial and Parallel Supply Chains Operating in a Non-Linear Environment", *Mathematical Methods in Economics 2009: 27th International Conference*, 2009.

Adaptive algorithm of parameter adjustment for approximate solving of the p-median problem

Marek Kvet[1]

Abstract. This contribution deals with the algorithm suggested for parameter adjustment of the approximate approach to the p-median problem. Large instances of this problem can be often met when a real public service system is designed and the discomfort of its customers is minimized. If a large instance of the problem is described by a location-allocation model then the model size often exceeds any acceptable limit for available software tools. Presented approximate approach enables us to take the advantage of the set covering problem to solve large p-median instances using common decision support tools. The covering approach is based on specific model reformulation where the distance between a customer and the nearest located service center is estimated by some upper and lower bound. To get the bounds it is necessary to partition the range of values from the distance matrix into several zones. The zone adjustment depends on the value of the model parameter. The main goal of this paper is to present the algorithm suggested for finding the optimal value of the parameter to improve the accuracy of the distance estimation.

Keywords: large p-median problems, approximate approach, covering model, parameter adjustment, adaptive algorithm

JEL Classification: C61
AMS Classification: 90C06

1 Introduction

Large instances of the p-median problem and associated solving approaches form a background of many public system design problems where the quality criterion of the design takes into account the average distance between served objects and the source of provided service [2], [5], [6], [9]. The family of public service systems includes medical emergency system, fire-brigade deployment, public administration system design and many others. To obtain good decisions on service center location in a served area, a mathematical model of the problem can be formulated and some of mathematical programming methods used to get the optimal solution of the problem.

The particular location-allocation models are characterized by considerably big number of possible service center locations which must be taken in consideration. The number of served objects takes the value of several thousands and the number of possible facility locations can take this value as well [7]. For some integer programming algorithms, the number of possible service center locations impacts the computational time necessary for finding the optimal solution. Concerning the problem size it is obvious that these models constitute such mathematical programming problems that resist to any attempt for fast solution.

On the other hand, it is generally known that large covering problems are easily solvable by common optimization software tools even if they belong to the family of NP-hard problems. The aim of suggested approach is to take the advantage of the covering problem to solve real-sized instances of the p-median problem. The keystone of the approximate approach is based on specific model reformulation reported in [4], [7] and [8]. It enables us to solve real-sized instances in admissible time.

It is obvious that the approximate approach may cause loss of accuracy. Nevertheless the computational time could be much smaller. The other advantage is the fact we can employ a common integer programming solver instead of the specialized one to get the optimal solution. It is very important to note that the accuracy of the solution depends on the value of model parameter. Therefore the main goal of our survey is to find out an adaptive algorithm that provides appropriate value of mentioned parameter to achieve good accuracy of the solution.

In this paper we want not only to present suggested adaptive algorithm as a sequence of some operations that may work. We also give an overview of numerical experiments to prove or disprove usefulness of suggested technique. Furthermore we want to explore a relation between the accuracy and the value of model parameter. The other relation to explore is the relation between the optimal value of mentioned parameter and the size of solved p-median instance. We also compare this algorithm to other approaches of model parameter adjustment.

[1] The University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitná 1, 010 26 Žilina, Slovak republic, e-mail: Marek.Kvet@fri.uniza.sk

2 Model reformulation

Generally, the p-median problem can be formulated as a task of selection of at most p network nodes so that the sum of network distances from each node to the nearest located service center is minimal. Accordingly to the notation used in public service system designing we denote J a set of served customers and the symbol I denotes the set of possible service center locations which are also network nodes. The network distance between the possible service center location $i \in I$ and the customer $j \in J$ is given by d_{ij}. In the common location-allocation model published in [7], we are supposed to decide which served object is assigned to which located service center and also where the service centers should be located.

According to [4] and [7], the idea of model reformulation consists of excluding the assign variables from the model. Concerning particular structural conditions it is obvious that each customer can be assigned to only one located service center. It means that only one distance from each matrix column becomes relevant and increases the objective function value. Thanks to this feature we can estimate the distances and formulate a covering model of the p-median problem. To this purpose, we partition the range $<0, max\{d_{ij}: i \in I, j \in J\}>$ of all possible distances of the former location-allocation problem into $r+1$ zones. The zones are separated by the ascending sequence $D_1, D_2, \ldots D_r$, where $0 = D_0 < D_1$ and $D_r < D_m = max\{d_{ij}: i \in I, j \in J\}$. We introduce a numbering of these zones so that the zone k corresponds to the interval $(D_k, D_{k+1}>$, the zone one corresponds to the interval $(D_1, D_2>$ and so on till the r-th zone which corresponds to the interval $(D_r, D_m>$. The width of the k-th interval is denoted by e_k for $k = 0, \ldots, r$. The partitioning of values included in the distance matrix is shown on Figure 1.

Figure 1 Partitioning of the values from the distance matrix

In addition to the bivalent variable $y_i \in \{0, 1\}$, which takes the value of 1 if a facility should be located at the place $i \in I$ and which takes the value of 0 otherwise, we introduce auxiliary zero-one variables $x_{jk} \in \{0, 1\}$ for each $j \in J$ and $k = 0, \ldots, r$. The variable x_{jk} takes the value of 1 if the distance of the customer $j \in J$ from the nearest located center is greater than D_k and it takes the value of 0 otherwise. Then it is easy to show that the expression $e_0x_{j1} + e_1x_{j2} + e_2x_{j3} + e_3x_{j4} + \ldots + e_{r-1}x_{jr}$ is the lower approximation of d_{ij} and the expression $e_0x_{j0} + e_1x_{j1} + e_2x_{j2} + e_3x_{j3} + \ldots + e_rx_{jr}$ is the upper approximation of d_{ij}. It means that if the distance d_{ij} falls to the interval $(D_k, D_{k+1}>$, it is estimated by the lower bound D_k and the upper bound D_{k+1} respectively with a possible deviation e_k. Furthermore, it is necessary to define a bivalent constant $a_{ij}^k \in \{0, 1\}$ for each triple $<i, j, k> \in I \times J \times \{0, \ldots, r\}$. This constant takes the value of 1 if the distance between the customer j and the possible service center location i is less or equal to D_k, Otherwise the constant a_{ij}^k takes the value of 0. According to [4], [7], [8] we can formulate a covering model connected with the upper bound in the form of (1) – (5).

$$Minimize \quad \sum_{j\in J}\sum_{k=0}^{r} e_k x_{jk} \tag{1}$$

$$Subject\ to: \quad x_{jk} + \sum_{i\in I} a_{ij}^k y_i \ge 1 \quad for\ j \in J\ and\ k = 0,\ldots,r \tag{2}$$

$$\sum_{i\in I} y_i \le p \tag{3}$$

$$x_{jk} \ge 0 \quad for\ j \in J\ and\ k = 0,\ldots,r \tag{4}$$

$$y_i \in \{0,1\} \quad for\, i \in I \tag{5}$$

The objective function (1) gives the upper bound of the sum of original distances. The constraints (2) ensure that the variables x_{jk} are allowed to take the value of 0 if there is at least one center located in radius D_k from the customer j. The constraint (3) limits the number of located facilities by p. This covering approach is reported in more details in [4], [7] and [8]. If we want to solve the p-median problem using this approximate approach, we have to set the number of dividing points r and also the width e_k of each interval given by a pair of dividing points. So the only problem is to set the dividing points in an appropriate way to minimize total deviation.

3 Selection of dividing points using model parameter

It is obvious that only limited number of dividing points can keep the model (1) - (5) in a solvable extent. This restriction impacts a deviation of the approximate solution from the exact one. The only problem is to find the appropriate way of the dividing points selection.

As we have shown in [7] or [8], the elements of the distance matrix $\{d_{ij}\}$ form a finite ordered set of values $d_0 < d_1 < \ldots < d_m$ where $D_0 = d_0$ and $D_m = d_m$. Each value d_h is connected with a frequency N_h of its occurrence in the matrix $\{d_{ij}\}$. If there are only r different values between d_0 and d_m, we could determine the dividing points $D_1, D_2, \ldots, D_r$ so that they would be equal to these values. Then we can obtain the exact solution solving the covering problem described by the model (1) - (5).

Otherwise the distance between a customer and his nearest located service center can be only estimated knowing that it belongs to the interval $(D_k, D_{k+1}>$ given by a pair of dividing points. If we were able to anticipate the frequency n_h of each d_h in the optimal solution, we could obtain the dividing points in an exact way – by solving the following mathematical model minimizing the deviation.

$$\text{Minimize} \quad \sum_{t=1}^{m}\sum_{h=1}^{t}(d_t - d_h)n_h x_{ht} \tag{6}$$

$$\text{Subject to:} \quad x_{(h-1)t} \le x_{ht} \quad \text{for } t = 2,\ldots,m \text{ and } h = 2,\ldots,t \tag{7}$$

$$\sum_{t=h}^{m} x_{ht} = 1 \tag{8}$$

$$\sum_{t=1}^{m-1} x_{tt} = r \tag{9}$$

$$x_{ht} \ge 0 \quad \text{for } t = 1,\ldots,m \text{ and } h = 1,\ldots,t \tag{10}$$

The decision variable x_{ht} takes the value of 1 if the distance d_h belongs to the interval which ends by the dividing point d_t. The link-up constraints (7) ensure that the distance d_{h-1} can belong to the interval ending with d_t only if each distance between d_{h-1} and d_t belongs to this interval. The constraint (8) assures that each distance d_h belongs to some interval and the constraint (9) enables only r dividing points. After the problem (6) - (10) is solved, the nonzero values of x_{tt} indicate the distances which correspond with the optimal dividing points.

Remember that we should be able to anticipate the frequency n_h of each d_h in the optimal solution. The mentioned sequence N_h of occurrence frequencies does not provide us the information, because it reports only on the elements contained in the matrix $\{d_{ij}\}$. Consider, if the parameter $p > 2$, then it is clear that the biggest value from the j-th column will be never included in the optimal solution of the p-median problem. It is natural that the customer is assigned to the nearest located service center. Therefore we start here from the hypothesis that the frequency n_h of d_h may be proportional to N_h and to some weight which decreases with increasing value of d_h. According to [7] and [8] we formulate this hypothesis in the following form.

$$n_h = N_h e^{-d_h/T} \tag{11}$$

The symbol T represents a positive parameter called "*temperature*" and the symbol N_h denotes mentioned occurrence frequency where only $|I| - p + 1$ smallest distances of each matrix column are included.

Now, the whole principle of the approximate covering approach to the p-median problem is clear. The main problem needed to be solved is the value of parameter T.

4 Model parameter adjustment

The first way to find out the optimal value of "*temperature*" is presented in [8]. The principle of this method is based on a finite set of possible values of T. For each solved instance of the p-median problem we have tested each value of T from the mentioned set. As the optimal value of T we have taken the value for which the accuracy of the covering solution was the highest. The usefulness of this approach depends on the values included in the finite set. It is obvious that this method can give us an estimation of the relation between the ratio of $|I|$ to p and the optimal value of T, but the estimation is not very punctual. On the other hand it also shows that the accuracy of the solution depends on the value of mentioned parameter.

The adaptive algorithm suggested for finding the optimal value of T is based on an iterative approach. The basic idea of the algorithm is built on computing the dividing points for the actual value of T, solving the covering model and comparing the solution to the best found. According to the results of the actual iteration and to the comparison of the solution to the previous one, the parameters of the model are changed to achieve better solution in the next step. The keystone of mentioned algorithm is described by the following scheme.

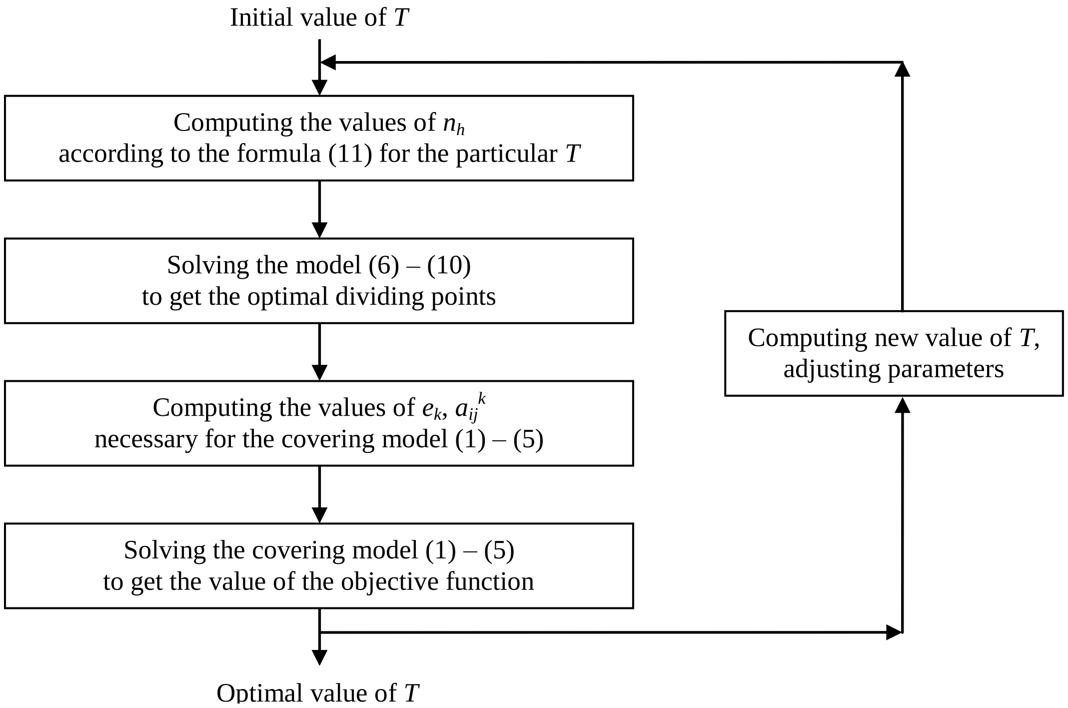


Figure 2 Scheme of the adaptive algorithm

The adaptive algorithm seems to be very simple. The only problem is to answer the question when the looping process should stop and how the model parameters should be changed during the iterations.

In our implementation of the algorithm, there are two criteria to stop the looping process. The first criterion is the maximal total number of iterations and the second one is the maximal number of consecutive iterations without any actualization of the best found solution.

The change of the parameter T depends on a probability denoted as pr and a random number $rnd \in <0, 1)$. If the generated random number rnd is smaller or equal than the value of pr, then the value of T in the next iteration (T_{next}) is computed from the actual value (T_{act}) according to the formula (12), otherwise according to the formula (13). The symbol *del* denotes positive value of the step size. In the other words, with probability equal to pr we increase the value of "*temperature*", otherwise we decrease the value of this parameter. The initial values of T, pr and *del* are parameters of this adaptive algorithm.

$$T_{next} = T_{act} + del \tag{12}$$

$$T_{next} = \frac{T_{act}}{del + 1} \tag{13}$$

The value of parameter *del* is also changed during the iteration process. Its new value (del_{next}) depends on its actual value (del_{act}) and also on w that is other parameter of the algorithm. The relation between w and *del* is the following:

$$del_{next} = 2 * |w| * del_{act} \tag{14}$$

The parameter w does not influence only the value of *del*. It is used also to change the value of the probability pr. If the value of w is smaller than -1, then we set the value of pr to 0. If the value of w exceeds 1, then we set the value of pr to 1. Otherwise we compute the probability according to the formula (15).

$$pr = \frac{w+1}{2} \tag{15}$$

The initial value of w is up to the concrete implementation of this algorithm. If we have at least two solutions to compare (the total number of iterations is higher than 1), then we regulate the value of w in each iteration according to the following relation.

$$w_{next} = b * w_{act} + c * sign\ ((F(T_{act}) - F(T_{prev})) * (T_{act} - T_{prev})) \tag{16}$$

The symbol w_{next} represents the next value of w and w_{act} is used for the actual value of this parameter. Arguments b and c are parameters of the algorithm. Parameter b is called *forgetting parameter* and the second one is called *intensity of learning*. $F(T_{act})$ represents the objective function value of the actual solution and the symbol $F(T_{prev})$ is used for the objective function value of the previous solution (solution of the previous iteration). Similarly, T_{act} is the actual value of "*temperature*" and T_{prev} represents the value of T in the previous iteration. The main principle of suggested algorithm comes from the adaptive algorithm of random searching described in [3].

5 Numerical experiments

We have suggested and realized a sequence of numerical experiments to prove the usefulness of suggested adaptive algorithm and to find out the relation between the size of the problem and appropriate value of parameter T.

All the experiments have been performed on a computer equipped with the Intel Core 2 6700 processor with parameters: 2.66 GHz and 3 GB RAM. As a source of data we used the *OR–Lib* set of the p-median problem instances [1]. The results of our experiments are given in Table 1. Each row represents one solved instance of the problem. The first column is dedicated to the name of used benchmark file. The size of the p-median instance is defined by $|I|$. The number of possible facility locations is equal to the number of served objects in all the solved instances. The column p is dedicated to the maximal number of located facilities. The ratio of $|I|$ to p is given in the fourth column of the table. The exact solution was obtained by the universal IP-solver XPRESS-IVE [10] which is used very often to solve various optimization problems. The next two columns contain the results of the first method of parameter T adjustment. These results are published also in [8]. The number of dividing points was set to 20. The optimal value of T was selected from the finite set of values *{1, 5, 10, 100, 1000}*. Since the approximate covering model provides only estimation of the original objective function, the table does not contain the approximate value. The real value of the objective function is given in the *RealObjF* column. The best value of "*temperature*" is given in *BestT*. The right part of the table contains the results of presented adaptive algorithm. We have made three sets of tests where the initial value of *del* was set to 1, 10 and 100. The other parameters of the algorithm were: initial value of w was set to 0, the value of *forgetting parameter* was equal to 0,5 and the *intensity of learning* was set to 0,5. The initial value of pr was equal to 0,5. The maximal number of iterations was set to 300 and the maximal number of iterations without any change of the best solution was 50.

File	|I|	p	|I| / p	Exact solution	T ∈ {1, 5, 10, 100, 1000}		Initial del = 1		Initial del = 10		Initial del = 100	
					RealObjF	Best T	RealObjF	Best T	RealObjF	Best T	RealObjF	Best T
pmed8	200	100	2,000	**933**	**933**	1	**933**	1,000	**933**	1,000	**933**	1,000
pmed31	700	292	2,397	**1289**	**1289**	1	**1289**	1,000	**1289**	1,000	**1289**	1,000
pmed5	100	33	3,030	**1434**	**1441**	10	**1441**	6,937	**1436**	11,000	**1436**	8,504
pmed16	400	128	3,125	**1750**	**1750**	1	**1750**	1,000	**1750**	1,000	**1750**	1,000
pmed16	400	112	3,571	**2035**	**2035**	1	**2035**	1,000	**2035**	1,000	**2035**	1,000
pmed16	400	96	4,167	**2367**	**2375**	5	**2368**	2,750	**2368**	3,285	**2368**	3,201
pmed29	600	120	5,000	**3016**	**3016**	1	**3016**	1,000	**3016**	1,000	**3016**	1,000
pmed12	300	54	5,556	**3085**	**3100**	10	**3087**	5,358	**3087**	5,177	**3091**	4,008
pmed12	300	45	6,667	**3475**	**3485**	10	**3483**	5,875	**3485**	10,008	**3482**	8,593
pmed8	200	20	10,000	**4459**	**4482**	100	**4500**	9,797	**4459**	10,893	**4459**	12,851
pmed12	300	22	13,636	**4963**	**4989**	10	**4972**	13,047	**4972**	12,149	**4972**	12,499
pmed12	300	15	20,000	**5761**	**5765**	100	**5761**	13,406	**5761**	12,754	**5761**	6,364
pmed12	300	10	30,000	**6645**	**6651**	100	**6645**	6,453	**6645**	6,403	**6645**	4,735
pmed16	400	5	80,000	**8068**	**8068**	100	**8069**	5,092	**8068**	15,008	**8068**	7,121
pmed21	500	5	100,000	**9380**	**9390**	1000	**9380**	6,125	**9380**	11,000	**9380**	25,665

Table 1 Results of numerical experiments

The results of the first tested method show there is a relation between the ratio of $|I|$ to p and the best value of T. The higher is the ratio of $|I|$ to p, the higher should be the value of T. It is very important to note that these results are not enough valid, because the finite set of possible values of T was too small and therefore the results can not be used to make some relevant conclusions.

The results of our suggested adaptive algorithm show that the mentioned relation between the problem size and the value of T is not as clear as it seems to be. It is obvious that there is some relation between the problem size and the optimal value of T, but it is interesting enough to become a topic of a future possible research which has to contain much more experiments with more values of the explored ratio. The numerical experiments prove that the presented algorithm is very useful. It provides very good solution in admissible time. The deviation between the optimal solution and the best solution obtained by the approximate covering technique is very small. These features make the adaptive algorithm very appropriate to be applied in the approximate approach to the p-median problem.

6 Conclusion

The aim of this contribution was to present an adaptive algorithm of parameter adjustment for the approximate approach to large instances of the p-median problem. The presented results of numerical experiments prove that the approximate covering models constitute a promising approach to large-scaled p-median problems which have been resisted to recent attempts to their solving. As we have shown, a very satisfactory accuracy of the covering solution can be achieved by the appropriate selection of the parameter T. Therefore the presented algorithm can be considered very useful. It provides very good accuracy of the solution by finding appropriate value of the model parameter.

The presented results prove that there is a relation between the p-median problem size and the most appropriate value of the parameter T. It is obvious that if we consider small value of p, then we have to take into account also high values of the distances. So the expected frequency of higher distances must be higher. On the other hand, if the value of p is very high, then it is possible to show that the small distances are more frequently used in comparison to the higher ones. Since this problem is very interesting, it will become a topic of a future possible research.

Acknowledgements

This work was supported by the research grant VEGA 1/0361/10 „*Optimal design of public service systems under uncertainty*“.

References

[1] Beasley J. E.: OR Library: *Distributing Test Problems by Electronic Mail*, Journal of the Operational Research Society, 41(11), 1069-1072

[2] Doerner, K.F. et al.: *Heuristic solution of an extended double-coverage ambulance location problem for Austria*. Central European Journal of Operations Research, 2005, Vol. 13, No 4, pp 325-340

[3] Janáček, J.: *Operační analýza I*, učebné texty, Alfa Bratislava, 1985, 202 p.

[4] Janáček, J.: *Approximate covering models of location problems*. In: Lecture Notes in Management Science. Proceedings of the 1st International Conference on Applied Operational Research – ICAOR ´08, Vol. 1, Sept. 2008, Yerevan, Armenia, pp 53-61

[5] Janáček, J., Jánošíková, Ľ.: *Computability of the emergency service system design problem*. In: Communications - Scientific Letters of the University of Žilina, 2008,Vol. 10, No. 2, ISSN 1335-4205, pp. 5-9

[6] Jánošíková, Ľ.: *Emergency medical service planning*. In: Communications –Scientific Letters of the University of Žilina, 2007, Vol. 9, No 2, ISSN 1335-4205, pp 64-68

[7] Kvet, M.: *Aproximácia matice vzdialenosti pre približné riešenie rozsiahlych úloh p-mediánu*. In: Proceedings of the conference INFOTRANS 2011, Pardubice, 2011, ISBN 978-80-7395-397-3, pp. 123-128

[8] Kvet, M.: *Static Zone Adjustment for Approximate Solving of the p-Median Problem*. In: Proceedings of the conference TRANSCOM 2011, Section 1 – Transport and Communications Technology, Žilina, June 27-29 The University of Žilina, 2011, ISBN 978-80-554-0369-4, pp 129-132

[9] Marianov, V., Serra, D.: *Location problems in the public sector*. In: Drezner Z (ed.) et al.: *Facility location. Applications and theory*. 2002, Berlin: Springer, pp 119-150

[10] XPRESS XPRESS-Mosel “*User guide*”, Dash Associates, Blisworth, 2005, UK, 99 p.

Analyzing Behaviour of Czech Economy: A Nonlinear DSGE Model Framework

Zdeněk Liščinský[1]

Abstract. In this paper, we analyze behaviour of nonlinear New Keynesian dynamic stochastic general equilibrium model of the Czech Republic. We solve the model by second-order perturbation method and estimate it using Bayesian techniques, specifically evaluate its likelihood function with particle filter. Our results show that the nonlinear model reveals richer dynamics as compared to linearized model: impulse response functions to structural shocks differ, depending on the distance of the initial state vector from steady state.

Keywords: DSGE models, nonlinear approximation, particle filter, impulse response functions

JEL classification: C11, C51
AMS classification: 91B51

1 Introduction

In this paper, we study the effect of nonlinearities on the performance of a DSGE model of the Czech Republic. In particular, we solve the model by second-order perturbation method and estimate it using Bayesian techniques, specifically evaluate its likelihood function with particle filter. Finally, we analyze its behaviour using impulse response functions and compare it with impulse responses obtained from the linearized version of the same model.

Most of the existing literature on DSGE models deals only with linearized solution of these models. This approach is sufficient in many applications, but there are at least two reasons to be interested also in their nonlinear features.

The first one is that nonlinear models can provide more accurate estimates of the structural parameters and considerably better data fit (measured by marginal likelihood), compared to their linearized counterparts – as documented by Fernández-Villaverde and Rubio-Ramírez [5], An and Schorfheide [3] or An [2].

The second reason is connected with macroeconomic dynamics of linearized models – they are, by construction, suitable only for analyzing impacts of small deviations from steady state. When a large shock occurs, they might deliver misleading results, as shown in Amisano and Tristani [1].

The rest of this paper is organized as follows: Section 2 briefly describes the model, Sections 3 and 4 outline solution and estimation method, respectively. Section 5 presents the results and Section 6 concludes.

2 The Model

For our purpose we employed a small-scale model based on An and Schorfheide [3]. In the interest of space we present only a brief description of the model; we refer the reader to the cited paper for more detailed discussion.

The model economy consists of a representative household, a continuum of intermediate goods producing firms, a final goods producing firm, and a monetary as well as a fiscal authority. The representative household consumes final goods and supplies labour. Monopolistically competitive firms in the intermediate goods sector face nominal rigidities in the form of quadratic price adjustment costs. The monetary

[1]Masaryk University, Department of Economics, Lipová 41a, 60200 Brno, e-mail: 63335@mail.muni.cz

authority sets nominal interest rate according to a Taylor type rule and the government spending is considered to be exogenous.

The model's equilibrium conditions are given by the following system of equations

$$\begin{aligned}
1 &= E_t\left[\exp\left(-\tau\left(\hat{c}_{t+1}-\hat{c}_t\right)+\hat{r}_t-\hat{\pi}_{t+1}-\hat{z}_{t+1}\right)\right], && \text{(1a)}\\
0 &= \frac{1-\nu}{\nu\phi\left(\pi^{ss}\right)^2}\left(1-\exp\left(\tau\hat{c}_t\right)\right)+\left(\exp\left(\hat{\pi}_t\right)-1\right)\left[\left(1-\frac{1}{2\nu}\right)\exp\left(\hat{\pi}_t\right)+\frac{1}{2\nu}\right]\\
&\quad -\beta E_t\left[\left(\exp\left(\hat{\pi}_{t+1}\right)-1\right)\exp\left(-\tau\left(\hat{c}_{t+1}-\hat{c}_t\right)+\hat{y}_{t+1}-\hat{y}_t+\hat{\pi}_{t+1}\right)\right], && \text{(1b)}\\
0 &= \exp(-\hat{g}_t)-\exp(\hat{c}_t-\hat{y}_t)-\frac{1}{2}\phi g^{ss}(\pi^{ss})^2\left(\exp(\hat{\pi}_t)-1\right)^2, && \text{(1c)}\\
\hat{r}_t &= \rho_r\hat{r}_{t-1}+\left(1-\rho_r\right)\psi_1\hat{\pi}_t+\left(1-\rho_r\right)\psi_2\left(\hat{y}_t-\hat{g}_t\right)+\varepsilon_{r,t}, && \text{(1d)}\\
\hat{g}_t &= \rho_g\hat{g}_{t-1}+\varepsilon_{g,t}, && \text{(1e)}\\
\hat{z}_t &= \rho_z\hat{z}_{t-1}+\varepsilon_{z,t}, && \text{(1f)}
\end{aligned}$$

where variables with hats are deviations from the steady state. The structural shocks ε_r, ε_g and ε_z are Gaussian with mean zero and standard deviation σ_r, σ_g and σ_z, respectively. Parameters of the model are described in Table 1.

Parameter	Description	Domain
β	Discount factor	$(0,1)$
τ^{-1}	Elasticity of intertemporal substitution	$\mathbb{R}^+$
ν^{-1}	Elasticity of demand for intermediate goods	$\mathbb{R}^+$
ϕ	Measure of price stickiness	$\mathbb{R}^+$
ψ_1	Elasticity of interest rate to inflation	$\mathbb{R}^+$
ψ_2	Elasticity of interest rate to output	$\mathbb{R}^+$
ρ_r	Interest rate smoothness	$[0,1)$
ρ_g	Persistence of government spending shock	$[0,1)$
ρ_z	Persistence of technology shock	$[0,1)$
π^{ss}	Steady state inflation rate	$\mathbb{R}^+$
γ^{ss}	Steady state growth rate of technology	$\mathbb{R}^+$

Table 1: Model Parameters

We use 3 time series to estimate the model: growth rate of seasonally adjusted real output per capita (YGR_t), CPI inflation per quarter ($INFL_t$), and 3-month PRIBOR interest rate (INT_t). The observed data are linked with the model variables via measurement equations

$$\begin{aligned}
YGR_t &= \gamma^{ss}+\hat{y}_t-\hat{y}_{t-1}+\hat{z}_t+w_{y,t}, && \text{(2a)}\\
INFL_t &= \pi^{ss}+\hat{\pi}_t+w_{\pi,t}, && \text{(2b)}\\
INT_t &= \pi^{ss}+100(\beta^{-1}-1)+\gamma^{ss}+\hat{r}_t+w_{r,t}, && \text{(2c)}
\end{aligned}$$

where w_y, w_π and w_r denote Gaussian measurement errors with mean zero and standard deviations σ_{my}, $\sigma_{m\pi}$ and σ_{mr}, respectively.

3 Model Solution

This section presents a brief description of the perturbation method used to solve the model. The model equations (1) can be expressed in the compact form

$$E_t\left[\mathbf{f}\left(\mathbf{y}_{t+1},\mathbf{x}_{t+1},\mathbf{y}_t,\mathbf{x}_t\right)\right]=\mathbf{0}, \tag{3}$$

where $\mathbf{x}_t=[\hat{y}_{t-1},\hat{r}_{t-1},\varepsilon_{r,t},\hat{g}_t,\hat{z}_t]'$ is the vector of predetermined (state) variables and $\mathbf{y}_t=[\hat{c}_t,\Delta\hat{y}_t,\hat{\pi}_t]'$ is the vector of control variables.

Following Gomme and Klein [6], we introduce perturbation parameter σ which scales the matrix $\mathbf{J}$ containing standard deviations of structural shocks; particularly, setting $\sigma = 0$ leads to non-stochastic system. The exact solution to (3) is then given by

$$\mathbf{y}_t = \mathbf{g}\left(\mathbf{x}_t, \sigma\right) \tag{4a}$$
$$\mathbf{x}_{t+1} = \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma \mathbf{J} \mathbf{v}_{t+1}, \tag{4b}$$

where $\mathbf{v}_{t+1} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_{n_v})$ denotes the vector of innovations. As functions $\mathbf{g}$ and $\mathbf{h}$ cannot be computed analytically, we approximate them by Taylor expansion up to the second order around non-stochastic steady state:

$$\tilde{\mathbf{g}}\left(\mathbf{x}_t, \sigma\right) = \frac{1}{2}\sigma^2 \mathbf{g}_{\sigma\sigma} + \mathbf{G}_x \mathbf{x}_t + \frac{1}{2}\left(\mathbf{I}_{n_y} \otimes \mathbf{x}_t'\right) \mathbf{G}_{xx} \mathbf{x}_t \tag{5a}$$
$$\tilde{\mathbf{h}}\left(\mathbf{x}_t, \sigma\right) = \frac{1}{2}\sigma^2 \mathbf{h}_{\sigma\sigma} + \mathbf{H}_x \mathbf{x}_t + \frac{1}{2}\left(\mathbf{I}_{n_x} \otimes \mathbf{x}_t'\right) \mathbf{H}_{xx} \mathbf{x}_t, \tag{5b}$$

where $\mathbf{g}_{\sigma\sigma}$, $\mathbf{G}_x$, $\mathbf{G}_{xx}$ and $\mathbf{h}_{\sigma\sigma}$, $\mathbf{H}_x$, $\mathbf{H}_{xx}$ are the derivatives[1] of $\mathbf{g}$ and $\mathbf{h}$, respectively, all evaluated at steady state, and the symbol $\otimes$ denotes Kronecker product. To compute these unknown matrices we can proceed in the following way: at first we plug the proposed solution (4) into (3) and define function F

$$F\left(\mathbf{x}_t, \sigma\right) \equiv E_t\left(\mathbf{g}\left(\mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma \mathbf{J} \mathbf{v}_{t+1}, \sigma\right), \mathbf{g}\left(\mathbf{x}_t, \sigma\right), \mathbf{h}\left(\mathbf{x}_t, \sigma\right) + \sigma \mathbf{J} \mathbf{v}_{t+1}, \mathbf{x}_t\right). \tag{6}$$

It is clear that $F\left(\mathbf{x}_t, \sigma\right) = \mathbf{0}$ and therefore also $\mathrm{D}(F) = \mathbf{0}$. Based on this we can obtain matrices $\mathbf{H}_x$ and $\mathbf{G}_x$ which characterize linear approximation of (4). Second-order terms $\mathbf{g}_{\sigma\sigma}$, $\mathbf{G}_{xx}$, $\mathbf{h}_{\sigma\sigma}$ and $\mathbf{H}_{xx}$ can be consequently computed from $\mathrm{H}(F) = \mathbf{0}$. We refer to Gomme and Klein [6] or Schmitt-Grohé and Uribe [8] for detailed description of this solution method.

4 Estimation procedure

The measurement equations (2) and the approximate model solution can be cast in the following state-space representation

$$\mathbf{y}_t^o = G\left(\mathbf{s}_t, \mathbf{w}_t; \theta\right) \tag{7a}$$
$$\mathbf{s}_{t+1} = H\left(\mathbf{s}_t, \mathbf{v}_{t+1}; \theta\right), \tag{7b}$$

where $\mathbf{s}_t = [\mathbf{x}_t', \mathbf{y}_t']'$ collects the model (predetermined and control) variables, $\mathbf{y}_t^o$ is the vector of observables, $\mathbf{w}_t = [w_{y,t}, w_{\pi,t}, w_{r,t}]'$ and vector θ contains model parameters.

Our goal is to obtain posterior distribution of parameters according to the Bayes rule in the form

$$p\left(\theta \,|\, \mathbf{y}_{1:T}^o\right) \propto p\left(\mathbf{y}_{1:T}^o \,|\, \theta\right) p\left(\theta\right), \tag{8}$$

where $p\left(\mathbf{y}_{1:T}^o \,|\, \theta\right)$ denotes likelihood function, $p\left(\theta\right)$ denotes prior distribution and $\mathbf{y}_{1:T}^o = \{\mathbf{y}_t^o\}_{t=1}^T$ collects all observed data up to time T.

In order to compute the likelihood function, we can rewrite it as

$$p\left(\mathbf{y}_{1:T}^o \,|\, \theta\right) = \prod_{t=1}^{T-1} p\left(\mathbf{y}_{t+1}^o \,|\, \mathbf{y}_{1:t}^o, \theta\right) = \prod_{t=1}^{T-1} \int p\left(\mathbf{s}_{t+1} \,|\, \mathbf{y}_{1:t}^o, \theta\right) p\left(\mathbf{y}_{t+1}^o \,|\, \mathbf{s}_{t+1}, \theta\right) \mathrm{d}\mathbf{s}_{t+1}. \tag{9}$$

If the system (7) is linear and the shocks are normally distributed, this integral can be solved analytically, which leads to Kalman filter recursion. In the case of nonlinear system we approximate (9) by the standard particle filter (i.e. the specific version of the filter which uses the state transition distribution as the proposal distribution.[2]) The algorithm can be summarized as follows

0. (Initialization, $t = 0$): Draw N particles $\left\{\hat{\mathbf{s}}_0^{(i)}\right\}_{i=1}^N$ from $p\left(\mathbf{s}_0\right)$ and set $w_0^{(i)} = \frac{1}{N}$ for all i.

[1] We define Hessian matrix of twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$ as $mn \times n$ matrix $\mathrm{H}(f) = \partial^2 f(x)/\partial x \partial x' = \mathrm{D}\left(\mathrm{vec}\left(\mathrm{D}\left(f\right)\right)\right)$, where $\mathrm{D}(f) = \partial f(x)/\partial x'$ is $m \times n$ matrix of first-order derivatives.

[2] Doucet et al. [4] provides comprehensive review of various algorithms as well as relevant theoretical background.

1. (Prediction): Draw N particles $\left\{\mathbf{s}_{t+1}^{(i)}\right\}_{i=1}^{N}$ from $p\left(\mathbf{s}_{t+1} \,|\, \hat{\mathbf{s}}_t^{(i)}, \theta\right)$.

2. (Filtration): Assign a weight $w_{t+1}^{(i)}$ to each particle $\mathbf{s}_{t+1}^{(i)}$: $w_{t+1}^{(i)} = w_t^{(i)} p\left(\mathbf{y}_{t+1}^{o} \,|\, \mathbf{s}_{t+1}^{(i)}, \theta\right)$, $i = 1, \ldots, N$.

3. Resample with replacement the draws $\mathbf{s}_{t+1}^{(i)}$ using probabilities $\tilde{w}_{t+1}^{(i)} = w_{t+1}^{(i)} / \sum_{j=1}^{N} w_{t+1}^{(j)}$ to obtain new sample $\left\{\hat{\mathbf{s}}_{t+1}^{(i)}\right\}_{i=1}^{N}$ approximately distributed according to $p\left(\mathbf{s}_{t+1} \,|\, \mathbf{y}_{1:t+1}\right)$. If $t < T$, set $t = t+1$ and go to 1.

The main object of interest is the sequence of unnormalized weights computed in Step 2. It can be shown that

$$\sum_{t=0}^{T-1} \sum_{j=1}^{N} w_{t+1}^{(j)} \approx p\left(\mathbf{y}_{t+1}^{o} \,|\, \mathbf{y}_{1:t}^{o}, \theta\right). \tag{10}$$

5 Results

Our results are based on quarterly Czech data (described in Section 2), the sample covers the period from 1999Q1 to 2010Q4. For both linear and quadratic approximation we simulated 100,000 draws from posterior density using Random walk Metropolis-Hastings algorithm and we discarded first 50,000 simulations to get rid of effect of initial conditions. We used 40,000 particles in the particle filtering step. All computations were implemented in Matlab and Dynare++.

We estimated all parameters except for the discount factor β, which was calibrated to value 0.995. The results of the estimation are summarized in Table 2. (Note that in the linear approximation the parameters ν and ϕ are not separately identifiable, therefore we expressed the model in terms of κ, defined as $\kappa = \frac{1-\nu}{\nu\phi\pi^{ss2}}$.)

A brief look at the means of posterior distributions suggests that the estimates obtained from linear and quadratic approximation do not differ significantly. However, the estimates based on the nonlinear model are more precise: standard deviations of all parameters except for steady state inflation π^{ss} are lower. This result is consistent with previous findings in the literature.

Finally, we turn to impulse response functions. It is a well-known fact that impulse responses of linear models are symmetric (i.e. the opposite shock of the same magnitude has exactly the opposite effect on the response), scalable (a shock of double magnitude has twice as much effect) and don't depend on initial conditions.[3] In order to explore the dynamics of nonlinear model, we focus on a situation when a shock hits the economy which is already far from its steady state. Specifically, the largest deviation in our data sample occurs in 2007Q4 when the output was 4.1 % above its steady state value. We chose the filtered state vector on that date as the starting point for calculating impulse responses.

The resulting impulse response functions are depicted in Figure 1. The red solid line represents the response of linearized model to one unit technology shock ($\sigma_z = 0.25\,\%$) and the blue dashed line shows the response of the nonlinear model to the same shock, but starting from the state vector described above. We can see that the amplitude and persistence of the responses to the shock are higher in the case of the nonlinear model.

6 Conclusion

In our contribution, we estimated a small New Keynesian model of the Czech economy approximated by second-order perturbation method and we analyzed its behavior using impulse response functions. Our results show that the nonlinear model reveals richer dynamics as compared to linearized model: impulse response functions to structural shocks differ depending on the distance of the initial state vector from the steady state.

[3] For more details see e.g. Koop, Pesaran and Potter [7].

	Prior Distribution			Linear/Kalman				Quadratic/Particle			
Param.	Density	Mean	Std	Mean	Std	90% Interval		Mean	Std	90% Interval	
τ	Gamma	2.000	0.500	**2.060**	0.415	1.549	2.589	**1.907**	0.346	1.398	2.330
ν	Beta	0.100	0.050	**0.098**	0.053	0.018	0.181	**0.091**	0.045	0.023	0.152
κ	Gamma	3.000	0.500	**2.977**	0.181	2.726	3.211	**3.008**	0.128	2.829	3.184
ψ_1	Gamma	1.500	0.250	**1.473**	0.090	1.343	1.611	**1.530**	0.035	1.478	1.582
ψ_2	Gamma	0.500	0.250	**0.380**	0.138	0.181	0.603	**0.662**	0.125	0.474	0.881
ρ_r	Beta	0.800	0.100	**0.848**	0.050	0.777	0.912	**0.862**	0.024	0.825	0.902
ρ_g	Beta	0.850	0.100	**0.805**	0.111	0.648	0.994	**0.895**	0.046	0.819	0.970
ρ_z	Beta	0.800	0.100	**0.914**	0.022	0.983	0.944	**0.936**	0.007	0.925	0.947
π^{ss}	Gamma	3.000	0.500	**2.795**	0.419	2.202	3.502	**3.201**	0.554	2.258	3.917
γ^{ss}	Gamma	0.600	0.250	**0.419**	0.131	0.208	0.632	**0.742**	0.117	0.552	0.940
$100\sigma_r$	InvGam	0.300	0.300	**0.148**	0.046	0.070	0.226	**0.193**	0.036	0.136	0.248
$100\sigma_g$	InvGam	0.500	0.300	**0.166**	0.214	0.009	0.380	**0.121**	0.115	0.010	0.274
$100\sigma_z$	InvGam	0.500	0.300	**0.263**	0.016	0.239	0.286	**0.251**	0.009	0.237	0.264

Table 2: Parameter Estimates

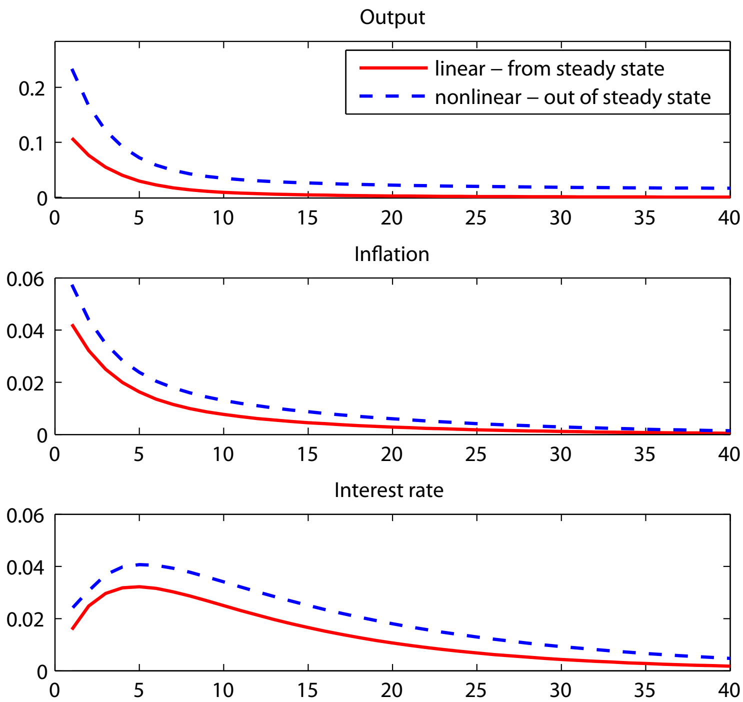


Figure 1: Impulse response functions to technology shock

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009.

References

[1] Amisano, G. and O. Tristani (2007): Euro area inflation persistence in an estimated nonlinear DSGE model, *Working Paper Series* 754, European Central Bank.

[2] An, S. (2006): Bayesian Estimation of DSGE Models: Lessons from Second-Order Approximations, *Working paper*, available at `http://www.mysmu.edu/faculty/sungbae/research/`

[3] An, S. and F. Schorfheide (2006): Bayesian Analysis of DSGE Models, *Econometric Reviews* 26, 113–172.

[4] Doucet. A., N. de Freitas and N. Gordon (2001): *Sequential Monte Carlo Methods in Practice*, Springer Verlag.

[5] Fernández-Villaverde, J. and J. Rubio-Ramírez (2005): Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood, *Journal of Applied Econometrics* 20, 891–910.

[6] Gomme, P. and P. Klein (2008): Second-order approximation of dynamic models without the use of tensors, *Working paper*, available at `http://alcor.concordia.ca/~pgomme/secondorder.pdf`

[7] Koop, G., M. H. Pesaran and S. Potter (1996): Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* 74, 119–147.

[8] Schmitt-Grohé, S. and M. Uribe (2004): Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function, *Journal of Economic Dynamics and Control* 28, 755–775.

Weight generation and construction of preferences by Monte Carlo method in AHP with inconsistent interval comparison matrices

Ladislav Lukáš[1]

Abstract. Paper presents Monte Carlo method and its numerical implementation for generation of weights and construction of preferences used in AHP to support multiple criteria decision making. However, in case of inconsistent interval comparison matrices the problem of weight generation is much more difficult. An approach based on consistency index is widely used in such case. Our procedure constructs a suitable measure of consistency as a relation between consistency index and some spectral characteristics depending on right eigenvector of comparison matrix. Using Monte Carlo method matrix samples are generated from the interval comparison matrix. The spectral-problems for such matrices are solved and the leading eigenvalues extracted. These subsequently serve for estimation of weights extremal bounds. The computer implementation is presented in detail, and numerical results discussed, too.

Keywords: Analytic hierarchical process, inconsistent interval comparison matrix, generation of weights, spectral method, criteria for decision-making.

JEL Classification: C02, C65, C81, D01, D81
AMS Classification: 15A18, 65G20, 90C29, 91B06

1 Introduction

The analytic hierarchy process (AHP) is one of the popular methods for decision-making in practice. Basic AHP procedure consists of two steps – first one, forming pairwise comparison matrix, and the second one, generation of weights for selected objectives applied, which lead next for preferences formulation and thus ordering of objectives. There exist a huge number of papers and books relating the subject, which range from pure theoretical works to applications as well. First, we refer the seminal book of prof. Saaty [5]. For further details we refer [3] and [6], and [2], [4] and [7] as taking special aspects being devoted to inconsistency of comparison matrices.

2 Weight generation

Since the early stages of AHP generation of weights plays very important role. In its original form, the AHP adopts a simple scale {1, 3, 5, 7, 9} for measuring comparison relations between two objectives ranging from equal importance till absolutely more importance.

Let us write a matrix of pairwise comparisons of n objectives in the following well-known form

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \vdots & a_{nn} \end{bmatrix}, \tag{1}$$

where $a_{ij} > 0$, $a_{ji} = (a_{ij})^{-1}$, $i,j=1,..,n$, $j < i$, i.e. the elements a_{ij} and a_{ji} are standing in the reciprocity each other.

We know that matrix $\boldsymbol{A}$ is consistent if it can be transformed into the following form

$$\boldsymbol{W} = \begin{bmatrix} 1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & 1 & \cdots & w_2/w_n \\ \vdots & \vdots & & \vdots \\ w_n/w_1 & w_n/w_2 & \vdots & 1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \begin{bmatrix} 1/w_1 & 1/w_2 & \ldots & 1/w_n \end{bmatrix}, \tag{2}$$

[1] Univ.of West Bohemia, Faculty of Econ., Dept. of Econ. and Quantitative Methods, Husova 11, 30614 Pilsen, lukasl@kef.zcu.cz.

where $w_i > 0$ represents an estimated weight of i-th objective, $i=1,..,n$. So, we may write $\boldsymbol{W} = \boldsymbol{w}\boldsymbol{\omega}^{\mathrm{T}}$, with vector $\boldsymbol{w}$ containing w_i and vector $\boldsymbol{\omega}$ their reciprocals. Note that both matrices $\boldsymbol{A}$, $\boldsymbol{W}$ are positive. Further, the matrix theory assures that matrix $\boldsymbol{W}$ has rank($\boldsymbol{W}$) = 1, so the only one non-zero eigenvalue λ_W exists, and it holds

$$\lambda_W = \mathrm{tr}(\boldsymbol{W}) = n. \tag{3}$$

Theory of AHP provides widely used measure for inconsistency of a pairwise comparison matrix $\boldsymbol{A}$ known as consistency index c_{I}

$$c_{\mathrm{I}} = (\lambda_{\max} - n)/(n - 1), \tag{4}$$

where $\lambda_{\max}$ is the greatest eigenvalue of a spectral problem

$$\boldsymbol{A}\boldsymbol{w} = \lambda\boldsymbol{w}. \tag{5}$$

Hence, solution of eigenproblem is closely related with construction of weights. The vector of searched objective weights $\boldsymbol{w}$ is nothing else but the right eigenvector of matrix $\boldsymbol{A}$ corresponding to $\lambda_{\max}$. From practical point of view it is useful to norm the right eigenvector $\boldsymbol{w}$ by relation $\sum_{i=1}^{n} w_i = 1$, thus providing natural interpretation of weights.

However, c_{I} given by (4) depends upon n which complicates versatility of consistency index in many practical situations. So, the theory of AHP provides another quantity called consistency ratio of a comparison matrix given by following expression

$$c_{\mathrm{R}} = c_{\mathrm{I}} / r_{\mathrm{I}}(n), \tag{8}$$

where $r_{\mathrm{I}}(n)$ denotes a mean value of c_{I} which is regarded as random variable calculated by formula (4) from random matrices formed by comparing fractions $w_{\mathrm{i}}/w_{\mathrm{j}}$, i.e. $r_{\mathrm{I}}(n) = \mathrm{E}(c_{\mathrm{I}})$, where E is an expectation operator.
The quantity $c_{\mathrm{R}} \in [0,1]$ expresses suitable consistency measure of any comparison matrix, ranging from the ideal case with perfect transversality of comparison relations giving $c_{\mathrm{R}} = 0$, till a random matrix lacking any detectable transversality of comparison relations, which produces $c_{\mathrm{R}} = 1$.

In [1] we have already investigated some spectral properties of interval pairwise comparison matrices. Entries of such matrix are interval numbers, and it takes the following form

$$\boldsymbol{A} = \begin{bmatrix} 1 & [l_{12}, u_{12}] & \cdots & [l_{1n}, u_{1n}] \\ [l_{21}, u_{21}] & 1 & \cdots & \\ \vdots & \vdots & \vdots & \vdots \\ [l_{n1}, u_{n1}] & [l_{n2}, u_{n2}] & \cdots & 1 \end{bmatrix}, \tag{9}$$

where $[l_{ij}, u_{ij}]$ denotes a closed interval of real positive numbers, $l_{ij}, u_{ij} > 0$, such as $l_{ij} \le u_{ij}$, $l_{ij} = 1/u_{ji}$, and $u_{ij} = 1/l_{ji}$, $i,j=1,2,...,n$, which expresses a range of importance between i-th and j-th objectives.

Computational procedure for weight generation based upon statistical sampling technique with consistency measure control of matrices $\mathcal{A}$ sampled from $\boldsymbol{A}$ is described in [1], too. The pairwise interval comparison matrix $\boldsymbol{A}$ takes the form (9), and procedure for generation interval weights $[w_i^{\mathrm{L}}, w_i^{\mathrm{U}}]$, $i=1,2,...,n$, has five steps:

1) Build crisp matrices $\mathcal{A} = [\alpha_{ij}]$ by Monte Carlo sampling of $\boldsymbol{A}$ assuming uniform distribution of values α_{ij} over $[l_{ij}, u_{\mathrm{ij}}]$.
2) Find $\tilde{\lambda}_{\max}$ solving spectral problem of $\mathcal{A}$, and calculate sample consistency index γ_{I} by (4).
3) Accept such $\mathcal{A}$ passing adjusted threshold on its c_{R} evaluated by (8), usually $c_{\mathrm{R}} < 0.1$, with $c_{\mathrm{I}} = \gamma_{\mathrm{I}}$.
4) Solve n-tuples of LP problems for crisp matrix $\mathcal{A}$ dependent weight interval $[{}_{\mathrm{d}}w_i^{\mathrm{L}}, {}_{\mathrm{d}}w_i^{\mathrm{U}}]$.
5) Estimate weight interval $[w_i^{\mathrm{L}}, w_i^{\mathrm{U}}]$ from set of intervals $[{}_{\mathrm{d}}w_i^{\mathrm{L}}, {}_{\mathrm{d}}w_i^{\mathrm{U}}]$, $d=1,2,...,h$, generated by selected crisp matrices.

3 Computational results

We have selected two different examples of interval comparison matrices given in [7], and analyzed already in [1], too. Now, we run rumerical experiments based on Monte Carlo approach to further investigate spectral behavior of $\lambda_{\max}$ and c_{I}, which play crucial role in construction of weight intervals $[w_i^{\mathrm{L}}, w_i^{\mathrm{U}}]$, with special respect to – influence of interval consistency and reciprocity. All computations were performed by sw Mathematica 7.0, WolframResearch Inc.

3.1 First example – consistent interval pairwise comparison matrix

The interval matrix $\boldsymbol{A}$ is given by (10) and can be represented by a couple of crisp matrices $(\mathcal{A}_1, \mathcal{A}_2)$

$$\boldsymbol{A} = \begin{bmatrix} 1 & [2,5] & [2,4] & [1,3] \\ \left[\frac{1}{5},\frac{1}{2}\right] & 1 & [1,3] & [1,2] \\ \left[\frac{1}{4},\frac{1}{2}\right] & \left[\frac{1}{3},1\right] & 1 & \left[\frac{1}{2},1\right] \\ \left[\frac{1}{3},1\right] & \left[\frac{1}{2},1\right] & [1,2] & 1 \end{bmatrix},\ \mathcal{A}_1 = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 0.2 & 1 & 1 & 1 \\ 0.25 & 0.3\dot{3} & 1 & 0.5 \\ 0.3\dot{3} & 0.25 & 1 & 1 \end{bmatrix},\ \mathcal{A}_2 = \begin{bmatrix} 1 & 5 & 4 & 3 \\ 0.5 & 1 & 3 & 2 \\ 0.5 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix}. \qquad (10)$$

Using Monte Carlo approach, we build two sets of crisp matrices, denoted $\{{}_0\mathcal{B}_k\}$ and $\{{}_1\mathcal{B}_k\}$, $k=1,..,K=100$, sampled from $\boldsymbol{A}$ (10), where prefix 0 denotes a non-preserving reciprocity case, whereas 1 a reciprocity preserving case. Calculated $\lambda_{\max}$ are given on Fig. 1, and c_{I} on Fig. 2.

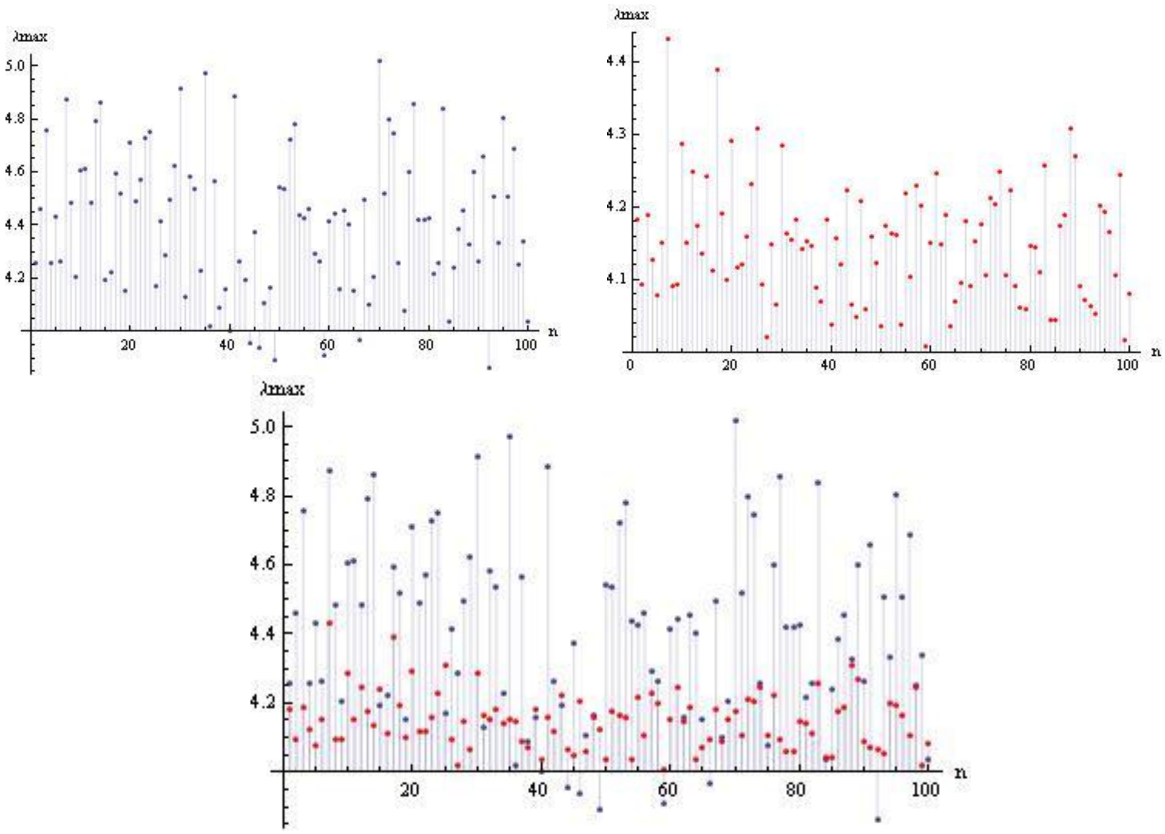


Fig. 1 Eigenvalues $\lambda_{\max}$ of matrices ${}_0\mathcal{B}_k$ ~ left, ${}_1\mathcal{B}_k$ ~ right, $k=1,..,K$, both ~ bellow

Basic characteristics are summarized on Tab. 1, which contain also a span of values given by max – min.

		μ	σ^2	σ	min	max	max – min
Case 0	$\lambda_{\max}$	4.41121	0.073252	0.270651	3.86034	5.01706	1.15672
	c_{I}	0.137069	0.00813911	0.090217	-0.0465546	0.339022	0.3855766
Case 1	$\lambda_{\max}$	4.14937	0.00651079	0.0806895	4.00734	4.43075	0.42341
	c_{I}	0.04979	0.000723421	0.0268965	0.00244724	0.143582	0.14113476

Table 1 Characteristics of $\lambda_{\max}$ and c_{I} of matrices ${}_0\mathcal{B}_k$ ~ case_0, ${}_1\mathcal{B}_k$ ~ case_1, $k=1,..,K$.

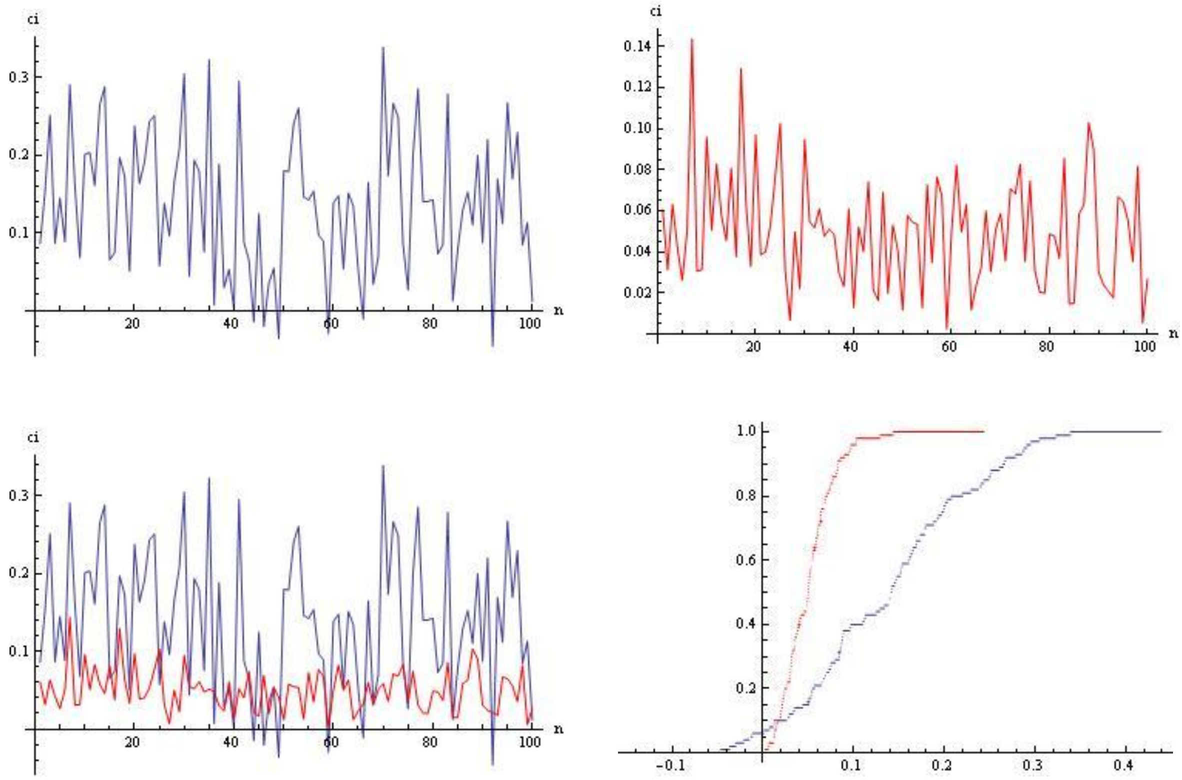


Fig. 2 Consistency index c_I of ${}_0\mathcal{B}_k$ ~ top-left, ${}_1\mathcal{B}_k$ ~ top-right, k=1,..,K,
both ~ bellow-left, EDF both ~ bellow-right

The most interesting is comparison of empirical distribution functions (EDF) calculated for values c_I of both sets of crisp matrices $\{{}_0\mathcal{B}_k\}$ and $\{{}_1\mathcal{B}_k\}$. The Monte Carlo generated matrices $\{{}_1\mathcal{B}_k\}$ which preserves reciprocity give significantly narrower span of values then provide matrices $\{{}_0\mathcal{B}_k\}$.

3.2 Second example – inconsistent interval pairwise comparison matrix

Now, the interval matrix $\boldsymbol{A}$ is given by (11) being represented by a couple of crisp matrices ($\mathcal{A}_1$, $\mathcal{A}_2$) as well.

$$\boldsymbol{A} = \begin{bmatrix} 1 & [1,2] & [1,2] & [2,3] \\ \left[\frac{1}{2},1\right] & 1 & [3,5] & [4,5] \\ \left[\frac{1}{2},1\right] & \left[\frac{1}{5},\frac{1}{3}\right] & 1 & [6,8] \\ \left[\frac{1}{3},\frac{1}{2}\right] & \left[\frac{1}{5},\frac{1}{4}\right] & \left[\frac{1}{8},\frac{1}{6}\right] & 1 \end{bmatrix},\ \mathcal{A}_1 = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 0.5 & 1 & 3 & 4 \\ 0.5 & 0.2 & 1 & 6 \\ 0.3\dot{3} & 0.2 & 0.125 & 1 \end{bmatrix},\ \mathcal{A}_2 = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 1 & 5 & 5 \\ 1 & 0.3\dot{3} & 1 & 8 \\ 0.5 & 0.25 & 0.1\dot{6} & 1 \end{bmatrix}. \qquad (11)$$

Basic characteristics are summarized on Tab. 2, which contain the same type of quantities in order to make a comparison as simple as possible.

		μ	σ^2	σ	min	max	max – min
Case 0	λ_{max}	4.63898	0.0292195	0.170937	4.31179	5.06828	0.75649
	c_I	0.212993	0.00324661	0.056979	0.103932	0.356093	0.252161
Case 1	λ_{max}	4.56443	0.00880797	0.0938508	4.35427	4.77612	0.42185
	c_I	0.188144	0.000978663	0.0312836	0.11809	0.258708	0.140618

Table 2 Characteristics of λ_{max} and c_I of matrices ${}_0\mathcal{C}_k$ ~ case_0, ${}_1\mathcal{C}_k$ ~ case_1, k=1,..,K.

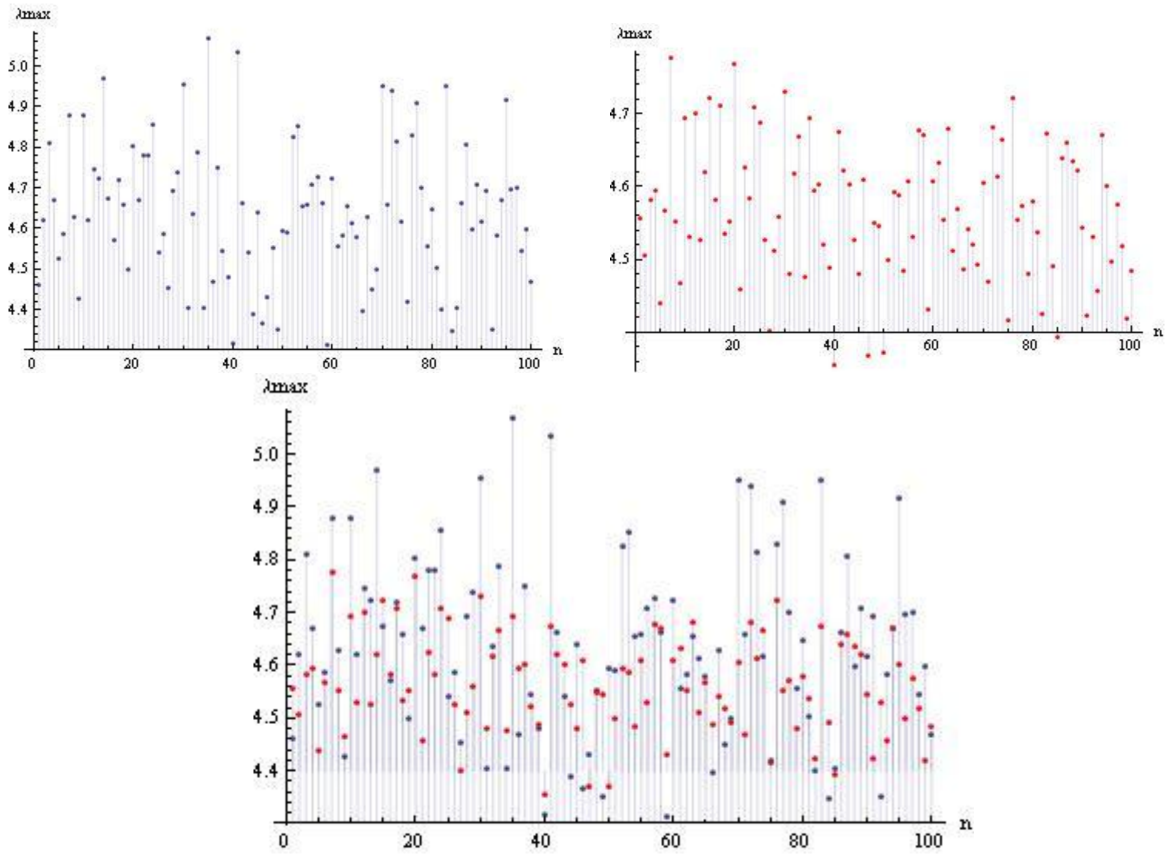


Fig. 3 Eigenvalues $\lambda_{\max}$ of matrices ${}_0\mathcal{C}_k$ ~ left, ${}_1\mathcal{C}_k$ ~ right, k=1,..,K, both ~ bellow

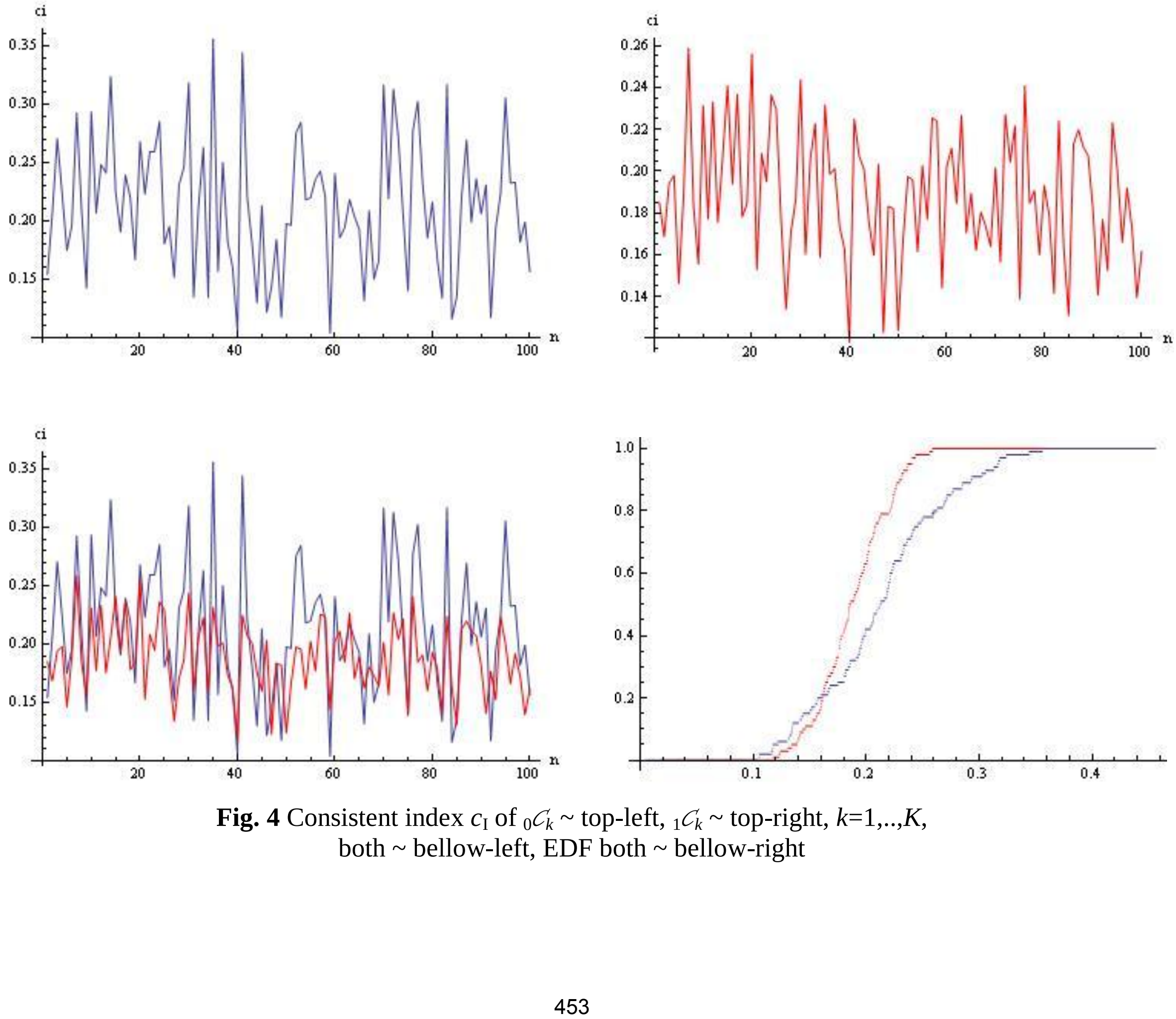


Fig. 4 Consistent index c_I of ${}_0\mathcal{C}_k$ ~ top-left, ${}_1\mathcal{C}_k$ ~ top-right, k=1,..,K,
both ~ bellow-left, EDF both ~ bellow-right

Numerical experiments with inconsistent pairwise interval comparison matrix $\boldsymbol{A}$ (11) are based on similar Monte Carlo approach. However, corresponding two sets of crisp matrices are denoted $\{{}_0\mathcal{C}_k\}$ and $\{{}_1\mathcal{C}_k\}$, k=1,..,K=100, now. The meaning of both prefixes is exactly the same, i.e. 0 for non-preserving case, and 1 for preserving reciprocity of crisp matrix. Calculated $\lambda_{\max}$ are given on Fig. 3, and c_{I} on Fig. 4.

At the first glance we may detect differences with corresponding results of the first example, which presented the consistent interval comparison matrix. The most impressive result is that EDF-s of c_{I} calculated for sets $\{{}_0\mathcal{C}_k\}$ and $\{{}_1\mathcal{C}_k\}$ are much closer each other than on the Fig. 2 bellow-right.

4 Conclusion

We have analyzed two examples of interval pairwise comparison matrices of AHP using Monte Carlo approach. The results of spectral problems, and calculated maximal eigenvalues and consistency indeces in particular, have shown interesting behaviour. Computational analysis justifies differences between consistent and inconsistent matrices being anticipated. In particular, the inconsistent matrices yield higher consistency indeces, which corresponds to natural estimations. However, various generalizations of AHP and their applications are still a fruitfull field of research.

Acknowledgements

The author is grateful for financial support by project LC06075, Ministry of Education, Czech Republic.

References

[1] Beck, J., Lukáš, L.: Analytic hierarchy process and weight generation for inconsistent interval comparison matrices by eigenvalue method. *Proceedings of the 28th Int.Conf. on Mathematical Methods in Economics* 2010, Part I, Publisher Univ. of South Bohemia in České Budějovice, České Budějovice, 2010, pp. 36-42, ISBN 978-80-7394-218-2.

[2] Dadkhah, K.M., Zahedi, F.: A mathematical treatment of inconsistency in the analytic hierarchy process. *Math.Comput.Modelling*. Vol.17, No.4/5, pp.111-122, Pergamon Press 1993.

[3] Jablonský, J.: Measuring the efficiency of production units by AHP models. *Mathematical and Computer Modelling* 46 (2007), 1091-1098.

[4] Pelaez, J.I., Lamata, M.T.: A New Measure of Consistency for Positive Reciprocal Matrices. *Computers and Math. with Appl*. 46, pp.1839-1845, Elsevier 2003.

[5] Saaty, T. L.: *The Analytic Hierarchy Process*. New York, Mc Graw Hill 1980.

[6] Saaty, T. L., Ozdemir, M. S.: “Why the Magic Number Seven Plus or Minus Two”. *Math. and Comput. Modelling* 38, pp.233-244, Pergamon Press 2003.

[7] Wang, Y-M., Yang, J-B., Xu.D-L.: Interval weight generation approaches based on consistency test and interval comparison matrices. *Appl. Math. and Comput*. 167 (2005) 252-273, Elsevier 2005.

The Economic Application of Game Theory on the Case of Least Development Countries (LDCs)

Ingrid Majerová[1]

Abstract. Game theory is a mathematical discipline that studies the choice strategies by interacting rational participants. The basis of the game theory analysis is to determine which strategy is the best response to the strategies chosen by the other participants. Game theory is a scientific metaphor, it´s not only about finding a strategy in mere games but also in other areas of human activity, particularly in economics and defence, and in transport, agriculture ect. There are many variants of game theory: a game with two or more participants, with zero or non-zero sum, with perfect or imperfect information, with dominant strategy equilibrium or Nash equilibrium. In some games, one participant always wins and the other loses, in another ones both of them can get. The paper aims at the game theory application in international relations, the possibility of applying the above-mentioned groups of games in trade policy and international economics is analyzed, emphasizing the group of the Least Developed Countries that are unable to reap the benefits of international trade due to inadequate knowledge of trade issues. The method of game tree and matrix are used in this analysis.

Keywords: game theory, Least Development Countries, international trade, game three, matrix.

JEL Classification: C70, C71, C72
AMS Classification: 91A80

1 Introduction

"It is not that there exists any fundamental reason why mathematics should not be used in economics."

Neumann and Morgenstern [19]

The above mentioned idea lays the basis for writing this article due to the fact that the mathematical approach accompanies the modern economics "on each step". The game theory itself as one of mathematical methods, is not only an integral part of economics but also of other disciplines, such as defense, biology, medicine, environment, law, telecommunications and transport, media, agriculture and so on or everyday life. Each of these disciplines is subjected to daily situations, which must opt for an appropriate strategy so that it achieved the best results. If this decision is influenced by the decisions of other participants, we can speak about a principle of game theory.

As is described in section 2, there is no uniform definition of what game theory is. The author focuses on those most concise and most popular used in practice. In contrast to an inconsistent definition there is an exact division of game theory groups, which are also discussed (a game about two or more actors, with zero or non-zero-sum, with perfect or imperfect information, with dominant strategy equilibrium or Nash equilibrium). There is used method of description in this section.

Although the foundation of modern game theory was laid by John von Neumann at the beginning of last century, the first mention of game theory, having a base in the strategic decisions, we could have already found two and a half thousand years ago by China philosopher and army general Sun Tzu. The development of game theory, from the first references to the mathematical approach, is discussed in section 3, where the method of description is chosen as well.

The core is the fourth section, in which the author seeks to meet the stated aim – namely the game theory application in international relations and the possibility of applying the mentioned games groups in trade policy and international economics on the example of the Least Developed Countries (LDCs). Here is used the method of analysis of the status of LDCs and after, by the application of game theory, matrix and the game tree method.

[1] Silesian University in Opava, School of Business Administration in Karviná. Department of Economics, Univerzitní náměstí 1943, Karviná. majerova@opf. slu.cz.

2 Definition of Game Theory and its Division according to various Criteria

Before describing the development of game theory and applying this theory to issues related to Least Development Countries, it is desirable to define briefly the base of game theory and its various types. As mentioned in the introduction, there is no single universally accepted definition of what the game theory is. In various literatures can find various concepts of game theory, but the nature is the same everywhere – game theory is a relationship between two or more participants (those participants are not only individuals, but also firms, states, various interest groups of people or integration groups) and bargaining of their best positions.

The following overview at least some of the well-known definitions are shown, which indicate that the theory of game is:

- (relatively) rigorous analysis of situations of strategic interdependence (Dimand and Dimand [6]);
- mathematical analysis of conflict situations (Dlouhý and Maňas [8]);
- systematic study of the relationship between rules, choice and outcome in competitive situations (Dobre [9]);
- study of multiperson decision problems (Gibbons [11]);
- study of the choice of strategies by interacting rational agents – interactive decision theory (McCain [17]);
- study of mathematical models of conflict and cooperation between intelligent rational decision-makers (Myerson [18]);
- bag of analytical tools designed to help us understand phenomena that we observe when decision-makers interact (Osborne and Rubinstein [20]).

To define game theory we can also adopt two approaches: first, an analytical approach and second behavioral approach (Dobre [9]). The *analytical approach* lies in the fact that game theory is the games analysis played by non-empirical "ideal" participants (players), who don´t correspond to real players. On the other hand *behavioral approach* of game theory studies the real participants confronted with well-defined games and their conflicts and strategies[2]. The strategy in game theory defines Davis [5] as a complete action plan that describes what a participant will do under all possible circumstances.

The basis of the game theory analysis is to determine which strategy is best response to the strategies chosen by other participants. The best response is defined by McCain [17] as the strategy that gives that player the maximum payoff, given the strategy the other player(s) has chosen or can be expected to choose.

The important elements of game theory are players, actions (strategies), payoff and information, by Rasmusen [23] so called "PAPI". This is the base of the following three types of a game theory division: the first is a combination of that provides Dobre [9], Dlouhý and Maňas [8], the second – taking into account the information – is used by Gibbons [11], Fundenberg and Tirol [10] and the third type of structure (collaboration and information) was analyzed by Osborne and Rubinstein [21].

3 Historical development of Game Theory

Although the modern mathematical approach to game theory starts with the work of John von Neumann "Zur Theorie der Gesellschaftspiele"[3] (see Luce and Raiffa [13]), the history of this mathematical-economic discipline has deeper roots. Let me mention at least the most important authors or works on game theory.

Peters [22] mentions the origins of game theory from two and a half thousand years ago, when the Chinese general and philosopher Sun Tzu (544-496 BC) wrote his work The Art of War, one of the oldest and most successful books on military strategy in the World. This work was based on his life experiences and Sun described in 13 chapters the importance of positioning in military strategy. Another historically important work that deals with game theory is Talmud[4]. Aumann and Maschler became the first who indroduced the one problem of game theory in Babylonian Talmud, and so the coalitional bankruptcy problem. (Orbay and Orbay [20]) Aumann and Maschler [1] described three examples: the first one is about three creditors with debt claims 100, 200 and 300 and with estate possibility 100, 200 and 300 as well. The recommendation of Talmud is – the smallest estate (100) equals the smallest debt division while estate of 200 and 300 means uneven division – for estate 200 it´s "mysterious" (50, 75, 75) and for estate 300 it´s proportional (50, 100, 150). The second example is about two holders of one garment, while one claims the whole garment, the second one claims "only" half – the result is

[2] In another interpretation, we consider the second approach, a behavioral game theory.
[3] First it was published in *Mathematische Annalen* **100** (1928), 295-320.
[4] Talmud is the 2 thousand-years old Jewish document – a base for civil, criminal and religious law, which has two kinds of origins – Babylonian and Jerusalem.

that the one awarded ¾, the other ¼ because the remaining half of the second owner is evenly divided between the two. The third example relates to the Jerusalem Talmud and describes the coalition formation – *coalition game* – of the second and third creditor against the first one (from the first example). The creditors form coalition, when estate is less or equal to half of the total debt. Aumann and Maschler showed that the solution based on the garment division principle is the unique solution to all bankruptcy problems and coincides with the nucleus of these coalitional games. (Orbay and Orbay [20])

In the middle of 17th century (exact in year 1654) the base of mixed strategy was described - based of probability concept - as one of the game theory pillars by Pierre de Fermat and Blaise Pascal (Hykšová [12]). At the beginning of 18th century, specifically in 1713 the first known minmax discussion of two-players game theory was occurred in letters from James Waldegrave to Pierre-Remond de Montmort. Waldgrave described minmax mixed strategy solution (equilibrium) of the card game le Her[5] (more Hykšová [12]). This area of game theory was also further analyzed by Montmort and Nicholas Bernoulli. In the same year (1713) these two gentlemen began (and from the year 1728 with participation of Gabriel Cramer) to discuss problem, which is in game theory known as *St. Petersburg Paradox*. Thus it was named in 1731 (published 1738) by Daniel Bernoulli, Nicholas Bernoulli´s cousin. Like in Waldgrave´s explanation of card game, in describing of paradox two-players are participated. The aim was to determine the value of players´ expectation. The Paradox was that no one would be willing to purchase (the right to play this game) at a modestly high price even though it has an infinite expected value. (Ciecka [3])

More than one hundred years after Waldgrave correspondence Antoine Augustine Cournot describes mathematic principle of imperfect competition, particularly the duopoly model which solution (duopoly participants´ strategies and payoff in the form of profit) was explained later as *Nash equilibrium* (of a pure strategy). At the end of 19th century the game theory was first used for other science than for economical solution. In 1871 Charles Darwin discussed the equilibrium sex ratio by using the first game theoretic argument in evolutionary biology. (Walker [27]) Like Cournot also Francis Ysidro Edgeworth applied game theory to the microeconomic model, and in 1881 on the problem of outcome trade determination between individuals[6].

Another important contribution to development of game theory brought Ernst Zermelo in 1913 on the example of chess and later known as *Zermelo´s theorem*. This theorem (on the base of proofs of Denes König and Laszlo Kalmar, see Schwalbe and Walker [24]) is the concept of winning[7]. In 1921 Emile Borel first published articles on issue of symmetric two-player zero-sum games, more Dimand and Dimand [7]. He was the first who attempted to mathematize the game of strategy, he introduced the concept of method of lay in the sense of today pure strategy and he was looking for a solution in mixed strategies in the sense of today pure strategy and he was looking for a solution in mixed strategies in the sense of today minimax solution. (Hykšová [12])

Year 1928 is an important milestone in game theory, because since this year, thanks to John von Neumann´s work, is dating existence of game theory as the unique field of science. Von Neumann continued to Zermelo, König and Kalmar and mathematically proved their solution, set equilibrium in two-player zero-sum game in mixed strategy, known as *Von Neumann theorem*. The result of this theorem is based on matrix method and says that mixed strategies always exist. When players choose strategies *a, b,* than exist more results ($c_1,\dots c_n$), ($d_1,\dots d_m$) and expected payoff for first player is $p_1(a, b)$, so for existing mixed strategy relation (1) is valid:

$$p_1\left(a^*, b^*\right) = \max_a \min_b p_1\left(c_i, d_j\right) \le \min_b \max_a p_1\left(c_i, d_j\right) \tag{1}$$

About 16 years later, von Neumann published with Oscar Morgenstern a book Theory of Games and Economic Behavior (see [19]). There were given the general description of strategy game, zero-sum game for two and more players, none-zero sum game, the non-cooperative and cooperative (coalition) game as well and concept of utility theory. Another important leader in the field of game theory is John Forbes Nash, who in early 50´of last century began dealing with the game theory, especially this non-cooperative. The result was the above mentioned *Nash equilibrium*: it´s a solution concept of a mixed game for two and more players, in which is assumed that each player knows the equilibrium strategies of the other, and no player can benefit by changing only his own strategy unilaterally. The concept of Nash equilibrium is used to analyze the outcome of the strategic interaction of several decision makers in a wide variety of applications.

[5] Le Her is agambling game from 18th century played standard with 52 cards by two people.
[6] In *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* C. Kegan Paul & Co., London, 1881.
[7] There are many interpretation of this theorem, for them see Schwalbe and Walker [24].

4 The Game theory and its Application for Problem Solution in LDCs

From the text it is clear that game theory and its various applications found their place in various scientific disciplines and belongs to the components of daily (not only) human activities[8]. Increasing use has also game theory in international economics, cooperation and international relations (see Correa [4], Snidal [25] or Bennett [2]). As the globalized world has more than 200 economies, and that two thirds are of developing countries, it is obvious that this fact affects international relations, in our case relations based on the operation of game theory. This section is devoted to detailed application of game theory, and in relation to the Least Developed Countries, known as the LDCs.

4.1 Short Definition of LDCs

Majerová [14] shows, that the Least Development Countries, LDCs is a very specific group of developing economies with a high degree of backwardness. All these economies are geographically located in the "poverty belt"[9], with a population of more than 818 million (approximately 12 per cent of world population), who produce fewer than 4 per cent of global GDP and their share in world trade is only about 6 per cent (in 2010). For the first time the term LDCs was used in 1971 and from the 90th of the last century, they have paid attention to remove their backwardness and thus to achieve better results of global development of the world economy. The original number of these countries was 25, gradually expanded and has stabilized at 49.

Countries wishing to be included in the LDCs group, must meet certain criteria of the UN (low national income, a high degree of economic vulnerability and low human activity, more Majerová [14]) and are characterized by the following characteristics [26]:

- low level of GDP per capita (with the paradoxically high rate of growth of GDP)[10];
- low productivity of agricultural production;
- high indebtedness;
- lack of diversification;
- large number of people living below poverty line;
- large number of people infected with contagious diseases such as HIV / AIDS, malaria, tuberculosis and diarrheal diseases (and dying for their consequences)
- higher degree of dependence on commodity exports, and food and fuel imports.

The above characteristics can be the base for game theory analysis on the example of those countries, whether from the perspective of international trade, trade policy, providing government subsidies or drugs sale for infectious diseases.

4.2 Game Theory Application for some Problem of LDCs

In this section three situations will be analyzed that may arise in the LDCs in economic terms and then look for solutions, all with the game theory application. These situations may include:

- application of game theory to the fundamental contradiction of *international economics* – specifically, liberalization or protection;
- application of game theory to *microeconomics* – specifically *trade policy* and the intersection of LDCs to developed country markets;
- application of game theory to *health issues* – specifically the possibility of providing cheaper drugs for AIDS LDCs.

Liberalization or Protection

Protectionism principle is as old as humanity itself, as well as efforts to eliminate it through the liberalization of international trade. It´s widely considered that the global market liberalization is a way of LDCs´ development, but we show that from the perspective of game theory, this view is not as clear, since the introduction of protectionist measures (typically a dominant strategy) leads to a Nash equilibrium, although it is inefficient.

Now imagine that the LDCs will enter into international trade with the developed world (DeWo) and both participants may choose the strategy of liberalization (L) or protectionism (P). Establish the following payoff matrix (Figure 1). Numbers´ values are fictitious income from mutual trade.

---

[8] Maňas [15] describes specific examples related to the Czech Republic.
[9] This is an area between the tropic of Cancer and Capricorn, exception of Afghanistan.
[10] In 2009 LDCs reached 4.3 per cent of GDP growth, compared with 2.3 per cent in emerging and other developing economies, and -3.2 in developed countries.

		DeWo	
		P	**L**
LDCs	P	(5,5)	(9,4)
	L	(4,9)	(8,8)

Figure 1 Payoff Matrix

From the matrix we can see, that if both participants choose liberalization, they both gain by trade (8,8), if they use f. e. tariffs, they both lose (5,5), although reaching (inefficient) equilibrium. In situation, when DeWo liberalize trade, it is better for LDCs to introduce tariffs (9,4) and when DeWo uses protectionism, for LDCs is still better to protect their markets (payoff of 5 is higher than payoff of 4). His is the game with dominant strategy equilibrium.

Monopoly and Rookie

Another example of application of game theory to the LDCs has the microeconomic basis – let's assume the existence of the monopoly economy (Mone) and the potential entrant "rookie" to the market (LDCs). One problem is that LDCs doesn´t know the Mone´s marginal costs – they could be higher or lower than of LDCs´costs. If they are higher, Mone won´t fight with LDCs due to inability to reduce its prices. But if Mone´s marginal costs are lower, it´ll fight with LDCs and win.

This situation can be illustrated by game trees, where are shown payoffs of players in parentheses which values correspond to the above conclusions. The left game tree (Figure 2) shows passivity of Mone by higher marginal costs, the right game tree (Figure 3) reflects an active "fight" monopoly economy with just entered competitor from LDCs.

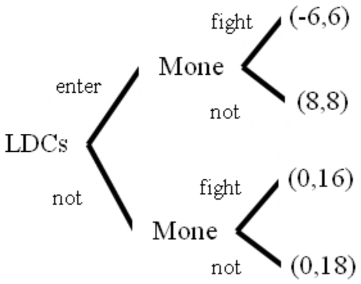


Figure 2 Higher marginal Costs

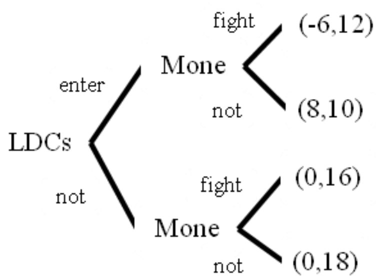


Figure 3 Lower marginal Costs

The above mentioned cases are "only" models, in the real world is more common (although it is very costly) for Mone to fight with LDCs, because he knows that as a traditional producer pushes LDCs from the market and recover the monopoly – so the result is always loss for LDCs. Here the state can also help through subsidization for Mone. On the other hand, LDCs can get support and in this case the LDCs always enter the market dominated by Mone. This is the case of game theory with imperfect information.

AIDS Drugs to LDCs

The spread of infectious diseases such as AIDS, malaria and tuberculosis is a very disturbing problem of LDCs, in particular due to the unavailability of appropriate drugs. The problem isn´t the lack of them, but the price – at least many disabled people in developing countries with income less than $ 1 per day cannot afford to pay for the treatment of thousands dollars. The solution is to reduce the price of drug for these patients, but this is accompanied by certain conditions of pharmaceutical companies. One of these conditions, to maintain their profits, it can be sales of generic drugs[11], subject to the prohibition of the (usually speculative) back-spread to the advanced (and wealthy) economies. This is a case of one-shot game.

Supposing that in the LDCs the pharmaceutical company (Phar) begins to sell drugs to AIDS treat, which can sell for double kind of the price - high (H) and low (L). LDCs have only two strategies – to accept restrictions on resale (yes, Y), or not accept (or not, N). The payoffs are described by four numbers – number 0 being the unchanged situation, -10 means a deterioration of the current situation, 10 means better situation and 20 is still

[11] Generic drugs are the equivalents of original drugs, which may come on the market after the expiration of their patent protection and are up to 80 percent cheaper than the originals.

better than 10. There are three situations – first, when Phar not wait for the LDCs adopted restrictions and offer drugs, the other when both participants will “play” at the same time and the third, when Phar comes into play only after the LDCs adopted the required measures.

Results from the first two mentioned situations are the same - if Phar introduces high drug prices, it is indifferent for both players whether LDCs implement or load restriction - payoff is always 0,0 (status qou). It is different in case of sale of drugs at low prices that Phar payoff is 10 in the case of the restrictions introduction and if LDCs don´t impose them, the payoff is -10. On the other hand for LDCs payoff is 10, if they introduce restrictions, if not, the payoff is 20 (because their profits from resale to other countries). These strategies and payoff matrix shows the game tree in Figure 4.

		LDCs	
		Y	**N**
Phar	H	(0,0)	(0,0)
	L	(10,10)	(-10,20)

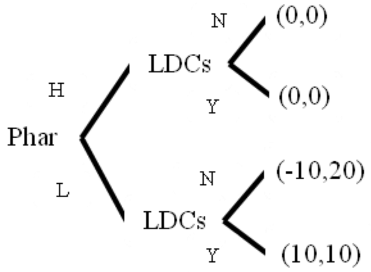


Figure 4 Payoffs by Phar´s first Start or simoultaneous Start of both Players

If Phar waits for the LDCs´ measures adoption, the structure of the game tree changes (while matrix structure remains unchanged). Their non-acceptance means status quo for both players in the case of high prices, while by the low prices is payoff (20,-10) for the LDCs and Phar. In the case of the restriction measures adoption for resale, then means the status quo by high prices and in the case of low price improvement for both (10,10) again. These situations are shown in a matrix and in the graph tree (Figure 5).

		LDCs	
		Y	**N**
Phar	H	(0,0)	(0,0)
	L	(10,10)	(-10,20)

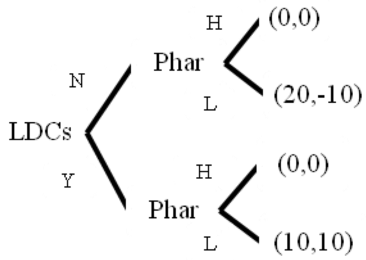


Figure 5 Payoffs by LDCs´ first Start

This type of game is an example of how important it is to be the first or another player and how important cooperation between various actors is although this can be sometimes enforced and the players must play according for them not always advantageous rules for them.

Conclusion

Game Theory accompanied mankind since its beginnings - not only from a historical perspective through the strategies of war (Sun Tzu) and legal solutions (Talmud) or economic (duopoly) problems, but also "from cradle to grave", through children's games (Rock, paper, scissors), student entertainment (Game of chicken) to behavior of adult (Battle of Sexes or Stag hunt).

Although no unified definition of game theory exists, its base is clear – it´s mathematical discipline that treat the question of optimal behavior of those who participate in it and determine the resulting equilibrium. In this contribution behavioral approach to the game theory was selected, which is based on research how individual players make choice and solve mutual conflicts. For this article was selected one area of the game theory activity and economic area were selected. After that, the issue of international trade, specifically in relation to the Least Development Countries LDCs was discussed. Those countries are generally considered as “the weakest player” in the field of international trade that is no able to reap the benefits of international trade due to inadequate knowledge in trade issues.

Three examples of the economic application of the game theory on the case of LDCs were described – choice between free trade and protectionist policies, the entry "rookie" LDCs companies to established monopoly markets and sale of generic AIDS drugs in LDCs. Regarding the first example, that if both participants choose liberalization, they both gain by trade, if they use protection, they both lose, although reaching (inefficient) equilibrium. In situation, when other participant liberalizes or protects trade, it is better for LDCs to use protectionism. The result of the Monopoly and rookie game is “fight” of the monopoly with LDCs competition (associated with considerable financial costs) and losing rookie, but that can be reversed in the win on the basis of government subsidies. Pharmaceutical game has several solutions, depending on the order in which each participant enters to it – when the pharmaceutical company comes into play as a first, it may be able to monitor rules but drugs valuation may be different than when it enters the game as the second player. In any case, the optimum is reached, when the drugs are supplied to the LDCs at lower prices in compliance with the conditions.

Acknowledgements

This contribution was produced with the support of an Internal grant system of the Silesian University in Opava, registration number IGS/3/2011, project title: “Economic and Political Development of LDCs and their Development Prospects in the global Economy”.

References

[1] Aumann, R., Maschler, M. Game theoretic analysis of a bankruptcy problem from the Talmud. *Journal of Economic Theory* **36** (195-213).

[2] Bennett, P. G. Modeling Decisions in International Relations” Game Theory and Beyond. *Meshon International Studies Review* **39** (1995), 19-52.

[3] Ciecka, J. Daniel Bernoulli on the Measurement of Risk. *Journal of Legal Economics* (2010) [online]. URL: http://www.allbusiness.com/science-technology/mathematics/14622371-1.html.

[4] Correa, H. Game Theory as an Instrument of the Analysis of International Relations. *Ritsumeikan Annual Review of International Studies* **14** (2001), 1-22.

[5] Davis, M. D. *Game Theory: a nontechnical Introduction.* Dover Publications, Mineola, 1997.

[6] Dimand, M. A., Dimand, R. W. A History of Game Theory, Volume I.: From Beginnings to 1945. In *Routledge Studies in History of Economics*. London and New York, Routledge, 1996.

[7] Dimand, R. W., Dimand, M. A. The Early History of the Theory of Strategic Games from Waldegrave to Borel. In *Toward History of Game Theory*. Duke University Press, Durham, 1992.

[8] Dlouhý, M., Maňas, M. *Games and Economic Decisions*. Oeconomica, Praha, 2009.

[9] Dobre, C. I. Implication of Game Theory to International Trade. *Revista Tinerior Economisti* **1** (2001), 132-137.

[10] Fundenberg, D., Tirol. J. *Game Theory.* MIT Press, Cambridge, 2000.

[11] Gibbons, R. *Game Theory for Applied Economists*. Princeton University Press, Princeton, 1992.

[12] Hykšová, M. Several Milestones in the History of Game Theory. In: *Jubiläen – Change oder Plage?* VII. Österreichisches Symposion zur Geschichte der Mathematik, Wien, 2004, 49-56.

[13] Luce, R. D, Raiffa, H. *Games and Decisions. Introduction and Critical Survey*. Dover Publications, Mineola, 1989.

[14] Majerová, I. *Rozvojové ekonomiky.* Kleinwächter, Frýdek-Místek, 2008.

[15] Maňas, M. *Games and Economic Decisions*. Oeconomica, Praha, 2003.

[16] Marschack, J. Neumann´s and Morgenstern´s new Approach to static Economics. *The Journal for Political Economy* **LIV** (1946), 97-115.

[17] McCain, R. A. *Game Theory. A Nontechnical Introduction to the Analysis of Strategy.* Drexel University, Philadephia, 2010.

[18] Myerson, R. B. *Game Theory: Analysis of Conflict.* Harvard University Press, Harvard, 1997.

[19] Neumann, J. von, Morgenstern, O. *Theory of Games and Economic Behavior.* Princeston University Press, Princeston, 1953.

[20] Orbay, B. Z., Orbay, H. Talmudic Division as a Cartel Rule. *Journal of Economics and Business* **55** (2003), 167-175.

[21] Osborne, M. J., Rubinstein, A. *A Course in Game Theory*. MIT Press, Cambridge, 1994.

[22] Peters, H. *Game Theory: A multilevel Appproach.* Springer Verlag, Heidelberg, 2008.

[23] Rasmusen, E. *Games and Information: an Introduction to Game Theory*. Blackwell Publishing, Oxford, 2007.

[24] Schwalbe, U. Walker. P. Zermelo and the Early History of Game Theory. *Games and Economic Behavior* **34** (2001), 123-137.

[25] Snidal, D. The Game Theory of International Politics. *World Politics* **38** (1985), 22-57.

[26] The Least Developed Countries Report 2010. *Towards a New International Development Architecture for LDCs.* United Nations. New York and Geneva, 2010.

[27] Walker, P. An Outline of the History of Game Theory. [online]. URL: http://william-king.www.drexel.edu/top/class/histf.html.

The fuzzy permutation model for the ordinal ranking problem solution

Jiří Mazurek[1]

Abstract. The aim of the article is to propose a new fuzzy permutation model for a group decision analysis with ordinal rankings of alternatives. In the problem's setting, a set of k decision makers rank n alternatives from the 1st place to the n^{th} place with regard to some criterion (or criteria). The goal is to find a consensus – an overall ranking of alternatives by a group. The fuzzy permutation model is proposed both for single and multiple criteria problems, and it can handle different weights of decision makers as well. Therefore, it provides a generalization of classic methods for ordinal consensus ranking problem solution such as Borda-Kendall's method of marks (BAK), maximize agreement heuristic (MAH) or consensus ranking model (CRM). Moreover, the proposed method is numerically simple and time-saving, and its use is demonstrated by examples. Results of the fuzzy permutation model are compared with results of BAK method and a hybrid distance-based consensus ranking model by Tavana et al.

Keywords: fuzzy permutation, group decision analysis, ordinal rankings, ordinal ranking problem, permutation

JEL Classification: D71
AMS Classification: 90B50, 91B08, 91B10

1 Introduction

The group decision analysis, which deals with objects ranked from the 1st to the n^{th} place under some criterion (or multiple criteria), is called *ordinal ranking problem*. The goal of a group decision analysis in such setting is to achieve a consensus ranking of objects or alternatives. The problem concerns many disciplines such as operational research, management, marketing, social science, politics, psychology, etc. Its history dates back to preferential elections in the 18th century, and at present it constitutes a part of the modern (multi-criteria) group decision analysis. As for the problem solution, there are two classes of methods:

- *ad-hoc* methods include Borda-Kendall's method of marks, Condorcet's simple majority rule, Copeland's method, etc.
- *distance-based* methods use distance (metric) functions on vector or matrix spaces and include maximize agreement heuristic (MAH) by Beck and Lin [1], consensus ranking model (CRM) by Cook and Kress [5], distance-based ideal-seeking consensus ranking model (DCM) by Tavana et al. [10] and others.

Ad-hoc methods are widely used for its simplicity, but are considered unstructured and not satisfying Arrow's welfare axioms by some authors [10]. Other common criticism of BAK-like methods stems from the fact that they use rank positions as the values of being ranked at those levels [6]. The fundamental limitation of distance-based methods rests in growing computational complexity as a number of alternatives grow. For n alternatives there exist $n!$ permutations, which have to be searched through to obtain a solution. The first model for a multiple criteria was proposed by Cook and Kress [5], and a fuzzy approach to ordinal consensus ranking model was pioneered by J. J. Buckley, see e.g. [3].

The aim of the article is to propose a new method for a group decision analysis with ordinal information about alternatives, which is based on a concept of fuzzy permutations. The method is both mathematically simple, computationally affable and time undemanding. It provides a generalization of ordinal consensus ranking problems with crisp rankings (permutations) of alternatives. The use of the method is demonstrated in Section 5, and results are compared to those by Borda-Kendall's method and distance-based DCM method (for a detailed method's description see [7] or [10]:

[1] Silesian University in Opava, School of Business Administration in Karviná, Department of Mathematical Methods in Economics. Univerzitní náměstí 1934, Karviná 733 40. e-mail: mazurek@opf.slu.cz.

- *Borda-Kendall's method* assigns each alternative a number of points corresponding to its position given by each DM. The best alternative is the alternative with the lowest total count (mark) or with the lowest average (which is equivalent).
- In *DCM method* with n DMs and m alternatives, a single preference matrix (A) is constructed with elements a_{ij} equal to a number of times an alternative i is preferred to an alternative j by a set of DMs. Then ideal consensus matrix (B) for a set of DMs is constructed for each feasible permutation of alternatives. In this matrix, all elements b_{ij} are equal to 0 or n (all DMs are in accord, hence the name 'ideal consensus matrix'). Subsequently, a distance of A and B is computed using l_1 metric:

$$d(A,B) = \sum_{i=1}^{m} \sum_{j=1}^{m} \left| a_{ij} - b_{ij} \right|$$

The consensus is represented by a matrix, which minimizes the distance to the matrix A.

The paper is organized as follows: in Sections 2 and 3 fuzzy permutations and a dominance relation are introduced, Section 4 provides the model's description and Section 5 includes numerical examples.

2 Permutations and fuzzy permutations

The permutation of a finite set A is defined as a bijection from A to itself: $\pi : A \to A$. Each permutation (ranking[2]) of n objects can be represented by a square binary matrix of order n with exactly one value of 1 in each row and column – a *permutation matrix*. For instance, the permutation π = (1,3,2) is represented by a matrix π with rows corresponding to alternatives and columns to positions:

$$\pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

The set of permutation matrices S_n of order n is a subset of a set of bistochastic matrices (see Theorem 1 thereinafter), which are defined as square matrices $A = (a_{ij})$ with the sum of all rows and columns equal to one [9]:

$$\sum_{i=1}^{n} a_{ij} = 1 \text{ and } \sum_{j=1}^{n} a_{ij} = 1 \tag{1}$$

Relations (1) pose $2n - 1$ constraints for a_{ij} values, hence the number of degrees of freedom is $(n-1)^2$. The set of bistochastic matrices is often denoted as B_n, where B_n stands for Birkhoff's polytopes of order n.

Theorem 1 (Birkhoff-von Neumann): *The set of bistochastic matrices of order n is the convex hull*[3] *of the set of permutation matrices of order n.*

Proof : see [2] or [8].

Theorem 1 enables to define fuzzy permutations, which are generalization of crisp permutations:

Definition 1. *Let* π_i *be a permutation matrix of n objects given by i-th decision maker with a weight* $w_i \geq 0$, $i \in \{1, 2, ..., K\}$, $\sum_{i=1}^{K} w_i = 1$. *Then, a fuzzy permutation* $\Pi(\pi_1, ..., \pi_K ; w_1, ..., w_K)$ *is given as:*

$$\Pi(\pi_1, ..., \pi_K ; w_1, ..., w_K) = \sum_{i=1}^{K} \pi_i \cdot w_i \tag{2}$$

Hence, a fuzzy permutation is defined as the weighted arithmetic mean of permutation matrices (where weights are normalized). From Theorem 1 it follows that $\Pi(\pi_1, ..., \pi_K ; w_1, ..., w_K)$ is a bistochastic matrix. When decision makers express their rankings of n objects, these preferences can be converted into permutation matrices and aggregated by relation (2). Therefore, a fuzzy permutation contains information about DMs' crisp rankings, and this ranking is 'fuzzy' in the sense that each alternative is assigned to all positions, but with generally different confidence.

[2] Here, terms 'ranking' and 'permutation' are used as synonyms.

[3]The convex hull $H(X)$ of a set X is given as [9]: $H(X) = \left\{ \sum_{i=1}^{n} \theta_i a_i \,/\, a_i \in X, \theta_i \in R, \theta_i \geq 0, \sum_{i=1}^{n} \theta_i = 1, n = 1, 2, ... \right\}$.

Interestingly, with respect to its algebraic structure, the set S_n is a group under matrix multiplication, while the set B_n is only a monoid:

Proposition 1. *The set of bistochastic matrices (B_n) of order $n \geq 2$ is a monoid under matrix multiplication.*

Proof: To be a monoid, the set B_n has to be closed under matrix multiplication, it must satisfy associativity and a unitary element I must exist for all $A \in B_n$. Because matrix multiplication is associative and the identity matrix $I \in B_n$, it suffices to show that B_n is closed under multiplication: Let $A \in B_n$, $B \in B_n$ and $C = A \cdot B$. Then the sum of all elements in the k-th row of the matrix C is: $\sum_{j=1}^{n} a_{kj} \cdot \left(\sum_{i=1}^{n} b_{ji} \right) = \sum_{j=1}^{n} a_{kj} \cdot 1 = 1$, the proof for columns is analogous. As B_n contains singular matrices as well, it is not a group.

Proposition 2. *The arithmetic mean and weighted arithmetic mean of fuzzy permutations is a fuzzy permutation.*

Proof : The proposition is obvious.

Proposition 2 is important for the aggregation of decision makers' preferences when multiple criteria are involved. It is used in the fuzzy permutation model with multiple criteria in Section 4.2.

3 Ordering of alternatives

The goal of the ordinal consensus ranking problem is to find the best alternative after decision makers express their rankings. To accomplish this, a pair-wise dominance relation is introduced by following two definitions:

Definition 2: *Let Π_{ij} be the fuzzy ranking of an alternative i at a position j. Then, a cumulative fuzzy ranking H_{ij} of an alternative i from the 1st to the jth position is given as:*

$$H_{ij} = \sum_{k=1}^{j} \Pi_{ik} \tag{3}$$

Definition 3: *An alternative r dominates an alternative s $(r \succ s)$ if and only if all cumulative fuzzy rankings H_{rj} of an alternative r are higher or equal to cumulative fuzzy rankings H_{sj} of an alternative s*:

$$(r \succ s) \Leftrightarrow H_{rj} \geq H_{sj}, \; j \in \{1, 2, ..., n\} \tag{4}$$

The best alternative is the alternative which is not dominated by any other alternative. The fuzzy permutation from Definition 1 and the dominance relation from Definition 3 constitute fundamental parts of the fuzzy permutation model for ordinal consensus problem solution described in Section 4.

4 The model

4.1 A single criterion fuzzy permutation model

The fuzzy permutation model for ordinal consensus problem with one criterion, where DMs are assigned weights $w_i \geq 0$, $i \in \{1, 2, ..., K\}$, $\sum_{i=1}^{K} w_i = 1$, proceeds in the following steps:

1. DMs express their (crisp) rankings of alternatives, which are converted into permutation matrices π_i.
2. DMs' rankings are turned into a fuzzy ranking (permutation) by relation (2).
3. The cumulative function H_{ij} is evaluated for each alternative by formula (3).
4. The dominance relation (4) is used to alternatives pair-wise comparison.
5. The best alternative (alternatives), which is non-dominated, is chosen.

The use of the model is illustrated in Section 5.1.

4.2 A multiple criteria fuzzy permutation model

The single criterion model can be easily extended to multiple (m) criteria model, where DMs and criteria are assigned weights $w_i \geq 0$, $\sum_{i=1}^{K} w_i = 1$ and $v_i \geq 0$, $\sum_{i=1}^{m} v_i = 1$ respectively. It proceeds in the following steps:

1. Each DM expresses his crisp ranking of alternatives according to each criterion j as permutation matrices π_i^j, where $i \in \{1, 2, ..., K\}$ and $j \in \{1, 2, ..., m\}$.
2. DMs' rankings are turned into a fuzzy ranking Π_j for each criterion j with the use of relation (2):

$$\Pi_j\left(\pi_1^j, ..., \pi_K^j; w_1, ..., w_K\right) = \sum_{i=1}^{K} w_i \pi_i^j \tag{5}$$

3. A final fuzzy ranking Π of all alternatives is obtained as a weighted average over all criteria (all Π_j):

$$\Pi\left(v_1, ..., v_m; \Pi_1, ..., \Pi_m\right) = \sum_{j=1}^{m} v_j \Pi_j \tag{6}$$

4. The cumulative function H_{ij} is evaluated for each alternative by formula (3).
5. The dominance relation (4) is used to alternatives pair-wise comparison.
6. The best alternative (alternatives), which is non-dominated, is chosen.

The use of the model is illustrated in Section 5.2.

5 The numerical section

5.1 A single criterion model: a numerical example

Four decision makers with equal weights w_i = 0.25, $i = \{1, 2, 3, 4\}$ evaluate four projects A, B, C and D, and rank them from the best (the 1st place) to the worst (the 4th place), see Table 1. These rankings were converted into a matrix format in Table 2. The fuzzy permutation obtained via (2) is shown in Table 3. A cumulative fuzzy ranking H_{ij} of each alternative is presented in Table 4. Using the dominance relation (4) we get:

$$A \succ B,\ A \succ C, A \succ D, A \succ B, B \succ C, B \succ D, C \succ D.$$

Therefore, the final consensus ranking is (A, B, C, D), and the best alternative is A.

For a comparison two other methods for ordinal consensus ranking problem were chosen: Borda-Kendall's method of marks as a representative of *ad-hoc* methods and DCM method by Tavana et al. as a representative of *distance-based* methods. Table 5 provides results obtained by Borda-Kendall's method, which also chooses the alternative A as the best. Results of DCM are presented in Table 6. However, using DCM method five (equally good) different solutions to the problem were found:

(A,B,C,D), (A,B,D,C), (A,C,B,D), (B,A,C,D) (B,A,D,C).

All these rankings have the same overall distance (18) to the rankings given by DMs. The best alternatives are A and B. In general, the dominance relation (4) provides only partial quasi-order on a set of alternatives, as some alternatives might not be comparable.

DM/position	1.	2.	3.	4.
DM1	A	C	B	D
DM2	B	A	D	C
DM3	C	A	B	D
DM4	D	B	A	C

Table 1. Decision makers' rankings of alternatives A, B, C and D.

1	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	0
0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1
0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0

Table 2. DMs' rankings (from DM1 on the left to DM4 on the right) in a matrix format.

Alternative	1.	2.	3.	4.
A	0.25	0.5	0.25	0
B	0.25	0.25	0.5	0
C	0.25	0.25	0	0.5
D	0.25	0	0.25	0.5

Table 3. The fuzzy ranking (permutation) of alternatives A, B, C and D.

Alternative	1.	2.	3.	4.
A	0.25	0.75	1	1
B	0.25	0.5	1	1
C	0.25	0.5	0.5	1
D	0.25	0.25	0.5	1

Table 4. The cumulative fuzzy ranking (H_{ij}) of alternatives A, B, C and D.

Alternative	A	B	C	D
Sum of marks	8	9	11	12
Average mark	2,00	2,25	2,75	3,00

Table 5. Borda-Kendall's method of marks. The best alternative is A.

(A,B,C,D)	single preference matrix				total agreement matrix				absolute difference				sum
	A	B	C	D	A	B	C	D	A	B	C	D	
A	0	2	3	3	0	4	4	4	0	2	1	1	
B	2	0	2	3	0	0	4	4	2	0	2	1	
C	1	2	0	2	0	0	0	4	1	2	0	2	
D	1	1	2	0	0	0	0	0	1	1	2	0	18

Table 6. DCM method: the evaluation of the consensus ranking (A,B,C,D), one out of five solutions.

5.2 A multiple criteria model: a numerical example

In this example, let's consider four DMs (DM1 to DM4) with weights w_1 = 0.125, w_2 = 0.250, w_3 = 0.250 and w_4 = 0.375, and criteria C_1 and C_2 with weights v_1 = 0.25 and v_2 = 0.75, and four alternatives A, B, C and D. With regard to the criterion C_1, DMs‘ rankings are given in Table 1 and the fuzzy ranking with regard to this criterion is presented in Table 7. DMs‘ rankings according to the criterion C_2 are given in Table 8 and the fuzzy ranking with regard to this criterion is presented in Table 9. In Table 10 the final fuzzy ranking is shown, and Table 11 contains the cumulative fuzzy ranking. From the dominance relation (4) we obtain:

$$A \succ B,\ D \succ C$$

So the best alternatives, which are not dominated by any other alternative, are A and D.

Alternative	1.	2.	3.	4.
A	0.125	0.5	0.375	0
B	0.25	0.375	0.375	0
C	0.25	0.125	0	0.625
D	0.375	0	0.25	0.375

Table 7. Fuzzy ranking (permutation) of alternatives A, B, C and D according to the criterion C_1.

DM/position	1.	2.	3.	4.
DM1	B	C	A	D
DM2	A	B	D	C
DM3	C	A	D	B
DM4	D	A	B	C

Table 8. DM's rankings of alternatives A, B,C and D according to the criterion C_2.

Alternative	1.	2.	3.	4.
A	0.25	0.625	0.125	0
B	0.125	0.25	0.375	0.25
C	0.25	0.125	0	0.625
D	0.375	0	0.5	0.125

Table 9. The fuzzy ranking of alternatives A, B, C and D according to the criterion C_2.

Alternative	1.	2.	3.	4.
A	0.219	0.594	0.188	0
B	0.156	0.281	0.375	0.188
C	0.250	0.125	0	0.625
D	0.375	0	0.438	0.188

Table 10. The final fuzzy ranking of alternatives A, B, C and D.

Alternative	1.	2.	3.	4.
A	0.219	0.813	1	1
B	0.156	0.437	0.813	1
C	0.250	0.375	0.375	1
D	0.375	0.375	0.813	1

Table 11. The cumulative fuzzy ranking (H_{ij}) of alternatives A, B, C and D.

6 Conclusions

The aim of the article was to introduce a new method for ordinal consensus ranking problem, which utilizes a concept of fuzzy permutations. The presented method is computationally simple and time efficient. Moreover, it allows for a multiple criteria and different weights of decision makers, thus providing a generalization to classic methods such as BAK, MAH, or DCM. The use of the fuzzy permutation model was demonstrated with numerical examples and its results were compared to results obtained by classic methods.

Acknowledgements

The work was supported by GAČR project No. 402090405.

References

[1] Beck, M. P., Lin, B. W. Some heuristics for the consensus ranking problem. *Computers and Operations Research* 1983; Vol. 10, Issue 1, p. 1–7.

[2] Birkhoff, G. Three observations on linear algebra, *Univ. Nac. Tacum´an Rev. S. A* 5, 1946, p. 147–151.

[3] Buckley, J. J. The multiple judge, multiple criteria ranking problem: A fuzzy set approach. *Fuzzy Sets and Systems***,** 1984, Vol. 13, Issue 1, p. 25-37.

[4] Cook, W. D. Distance-Based and Ad Hoc Consensus Models in Ordinal Preference Ranking. *European Journal of Operational Research*, 2006, Nr. 172, p. 369-385.

[5] Cook, W. D., Kress, M. A multiple criteria decision model with ordinal preference data. *European Journal of Operational Research*, 1991, Vol. 54, p. 191-198.

[6] Cook, W. D., Kress, M., Seiford, L. M. A general framework for distance-based consensus in ordinal ranking models. *European Journal of Operational Research*, 1996, Vol. 96, p. 392-397

[7] Mazurek, J. Group Decision Making in Ordinal Consensus Ranking Problem. In: *Proceedings of the 28th International Conference on Mathematical methods in Economy*, České Budějovice, 2010.

[8] Neumann von, J. A certain zero-sum two-person game equivalent to an optimal assignment problem, *Ann. Math. Studies*, 1953, Vol. 24, p. 5–12.

[9] Serre, D. *Matrices: Theory and Application*, 2nd ed., Springer, 2010.

[10] Tavana, M., LoPinto, F., Smither, J. W. A Hybrid Distance-Based Ideal-Seeking Consensus Ranking Model. *Journal of Applied Mathematics and Decision Sciences*, 2007, 18 pages.

Transition of votes for consecutive elections – ecological regression modelling

Mariusz Mazurkiewicz[1], Jacek W. Mercik[2], František Turnovec[3]

Abstract. The transition of votes is observed when a voter who voted for a given party in a previous election, votes for another party in current election. Ecological regression technique [1, 2, 3, 4, 5, 6] gives an opportunity to obtain quantitative description of electoral behaviour from aggregated data under homogeneity assumption. Aggregated data available from public statistics contains the information about individual voting behaviour indirectly only. Data aggregation bases usually on geographic units such as counties and constituencies. Regional decomposition is an estimation process, used for the description of a voter's behaviour. It extends ecological regression to larger regions or even for the whole country. For the case 2*x*2 (two consecutive elections, two parties) the results of econometric approach are almost obvious. For the case n*x*2 (two consecutive elections, *n* parties) the statistical approach is necessary. We suggest that in order to estimate the parameters of ecological regression, the combined regional decomposition of transition matrix and the maximum likelihood function can be used. Moreover, the assumption of random distribution of residuals is tested too. At the end, we examine the proposed methodology using data from general parliamentary elections in Poland (from 2005 and 2007 elections) using statistical sample containing over 25 000 results from electoral districts.

Keywords: ecological regression, electorate flows, transition of votes, homogeneity of electorate, decomposition.

JEL Classification: C44
AMS Classification: 91-08

1 Introduction

Voting behaviour expressed by voting results may be described in full with voting data at the individual level. Unfortunately, this type of statistical information is not available in general. On the other side, there are aggregated data available from published statistics. Our main goal is to obtain the proportions of voters who voted for the same party or who changed their preferences in a party's choice between consecutive elections. Ecological regression model [3] is one of the statistical tools which give solution to this problem. The method is popular and well known especially in the case of two-party system. There are also many improvements, modifications and extensions of ecological regression method [2,5,7]. In our approach, we consider two consecutive elections in one country. We would like to obtain proportions from aggregated election data and estimate individual level transition coefficient from electoral data aggregated over various voting districts.

For the case 2x2 (two consecutive elections, two parties) the results of econometric approach are almost obvious. We suggest that in order to estimate the parameters of ecological regression for the case nx2 (two consecutive elections, n parties) the combined regional decomposition of transition matrix and the maximum likelihood procedure can be used.

2 The 2 x 2 case

The simplest case in ecological regression approach is the 2 x 2 case. This case occurs when there are only two competitive parties in two consecutive elections: party *1* and party *2*.

Let N_{ij} ($i,j = 1,2$) describe the number of those who vote for party i in the first election and party *j* in the second election. In table 1 the situation is presented in more convenient way where $N_{.j}$ and $N_{i.}$ are marginal values for the first and the second elections respectively: $N_{.j}$ – part of electorate, which votes for the party *j* ($j=1,2$) in the second elections, and $N_{i.}$ – part of electorate voting for the party *i* in the first election.

[1] Wroclaw University of Technology, mariusz.mazurkiewicz@pwr.wroc.pl
[2] Wroclaw University of Technology, jacek.mercik@pwr.wroc.pl
[3] Charles University in Prague, frantisek.turnovec@mbox.fsv.cuni.cz

N_{11}	N_{12}	$N_{1.}$
N_{21}	N_{22}	$N_{2.}$
$N_{.1}$	$N_{.2}$	

Tab. 1. Distribution of the electorate between two parties in two consecutive elections.

For two parties and two consecutive elections all marginal values are known as a result of elections and all cellular values could be obtained as a solution of the system of equations. However, the 2x2 case is a simplification; it is not too difficult to show that even such a model is useful in practical analysis. If party 1 is a fraction of electorate taking part in the elections and party 2 means the part of the electorate not participating in these elections the above model is proper to use for evaluation of electorate's flows from absence to participation and vice versa. The participation or non-participation is frequently under investigation of policy makers. The quantitative description of electorate's flows gives complementary information about dynamic of the electorate in the sense of politically active part of the society.

In real life, the size of electorate is changing slightly for two consecutive elections: $N_{.1} + N_{.2} \cong N_{1.} + N_{2.} \cong N$

Let x_i denote proportion of votes obtained by party i in the first election. Thus $x_i = \frac{N_{i.}}{N}$, and $y_j = \frac{N_{.j}}{N}$ where y_j denotes proportion of votes obtained by party j in the second election. Let t_1 denote the proportion of voters who voted for the same party in the first election and in the second election. Let t_2 denote the proportion of voters who switched to another party between two consecutive elections. The results of the above transformation are presented in table 2.

			Total
	t_1x_1	$(1-t_1)x_1$	x_1
	t_2x_2	$(1-t_2)x_2$	x_2
Total	y_1	y_2	*1*

Tab. 2. Transition of the electorate between two parties in two consecutive elections.

The proportion of voters who vote for party 1 in the second elections is a linear combination (called an ecological regression) of x_1: $y_1 = t_1x_1 + t_2x_2 = t_1x_1 + t_2(1-x_1) = (t_1 - t_2)x_1 + t_2$ The solution is well known [1,4,5].

Generally, the relation between proportions of votes in the first elections and in the second elections is described by the system of the following regression functions:

$$y_1 = t_{11}x_1 + t_{12}x_2 + \varepsilon_1$$
$$y_2 = t_{21}x_1 + t_{22}x_2 + \varepsilon_2$$

where: $\varepsilon_1, \varepsilon_2$ are unobserved disturbances. Thus, normalized 2x2 case of transition of votes between elections is presented as in table 3.

	t_{11}	t_{12}
	t_{21}	t_{22}
Total	1	1

Tab. 3. Coefficient of transition of votes between parties in two consecutive elections.

3 The n x 2 case

Voting results in two consecutive elections can be expressed as a cross-tabulation of election data. Let N denote the size of the whole electorate. The situation can be described in general like in table 4.

Election I	Election II				Total (election I)
	party 1	party 2	…	party q	
party 1	N_{11}	N_{12}	…	N_{1q}	$N_{1\bullet}$
party 2	N_{21}	N_{22}		N_{2q}	$N_{2\bullet}$
…	…	…	…	…	…
party p	N_{p1}	N_{p2}	…	N_{pq}	$N_{p\bullet}$
Total (election II)	$N_{\bullet 1}$	$N_{\bullet 2}$	…	$N_{\bullet q}$	N

Tab. 4. Distribution of the electorate in multiparty system during two consecutive elections.

In table 4 contingency tables notation is used. Marginal frequencies in the last row correspond to aggregated election results for the second elections. Marginal frequencies in the last column correspond to aggregated elections results for the first elections. Cellular frequencies N_{ij} denote the number of voters who voted for party i during the first elections and for party j during the second elections.

Results of both elections are also described by the vector of votes given for parties taking part in first or second elections. The aim is description of a transition-voting process in two consecutive elections with the probabilities of changes of political preferences. Estimating procedure for transition probabilities is necessary when we analyse behaviour of the whole electorate. A classical ecological approach bases on the assumption that the results of the second election are a linear function of results of the first election:

$$TX + \varepsilon = Y \tag{1}$$

T is a transition matrix, X is a vector of results of the first election, Y is a vector of results of the second election, and ε is a vector of unobserved disturbances.

$$T = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1k+1} \\ t_{21} & t_{22} & \dots & t_{2k+1} \\ \vdots & \vdots & \dots & \vdots \\ t_{k+11} & t_{k+12} & \dots & t_{k+1k+1} \end{bmatrix}; X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k+1} \end{bmatrix}; Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{k+1} \end{bmatrix},$$

where:
x_j – share of votes for party j obtained during the first elections;
y_j – share of votes for party j obtained during the second elections;
k – total number of parties (k=max(p,q));
x_{k+1} – abstention during the first elections;
y_{k+1} – abstention during the second elections;
t_{ij} – transfer coefficient of votes transferred from party j during the first elections to votes for party i in the second elections, i,j=1,2,..,k;
t_{k+1j} – transfer of votes for party j in the first elections to abstention in the second elections;
t_{ik+j} – transfer of votes from abstention in the first elections to votes for party i in the second elections.

Extended form of general equation $TX + \varepsilon = Y$ is given by the system of equations:

$$\begin{cases} t_{11}x_1 + t_{12}x_2 + \dots + t_{1k+1}x_{k+1} + \varepsilon_1 = y_1 \\ t_{21}x_1 + t_{22}x_2 + \dots + t_{2k+1}x_{k+1} + \varepsilon_2 = y_2 \\ \vdots \\ t_{k+11}x_1 + t_{k+12}x_2 + \dots + t_{k+1k+1}x_{k+1} + \varepsilon_{k+1} = y_{k+1} \end{cases},$$

where coefficients t_{ij} fulfil conditions $\sum_{i=1}^{k+1} t_{ij} = 1$ for all j=1,2,..,k+1. In the above system of equations transfer coefficients are unknown. Transfer coefficients should be estimated from voting results for considered aggregation level.

First of all, any estimation approach needs a statistical sample. In the case of ecological regression there is one way to obtain a statistical sample: by using aggregated voting results only. Usually, the voting results data are available, even on basic electoral districts level. Assumptions that data from every electoral district are homogenous are obviously not held. There is a necessity to divide electoral districts into homogenous groups - in the sense of electoral behaviour. This approach will be called decomposition. The main idea is to construct the decomposition with respect to reasonable assumption that small regions are more homogeneous than large regions. It is very convenient approach, because usually electoral data are divided into geographically selected voting districts. Homogeneity simply means that those transfer coefficients are the same or roughly the same for selected voting districts. This definition of homogeneity is difficult to evaluate in the sense of fitness, because transfer coefficients are unknown. Thus, the problem of formal definition of homogeneity of transfer electorate coefficients should be avoided.

4 Statistical sampling

The homogeneity assumptions play crucial role in estimation procedure. Simplified approach to use every electoral district's results as a sample is obviously not proper. There is necessity to increase level of homogeneity [6]. The main part of this idea is to divide a given region to m more homogeneous sub-regions. Let vector D contain shares of total number of votes for sub-regions. $D = [d_1, d_2, ..., d_m]^{-1}$, where $\forall r = 1,2,...,m \quad d_r \geq 0$ and $\sum_{r=1}^{m} d_r = 1$. In this approach shares d_r are constant for two consecutive elections. In each sub-region there exists a unique transition matrix. Matrices: $T^{(1)}$, $T^{(2)}$, …, $T^{(m)}$ are defined for each sub-region separately. Let $X^{(1)}, X^{(2)}, ..., X^{(m)}$ denote vectors of results from the first elections divided into sub-regions, and $Y^{(1)}, Y^{(2)}, ..., Y^{(m)}$ vectors for the second elections. Main classical regression assumptions that transition coefficients t_{ij} are constant in the sense of conditional expectation $E\left(T^{(k)} \middle| X^{(k)}\right)$, where k is the number of sub-region, plays a crucial role here.

Therefore, for all sub-regions conditions $T^{(r)}X^{(r)} + \varepsilon^{(r)} = Y^{(r)}$ are fulfilled for r=1, 2, m, where $T^{(r)} = \begin{bmatrix} t_{11}^{(r)} & t_{12}^{(r)} & ... & t_{1k+1}^{(r)} \\ t_{21}^{(r)} & t_{22}^{(r)} & ... & t_{2k+1}^{(r)} \\ \vdots & \vdots & ... & \vdots \\ t_{k+11}^{(r)} & t_{k+12}^{(r)} & ... & t_{k+1k+1}^{(r)} \end{bmatrix}; X^{(r)} = \begin{bmatrix} x_1^{(r)} \\ x_2^{(r)} \\ \vdots \\ x_{k+1}^{(r)} \end{bmatrix}; Y^{(r)} = \begin{bmatrix} y_1^{(r)} \\ y_2^{(r)} \\ \vdots \\ y_{k+1}^{(r)} \end{bmatrix}$. Decompositions of election results fulfil conditions $X = \sum_{r=1}^{m} d_r X^{(r)}$ and $Y = \sum_{r=1}^{m} d_r Y^{(r)}$ for the first and the second elections. In this approach, the common transition matrix T, describing general voting transitions is a function of sub-regional matrices $T = f(T^{(1)}, T^{(2)}, ..., T^{(m)})$, where elements of matrix T are weighted averages $t_{ij} = \frac{\sum_{r=1}^{m} d_r t_{ij}^{(r)} x_j^{(r)}}{x_j}$, where i,j=1,2,…,k+1, $x_j = \sum_{r=1}^{m} d_r x_j^{(r)}$.

Decomposition of transition matrix extends area of application of ecological regression estimators. Such decomposition is sufficient to build common transition matrix for the whole country or a big region with respect to homogeneity assumption. From statistical point of view, quality of statistical sample in the event of decomposition is increasing in general. There are some problems in direct evaluation of influence of decomposition. Indirect evaluation could be made by investigation of standard deviation and other quality measures of obtained estimators. It is possible to compare standard deviations in two cases - with and without decomposition. This comparison will be made at the end of the paper.

5 Estimation procedure for transfer coefficient

In this part of the paper we assume that a homogenous region is given. Let matrix equation $TX + \varepsilon = Y$ be valid for this homogenous region. In this case, statistical sample contains voting results for consecutive elections for all electoral districts belonging to this area. Let assume that homogenous region is divided into n electoral districts. Thus, statistical data for two consecutive elections are given by:

$x_{11}, x_{12}, x_{13}, x_{14}, ..., x_{1n}$ - results for party number 1 (share) in the first elections for n electoral districts,

$x_{21}, x_{22}, x_{23}, x_{24}, ..., x_{2n}$ - results for party number 2 (share) in the first elections for n electoral districts,

......

$x_{k1}, x_{k2}, x_{k3}, x_{k4}, ..., x_{kn}$ - results for party number k (share) in the first elections for n electoral districts,

$x_{k+11}, x_{k+12}, x_{k+13}, x_{k+14}, ..., x_{k+1n}$ - absence in the first elections (share) for n electoral districts.

Results from the first election fulfil obvious condition $\sum_{j=1}^{k+1} x_{jl} = 1$ for each l=1,2,...,n.

$y_{11}, y_{12}, y_{13}, y_{14}, ..., y_{1n}$ - results for party number 1 (share) in the second elections for n electoral districts,

$y_{21}, y_{22}, y_{23}, y_{24}, ..., y_{2n}$ - results for party number 2 (share) in the second elections for n electoral districts,

......

$y_{k1}, y_{k2}, y_{k3}, y_{k4}, ..., y_{kn}$ - results for party number k (share) in the second elections for n electoral districts,

$y_{k+11}, y_{k+12}, y_{k+13}, y_{k+14}, ..., y_{k+1n}$ - absence in the second elections (share) for n electoral districts.

Results from the first election fulfil obvious condition $\sum_{j=1}^{k+1} y_{jl} = 1$ for each $l=1,2,...,n$ (summing inside each electoral districts).

Extended form of transition equation (1) for consecutive elections and for n homogenous electoral districts is as follows:

$$\begin{cases} y_{1i} = t_{11}x_{1i} + t_{12}x_{2i} + \ldots + t_{1k+1}x_{k+1i} + \varepsilon_{1i} \\ y_{2i} = t_{21}x_{1i} + t_{22}x_{2i} + \ldots + t_{2k+1}x_{k+1i} + \varepsilon_{2i} \\ \vdots \\ y_{ki} = t_{k1}x_{1i} + t_{k2}x_{2i} + \ldots + t_{kk+1}x_{k+1i} + \varepsilon_{k+1i} \\ y_{k+1i} = t_{k+11}x_{1i} + t_{k+12}x_{2i} + \ldots + t_{k+1k+1}x_{k+1i} + \varepsilon_{k+1i} \end{cases}, \tag{2}$$

where i=1,2,3,...,n denotes number of district.

We assume that disturbances $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}, ..., \varepsilon_{ki}, \varepsilon_{k+1i}$ are random variables identically normally distributed, N(0,σ), where only subset $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}, ..., \varepsilon_{ki}$ contains independent identical random variables normally distributed, *N(0,σ)*.

Under above assumption expectation for each result for the second election is given with the following formula

$$E(y_{ji}) = t_{j1}x_{1i} + t_{j2}x_{2i} + \ldots + t_{jk+1}x_{k+1i} \tag{3}$$

for every j=1,2,...,$k+1$ (parties), i=1,2,...,n (electoral districts).

Result as shares of voting for party j in the second election fulfill condition $y_{j1} + y_{j2} + y_{j3} + \ldots + y_{jn} = y_j$, where: y_j denotes sum of shares of votes obtained by party number j in the second election in n analysed electoral districts. For single electoral districts there is more useful condition $y_{1i} + y_{2i} + y_{3i} + \ldots + y_{k+1i} = 1$.The left side of the equation is simply total share of votes in *i-th* single electoral district held for all $i=1,2,...,n$.

We distinguish y_{k+1i} share now. This share can be clearly interpreted as the percentage of absenteeism. For distinguished share yk+1i there exists explicit dependence equation $y_{k+1i} = 1 - y_{1i} + y_{2i} + y_{3i} + \ldots + y_{ki}$ for each electoral districts $i=1,2,...,n$. For random disturbances, one of random disturbances ε_{k+1} is dependent on remaining disturbances $\varepsilon_1, \varepsilon_2, \varepsilon_3, ..., \varepsilon_k$. Distribution of sample $y_{1l}, y_{2l}, y_{3l}, y_{4l}, ..., y_{kl}, y_{k+1l}$ where l denotes fixed number of voting districts is given by the multivariate joint Gaussian distribution for non-independent variables with some fixed covariance matrix.

Probability multivariate density of random vector $y_{1l}, y_{2l}, y_{3l}, y_{4l}, ..., y_{kl}, y_{k+1l}$ is expressed by conditional probability density

$$f\left(y_{1l}, y_{2l}, y_{jl}, ..., y_{kl}, y_{k+1l}\right) = f\left(y_{1l}, y_{2l}, y_{3l}, y_{4l}, ..., y_{kl}\right) \cdot f\left(y_{k+1l} \middle| y_{1l}, y_{2l}, y_{3l}, y_{4l}, ..., y_{kl}\right) \tag{4}$$

Conditional density function $f\left(y_{k+1l} \middle| y_{1l}, y_{2l}, y_{3l}, y_{4l}, ..., y_{kl}\right)$ is fixed because $y_{k+1l} = 1 - y_{1l} + y_{2l} + y_{3l} + \ldots + y_{kl}$ determines value of share of absence y_{k+1l} as direct linear function of shares for remaining parties.

Voting results, for all electoral districts for the second election in one homogenous region, form a statistical sample $y_{11}, y_{11}, ..., y_{k1}, y_{k+11}, y_{12}, y_{22}, ..., y_{k2}, y_{k+12}, ..., y_{1n}, y_{2n}, ..., y_{kn}, y_{k+1n}$. For this sample, the likelihood function will be created under the assumption that results of voting are independent across n electoral districts, but there is no independence inside each electoral district.

The part of likelihood function related to the absence in each electoral district is fixed for the whole region. It means that procedure of maximizing likelihood function could exclude conditional part of equation (4).

The non-fixed part of likelihood function depends on voting results and transfer coefficients from equation (2) $L_1\left(t_{ij}, x_{ij}\right) = \prod_{j=1}^{n} f\left(y_{11}, y_{12}, ..., y_{kn}\right)$ where function $f\left(y_{11}, y_{12}, ..., y_{kn}\right)$ is joint multivariate normal distribution of independent random variables with marginal univariate distributions **N**($t_{j1}x_{1i} + t_{j2}x_{2i} + \ldots + t_{jk+1}x_{k+1i}$,σ), where the means are expressed by equation (3) for j=1,2,...,k; i=1,2,..,n. Likelihood function can be used in maximization procedures. In this case, conditional maximization procedure needs to be used with respect to constraints for transfer coefficients $\sum_{i=1}^{k+1} t_{ij} = 1 \textit{ and } t_{ij} \geq 0 \textit{ for all } i, j$

6 The empirical example

We examine maximum likelihood methodology using data from general parliamentary elections in Poland from the 2005 and 2007 elections. The statistical sample contains over 25 000 results from electoral districts. Territory of Poland is divided into 16 homogenous regions according to administrative division into voivodeships. From practical point of view this division is very convenient and obvious.

Maximal likelihood methodology delivers transition matrix T for the whole country. In estimation process there are obtained the matrices for 16 sub-regions (voivodeships). Therefore, the idea of regional decomposition provides one common transition matrix for the whole country.

Election 2007	Election 2005									
	LPR	PiS	SLD*	PO	PSL	Sam	PD*	SD*	Others	Absence
LPR	8,3%	-	0,6%	-	-	-	-	-	-	0,6%
PiS	23,4%	82,9%	10,6%	-	23,8%	33,7%	-	-	5,8%	7,5%
SLD	10,6%	-	21,7%	18,6%	10,3%	11,6%	1,3%	6,3%	-	5,2%
PO	9,4%	17,1%	30,0%	77,2%	-	-	98,7%	93,7%	78,2%	9,6%
PSL	15,5%	-	12,5%	1,8%	22,4%	19,7%	-	-	-	3,5%
Sam	5,0%	-	3,2%	-	6,5%	1,4%	-	-	-	0,5%
Others	8,6%	-	8,3%	2,4%	8,5%	4,7%	-	-	16,1%	0,1%
Absence	19,3%	-	13,0%	-	28,7%	28,8%	-	-	-	72,9%

Tab. 5. Transition matrix for the 2005 and 2007 general parliamentary elections in Poland.
* Parties SLD, PD and SD established a new coalition "SLD" in the 2007 elections.

At this stage of investigation the decomposition approach regarding the maximum likelihood methodology and the decomposition idea can only be evaluated by comparing the measures of fitting the predicted values obtained from the ecological regression model with the actual values (table 5). According to the results in table 6 one may see that the idea in the sense of improving quality of estimators is not rejected.

Measure	The decomposition approach	The whole country
Standard deviation	0,019	0,121
Maximum error	0,093	0,133

Tab. 6. Evaluation of estimation of coefficient of transition of votes between parties in two consecutive elections.

7 Conclusion

The decomposition idea extends the area of application of ecological regression modelling to the level of a country or, in general case, to the level of non-homogenous regions. The solution, as a weighted average of transition coefficients from local matrices, contains information from homogenous sub-regions. It is our next step to investigate probabilistic nature of obtained estimators at the local and the global level and the influence of constrains of the transition model on bias and variance of the obtained estimators.

References

1. Duncan, D., Beverley, D., 1953. *An Alternative to Ecological Correlation*. American Sociological Review vol. 18, 665-666.
2. Fűle, E., 1994. *Ecological Inference on Voting Data*. Department of Statistics, University of Lund.
3. Goodman, L., 1959. *Ecological Regression and the Behaviour of Individuals.* American Sociological Review vol. 18, No. 6 (December), 663-664.
4. Groofman, B., Merrill S., 2002. *Ecological Regression and Ecological Inference.* Prepared for presentation at the Ecological Inference Conference, 2002. Harvard University, Cambridge. http://course.wilkes.edu/merrill/.
5. King, G., 1997. *A Solution to the Ecological Inference Problem*. Princeton University Press.
6. Mazurkiewicz, M., Mercik J. Turnovec F., 2004, *Regional decomposition in Ecological Regression*, Proceedings of the 15th International Conference on Systems Science, Wroclaw 2004.
7. Mazurkiewicz M., Mercik J. Turnovec F., 2006 *Transfer of electorate's preferences during 2001 and 2005 parliamentary elections in Poland* (in Polish), [in:] „Badania operacyjne i systemowe 2006. Metody i techniki". (J. Kacprzyk and R. Budziński – eds). Warsaw, "Exit", 2006, 175-183.

The Competitiveness of Visegrad Four NUTS 2 Regions and its Evaluation by DEA Method Application

Lukáš Melecký[1], Michaela Staníčková[2]

Abstract. The paper deals with evaluating competitiveness level in Visegrad Four (V4) NUTS 2 regions by Data Envelopment Analysis application (DEA). The aim of the paper is to assess competitive potential of V4 NUTS 2 regions on the platform of CCR model of efficiency. It is convenient to use DEA method for regional competitiveness evaluation as it does not evaluate only one factor, but a set of different factors that determine the degree of economic development. DEA method is based on the inputs and outputs of indicators used. It evaluates the ability of regions to transform their inputs into outputs. Therefore, the region's effectiveness can be considered as a 'mirror' of competitiveness. The database of analysis consists of regional data available under the European growth strategies which is published by the European Statistical Office. The first part of the paper focuses on theoretical background of regional competitiveness and DEA methods. The application concentrates on the evaluation of V4 NUTS 2 region's efficiency in the frame of CCR input oriented model. Based on DEA method, finally the comparison of the level of competitive potential of V4 NUTS 2 regions is included.

Keywords: NUTS 2 region, efficiency, Data Envelopment Analysis, CCR input oriented model, competitive potential

JEL Classification: C61, O18, R15, R58
AMS Classification: 62H99, 90C99, 93B15, 93D25

1 Introduction

Measuring and evaluating competitiveness have come to the fore of economic interest recently. The definition of competitiveness is a problematic issue as there is a lack of mainstream view of understanding this term. Competitiveness is one of the fundamental criteria for evaluating economic performance and it also reflects the success in the broader comparison. The concept of competitiveness is understood at different levels - microeconomic, macroeconomic and especially regional. Anyway, there are some differences between these three approaches.

Nowadays, competitiveness is one of the most monitored characteristics of national economies and it is becoming a part of evaluation of their prosperity, welfare and living standards. The concept of competitiveness has quickly spread onto the regional level. Globally, regions are increasingly becoming the drivers of the economy. In general, one of the most striking features of regional economies is the presence of clusters, or geographic concentrations of linked industries [11]. Current economic fundamentals are threatened by shifting production activities to places that provide better conditions. *Regional competitiveness* is also affected by regionalization of public policy due to the shift of decision-making and coordination of activities to the regional level. Within governmental circles, the interest has grown into regional foundations of national competitiveness. Developing of new forms of regionally based policy interventions helps to improve competitiveness of every region and major city and hence the national economy as a whole. *Regions play an increasingly important role in the economic development of states* [7].

Evaluating competitiveness is no fewer complexes as the definition and understanding competitiveness concept itself. As there is a lack of mainstream view of competitiveness evaluation, there is space for alternative approaches. Evaluation of competitiveness in terms of differences between countries and regions should be measured through a complex of economic, social and environmental criteria that can identify imbalanced areas that cause main disparities. Competitiveness is most commonly evaluated by decomposition of aggregate macroeconomic indicators of international organizations – *Institute for Management Development* (IMD) and *World Economic Forum* (WEF). Decomposition of aggregate macroeconomic indicators of international organizations is the most commonly used approach at regional level, as well as *comprehensive* (mostly descriptive) *analysis*

[1] VŠB – Technical University Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava, lukas.melecky@vsb.cz.
[2] VŠB – Technical University Ostrava, Faculty of Economics, Department of European Integration, Sokolská třída 33, 701 21 Ostrava, michaela.stanickova.st@vsb.cz.

aimed at identifying the key factors of regional development, productivity and economic growth, for example [1,13,8]. Another approach is the evaluation by EU Structural indicators, which are used for the assessment and the attainment of the objectives of the Lisbon strategy, see [6, 9]. Finally, regional competitiveness can be explored by macro-econometric modelling with creation of an econometric panel data model, see [6]. The method of *Data Envelopment Analysis* (DEA) is the other approach for measuring regional efficiency and subsequent for measuring regional competitive potential.

2 Measuring Efficiency at Regional Level by DEA Method

Creating competitiveness evaluation system in terms of the European Union (EU) is greatly complicated by the heterogeneity of countries and regions and what is more their individual approach to the original concept of competitiveness. Finding sources of national competitiveness makes us focus primarily on lower territorial units. On the contrary, accepting the growing importance of regions in EU concept is necessary because it deserves increasing attention especially regarding the fact that the *economic performance of the regions is absolutely crucial for the competitiveness of the country* [10].

Evaluating regional competitiveness is determined by the selected territorial region level, especially in terms of the European Union through the Nomenclature of Territorial Units Statistics (NUTS). The reference period, availability and periodicity of data and selection of convenient specific factors do not play a minor role. Factors affecting regional competitiveness are therefore becoming the subjects of evaluation on qualitative or quantitative level. Selecting appropriate criteria for evaluating regional competitiveness is the key issue as these need to be universally acceptable. For evaluating regional competitiveness it is necessary to note that the data availability decreases in direct proportion to the lower territorial unit (NUTS). For evaluating regional competitiveness in the terms of the EU, the most appropriate territorial unit is NUTS 2 level which is in the centre of the interest of the European Commission in terms of fulfilling the objectives of EU Cohesion Policy.

There are economic, social and territorial disparities among the EU Member States and their regions. The disparities have a negative impact on the balanced development across the European Union and what is more they may weaken the competitiveness in the global context. These differences in regional efficiency are a subject of intense research. The performance analysis can be used for evaluation of regional development quality (efficiency with respect to the regional factor endowment). The Data Envelopment Analysis model is based on *Farrel model* for measuring the effectiveness of units with one input and one output, which expanded *Charnes*, *Cooper* and *Rhodes* (CCR model) and *Banker*, *Charnes* and *Cooper* (BCC model), see for example [2]. CCR is a method to measure decision making units' (DMU) efficiency. Since DEA in its present form was first introduced in 1978, researchers in a number of fields have quickly recognized that it is an excellent and easily used methodology for modelling operational processes for performance evaluations. DEA's empirical orientation and the absence of a need for the numerous a priori assumptions that accompany other approaches (such as standard forms of statistical regression analysis) have resulted in its use in a number of studies involving efficient frontier estimation in the governmental and non-profit sector, in the regulated sector, and in the private sector [4].

In the original study [see 2] is DEA described as a 'mathematical programming model applied to observational data (that) provides a new way of obtaining empirical estimates of relations - such as the production functions and/or efficient production possibility surfaces – that are cornerstones of modern economics. Formally, DEA is a relatively new "data oriented" approach for evaluating the performance of a set of peer entities, called Decision Making Units (DMUs), which convert multiple inputs into multiple outputs. DEA is a specialized model tool used for evaluating effectiveness, efficiency and productivity of homogenous group (DMU). The definition of a DMU is generic and flexible. DEA applications have used DMUs of various forms to evaluate the performance of entities, such as banks, hospitals, universities, cities, courts, business firms, and others, including the performance of countries, regions, etc. Because it requires very few assumptions, DEA has also opened up possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMUs [4].

DEA is *multi-criteria decision making method* because of the wide range of inputs and outputs for evaluation of DMU. DEA is convenient to determine efficiency of DMU which are mutually comparable. This means that we use same inputs, produce same outputs, but their performances are different. Homogenous production unit is a set of unit's producing identical or equivalent effects, which will be referred as the outputs of these units. To create such effects, each unit uses inputs that are contrary in their nature minimization, i.e. lower value of these inputs leads to higher performance of these units. An efficiency analysis compares the actual output of a DMU with the maximal output estimated by a production function. The best-practice units of a comparison group are used as a reference for the evaluation of the other group units. The aim of this method is to examine DMU on the effective and no effective by the size and quantity of consumed resources by the produced output or other type of output [3].

DEA method is thus convenient for comparing the competitiveness of regional units. For evaluation of regional units we can define a number of similar indicators, which are based on different data and results may not be and they typically are not consistent. A larger number of inputs, outputs should be used for measuring and evaluating the overall effectiveness of unit as well.

In *DEA models aimed at inputs* the efficiency coefficient of efficient units (located on the efficient frontier package) always equals 1, while the efficiency coefficient of inefficient units is less than 1. In *DEA models aimed at outputs* is the efficiency coefficient of efficient units (located on the efficient frontier package) always equals 1, while the efficiency coefficient of inefficient units is greater than 1. In the basic DEA models are efficient units depending on the unit rate of effectiveness. Depending on the chosen model, but also on the relationship between number of units and number of inputs and outputs, can be number of effective units relatively large.

3 Application of DEA Method in Analysis of V4 NUTS 2 Regions Efficiency

Data Envelopment Analysis is applied to all 35 NUTS 2 regions in Visegrad Four countries (V4). The application of multi-criteria decision making methods for evaluating the competitiveness and comparing the competitiveness of regional units seems to be very convenient. DEA evaluates the efficiency of the regions with regard to their ability to transform inputs into outputs. In other words, what results the region can achieve spending a relatively small number of inputs (resources). This fact is vital for us to perceive the efficiency like a "mirror" of competitiveness. We do not evaluate one factor, but a number of factors that determine the degree of economic development (efficiency) of the regions, same as for example [10].

Methodology and data used in this paper are related to the several fundamental aspects that have an influence on the presented outputs, and ultimately on the results of evaluating regional competitiveness. They are mainly based on: (1) the chosen level of territorial region, (2) the duration of the period, (3) the periodicity of data of the programming period, and (4) the selection of the appropriate indicators for analysis of the available regional data. The analysis of the regional efficiency in V4 NUTS 2 regions starts from a database of indicators monitored by Eurostat – indicators of Regions and cities statistics, EU structural (Lisbon) indicators and indicators of Strategy Europe 2020. All data used in analysis is available at [5]. Database analysis consists of *five indicators* – three of them are inputs and two outputs. The *reference period* is determined by an early adoption and the current start of the Lisbon strategy in 2000 and the availability of selected indicators at regional level which ends in 2009 for most indicators.

CCR input oriented model assuming *constant returns to scale* (CRS) have been used in this paper; CCR model evaluates the effectiveness of units for any number of input and output. The efficiency coefficient is the ratio of weighted sum of output and weighted sum of input. They are searched such weights (coefficients) to calculate the efficiency coefficient in interval 0 – 1. The unit with the efficiency coefficient equal to 1 is effective, the efficiency coefficient less than 1 indicates an inefficient unit, and determines the degree of reducing the input needed to ensure the effectiveness of the unit, as explained [12].

It is necessary to note criteria for selecting inputs and outputs used in DEA model. The first input is Gross domestic expenditure on research and development (GERD) which measures the key R&D investments that supports future competitiveness and results in a higher GDP. GERD represents one of the major drivers of economic growth in a knowledge-based economy. Trends in the GERD indicator provide key indications of the future competitiveness and wealth of the EU. It is quite obvious that the overall performance of the economy affects the number of people employed in various sectors of the economy, their skills and their length of working age (20-64 years) that is why we selected the criterion of Employment rate, so this is the second input. The third included input is a number of students by tertiary education that presents a new indicator targeted in Strategy Europe 2020.

There are two outputs in the case of the presented DEA model. Reflected outputs are measured by Gross domestic product (GDP) in purchasing parity standards and Labour productivity per person employed. The Gross domestic product is the most important macroeconomic aggregate, and if it is measured per region, we can take into account the limited number of inputs due to which it was achieved. Similarly, we may deal with the labour productivity as it shows us how much production economically active people have created, or employed persons in the national economy. In terms of regions, we also take into account which national economy sectors were involved the most or the least in the production.

In Figure 1 and Table 1, placed in annexes, apparently the best results are traditionally achieved by economically powerful regions which were "efficient" during the whole referred period. It means that the outputs achieved were greater than those incurred inputs. The resulting efficiency index achieved by DEA is equal to 1 for an ‘efficient’ region within the whole period 2000 to 2009. In Table 1, there are ‘efficient’ NUTS 2 regions coloured by grey colour. These are NUTS 2 regions in the Czech republic - CZ01 (Praha), CZ02 (Střední Čechy) and CZ04 (Severozápad). NUTS 2 regions PL22 (Sląskie), PL42 (Zachodniopomorskie) and PL43 (Lubuskie) belong among the most efficient NUTS 2 regions in Poland. In Slovakia, the most efficient NUTS 2 region is SK01 (Bratislava region). There is no region which efficient coefficient is equal to 1 in Hungary.

The above listed efficient regions are followed by a group of 10 regions which are also highly efficient. These regions achieved efficiency equal to 1 at least in one year during the referred period (coloured also by grey colour in Table 1). These regions are in Hungary - HU10 (Közép-Magyarország), HU21 (Közép-Dunántúl), HU22 (Nyugat-Dunántúl), HU23 (Dél-Dunántúl) and HU31 (Észak-Magyarország) and in Poland - PL33 (Swietokrzyskie), PL51 (Dolnoslaskie) and PL52 (Opolskie). Other regions with efficiency index less than 1 are classified as ‘ineffective’, i.e. these regions are considered non-competitive. In Table 1, there are ‘ineffective’ NUTS 2 regions coloured by light grey colour. In Figure 1 (in annexes), we can follow the dynamic character of efficiency index (by CCR input oriented model) of the individual NUTS 2 regions within the V4 countries in the period from 2000 to 2009. Taking into account “reverse” character of the efficiency index, similar dynamic character has been achieved also for CCR output oriented model (the corresponding graphs are not presented here).

4 Conclusion

The aim of the paper was to present a specific approach of multi-criteria evaluation for NUTS 2 regions competitiveness in Visegrad Four countries during period 2000 – 2009. The analysis was based on the applications of DEA input oriented CCR model that calculates the efficiency index of the region. It was found out that in 35 NUTS 2 regions of V4 countries, there were 7 efficient regions within the whole referred time period. Moreover, 4 other regions were effective in more than one year of the referred period and 4 more regions were effective at least in one of the years during period 2000-2009. The rest of 20 NUTS 2 regions of V4 countries belong to the category of inefficient regions and thus with less competitive potential. Application of DEA method presents, in absence of the mainstream approach of regional competitiveness evaluation, convenient way for cross-regional comparison of competitiveness in V4 countries. Presented approach (CCR input oriented model) is, of course, only one specific example of DEA methods. It is possible to evaluate competitiveness through other DEA methods, see e.g. [2, 3, 6, and 12].

Acknowledgements

This paper was created within Grant Competition of Faculty of Economics, VŠB-Technical University of Ostrava. Project registration number is SP2011/124.

References

[1] Blažek, L., Viturka, M. et al.: *Analýza regionálních a mikroekonomických aspektů konkurenceschopnosti*. Ekonomicko-správní fakulta MU Brno, Brno, 2008.

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

[3] Cook, W. D., Zhu, J.: *Data Envelopment Analysis: Modelling Operational Processes and Measuring productivity*. Kluwer Academic Publishers, Boston, 2008.

[4] Cooper, W. W., Seiford, L. M., Zhu, J.: *Data Envelopement Analysis: Models and Interpretations*. Kluwer Academic Publisher, Boston, 2004.

[5] Eurostat [online]. *Statistics*. 2011. [cit. 2011-05-09]. Available on WWW: <http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/themes>.

[6] Hančlová, J., Kubicová, I., Macháček, M., Melecký, A., Melecký, L., Melecký, M., Nevima, J., Ramík, J.: *Makroekonomické modelování české ekonomiky a vybraných ekonomik EU*. VŠB-TU Ostrava, Ostrava, 2010.

[7] Krugman, P.: Competitiveness: A Dangerous Obsession. *Foreign Affairs* (1994), 28-44.

[8] Martin, R. [online]. *A Study on the Factors of Regional Competitiveness*. 2003. <http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/3cr/competitiveness.pdf >.

[9] Melecký, L., Nevima, J.: Regionální konkurenceschopnost zemí V4 a její hodnocení prostřednictvím strukturálních ukazatelů EU. In: *Evropská unie po českém předsednictví*, Ostrava, 2009, 1 - 16.

[10] Nevima, J., Ramík, J.: Application of DEA for evaluation of regional efficiency of EU regions. In: *Mathematical Methods in Economics 2010*, České Budějovice, 2010, 477-482.

[11] Porter, M. E.: The Economic Performance of Regions. *Regional Studies* 37 (2003), 549-578.

[12] Zhu, J.: *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver*. Kluwer Academic Publishers, Boston, 2002.

[13] Žítek, V., Kunc, J., Tonev, P.: *Vybrané indikátory regionální konkurenceschopnosti a jejich vývoj.* Ekonomicko-správní fakulta MU Brno, Brno, 2006.

Annexes

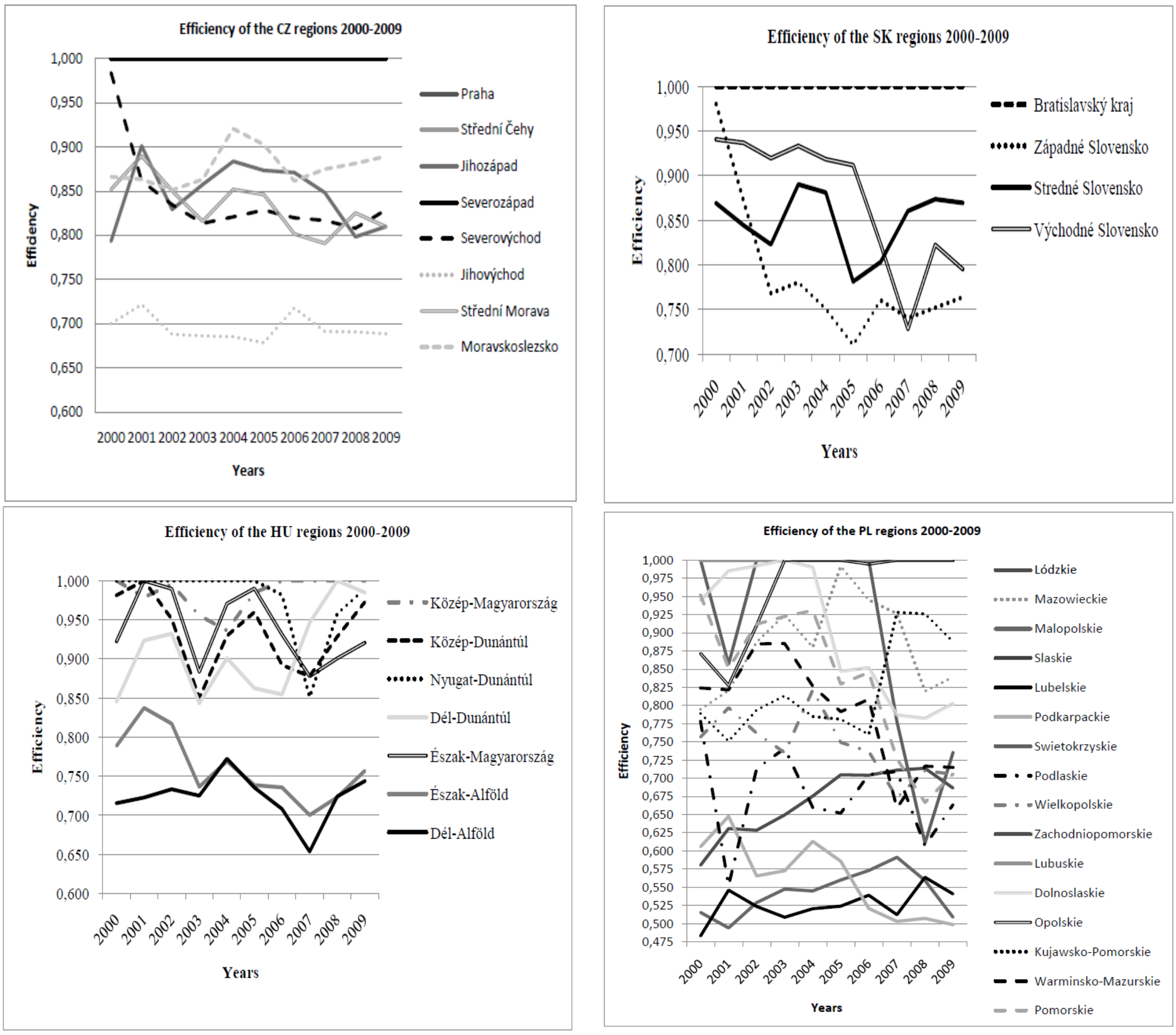


Figure 1 Efficiency of V4 NUTS 2 regions (CCR, Input oriented model)
Source: Own calculation and elaboration, 2011

Rank	Code	Region	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	CZ01	Praha	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
2	CZ02	Střední Čechy	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
3	CZ03	Jihozápad	0,793	0,901	0,829	0,858	0,884	0,873	0,871	0,848	0,798	0,810
4	CZ04	Severozápad	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
5	CZ05	Severovýchod	0,984	0,862	0,835	0,814	0,821	0,829	0,820	0,817	0,808	0,829
6	CZ06	Jihovýchod	0,700	0,722	0,688	0,686	0,685	0,678	0,718	0,691	0,691	0,689
7	CZ07	Střední Morava	0,852	0,890	0,850	0,816	0,852	0,846	0,802	0,791	0,825	0,810
8	CZ08	Moravskoslezsko	0,866	0,863	0,851	0,863	0,921	0,903	0,862	0,875	0,881	0,890
9	HU10	Közép-Magyarország	1,000	0,979	0,995	0,956	0,936	0,985	1,000	1,000	1,000	1,000
10	HU21	Közép-Dunántúl	0,981	1,000	0,951	0,850	0,929	0,960	0,892	0,878	0,928	0,972
11	HU22	Nyugat-Dunántúl	1,000	1,000	1,000	1,000	1,000	1,000	0,982	0,851	0,958	0,989
12	HU23	Dél-Dunántúl	0,846	0,924	0,932	0,843	0,901	0,863	0,854	0,946	1,000	0,985
13	HU31	Észak-Magyarország	0,922	1,000	0,989	0,883	0,970	0,990	0,931	0,878	0,901	0,920
14	HU32	Észak-Alföld	0,789	0,837	0,817	0,736	0,769	0,739	0,736	0,700	0,723	0,757
15	HU33	Dél-Alföld	0,715	0,723	0,733	0,725	0,772	0,735	0,708	0,654	0,724	0,744
16	PL11	Lódzkie	0,580	0,631	0,628	0,650	0,675	0,705	0,704	0,711	0,714	0,687
17	PL12	Mazowieckie	0,795	0,821	0,887	0,925	0,880	0,992	0,945	0,926	0,820	0,839
18	PL21	Malopolskie	0,515	0,494	0,529	0,548	0,545	0,560	0,573	0,591	0,559	0,509
19	PL22	Slaskie	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
20	PL31	Lubelskie	0,483	0,546	0,524	0,509	0,521	0,524	0,539	0,512	0,564	0,541
21	PL32	Podkarpackie	0,606	0,648	0,566	0,573	0,613	0,586	0,521	0,503	0,507	0,499
22	PL33	Swietokrzyskie	1,000	0,857	1,000	1,000	1,000	1,000	1,000	0,778	0,611	0,735
23	PL34	Podlaskie	0,778	0,552	0,713	0,741	0,660	0,652	0,705	0,709	0,607	0,664
24	PL41	Wielkopolskie	0,757	0,797	0,762	0,735	0,821	0,749	0,737	0,673	0,710	0,706
25	PL42	Zachodniopomorskie	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
26	PL43	Lubuskie	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
27	PL51	Dolnoslaskie	0,945	0,985	0,992	1,000	0,990	0,847	0,852	0,787	0,782	0,803
28	PL52	Opolskie	0,872	0,828	0,911	1,000	1,000	1,000	0,995	1,000	1,000	1,000
29	PL61	Kujawsko-Pomorskie	0,789	0,751	0,794	0,814	0,785	0,781	0,760	0,928	0,926	0,888
30	PL62	Warminsko-Mazurskie	0,824	0,822	0,885	0,885	0,827	0,792	0,809	0,659	0,717	0,715
31	PL63	Pomorskie	0,952	0,851	0,912	0,923	0,931	0,830	0,846	0,727	0,667	0,706
32	SK01	Bratislavský kraj	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
33	SK02	Západné Slovensko	0,980	0,871	0,768	0,780	0,751	0,710	0,760	0,741	0,752	0,764
34	SK03	Stredné Slovensko	0,869	0,844	0,823	0,890	0,881	0,781	0,803	0,860	0,874	0,869
35	SK04	Východné Slovensko	0,940	0,937	0,919	0,933	0,918	0,912	0,823	0,728	0,822	0,795

Table 1 Application of DEA for V4 NUTS 2 regions (CCR, Input oriented model)

Source: Own calculation and elaboration, 2011

Okun’s Law in Transition Economies – the Case of the Czech and Slovak Republics 1995-2010

Elena Mielcová[1]

Abstract. The main aim of this article is to study the Okun’s Law - the mutual relationship between the gross domestic product (GDP) and unemployment rate time series – on data from two transition countries, in this case on data from the Czech and Slovak Republics in 1995-2010. The analysis is based on two versions of Okun’s law – the difference and the gap version of the Okun’s law. Estimations of Okun’s law coefficients show the persistence of basic rule – the GDP growth causes the unemployment rate decrease. However, estimations based on the Czech and Slovak data sets give different results comparing to those from the US time series.

Keywords: Okun’s law, GDP, unemployment rate.

JEL Classification: E01, E17, E24, O11.
AMS Classification: 62P20, 91B64

1 Introduction

The relationship between gross domestic product (GDP) growth and unemployment rate change time series – so-called Okun’s Law – was originally studied by the economist Arthur Okun in early 1960s [13]. Okun revealed the negative correlation between the two variables and stated that each percentage point of real output growth above 4 percent was associated with a fall in the unemployment rate of 0.07 percentage point [14]. This relation was re-estimated and studied by other scientists on different time series; however the relation shows persistency in results over the long-term period. Now, students of economics learn that for 1 % excess of the natural unemployment rate, a 2-3 % GDP gap can be predicted [1]. These numbers are based on empirical analysis from the data of US economy. Empirical analysis of Okun’s law on the US data can be find in [5,7,15].

Similar analyses based on data from other countries support the persistency in Okun’s Law, however results are different [9]. For example, estimation based on Spanish regional data [17] reveals that for 1 % decrease of the natural unemployment rate is associated with 0.38-1.41 % increase in real output growth. Regional differences are bound with productivity growth. Re-estimation of the difference Okun’s model on data from France, the USA and the Czech Republic for the period of 1996-2009 reveals that the 1 percentage point decrease in the unemployment rate is bound with 2.7 % increase of real GDP in France, with 1.8 % increase of real GDP in the USA, and with 10.1 % increase of real GDP in the Czech Republic for the time period of the same length [12]. The last result indicates that the magnitude of Okun’s coefficient for transition countries could be much higher that those of developed countries. Moreover, the magnitude of Okun’s coefficient should decrease over time as countries are at the end in process of transition. Thus, the main aim of this article is to check if this observation is correct using data from the Czech and the Slovak Republic for the period of 1995-2010, as two representatives of transition countries.

Since the Velvet Revolution in 1989, the former Czechoslovakia started the period of huge changes. The socialist country with centrally planned economics and official zero unemployment rate started transition into the free-market economy of the western type. Moreover, in 1993, the country was divided into two independent countries – the Czech Republic and the Slovak Republic. Both countries continued transformation; the transformation was officially over in 2006 (“developed country” mark according to the World Bank [18]). The Czech and the Slovak Republics has joined the European union in 2004, moreover the Slovak Republic has became a member of EMU of the European Union in 2009. Hence both countries have mutual history, and similar development. Thus, we can expect similar results in estimations of basic economic rules. The only disadvantage is in the lack of data from the first years of transition period – the quarterly time series of basic macroeconomic indicators are available from 1995. Even though the data set is big enough to basic estimations.

The main aim of this article is to study the Okun’s Law - the mutual relationship between the gross domestic product and unemployment rate time series – on data from two transition countries, the Czech and Slovak Republics. The analysis is based on two versions of Okun’s law – the gap and the difference version of the Okun’s

[1] Silesian University /Faculty of Business Administration, Department of Mathematical Methods in Economics; CERGE-EI Affiliate Fellow; Univerzitní náměstí 1943/3, 733 40 Karviná, Czech Republic; mielcova@opf.slu.cz.

law. This article is organized as follows: The next part is devoted to explanation of different Okun's law estimation relations. Data description and results of analysis are given in the third part, followed by conclusion and list of references.

2 Okun's Law – Estimation Relations

As mentioned above, the relationship between GDP growth and unemployment rate change time series – so-called Okun's Law – was originally studied by the economist Arthur Okun in early 1960s [13]. There are two basic versions of Okun's law. The difference version of Okun's law relates quarterly changes in the unemployment rate Δu_t to real output growth $\frac{\Delta y_t}{y_t}$ [8,17]:

$$\Delta u_t = a + b\frac{\Delta y_t}{y_t} + \varepsilon_t \tag{1}$$

where u_t is the observed unemployment rate and y_t is the natural logarithm of observed real output; differences are derived such that $\Delta u_t = u_t - u_{t-1}$, $\Delta y_t = y_t - y_{t-1}$, disturbances are denoted by ε_t. Parameter b (often called the Okun's coefficient) is expected to be negative. The ratio $-\frac{a}{b}$ gives the output growth rate under a condition of stable unemployment rate. Okun's results based on the quarterly US data 1948-1960 data were $a = 0.30$ and $b = -0.07$ [8].

The original gap version of Okun's law relates unemployment rate level u_t to output gap – the difference between the (log of) level of potential output y_t^* and the (log of) level of observed real output y_t [8]:

$$u_t = u_t^* + b'(y_t - y_t^*) + \varepsilon_t \tag{2}$$

Constant term – variable u_t^* determines the level of natural unemployment rate – the unemployment rate related with full employment level. Regression coefficient b' is expected to be negative. Okun expected that the natural unemployment rate was 4 percent [8]; the coefficient b' is usually taken to have a value in the vicinity of (-3) [14].

Another possibility of the gap Okun's law model formulation can be of the form [17]:

$$y_t - y_t^* = \alpha + \beta(u_t - u_t^*) + \varepsilon_t \tag{3}$$

Where y_t^* denotes the (log of) level of potential output, and u_t^* is the natural level of unemployment rate. Coefficient β is negative and determines how much of the level of unemployment gap induce unit of output gap.

The main problem with the gap Okun's law model is the fact, that there are no observable values of y_t^* a u_t^*. The usual way to handle this problem is to generate trend series for both output and unemployment rate time series. This analysis uses four different detrending techniques – moving averages, an exponential smoothing method, the Hodrick-Prescott filter, and the Baxter-King Filter. These techniques are widely used to estimate the output and unemployment gap time series [2,3,6,10,19].

The Okun's law model does not cover the time changes in estimating coefficients. In order to capture the variation, Knotek in [8] proposed the technique of so-called "rolling regressions". A rolling regression estimates a particular relationship over many different moving sample periods; if the relationship is stable over time, then the estimated coefficients should be similar over time. This analysis cover the time series plot of coefficients using the "rolling regressions" technique.

Results of all three Okun's law model forms are given in the next part of this article. The trend of natural unemployment rate level is illustrated using rolling regressions technique.

3 Okun's Law - Estimation Results and Discussion

Estimations of all Okun's law relations for the case of the Czech Republic are based on quarterly data of unemployment rate and real GDP time series for the period of I/1995-IV/2010. Quarterly Czech macroeconomic data are available at the official site of the Czech Statistical Office (www.cso.cz) [4]. Estimations for the case of the Slovak Republic are based on quarterly data of unemployment rate and seasonally adjusted real GDP time series for the same period of I/1995-IV/2010. Quarterly Slovak macroeconomic data are available at the official site of the Statistical Office of the Slovak Republic, SLOVSTAT database (portal.statistics.sk)[16]. All estimations were done using gretl statistical software.

3.1 Difference version of Okun's law

The difference version of Okun's law relates unemployment rate changes to real output growth rate as given in equation (1): $\Delta u_t = a + b\frac{\Delta y_t}{y_t} + \varepsilon_t$. Results of OLS procedure reveals high autocorrelation of disturbances, the time series Prais-Winsten estimation gives reasonable results, as given in Table 1. Results of the difference Okun's law model for the Czech and Slovak Republic confirms the persistence of the negative relationship between the unemployment rate change and the output growth rate ($\mathbf{b}_{CR} = -0.897$, $\mathbf{b}_{sr} = -0.687$, both coefficients are statistically significant at 0.01 level).

Country	a	b	t-ratio	p-value	R^2	D-W	-a/b
Czech Republic	0.150	-0.985***	-7.871	$7.3.10^{-11}$	0.453	2.28	0.152
Slovak Republic	0.151	-0.693***	-5.423	$1.1.10^{-6}$	0.331	2.06	0.218

Table 1 Prais-Winsten estimation of the difference version of Okun's law model: $\Delta u_t = a + b\frac{\Delta y_t}{y_t} + \varepsilon_t$.

The output growth rate under a condition of stable unemployment rate is 0.152 and 0.218 for the Czech and Slovak Republics, respectively. Moreover, the relationship expresses, that each increase of GDP by approximately 1 % causes the decrease in unemployment rate by 1 percentage point in the Czech Republic. The Slovak data reveals that the 1 % decrease in unemployment rate is accompanied with 1.44 % increase in GDP. These numbers are different comparing with US estimations [8], however are close to results from Spain [17].

3.2 Gap version of Okun's Law – unemployment rate estimation

The originally proposed gap version of Okun's law relates the level of actual unemployment rate to output gap, as given in equation (2): $u_t = u_t^* + b'(y_t - y_t^*) + \varepsilon_t$. The constant term is expected to be so-called natural unemployment rate – the unemployment rate observed at a "full employment" level. In the original Okun's work the natural unemployment level was about 4 %.

Estimations of the gap model were done using four detrending methods – moving averages (4 periods), exponential smoothing (current weight 0.8), Hodrick-Prescott filter ($\lambda = 1600$), and Baxter-King filter (k=12). Results of Prais-Winsten estimations, together with values of R^2, Durbin-Watson statistics are given in Tables 2 and 3 for the Czech and Slovak Republics, respectively.

Detrending method	u*	b'	t-ratio	p-value	R^2	D-W
Moving averages	6.293	-6.902***	-4.578	$2.6.10^{-8}$	0.940	0.949
Exponential smoothing	5.840	-3.511***	-2.696	0.009	0.934	1.223
Hodrick-Prescott filter	6.138	-7.012***	-4.987	$5.2.10^{-6}$	0.945	0.976
Baxter-King filter	5.961	9.688	0.812	0.421	0.833	1.836

Table 2 Czech Republic, Prais-Winsten GLS estimation of the gap version of Okun's law model:
$u_t = u_t^* + b'(y_t - y_t^*)$

The first interesting result is that the expected level of natural unemployment rate is quite high - **u*** is expected to be around 6 % for the Czech Republic and it is up to 14 % in the Slovak Republic. These numbers are much higher comparing to those from the USA [8]. This result is caused by the high unemployment rate during the transformation.

Coefficient **b'** is negative with the exception of model estimation using Baxter-King filter for the Czech data set. Results of **b'** estimations have higher magnitude comparing to original USA results. Almost all estimated coefficients of **b'** are statistically significant at 0.001 level, thus the high amplitude of coefficients is caused by situation in both countries; the zero unemployment rate in the former Czechoslovakia was caused by overemployment in production facilities. Restructuralization of production facilities led to higher unemployment rate level with small shifts in output levels.

Detrending method	u*	b'	t-ratio	p-value	R^2	D-W
Moving averages	13.995	-7.705***	-4.274	$7.2.10^{-5}$	0.944	1.304
Exponential smoothing	13.270	-5.929***	-3.946	0.0002	0.939	1.430
Hodrick-Prescott filter	14.203	-7.464***	-4.601	$2.1.10^{-5}$	0.943	1.362
Baxter-King filter	14.939	-49.663***	-2.911	0.006	0.914	1.880

Table 3 Slovak Republic, Prais-Winsten estimation of the gap version of Okun's law model:

$$u_t = u_t^* + b'(y_t - y_t^*)$$

In general, the estimation of Okun's relationship does not cover time variation in coefficients. However, using rolling regressions can demonstrate a particular relationship over time period. Figure 1 gives estimated levels of unemployment rate for the Czech and Slovak Republics. Estimations of rolling regressions for the Czech data were done using data from the moving periods of the length 12 (3 years).

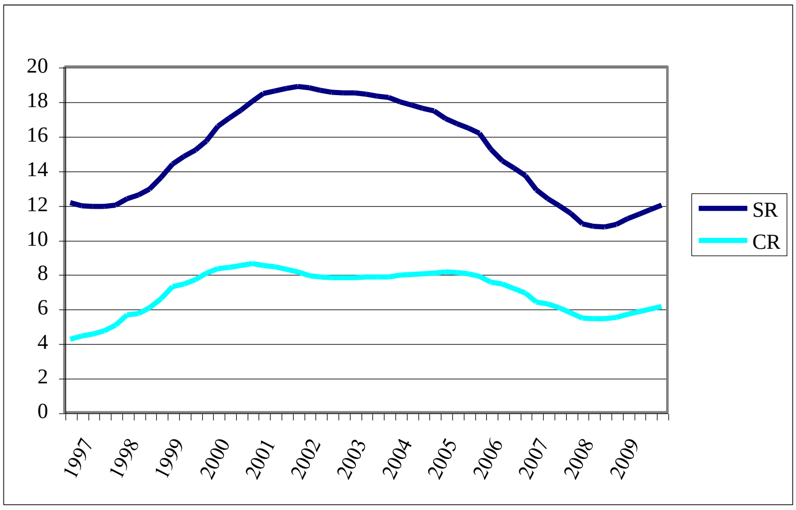


Figure 1 Level of natural unemployment rate – rolling regression estimation.

Levels of natural unemployment rate of the Czech and Slovak Republics were highly correlated (r=0.9). The increase in 1997-2001 was caused by transformation of countries' economies. From 2006 (official end of transformation), the level of natural unemployment rate decreased, however it started to increase in 2008 – the last upward shift was caused by start of world recession (evaluation of recession can be find in [11]).

3.3 Gap version of Okun's Law – GDP gap estimation

The last estimated Okun's model in this article – the gap version with GDP gap estimation is of the form $y_t - y_t^* = \alpha + \beta(u_t - u_t^*) + \varepsilon_t$. The only reasonable results of the model for both countries are that based on the moving average detrending method. This contradicts to results of originally estimated model based on Spanish data [17]. The β coefficient is negative, however it is very small comparing to Spanish results (in [17] magnitude of β varied from 0.38 to 1.41. Results indicate, that for 1 % excess of the natural unemployment rate, a 0.068 % and 0.048 % GDP gap can be predicted for the Czech Republic and the Slovak Republic, respectively. Results of the ordinary least squares estimation, together with values of R^2, Durbin-Watson statistics, and t-values together with the significance level (P-value) for estimated coefficient β are given in Table 4 and 5.

Detrending method	**alpha**	**beta**	**t-ratio**	**p-value**	$\mathbf{R^2}$	**D-W**
Moving averages	0.000	-0.068***	-6.409	$2.9.10^{-8}$	0.415	2.143
Exponential smoothing	0.018	0.061***	4.508	$3.0.10^{-5}$	0.247	0.257
Hodrik-Prescott filter	0.000	-0.007*	-1.748	0.085	0.047	1.526
Baxter-King filter	0.004	0.004	1.308	0.1989	0.043	0.144

Table 4 Czech Republic, ordinary least squares estimation estimation of the gap version of Okun's law model: $y_t - y_t^* = \alpha + \beta(u_t - u_t^*) + \varepsilon_t$

Detrending method	**alpha**	**beta**	**t-ratio**	**p-value**	$\mathbf{R^2}$	**D-W**
Moving averages	-0.001	-0.048***	-5.481	$9.6.10^{-7}$	0.341	2.075
Exponential smoothing	0.022	0.009	0.681	0.498	0.007	0.179
Hodrik-Prescott filter	0.000	-0.015***	-4.502	$3.0.10^{-5}$	0.246	1.745
Baxter-King filter	0.006	-0.010***	-5.088	$1.0.10^{-5}$	0.405	0.244

Table 5 Slovak Republic, ordinary least squares estimation S estimation of the gap version of Okun's law model: $y_t - y_t^* = \alpha + \beta(u_t - u_t^*) + \varepsilon_t$

4 Results Summary

The main aim of this article was to study the mutual relationship between the real GDP and unemployment rate time series – Okun's law – on data from the Czech and Slovak Republics in 1995-2010. Three forms of the Okun's model were estimated. Results summary:

- Results of the difference model $\Delta u_t = a + b\dfrac{\Delta y_t}{y_t} + \varepsilon_t$ indicate that the output growth rate under a condition of stable unemployment rate is 0.152 and 0.218 for the Czech and Slovak Republics, respectively. Estimations reveal that the 1 % decrease in unemployment rate is accompanied with 1.44 % GDP increase in the Slovak Republic and 1.01 % GDP increase in the Czech Republic.
- Results of the gap version model $u_t = u_t^* + b'(y_t - y_t^*) + \varepsilon_t$ denote that the natural unemployment rate is expected to be about 6 % for the Czech Republic and it is up to 14 % in the Slovak Republic. Moreover, the 1 % real output gap decrease is associated with 3.5-7 % unemployment rate increase under the natural level of unemployment in the Czech Republic and with 5.9-7.7 % unemployment rate increase in the Slovak Republic. The natural level of unemployment is shifting through the time; both Czech and Slovak levels of natural unemployment rates are highly correlated.
- Results of the gap version model $y_t - y_t^* = \alpha + \beta(u_t - u_t^*) + \varepsilon_t$ indicate, that for 1 % excess of the natural unemployment rate, a 0.068 % and 0.048 % GDP gap can be predicted for the Czech Republic and the Slovak Republic, respectively.
- Main differences between Slovak and Czech results are caused by different transformation impacts after the Velvet Revolution. Originally, Slovak industrial production was mainly based on machinery and heavy industry with high ratio of import-oriented arms industry, reduced in 1990s. That reduction caused higher unemployment rate levels in Slovakia, while the unemployment rate decrease was lower in the Czech Republic because of its high variety production facilities and higher previous development level. Even though the transformation in both countries has finished in 2006, differences between countries still remain.

The obtained results support basic negative relationship between the unemployment change rate and the output change rate. However, results differ from those estimated on data from developed economies. Results indicate that the way from the centrally planned to the free market economy is associated with untypical relations between main macroeconomic indicators.

References

[1] Abel, A. B., Bernanke, B. S.: *Macroeconomics (5th ed.).* Pearson Addison Wesley, 2005.

[2] Araujo, C. H. V., Areosa M. B. M., Guillén O. T. de C.: *Estimating Potential Output and Output Gap for Brazil.* Banco Central do Brasil, July 2004.

[3] Cerra V., Saxena S. C.: Alternative Methods of Estimating Potential Output and the Output Gap: An Application to Sweden. NBP Conference "Potential Output and Barriers to Growth", *IMF Working Paper* WP/00/59, 2010.

[4] Czech Statistical Office, <http:/www.cso.cz>, accessed 05/20/2011.

[5] Evans G. W.: Output and Unemployment Dynamics in the United States: 1950-1985. *Journal of Applied Econometrics* **4** (1989), 213-237.

[6] Gibbs D.: Potential Output: *Concepts and Measurement*. Labour Market Bulletin **1** (1995), 72-115.

[7] Gordon R. J.: Okun's Law and Productivity Innovations. *American Economic Review: Papers and Proceedings* **100** (2010), 11-15.

[8] Knotek II, E. S.: How Useful is Okun's law? *Federal Reserve Bank of Kansas City Economic Review*, Fourth Quarter 2007, 73-103.

[9] Lang D., de Peretti C.: A Strong Hysteretic Model of Okun's Law: Theory and a Preliminary Investigation. International Review of Applied Economic **23** (2009), 445-462.

[10] Maliszewski, W. S.: *Vietnam: Bayesian Estimation of Output Gap.* IMF Working Paper WP/10/149, 2010.

[11] Mazurek, J.: Comparison of Recession Magnitudes in EU Countries During the Global Financial Crisis – A New Perspective. *Unpublished manuscript.*

[12] Mielcova, E.: Economic Growth and Unemployment Rate of the Transition Country – The Case of the Czech Republic 1996-2009. *E+M Ekonomie a Management* **1**(2011), 29-37.

[13] Okun, A. M.: Potential GDP: Its Measurement and Significance. In: *Proceedings of the Business and Economics Statistics Section*, 1962, 98-104.

[14] Prachowny M. F. J.: Okun's Law: Theoretical Foundations and Revised Estimates. *The Review of Economics and Statistics* **75** (1993), 331-336.

[15] Silvapulle P., Moosa I. A., Silvapulle M. J.: Asymmetry in Okun's Law. *The Canadian Journal of Economics* **37** (2004), 353-374.

[16] Statistical Officeof the Slovak Republic, SLOVSTAT database, <http://portal.statistics.sk>, accessed 05/20/2011.

[17] Villaverde J., Maza A.: The Robustness of Okun's Law in Spain, 1980-2004 Regional Evidence. *Journal of Policy Modeling* **1** (2009), 289-297.

[18] Wikipedia: Czech Republic, http://en.wikipedia.org/wiki/Czech_Republic, accessed 05/20/2011.

[19] Yashiv E.: The Determinants of Equilibrium Unemployment. *The American Economic Review* **90** (2000), 1297-1322.

On the Predictability of Institutional Environment

Štěpán Mikula [1]

Abstract. The aim of this paper is to develop method for evaluating the predictability of the institutional environment. It is based on standard time-series analysis and suggests two measures of predictability. This method is applied on development of property rights protection level which is approximatively described by index based on expert evaluation from International Country Risk Guide (ICRG). Obtained measures of predictability are tested using econometric model with cross-sectional data and recognized as a significant predictor of economic performance.

Keywords: Institutional environment, predictability, time series, ARI process, economic performance.

JEL classification: O43
AMS classification: 91B84

1 Motivation, hypotheses and data

There is a broad consensus in recent literature on relationship between institutional environment and economic performance. Empirical research of this topic is mostly focused on relationship between economic performance (measured by GDP per capita) and the state of institutional environment. The other features of institutional environment are left unattended. This paper fills this gap by developing measure of predictability of institutional environment. Obtained measure is used in econometric model for explaining economic performance.

The institutional environment represents broad set of institutions, which are understood according to North [11] as "rules of the game". This paper deals with property rights protection which is recognized as the key aspect of institutional environment with respect to economic performance (see e.g. [1] or [8] for more general explanation).

The impact of state of property rights protection on economic performance is well investigated on both theoretical and empirical level, but there have not been any hypotheses of how (and if) its predictability influences economic performance. It can be assumed that lower predictability leads to following:

1. Lower predictability increases uncertainty about investment risk (e.g. by increased future risk of expropriation) and therefore decreases volume of investment and steady-state of economy (see [8]).
2. Higher uncertainty about future property rights protection motivates agents to allocate more resources for securing their investments instead of using them in a productive way.

These hypotheses indicates there could be positive relationship between predictability of property rights protection and economic performance – i.e. better predictability could lead to higher performance.

1.1 Data on property rights protection

Institutions are unmeasurable and therefore their level must be approximated by means of proxy variables. The protection of property rights is often described by index ($ICRG_{PRP}$) based on experts evaluation published by Political Risk Services (PRS) in International Country Risk Guide (see e.g. [7], [4], [1] or

[1]Faculty of Economics and Administration of Masaryk University, Department of Economics, Lipová 41a, Brno, e-mail: name@university.edu

[3]). The value of $ICRG_{PRP}$ is obtained as equally weighted sum of PRS's evaluation of bureaucracy quality (BQ), corruption (Cor), investment profile (IP) and law and order (LaO):[1]

$$ICRG_{PRP} = \frac{10}{16}BQ + \frac{10}{24}Cor + \frac{10}{48}IP + \frac{10}{24}LaO \tag{1}$$

Weights in equation (1) compensate different scales of evaluation of particular components and standardize $ICRG_{PRP}$ on continuous scale $[0, 10]$, where higher values indicate better protection of property rights. The index is available in the form of yearly data for period 1984–2010.

This period covers events which have significantly changed the map of the world. The number of countries in the world has risen for example due to the collapse of Soviet Union and split of the Yugoslavia and Czechoslovakia. The $ICRG_{PRP}$ time series were altered to reflect these events. The succeeding states inherited the values of its predecessors – for example the time series of $ICRG_{PRP}$ of the Czech Republic consist of values for Czechoslovakia (till 1993) and the Czech Republic (since 1993). If there is a gap in the time series then the last uninterrupted sequence is taken into account.

2 Evaluation of predictability

For the purpose of evaluation of predictability it is necessary to make hypotheses about the way how the agents predict future development of institutional environment. I have chosen the approach supported by North's [11] assumption of path-dependent evolution of institutions. This theory implies that the current state of institutions is not random but is dependent on (or even determined by) its past states.

In accordance with this theory we can describe development of institutional environment as ARI(p,d) process and we can even assume that rational agents predict its future development on the basis of this process.

There are almost certainly more predictors of development of property rights which agents take into account (e.g. probability of conflicts, revolutions, coup d'états, state bankruptcy etc.), but these predictors may significantly differ around the world. For example, the change of executive may be a significant predictor of abrupt change in property rights protection in Africa but it is probably entirely insignificant in well established western democracies. Besides the main goal of this paper is to develop a more general method of evaluation of predictability. These extra predictors would probably differ for various aspects of institutional environment. On the other hand it is necessary to keep in mind that method based solely on ARI(p,d) processes could undervalue the real predictability of institutional environment.

Identification of ARI(p,d) processes were carried out by function `auto.arima` from `forecast` package for R developed by Hyndman and Khandakar [6]. The automatic `auto.arima` function returns the best ARI(p,d) model described by equation:

$$y_t = c + \alpha t + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t \tag{2}$$

where c is an intercept, α is a drift coefficient and p order of AR process. The best model was chosen by minimizing Akaike information criterion (AIC). The augmented Dickey-Fuller unit-root test was used for testing stationarity.

At first the model was identified over the whole time series and then the one-step in-sample predictions were made.[2] The figure 1a depicts this approach by the actual and fitted (one step ahead in-sample predictions) values of $ICRG_{PRP}$ for Argentina.

Significant advantage of this approach is that the length of time series remains unchanged. This is very important from the historical point of view – particularly in this period. Shortening of time series would for example make impossible to evaluate predictability at the time when post-communist economies begin to transform. Nevertheless the drawback of that this approach is that it assumes that the agents know the future development of institutional environment which is reflected in estimated parameters of ARI(p,d) process. This could overvalue the predictability.

[1] The components of $ICRG_{PRP}$ are described in methodology of International Country Risk Guide which is available at `http://www.prsgroup.com/ICRG_Methodology.aspx`. Construction of index is widely explained in [7].

[2] Multiple steps ahead predictions were also computed for all versions of model but it turns out that resulting RMSEs and RMSNEs are well correlated and therefore bring no added value.

(a) in-sample predictions (b) out-of-sample predictions ($V1$) (c) out-of-sample predictions ($V2$)

Figure 1: Observed (solid lines) and predicted (dashed lines) values of $ICRG_{PRP}$ for Argentina

Disadvantages of in-sample predictions can be eliminated by recursive estimation of the time series. In this approach the initial period of 15 years[3]) is left to the representative agent to learn how the environment works and the values for the next years are predicted on the basis of all past years. It means that prediction for year 2000 is made using data from period 1984–1999, prediction for 2001 is made on basis of period 1984–2000 and so on. This approach was used in two versions:

- First version (V1) assumes that representative agent is able to rapidly change the way he make predictions. He is allowed to change order and parameters of the model every year. It means that the process is identified and estimated all over again for each predicted year.
- Second version (V2) allows the agent to change just estimated parameters – i.e. the model is identified on 15 years learning period and after that the order of ARI(p,d) process remains unchanged but parameters are estimated again for every year.

Recursively estimated time series show greater response to abrupt changes in observed values when out-of-sample predictions differ from observed values much more then in-sample predictions (see Figures 1a, 1b and 1c).

Differences between predictions and observed values reflect the predictability of property rights protection. Small differences mean that agents are able to predict future states with great accuracy and vice versa. Two measures were used to quantify predictability – standardly defined root mean square error (RMSE, see (3)) and root mean square negative error (RMSNE, see (4) and (5)).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \tag{3}$$

$$RMSNE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} \tag{4}$$

$$e_t = \begin{cases} y_t - \hat{y}_t & \text{if } (y_t - \hat{y}_t) < 0 \\ 0 & \text{if } (y_t - \hat{y}_t) \geq 0 \end{cases} \tag{5}$$

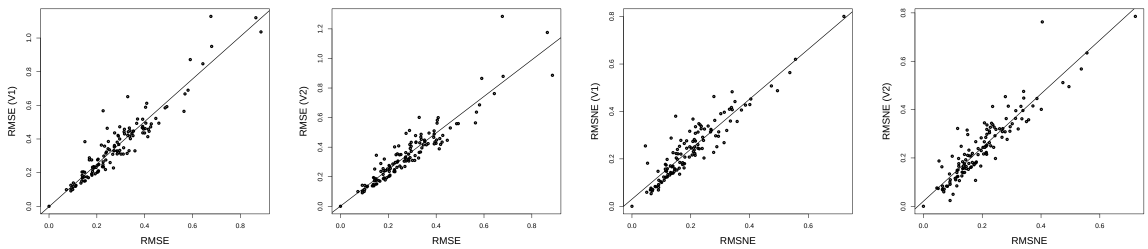
The main difference between RMSE and RMSNE is that RMSE takes into account all differences between fitted and observed values while RMSNE only negative ones. It means that RMSNE reflects only situations in which predicted values are higher then actual ones. It makes RMSNE measure of "unpleasant surprises".

[3]It implies that it is not possible to carry out recursive estimation for countries with time series shorter then 15 observations. This is the case of Armenia, Azerbaijan, Belarus, Estonia, Kazakhstan, Lithuania, Latvia, Moldova and Ukraine.

2.1 Predictability evaluated

Measures of predictability were calculated for 140 countries from in-sample predictions and for 130 countries from out-of-sample predictions.[4]

The relationship between predictability obtained by means of in-sample and out-of sample predictions is depicted in Figures 2a–2d and Table 1. These figures show that the measures are well correlated.[5] It is a highly important result which justifies the use of predictability measures based on in-sample predictions and therefore allows to include longer historical period (including already mentioned beginning of transformation) into the analysis.



(a) RMSE, V1 (b) RMSE, V2 (c) RMSNE, V1 (d) RMSNE, V2

Figure 2: Correlation of RMSE and RMSNE based on out-of-sample (V1 and V2) and in-sample predictions for corresponding period.

	RMSE	RMSNE	RMSE (V1)	RMSNE (V1)	RMSE (V2)	RMSNE (V2)
RMSE	1.000	0.922	0.940	0.849	0.930	0.880
RMSNE		1.000	0.819	0.923	0.831	0.914
RMSE (V1)			1.000	0.863	0.979	0.892
RMSNE (V1)				1.000	0.856	0.923
RMSE (V2)					1.000	0.930
RMSNE (V2)						1.000

Table 1: Correlation (Pearson's rho) of predictability measures. (Correlation between RMSE and RMSNE based on in-sample and out-of sample predictions are based on the same observations.)

Measures of predictability are negatively correlated with the average level of time series of $ICRG_{PRP}$ – countries with better property rights protection show lower values of RMSE and RMSNE and therefore higher predictability (see Table 2).

measure	in-sample	out-of-sample (V1)	out-of-sample (V2)
RMSE	-0.392^{***}	-0.242^{***}	-0.267^{***}
RMSNE	-0.393^{***}	-0.294^{***}	-0.287^{***}

Table 2: Correlation (Pearson's rho) of measures of predictability obtained from one-step predictions with the average level of $ICRG_{PRP}$ for corresponding period.

3 Predictability and economic performance

Impact of predictability is tested using slightly altered model developed by Mankiw, Romer and Weil [10], which explains economic performance by equation:

$$\log(GDP_{pw}) = const + \beta_1 \log\left(\frac{I}{GDP}\right) + \beta_2 \log(n+g+\delta) + \beta_3 \log(SCHOOL) + \varepsilon \tag{6}$$

[4] Data are available at `http://econ.muni.cz/~137451/research/datafiles/PRED.gnumeric` or `http://econ.muni.cz/~137451/research/datafiles/PRED.csv`.

[5] Figures depict "corresponding periods". It means that RMSE and RMSNE for in-sample predictions are calculated using only prediction errors from the period for which prediction errors from recursive estimations are available.

Where GDP_{pw} is GDP per worker in 2007, $\frac{I}{GDP}$ average share of investment on GDP, n is compound annual growth rate of population, g denotes technology growth, δ is the rate of depreciation ([10] assume that $g+\delta$ is equal to 0.05) and $SCHOOL$ is average atteinment of secondary school in population 15+.[6]

For the purpose of testing influence of predictability is the original model (6) extended by measures of predictability ($PRED$) :

$$\log(GDP_{pw}) = const + \beta_1 \log\left(\frac{I}{GDP}\right) + \beta_2 \log(n + \underbrace{g+\delta}_{=0.05}) + \beta_3 \log(SCHOOL) + \beta_4 \log(PRED) + \varepsilon \tag{7}$$

Model was estimated using OLS in basic (6) and extended (7) form for non-oil producing countries.[7] (The same periods were used to obtain $PRED$ and others independent variables.) Results of regressions are presented in Table 3.

$const$	$\log(I/GDP)$	$\log(n+g+\delta)$	$\log(SCHOOL)$	$\log(RMSE)$	$\log(RMSNE)$	$\bar{R}^2$
In-sample predictions ($n=90$)						
5.6622 *** (2.0648)	0.7274 *** (0.2625)	− 2.1995 *** (0.6896)	0.6960 *** (0.1560)			0.4991
5.6031 *** (2.0418)	0.7434 *** (0.2597)	− 2.0223 *** (0.6895)	0.6601 *** (0.1557)	− 0.4414 * (0.2558)		0.5104
5.3203 ** (2.0508)	0.7106 *** (0.2597)	− 2.0880 *** (0.6848)	0.6732 *** (0.1548)		− 0.3738 * (0.2160)	0.5105
Out-of-sample predictions ($n=83$)						
5.5074 ** (2.4030)	0.7180 ** (0.2808)	− 2.2293 *** (0.8045)	0.8334 *** (0.1908)			0.4610
V1:						
4.8295 ** (2.4010)	0.6768 ** (0.2780)	− 2.3048 *** (0.7948)	0.8321 *** (0.1882)	− 0.3609 * (0.2026)		0.4755
4.7705 ** (2.3589)	0.6515 ** (0.2747)	− 2.2176 *** (0.7872)	0.8330 *** (0.1856)		− 0.4468 ** (0.1911)	0.4899
V2:						
4.7714 ** (2.3818)	0.6472 ** (0.2773)	− 2.2866 *** (0.7889)	0.8397 *** (0.1870)	− 0.4201 ** (0.2036)		0.4824
4.5233 * (2.3682)	0.6177 ** (0.2758)	− 2.2866 *** (0.7814)	0.8399 *** (0.1853)		− 0.4554 ** (0.1887)	0.4920

Table 3: Estimated coefficients of model given by equations (6) and (7). (Brackets contain standard errors.)

Coefficients for measures of predictability are in all cases significant and their values seem to support hypothesis from the first section of this paper. Lower predictability represented by higher values of RMSE and RMSNE indeed decrease economic performance. Adding predictability measure also improves fit of the model which is reflected by increased values of adjusted coefficient of determination ($\bar{R}^2$). Therefore predictability seems to be a significant predictor of economic performance.

4 Conclusion

The goal of this paper was to develop method for evaluation of predictability of particular aspects of the institutional environment. Development of institutional environment is modeled using ARI(p,d) processes which reflect path-dependent nature of institutions evolution. Two measures of predictability were suggested – standardly defined root mean square error (RMSE) and root mean square negative error (RMSNE).

Developed method was applied to evaluation of the predictability of property rights protection development. Obtained measures were used for explaining economic performance and were recognized as significant predictors.

[6] GDP_{pw} is variable `rgdpl2wok` from Penn World Tables 7.0 [5]; $\frac{I}{GDP}$ is calculated from variable `ci` from Penn World Tables 7.0 [5]; Data on population are obtained from Maddison [9]. $SCHOOL$ is variable `ls` from Barro–Lee dataset [2].

[7] A country is considered as non-oil producing if its rents from oil do not exceed 5 % of GDP on average. Data on rents come from World Bank database: `http://data.worldbank.org/indicator/NY.GDP.PETR.RT.ZS`.

The results of this model do not provide no clue to decide which version of RMSE/RMSNE provides the best description of predictability. All versions are highly correlated and therefore there is just a little difference between their influence on economic performance. However results implies that estimations based on in-sample predictions assigns higher importance to RMSE and out-of-sample based estimations to RMSNE. This fact has no clear interpretation – mostly because of different historical periods which they cover.

The focus of further research of this method is testing robustness to different methods of model selection and testing of stationarity. Another promising way of development is extension of basic ARI(p,d) model by adding set of predictors which should differ according to evaluated aspect of institutional environment and regions.

Acknowledgements

I am grateful to my colleagues from Department of Economics for useful comments and support – especially to Pavel Herber and Miroslav Hloušek.

References

[1] Acemoglu, D., and Johnson, S.: Unbundling Institutions, *Journal of Political Economy* **113** (2005), 949–995.

[2] Barro, R. J. and Lee, J.–W.: A New Set of Educational Attainment in the World, 1950–2010, *NBER Working Paper No. 15902* (2010).

[3] Besley, T. and Persson, T.: The Origins of State Capacity: Property Rights, Taxation, and Development: An Empirical Investigation, *The American Economic Review* **99** (2009), 1369–1401.

[4] Hall, R. E., and Jones, C. .I: Why Do Some Countries Produce So Much Output Per Worker Than Others?, *The Quarterly Journal of Economics* **114** (1999), 83–116.

[5] Heston, A., Summers, R. and Aten, B.: *Penn World Table Version 7.0*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2011.

[6] Hyndman, R. J., and Khandakar, .Y: Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software* **23** (2008).

[7] Knack S., and Keefer, P.: Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures, *Economics & Politics* **7** (1995), 207–227.

[8] Lueck, D., Miceli, T.: Property Law. In: *Handbook of Law and Economics, vol. 1.* (A. M. Polinsky, and S. Shavell, eds.), Elsevier, Oxford, 2007, 183–257.

[9] Maddison, A.: *Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD*, `http://www.ggdc.net/MADDISON/Historical_Statistics/horizontal-file_02-2010.xls` [on-line] (2010).

[10] Mankiw N. G., Romer, D. and Weil, D. N.: A Contribution to the Empirics of Economic Growth, *The Quarterly Journal of Economics* **107** (1992), 407–437.

[11] North, D. C.: *Institutions, Institutional Change and Economic Performance*. Cambridge University Press, Cambridge, 1990.

Analyzing the Czech Output Gap Using a Small Open Economy DSGE Model

Tomáš Motl [1], Osvald Vašíček [2]

Abstract. This contribution uses a simple New Keynesian DSGE model to examine the Czech output gap from 1996 to 2010. We estimate the DSGE model in order to obtain trajectories of exogenous shocks. We test the estimated model for its data-matching properties to show that is sufficiently approximates the real economy. Based on the model, we decompose the deviation of output from trend into contributions of the exogenous shocks. This decomposition allows us to interpret the economic sources of the fluctuations in the output gap.

We show that since the year 2000, the foreign demand shock and the exchange rate shock have been increasingly important for the domestic business cycle. Overall, domestic shocks account for more than a half of the variance in the domestic output gap. We contrast the recession of 1997-1998, caused primarily by domestic shocks, to the recession of 2009, which was caused by a drop in the foreign demand. We also find that the Czech economy is sensitive to the exchange rate movements, as documented by year 2002.

Keywords: DSGE model, business cycle, small open economy, exogenous shocks.

1 Introduction

Since the beginning of the transition to market based economy, the Czech economy has experienced two recessions and several years of moderate growth around three per cent but also a few years of a rapid economic growth over six per cent. What were the causes of these events? We use a simple dynamic stochastic general equilibrium (DSGE) model to answer this question. The advantage of the DSGE model is that it provides a structural view of the economy. It decomposes data volatility into primitive drivers such as technological shocks, demand shocks or monetary policy shocks.

In this contribution, we estimate the trajectories of domestic and foreign technological shocks, demand shocks, monetary policy shocks and an exchange rate shock. We decompose the output gap fluctuations in terms of these shocks. Our analysis provides an insight into which of these shocks can explain recession, recovery or expansion in the domestic output. Based on the model, our assumptions and observed data, we shed light on the basic factors that could influence the stability and growth of the Czech economy.

2 The Model

The DSGE model used in this paper is a simple New Keynesian small open economy model presented in Justiniano and Preston (2010), which is closely related to a model in Monacelli (2005). The model consists of two economies – a small open economy and a large foreign economy. The large economy has a strong influence over the small one, whereas the small economy does not affect the large economy. Each of the two economies consists of households, firms and a monetary authority. The domestic firms are divided into producers and importers. Both types of firms operate on monopolistically competitive markets.

The following subsections present only a basic scatch of the model in order to familiarize the reader with the exogenous shocks as these shocks constitute basis of further analysis. Foreign variables are

[1]Masaryk University, Department of Economics, Lipová 507/41a, 602 00 Brno, e-mail: 175869@mail.muni.cz
[2]Masaryk University, Department of Economics, Lipová 507/41a, 602 00 Brno, e-mail: Osvald.Vasicek@econ.muni.cz

denoted by asterisk superscript.

2.1 Households

The domestic economy consists of a continuum of identical infinitely-lived households. The representative agent maximizes his lifetime welfare represented by a utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t \tilde{\varepsilon}_{G,t} \left[\frac{(C_t - hC_{t-1})^{1-\sigma}}{1-\sigma} - \frac{N_t^{1+\phi}}{1+\phi} \right] \quad s.t. \quad C_t P_t + E_t Q_{t,t+1} D_{t+1} = D_t + W_t N_t + T_t. \tag{1}$$

Here, C_t denotes household consumption, N_t denotes hours worked, σ is the inverse intertemporal elasticity of substitution between the present and future consumption and ϕ is the inverse Frisch elasticity of labour supply. D_t is a nominal income from one period bonds bought in time $t-1$ maturing in time t, and $Q_{t,t+1}$ is the price of the bond. The financial markets are complete here. W_t is the nominal wage, P_t is a price of the consumption basket and T_t are the lump-sum transfers. The parameter h denotes exogenous habit in consumption and $\tilde{\varepsilon}_{G,t}$ is a preference shock.

2.2 Producers

We assume domestic sector to be populated by a $(0,1)$ continuum of monopolistically competitive producers. Each producer makes a decision about how much to produce and how much to charge. Producers share the same production technology

$$Y_t(i) = A_t N_t(i),$$

where $N_t(i)$ denotes the amount of labour hired by the i-th producer and A_t is an economy specific technological shock. Firms are assumed to set prices a la Calvo such that in every period a fraction $(1-\theta_H)$ of domestic producers is able to set optimal price. The producers maximize profit given by

$$E_t \sum_{T=t}^{\infty} \theta_H^{T-t} Q_{t,T} y_{H,T}(i) \left[P_{H,t}(i) - P_{H,T} MC_T \right], \tag{2}$$

where $y_{H,T}(i)$ is the demand for the production of the i-th producer, $P_{H,t}(i)$ is the optimal price set by the i-th producer in period t and MC_T are marginal costs given as a derivative of the total costs.

2.3 Real exchange rate, law of one price relation

The real exchange rate is defined in log terms by

$$q_t = e_t + p_t^* - p_t = e_t + p_t^* - p_{F,t} + (1-\alpha)s_t = \psi_t + (1-\alpha)s_t$$

where q_t is the log of the real exchange rate, $e_t = \log \hat{e}_t$ is the log of the nominal exchange rate, p_t^* is the log of the foreign CPI, $s_t = p_{F,t} - p_{H,t}$ are log terms of trade and ψ_t is the log of the law of one price gap $\psi_t = e_t + p_t^* - p_{F,t}$. The uncovered interest parity condition requires that

$$i_t - i_t^* + E_t \pi_{t+1}^* - E_t \pi_{t+1} = E_t(q_{t+1}) - q_t + \varepsilon_{Q,t}$$

where $\varepsilon_{Q,t}$ is an exchange rate shock and $\pi_t = p_t - p_{t-1}$ is inflation.

2.4 Importers

In the domestic economy there is a $(0,1)$ continuum of monopolistically competitive importers. Each importer imports a unique good denoted by the index j bought for the price $\hat{e}_t P_{F,t}^*(j)$ which holds "at the docks". The importers are also assumed to set prices a la Calvo such that in every period a fraction $(1-\theta_F)$ of importers is able to set optimal price. The problem of the importers is to maximize

$$E_t \sum_{T=t}^{\infty} Q_{t,T} \theta_F^{T-t} C_{F,T}(i) \left[P_{F,t}(i) - \hat{e}_t P_{F,t}^*(i) \right]$$

given the demand for their goods $C_{F,T}(i)$.

2.5 Monetary policy

Monetary policy in the foreign economy described by a Taylor rule with a lagged interest rate and a focus on the output and the inflation:

$$i_t^* \quad = \quad \rho^* i_{t-1}^* + (1-\rho^*)(\psi_\pi^* \pi_t^* + \psi_y^* y_t^*) + \epsilon_{M,t}^*$$

Here, ρ^* is parameter of interest rates smoothing, ψ_π^* (ψ_Y^*) measures the magnitude of the reaction of the monetary authority to the inflation (output), and $\epsilon_{M,t}^*$ is an exogenous monetary policy shock. We assume that the monetary authority in the domestic economy does not focus on output.

2.6 Exogenous shocks

The model features seven exogenous stochastic shocks. We describe only the domestic shocks and the exchange rate shock. The foreign shocks are identical to their domestic counterparts.

- The shock in preferences ε_G increases the present household utility, which makes the households more impatient and causes them to shift their consumption from future to present. This shift increases present demand and output. This shock can be interpreted as a demand shock.
- The technological shock ε_A enters the firms' production function, increases the labor productivity and lowers the marginal costs of producers which in turn affect the inflation. This shock can be interpreted as a supply shock.
- The monetary policy shock ε_M enters the Taylor rule and relates to the nominal interest rate dynamics. If the monetary policy shock becomes positive, the nominal interest rate is above the level given by the Taylor rule.
- The exchange rate shock ε_Q affects the uncovered interest parity.

3 Estimation and data fit

We estimate the model using Bayesian techniques in order to acquire the estimated trajectories of structural shocks. In this section, we describe the data and discuss model fit. For the sake of space we do not discuss estimates of the model parameters, but we directly assess the data matching properties of the model.

3.1 The Data

The model is estimated on Czech and EA12 economy data. The sample ranges from 1996Q1 to 2010Q2. We measure quarterly domestic and foreign real GDP per capita, domestic and foreign consumer price inflation, domestic and foreign nominal interest rate and real exchange rate. The observed data are detrended using the Hodrick-Prescott filter and all variables are expressed in percentage deviation from the HP trend. To avoid the beginning-of-sample and end-of sample filtration problem, we estimate the HP trend using observations from 1994Q1 and append data till 2012Q2 using the Czech National Bank forecast as of 2010Q4. Except for nominal interest rates and real exchange rate, all time series were seasonally adjusted using X-12 ARIMA.

We equal the output gap to the detrended output. An alternative and perhaps better method would be to use dynamic steady state. Such method would not be subject to potential problems arising from using Hodrick-Prescott filter. However, due to practical difficulties with modelling dynamic steady state, it is common to use the Hodrick-Prescott filter. We check the estimated values of output gap with the ones reported by Czech National Bank and find a good correspondence. Specifically, the CNB reports tightening of the output gap in 2000-2001 and subsequent widening. The CNB also reports that the output gap became positive in the second half of 2005 and turned negative again in 2009Q1[1].

[1]CNB Inflation report III/2010

		Y	Y^*	π	π^*	I	I^*	Q
Data	1st	0.8805	0.8278	0.5359	0.8236	0.8366	0.8708	0.6931
	2nd	0.7019	0.6378	0.4858	0.6583	0.6876	0.6517	0.3097
Model	1st	0.8540	0.8518	0.4382	0.4982	0.7821	0.8200	0.6405
	2nd	0.6825	0.6508	0.1739	0.2675	0.5871	0.6362	0.4000

Table 1: Implied autocorrelation of 1st and 2nd order. Y – domestic output gap. Y^* – foreign ouput gap. π – domestic inflation. π^* foreign inflation. I – domestic interest rate. I^* – foreign interest rate. Q – real exchange rate.

	lag/lead of output gap	-2	-1	0	1	2
Foreign output gap	Data	0.3546	0.5326	0.6713	0.6227	0.5307
	Model	0.3732	0.4610	0.5245	0.4320	0.3216
Domestic inflation	Data	0.4235	0.4080	0.5374	0.4716	0.4243
	Model	0.0932	0.2166	0.4719	0.3825	0.2942
Foreign inflation	Data	-0.2597	-0.1306	0.0654	0.1500	0.2355
	Model	-0.2848	-0.3156	-0.3333	-0.2954	-0.2567
Domestic interest rate	Data	0.5474	0.4558	0.3153	0.1479	0.0092
	Model	0.3440	0.3987	0.4359	0.3481	0.2685
Foreign interest rate	Data	0.5746	0.6614	0.6727	0.5247	0.3465
	Model	-0.0170	-0.0160	-0.0340	-0.0797	-0.1137
Real exchange rate	Data	-0.2610	-0.2708	-0.2097	-0.0476	0.0922
	Model	-0.1513	-0.1150	-0.0297	-0.0621	-0.0731

Table 2: Implied cross-correlation of Czech output gap with other time series.

3.2 Data fit

To add to the credibility of the model, we assess the data matching properties of the model. In our strategy, we compare the performance of statistical models on both real data and artificial data generated by the estimated model. We estimate a VAR(1) model on original data and on artificial data and take fitted values.

For each time series, we compute model implied autocorrelation and cross-correlation with respect to the GDP. The results are summarized in tables 1 and 2. We can see that the DSGE model approximates well the autocorrelations in output, interest rates and real exchange rate, but fails to replicate the high autocorrelation of inflations. The cross-correlations between domestic output and foreign output, domestic interest rate and real exchange rate are similar for both original and artificial data. We find it particularly encouraging that the model matches the data in the international dimension well.

4 Shock decomposition

The outcome of our analysis is depicted in figure 1. The black line shows the values of the output gap. The columns denote the contributions of structural shocks to this deviation. As the contributions of the foreign monetary and foreign technological shock are negligible, foreign shocks are summed together for clarity. Our analysis presents one of the possible interpretations of the history, which is conditioned on the model use and the simplifying assumptions we make.

The early periods are strongly influenced by the initial values, which can be interpreted as the overall effect of unidentified past shocks. This fact makes the year 1996 harder to interpret. The supply shock had a decreasing tendency already since the last quarter of 1996 and was the main reason why output fell in 1997 and 1998. Likely explanation is that the supply side problems helped to trigger the monetary crisis, which then in turn unfolded the supply side problems in the Czech economy.

Which factors brought the recovery? The recovery in 1999 was the result of the vanishing negative effect of domestic technological shock and increasing positive influence of the foreign demand shock. The recovery was countered by the domestic demand shock. It is important to notice that foreign shocks play no role in the causes of the recession and only a limited role in the recovery – the 1997-1998 recession

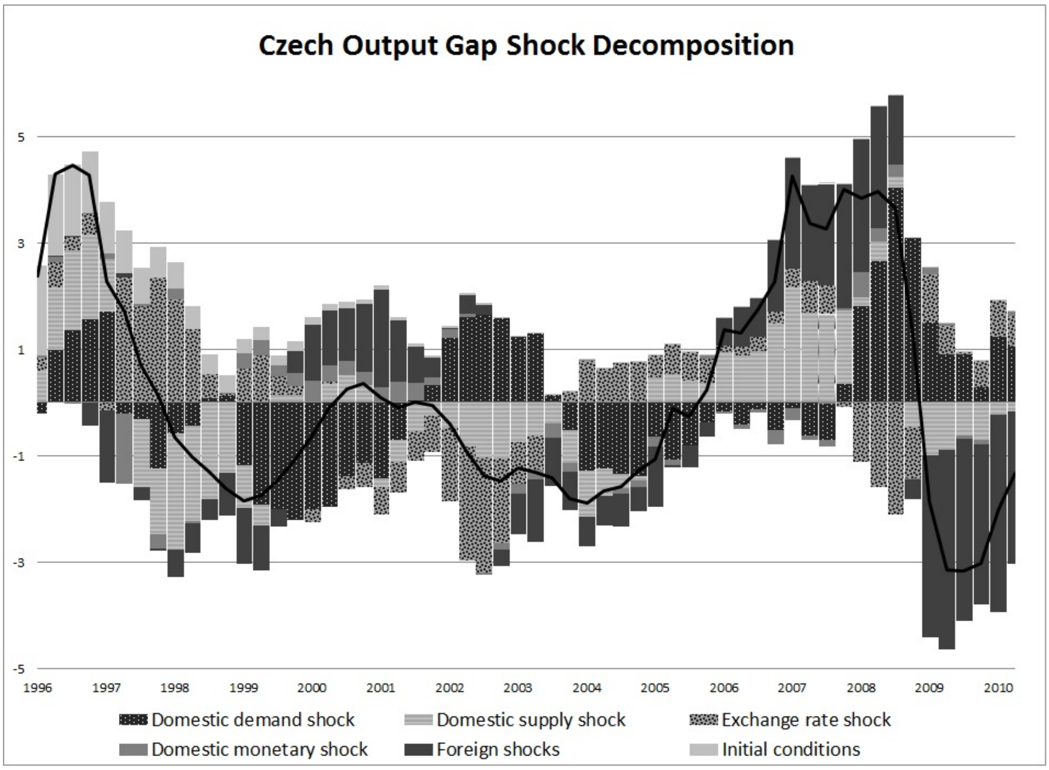


Figure 1: Shock decomposition of the Czech output gap.

and the 1999 recovery had primarily internal sources.

The period of 2000-2004 saw a rise in the importance of the shocks outside of the Czech economy, namely the foreign demand and exchange rate shock. The exchange rate shock pushed the output gap into negative values in 2001 and in 2002 it was the main driver of the output slowdown in 2002. The exchange rate appreciated from 36 CZK/EUR to 34 CZK/EUR in 18 months from the second quarter of 2000 on. In the beginning of 2002, the appreciation was even faster. The years 2001 and 2002 document the sensitivity of the Czech economy to the exchange rate movements.

The years 2005-2010 saw another period of economic growth and recession. Contrary to the previous period, however, the dynamics of the domestic output was given mainly by foreign shocks. This is in line with conventional wisdom, but we also find that domestic shocks still have had a significant role in the domestic business cycle.

The 2005-2007 period was very successful in terms of growth of the output, when the Czech economy grew by over 6 per cent annually. Most of the growth was driven by the growth in the potential ouput, as is obvious from the fact that the output gap turned positive only in late 2005. The causes of this development can be found both inside and outside of the Czech economy. Figure 1 shows that in 2005, closing of the output gap was likely caused by the demand factors, namely by the decreasing negative contributions of both domestic and foreign demand. Subsequently, since 2006 the foreign demand shock positively contributed to the output gap until the third quarter of 2008. We can see that the foreign demand was especially influential in the 2007-2008 period.

The year 2008 brought a slowdown in output growth to 2.8 per cent annually. We identify two main reasons. First, the exchange rate appreciation by some 15 per cent in one year depressed exports and lowered the output. as obvious from the exchange rate shock. Second, the foreign demand shock turned from a positive contribution in third quarter of 2008 to a significantly negative one in the first quarter of 2009 as the world-wide recession hit the foreign economy before it transpired into the domestic economy. The Czech crown in turn depreciated in 2008Q4 and the exchange rate shock decreased its negative contribution. Thus the main reason for positive, although decreasing output gap in 2008 was the domestic demand shock.

The foreign demand shock transferred the recession to the Czech economy as can be seen from its large negative contribution through the whole 2009 and early 2010. Domestic demand and technological shocks also partly contributed to the recession, but we identify the foreign demand shock as the main cause of the 2009 recession. This is in sharp contrast with the recession of 1997-199,8 which was caused by domestic shocks.

We construct variance decomposition of the domestic output gap to get an overview of the relative importance of particular shocks. The domestic demand shock is most influential with 36 per cent of the variance. The foreign demand shock and the exchange rate shock together account for 40 per cent of the variance. The foreign technological and monetary policy shocks have negligible influence on the Czech business cycle.

5 Conclusion

Based on the analysis in this paper, we provide several conclusions:

- The recession of 1997-1998 was caused entirely by domestic shocks. The domestic technological shock was the key factor that stood behind the negative GDP growth in 1997-1998. On the contrary, the recession of 2009 was almost entirely caused by a drop in foreign demand. Given the fact that foreign shocks were crucial for the Czech economy in the 2007-2010 period, it is likely that its recovery is dependent on the recovery of the foreign demand.
- Since 2000, the foreign shocks have had increasing importance in determining the domestic output. Most notably, the exchange rate shock was particularly important in 1997 and 2002. Starting in 2007, the foreign demand was the most significant outside factor that determined the domestic output, especially in the late 2008 and early 2009. In general, the Czech economy is sensitive to the foreign shocks.
- Although the Czech economy is sensitive to the foreign shocks, the domestic shocks have played significant role in determining the Czech business cycle. Domestic shocks explain approximately 60 per cent of the variance in the domestic output.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

References

[1] Iskrev, N.: Local identification in DSGE models, *Journal of Monetary Economics* **57(2)** (2010), 189–202.

[2] Justiniano, A. and Preston, B.: Monetary Policy and Uncertainty in an Empirical Small Open Economy Model, *Journal of Applied Econometrics* **25(1)** (2010), 93–128.

[3] Monacelli, T.: Monetary Policy in a Low Pass-through Environment, *Journal of Money, Credit and Banking* **37** (2005), 1047–1066

The algorithm for testing solvability of max-plus interval systems

Helena Myšková[1]

Abstract.

In this paper, we shall deal with the solvability of interval systems of linear equations in the max-plus algebra. The max-plus algebra is the algebraic structure in which classical addition and multiplication are replaced by $\oplus$ and $\otimes$, where $a \oplus b = \max\{a, b\},\ a \otimes b = a + b$.

The notation $\boldsymbol{A} \otimes \boldsymbol{x} = \boldsymbol{b}$ represents an interval system of linear equations, where $\boldsymbol{A} = \langle \underline{A}, \overline{A} \rangle$, $\boldsymbol{b} = \langle \underline{b}, \overline{b} \rangle$ are given interval matrix and interval vector, respectively, and a solution is from a given interval vector $\boldsymbol{x} = \langle \underline{x}, \overline{x} \rangle$. We can define several types of solvability of interval systems. In this paper, we define a T3, T5, T8 and T9 solvability and give necessary and sufficient conditions for them.

Keywords: max-plus algebra, interval system, solvability concepts

JEL classification: C02
AMS classification: 15A06; 65G30

1 Preliminaries

By *max-plus algebra* we understand a triple $(B, \oplus, \otimes)$, where

$$B = \mathbb{R} \cup \{\varepsilon\},\ a \oplus b = \max\{a, b\},\ a \otimes b = a + b,$$

where $\varepsilon = -\infty$.

Let m, n be given positive integers. Denote by M and N the sets of indices $\{1, 2, \ldots, m\}$, $\{1, 2, \ldots, n\}$, respectively. The set of all $m \times n$ *matrices* over B is denoted by $B(m, n)$ and the set of all column n-*vectors* over B by $B(n)$. For a given matrix $A \in B(m, n)$ and vector $x \in B(n)$ we have $[A \otimes x]_i = \max_{j \in N}\{a_{ij} + x_j\}$. We shall consider the *ordering* $\leq$ on the sets $B(m, n)$ and $B(n)$ defined as follows:

- for $A, C \in B(m, n):\ A \leq C$ if $a_{ij} \leq c_{ij}$ for all $i \in M,\ j \in N$,
- for $x, y \in B(n):\ x \leq y$ if $x_j \leq y_j$ for all $j \in N$.

It is easy to see that for each $A, C \in B(m, n)$ and for each $x, y \in B(n)$ the implication

$$\text{if } A \leq C \text{ and } x \leq y \text{ then } A \otimes x \leq C \otimes y$$

holds. We call this property the monotonicity of the operation $\otimes$.

For the given matrix interval $\boldsymbol{A} = \langle \underline{A}, \overline{A} \rangle$ with $\underline{A}, \overline{A} \in B(m, n),\ \underline{A} \leq \overline{A}$ and the given vector interval $\boldsymbol{b} = \langle \underline{b}, \overline{b} \rangle$ with $\underline{b}, \overline{b} \in B(m),\ \underline{b} \leq \overline{b}$ the notation

$$\boldsymbol{A} \otimes x = \boldsymbol{b} \tag{1}$$

represents the set of all systems of linear max-plus equations of the form

$$A \otimes x = b, \tag{2}$$

[1]Technical University in Košice/Faculty of Electrical Engineering and Informatics, Department of Mathematics, B. Nemcovej 32, 04001 Košice, Slovak Republic, e-mail: Helena.Myskova@tuke.sk

where $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$.

The set $\boldsymbol{A} \otimes x = \boldsymbol{b}$ will be called an *interval system* of max-plus linear equations. Each system of the form (2) is said to be a *subsystem* of system (1), if $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$.

Suppose that solutions of subsystems of (1) can be not arbitrary, but they are required to be from the given interval vector $\boldsymbol{x} = \langle \underline{x}, \overline{x} \rangle,\ \underline{x},\ \overline{x} \in B(n),\ \underline{x} \le \overline{x}$.

Denote by

$$\boldsymbol{A} \otimes \boldsymbol{x} = \boldsymbol{b} \tag{3}$$

an interval system with solutions $x \in \boldsymbol{x}$. We shall suppose for interval system (3) that all entries of vectors $x \in \boldsymbol{x}$ and $b \in \boldsymbol{b}$ are finite and all matrices $A \in \boldsymbol{A}$ are such that there exists at least one finite element in each row.

2 Solvability concepts

We shall consider over the solvability of interval system on the ground of the solvability of its subsystems. The solvability concepts of (1) have been studied by K. Cechlárová and R. A. Cuninghame-Green [1, 2] and by H. Myšková [4, 5, 6]. We dealt with several solvability concepts of (3) in [7]. Table 1 contains the list of all up to now defined types of solvability of interval system (3).

Solvability concept	**Condition**
T1 solvability	$(\exists A \in \boldsymbol{A})(\forall x \in \boldsymbol{x})(\exists b \in \boldsymbol{b}) : A \otimes x = b$
T2 solvability	$(\forall x \in \boldsymbol{x})(\exists A \in \boldsymbol{A})(\exists b \in \boldsymbol{b}) : A \otimes x = b$
T3 solvability	$(\forall x \in \boldsymbol{x}(\exists b \in \boldsymbol{b}))(\forall A \in \boldsymbol{A}) : A \otimes x = b$
T4 solvability	$(\exists b \in \boldsymbol{b})(\exists x \in \boldsymbol{x})(\forall A \in \boldsymbol{A}) : A \otimes x = b$
T5 solvability	$(\forall x \in \boldsymbol{x})(\forall A \in \boldsymbol{A})(\exists b \in \boldsymbol{b}) : A \otimes x = b$
weak T6 solvability	$(\exists b \in \boldsymbol{b})(\forall x \in \boldsymbol{x})(\exists A \in \boldsymbol{A}) : A \otimes x = b$
strong T6 solvability	$(\forall b \in \boldsymbol{b})(\forall x \in \boldsymbol{x})(\exists A \in \boldsymbol{A}) : A \otimes x = b$
weak T7 solvability	$(\exists b \in \boldsymbol{b})(\exists A \in \boldsymbol{A})(\forall x \in \boldsymbol{x}) : A \otimes x = b$
strong T7 solvability	$(\forall b \in \boldsymbol{b})(\exists A \in \boldsymbol{A})(\forall x \in \boldsymbol{x}) : A \otimes x = b$
T8 solvability	$(\forall A \in \boldsymbol{A})(\exists b \in \boldsymbol{b})(\forall x \in \boldsymbol{x}) : A \otimes x = b$
T9 solvability	$(\exists b \in \boldsymbol{b})(\forall A \in \boldsymbol{A})(\forall x \in \boldsymbol{x}) : A \otimes x = b$

Table 1. Solvability concepts

In the following we shall deal with these solvability concepts except of T4 solvability which has been studied in detail in [6] and requires another assumptions than the others.

3 Known results

Theorem 1. [8] *Interval system (3) is T5 solvable if and only if*

$$\underline{A} \otimes \underline{x} \ge \underline{b}, \tag{4}$$

$$\overline{A} \otimes \overline{x} \le \overline{b}. \tag{5}$$

Theorem 2. [7] *Interval system (3) is T2 solvable if and only if*

$$\overline{A} \otimes \underline{x} \ge \underline{b}, \tag{6}$$

$$\underline{A} \otimes \overline{x} \le \overline{b}. \tag{7}$$

To give a necessary and sufficient condition for T1 solvability of (3) we define a matrix $A^* = (a^*_{ij})$ as follows:

$$a^*_{ij} = \min\{\overline{b}_i - \overline{x}_j, \overline{a}_{ij}\} \tag{8}$$

for each $i \in M,\ j \in N$.

Theorem 3. [7] *Interval system (3) is T1 solvable if and only if interval system (3) is T2 solvable and*

$$A^* \otimes \underline{x} \geq \underline{b}. \tag{9}$$

Theorem 4. [7] *Interval system (3) is weakly T6 solvable if and only if interval system (3) is T2 solvable and*

$$\underline{A} \otimes \overline{x} \leq \overline{A} \otimes \underline{x}. \tag{10}$$

Theorem 5. [7] *Interval system (3) is strongly T6 solvable if and only if*

$$\underline{A} \otimes \overline{x} \leq \underline{b}, \tag{11}$$

$$\overline{A} \otimes \underline{x} \geq \overline{b}. \tag{12}$$

Theorem 6. [7] *Interval system (3) is weakly T7 solvable if and only (3) is weakly T6 solvable and for each $i \in M$ there exists $j \in N$ such that $\overline{a}_{ij} \otimes \underline{x}_j \geq b^*_i$ and $\underline{x}_j = \overline{x}_j$, where $b^*_i = \max\{[\underline{A} \otimes \overline{x}]_i, \underline{b}_i\}$.*

Theorem 7. [7] *Interval system (3) is strongly T7 solvable if and only if interval system (3) is strongly T6 solvable and for each $i \in M$ there exists $j \in N$ such that $\overline{a}_{ij} \otimes \underline{x}_j \geq \overline{b}_i$ and $\underline{x}_j = \overline{x}_j$.*

4 T3, T8 and T9 solvability

Definition 1. We say, that interval system (3) is

i) *T3 solvable* if for each vector $x \in \boldsymbol{x}$ there exists $b \in \boldsymbol{b}$ such that for each $A \in \boldsymbol{A}$ the equality $A \otimes x = b$ holds,
ii) *T8 solvable* if for each $A \in \boldsymbol{A}$ there exists $b \in \boldsymbol{b}$ such that for each $x \in \boldsymbol{x}$ the equality $A \otimes x = b$ holds,
iii) *T9 solvable* if there exists $b \in \boldsymbol{b}$ such that for each $x \in \boldsymbol{x}$ and for each $A \in \boldsymbol{A}$ the equality $A \otimes x = b$ holds.

For the given indices $i \in M,\ j \in N$ denote the vector

$$x^{(j)} = (\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{j-1}, \overline{x}_j, \underline{x}_{j+1}, \ldots, \underline{x}_n)$$

and the matrix $A^{(ij)} = (a^{(ij)}_{kl})$ where

$$(a^{(ij)}_{kl}) = \begin{cases} \overline{a}_{kl} & \text{for } k = i,\ l = j, \\ \underline{a}_{kl} & \text{otherwise.} \end{cases}$$

Lemma 1. Let $x \in \boldsymbol{x},\ A \in \boldsymbol{A}$. Then there exist $\alpha_i,\ \beta_{ij} \in B$ such that

i) $x = \bigoplus_{j \in N} \alpha_j \otimes x^{(j)}$,
ii) $A = \bigoplus_{j \in N} \beta_{ij} \otimes A^{(ij)}$.

Proof.

i) We prove that $\alpha_j = x_j - \overline{x}_j$. For each $k \in N$ we have

$$[\bigoplus_{j \in N} \alpha_j \otimes x^{(j)}]_k = \bigoplus_{j \in N} \alpha_j \otimes x^{(j)}_k = (\bigoplus_{j \neq k} \alpha_j \otimes x^{(j)}_k) \oplus (\alpha_k \otimes x^{(k)}_k) = (\bigoplus_{j \neq k} \alpha_j \otimes \underline{x}_k) \oplus (\alpha_k \otimes \overline{x}_k) =$$

$$(\bigoplus_{j \neq k} (x_j - \overline{x}_j + \underline{x}_k)) \oplus (x_k - \overline{x}_k + \overline{x}_k) = x_k$$

because of $x_j - \overline{x}_j + \underline{x}_k \leq \underline{x}_k \leq x_k$. So $x = \bigoplus_{j \in N} \alpha_j \otimes x^{((j)}$.

ii) In the same way we can prove that $\beta_{ij} = a_{ij} - \overline{a}_{ij}$.

$\square$

From the definitions of the T3, T5 and T8 solvability it is easy to see, that the T3 solvability implies the T5 solvability and the T8 solvability implies the T5 solvability . This means that T5 solvability is a necessary condition for the T3 solvability and T8 solvability, too.

Lemma 2. Interval system (3) is T3 solvable if and only if interval system (3) is T5 solvable and

$$\underline{A} \otimes x = \overline{A} \otimes x \tag{13}$$

for each $x \in \boldsymbol{x}$.

Proof. The T5 solvability implies that $\underline{A} \otimes x \geq \underline{A} \otimes \underline{x} \geq \underline{b}$ and $\overline{A} \otimes x \leq \overline{A} \otimes \overline{x} \leq \overline{b}$ which means that $\underline{A} \otimes x \in \boldsymbol{b}$ and $\overline{A} \otimes x \in \boldsymbol{b}$ for each $x \in \boldsymbol{x}$. Let $x \in \boldsymbol{x}$ be fixed. The existence of $b \in \boldsymbol{b}$ such that for each $A \in \boldsymbol{A}$ the equality $A \otimes x = b$ is satisfied means that the products $A \otimes x$ are the same for each $A \in \boldsymbol{A}$. This is equivalent to (13). $\square$

Lemma (2) does not provide an effective algorithm for testing the T3 solvability.

Theorem 8. *Interval system (3) is T3 solvable if and only if interval system (3) is T5 solvable and*

$$\underline{A} \otimes x^{(j)} = \overline{A} \otimes x^{(j)} \tag{14}$$

for each $j \in N$.

Proof. Let us suppose that $\underline{A} \otimes x^{(j)} = \overline{A} \otimes x^{(j)}$ for each $j \in N$ and interval system (3) is T5 solvable. Then

$$\underline{A} \otimes x = \underline{A} \otimes \bigoplus_{j \in N} \alpha_j \otimes x^{(j)} = \bigoplus_{j \in N} \alpha_j \otimes (\underline{A} \otimes x^{(j)}) = \bigoplus_{j \in N} \alpha_j \otimes (\overline{A} \otimes x^{(j)}) =$$
$$\overline{A} \otimes \bigoplus_{j \in N} \alpha_j \otimes x^{(j)} = \overline{A} \otimes x$$

for each $x \in \boldsymbol{x}$. By Lemma 2 interval system (3) is T3 solvable.
The converse implication is trivial. $\square$

Lemma 3. Interval system (3) is T8 solvable if and only if interval system (3) is T5 solvable and

$$A \otimes \underline{x} = A \otimes \overline{x} \tag{15}$$

for each $A \in \boldsymbol{A}$.

Proof. The T5 solvability and monotonicity of $\otimes$ imply that $A \otimes \underline{x} \in \boldsymbol{b}$ and $A \otimes \overline{x} \in \boldsymbol{b}$ for each $A \in \boldsymbol{A}$. Let $A \in \boldsymbol{A}$ be fixed. Then existence of $b \in \boldsymbol{b}$ such that for each $x \in \boldsymbol{x}$ the equality $A \otimes x = b$ is satisfied means that the products $A \otimes x$ are the same for each $x \in \boldsymbol{x}$. This is equivalent to (15). $\square$

Lemma 3 does not provide an effective algorithm for testing the T8 solvability.

Theorem 9. *Interval system (3) is T8 solvable if and only if interval system (3) is T5 solvable and*

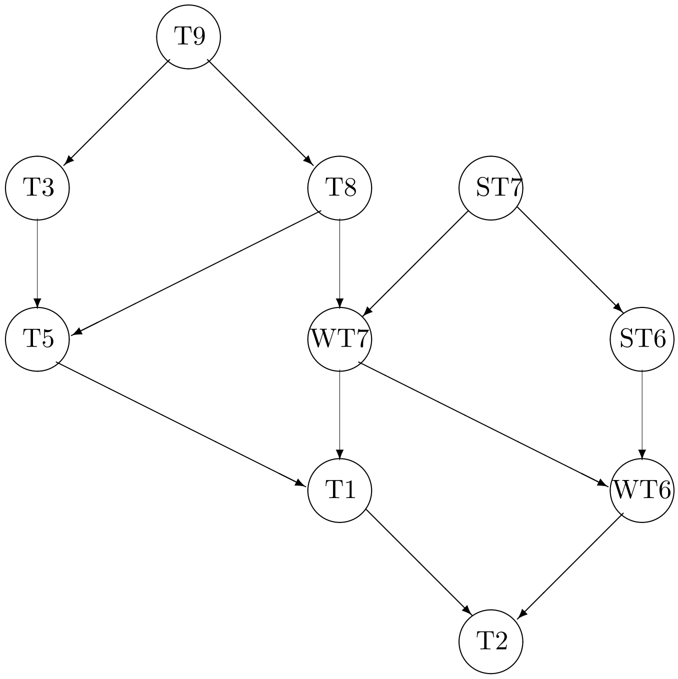
$$A^{(ij)} \otimes \underline{x} = A^{(ij)} \otimes \overline{x} \tag{16}$$

for each $i \in M,\ j \in N$.

Proof. Suppose that equality (16) is satisfied for each $i \in M,\ j \in N$ and interval system (3) is T5 solvable. By Lemma 1ii) we have

$$A \otimes \underline{x} = (\bigoplus_{i,j} \beta_{ij} \otimes A^{(ij)}) \otimes \underline{x} = \bigoplus_{i,j} (\beta_{ij} \otimes (A^{(ij)} \otimes \underline{x})) = \bigoplus_{i,j} (\beta_{ij} \otimes (A^{(ij)} \otimes \overline{x})) =$$
$$= (\bigoplus_{i,j} \beta_{ij} \otimes A^{(ij)}) \otimes \overline{x} = A \otimes \overline{x}$$

and by Lemma 3 interval system (3) is T8 solvable. $\square$



Used abbreviations:

T1 – T1 solvability
T2 –T2 solvability
WT6 – weak T6 solvability
WT7 – weak T7 solvability
T3 – T3 solvability
T5 – T5 solvability
ST6 – strong T6 solvability
ST7 – strong T7 solvability
T8 – T8 solvability
T9 – T9 solvability

Figure 1: Hasse diagram of the relations between different types of solvability of interval system

Theorem 10. *Interval system (3) is T9 solvable if and only if interval system (3) is T5 solvable and*

$$\underline{A} \otimes \underline{x} = \overline{A} \otimes \overline{x}.$$

Proof. If interval system (3) is T5 solvable and $\underline{A} \otimes \underline{x} = \overline{A} \otimes \overline{x} = b$ then $b = \underline{A} \otimes \underline{x} \leq A \otimes x \leq \overline{A} \otimes \overline{x} = b$. We get $A \otimes x = b \in \boldsymbol{b}$ for each $A \in \boldsymbol{A},\ x \in \boldsymbol{x}$, so there exists the vector b such that for each $A \in \boldsymbol{A},\ b \in \boldsymbol{b}$ the equality $A \otimes x = b$ is held.

The converse implication is trivial. $\square$

From the definitions of the solvability concepts it follow many implications, for example $T9 \Rightarrow T3 \Rightarrow T5 \Rightarrow T1 \Rightarrow T2,\ ST7 \Rightarrow ST6 \Rightarrow WT6 \Rightarrow T2$ and many other. Let us define on the set S of all solvability concepts a relation R such that S_i RS_j if and only if S_j implies S_i, where S_i and S_j are solvability concepts. The set S with the relation R is a partially ordered set, since the relation R is reflexive, antisymmetric and transitive. To describe the set of all solvability concepts by Hasse diagram we have to prove many non-implications.

Example 1. ($T1 \not\Rightarrow WT6$) Let us have

$$\boldsymbol{A} = \begin{pmatrix} \langle 3,5\rangle & \langle 2,6\rangle \\ \langle 5,7\rangle & \langle 7,9\rangle \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \langle 2,3\rangle \\ \langle 4,7\rangle \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} \langle 5,13\rangle \\ \langle 9,16\rangle \end{pmatrix}.$$

For checking the T1 solvability we compute the matrix A^* given by (8). We get $A^* = \begin{pmatrix} 5 & 6 \\ 7 & 9 \end{pmatrix}$. Since $A^* \otimes \underline{x} \geq \underline{b}$, the given interval system **is T1 solvable**.

We check the weak T6 solvability. Since $\underline{A} \otimes \overline{x} = (9,14)^T \leq \overline{b}$ and $\overline{A} \otimes \underline{x} = (10,13)^T \geq \underline{b}$, the given interval system is T2 solvable. As $\underline{A} \otimes x \not\leq \overline{A} \otimes \underline{x}$, this is **not weakly T6 solvable**.

Consequently, $T1 \not\Rightarrow ST6$ (if $T1 \Rightarrow ST6$ and $ST6 \Rightarrow WT6$, then $T1 \Rightarrow WT6$, a contradiction).

In this way, we can prove other non-implications. As a consequence we get Hasse diagram described in Figure 1.

Example 2. Check all solvability concepts for the given interval system $\boldsymbol{A} \otimes x = b$ with

$$\boldsymbol{A} = \begin{pmatrix} \langle 1,9\rangle & \langle 3,6\rangle & \langle 2,5\rangle \\ \langle 4,7\rangle & \langle 6,9\rangle & \langle 2,6\rangle \\ \langle 2,5\rangle & \langle 3,8\rangle & \langle 1,5\rangle \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} \langle 3,6\rangle \\ \langle 5,5\rangle \\ \langle 4,7\rangle \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} \langle 9,12\rangle \\ \langle 12,14\rangle \\ \langle 10,12\rangle \end{pmatrix}.$$

Strong T7 solvability:
For $i = 1$ there does not exist $j \in N$ such that $\overline{a}_{1j} \otimes \underline{x}_j \geq \overline{b}_1$ and $\underline{x}_j = \overline{x}_j$.
The given interval system *is not strongly T7 solvable.*

Strong T6 solvability: $\underline{A} \otimes \overline{x} \leq \underline{b}$ and $\overline{A} \otimes \underline{x} \geq \overline{b}$.
The given interval system *is strongly T6 solvable.* Then it *is weakly T6 solvable* and *T2 solvable*, too.

Weak T7 solvability: We have $b^* = \underline{b}$.
For $i = 1,\ j = 2$ we have $\overline{a}_{12} \otimes \underline{x}_2 \geq b_1^*$ and $\underline{x}_2 = \overline{x}_2$,
for $i = 2,\ j = 2$ we have $\overline{a}_{22} \otimes \underline{x}_2 \geq b_2^*$ and $\underline{x}_2 = \overline{x}_2$,
for $i = 3,\ j = 2$ we have $\overline{a}_{32} \otimes \underline{x}_2 \geq b_3^*$ and $\underline{x}_2 = \overline{x}_2$.
The given interval system *is weakly T7 solvable* which implies that this *is T1 solvable*, too.

T5 solvability:
Since $\underline{A} \otimes \underline{x} = (8, 11, 8)^T \not\geq \underline{b}$, the given interval system *is not T5 solvable.* Consequently, this is *not T3, not T8* and *not T9 solvable.*

Answer: The given interval system is **T1, T2, weakly T6, strongly T6** and **weakly T7 solvable.**

References

[1] Cechlárová, K.: Solutions of interval systems in max-plus algebra. In: *Proc. SOR 2001* (V. Rupnik, L. Zadnik-Stirn, S. Drobne, eds.), Preddvor, Slovenia, 321–326.

[2] Cechlárová, K., and Cuninghame-Green, R. A.: Interval systems of max-plus separable linear equations, *Linear Algebra and its Applications* **340** (2002), 215–224.

[3] Cuninghame-Green, R. A.: *Minimax Algebra.* Lecture notes in Economics and Mathematical systems 1966, Springer, Berlin, 1979.

[4] Myšková, H.: Interval systems of max-separable linear equations, *Linear Algebra and its Applications* **403** (2005), 263–272.

[5] Myšková, H.: Control solvability of interval systems of max-separable linear equations, *Linear Algebra and its Applications* **416** (2006), 215–223.

[6] Myšková, H.: On the algorithm for testing T4 solvability of max-plus interval systems, *Discrete Applied Mathematics*, in reviewing process.

[7] Myšková, H.: Interval max-plus systems of linear equations, *Linear Algebra and its Applications*, in reviewing process.

[8] Myšková, H.: Interval solutions in max-plus algebra, *Proceedings of 10th Internacional Conference APLIMAT 2011*, February 1–4, 2011, Bratislava, Slovakia, 143-150.

[9] Nachtigall, K.: Powers of Matrices over an Extremal Algebras with Applications to Periodic Graphs, *Math. Methods of Oper. Research* **46** (1997), 87–102.

[10] Zimmermann, K.: *Extremální algebra.* Ekonomicko-matematická laboratoř Ekonomického ústavu ČSAV, Praha 1976.

Czech labour market through the lens of a search and matching DSGE model

Daniel Němec[1]

Abstract. This contribution reveals some important structural properties of the Czech labour market in the last fifteen years and evaluates possible changes within this period. A small search and matching model incorporated into standard macroeconomic dynamic stochastic general equilibrium model is estimated using Bayesian techniques. The results show that search and matching aspect provides satisfactory description of employment flows in the Czech economy. Model estimates provide convincing evidence that wage bargaining process is determined mainly by the power of the unions and that the institutional changes of the Czech labour market in the last fifteen years had only little real impact on the matching effectiveness.

Keywords: search and matching model, Bayesian estimation, DSGE model, structural changes

JEL classification: C51, E24, J60
AMS classification: 91B40, 91B51

1 Introduction

The goal of my contribution is to reveal some interesting and important structural properties of the Czech labour market in the last fifteen years and to evaluate possible changes within this period. For this purpose, I use a small search and matching model incorporated into standard macroeconomic dynamic stochastic general equilibrium model (DSGE). Search and matching model is an important tool to model labour market dynamics. This model is a log-linear version of the model originally developed by Lubik [4]. Using real macroeconomic data I am able to estimate some key labour market indicators: the wage bargaining power of unions, the match elasticity of unemployed and the efficiency of the matching process.

The structure of my contribution is as follows. The next section provides a short description of the small search and matching DSGE model which is used for my analysis. Section 3 discusses used data, priors and estimation techniques. Section 4 presents the main results. Section 5 provides a deeper insight into model properties and its ability to match observed data. Section 6 concludes this contribution with some ideas regarding the possibilities of further research in this area.

2 The model

As mentioned previously, I use the model developed by Lubik [4]. It is a simple search and matching model incorporated within a standard DSGE framework. The labour market is subject to friction because a time-consuming search process for workers and firms. The wages are determined by the outcome of a bargaining process which serves as a mechanism to redistribute the costs of finding a partner.

Households Representative household maximizes its expected utility function

$$E_t \sum_{j=1}^{\infty} \beta^{j-t} \left[\frac{C_j^{1-\sigma} - 1}{1-\sigma} - \chi_j n_j \right], \tag{1}$$

[1] Masaryk University, Faculty of Economics and Administration, Department od Economics, Lipova 41a, 602 00 Brno, Czech Republic, e-mail: nemecd@econ.muni.cz

where C is aggregate consumption, $n \in [0,1]$ is a fraction of employed household members (determined by the matching labour market), $\beta \in (0,1)$ is the discount factor and $\sigma \geq 0$ is the coefficient of relative risk aversion. Variable χ_t represents an exogenous stochastic process which may be taken as a labour shock. The budget constraint is defined as

$$C_t + T_t = w_t n_t + (1 - n_t)b + \Pi_t, \tag{2}$$

where b is unemployment benefit financed by a lump-sum tax, T. Variable Π_t are profits from ownership of the firms and w is wage. There is no explicit labour supply because it is an outcome of the matching process. The first-order condition is thus simply

$$C_t^{-\sigma} = \lambda_t, \tag{3}$$

where λ_t is the Lagrange multiplier on the budget constraint.

Labour market The labour market is characterized by search frictions captured by a standard Cobb-Douglas matching function

$$m(u_t, v_t) = \mu_t u_t^{\xi} \nu_t^{1-\xi}, \tag{4}$$

where unemployed job seekers, u_t, and vacancies, ν_t, are matched at rate $m(u_t, \nu_t)$. Parameter $0 < \xi < 1$ is a match elasticity of the unemployed and μ_t is stochastic process measuring the efficiency of the matching process. Aggregate probability of filling a vacancy may be defined as

$$q(\theta_t) = m(u_t, \nu_t)/\nu_t, \tag{5}$$

where $\theta_t = \frac{\nu_t}{u_t}$ is a standard indicator of the labour market tightness. The model assumes that it takes one period for new matches to be productive. Moreover, old and new matches are destroyed at a constant separation rate, $0 < \rho < 1$, which corresponds to the inflows into unemployment. Evolution of employment, $n_t = 1 - u_t$, is given by

$$n_t = (1 - \rho)\left[n_{t-1} + \nu_{t-1} q(\theta_{t-1})\right]. \tag{6}$$

Firms As a deviation from the standard search and matching framework, the model assumes monopolistic firms. Demand function of a firm is defined by

$$y_t = \left(\frac{p_t}{P_t}\right)^{-1-\epsilon} Y_t, \tag{7}$$

where y_t is firm's production (and its demand), Y_t is aggregate output, p_t is price set by the firm, P_t is aggregate price index and ϵ is demand elasticity which will be not treated as a stochastic process in my empirical application. Production function of each firm is

$$y_t = A_t n_t^{\alpha}, \tag{8}$$

where A_t is an aggregate technology (stochastic) process and $0 < \alpha \leq 1$ introduces curvature in production. Capital is fixed and firm-specific. The firm controls the number of workers, n_t, number of posted vacancies, ν_t, and its optimal price, p_t, by maximizing the inter-temporal profit function

$$E_t \sum_{j=1}^{\infty} \beta^{j-t} \lambda_j \left[p_j \left(\frac{p_t}{P_t}\right)^{-(1+\epsilon)} Y_j - w_j n_j - \frac{\kappa}{\psi} \nu_j^{\psi} \right], \tag{9}$$

subject to the employment accumulation equation (7) and production function (8). Profits are evaluated in terms of marginal utility λ_j. The costs of vacancy posting is $\frac{\kappa}{\psi} v_t^{\psi}$, where $\kappa > 0$ and $\psi > 0$. For $0 < \psi < 1$, posting costs exhibit decreasing returns. For $\psi > 1$, the costs are increasing while vacancy costs are fixed for $\psi = 1$. The first-order conditions are

$$\tau_t = \alpha \frac{y_t}{n_t} \frac{\epsilon}{1+\epsilon} - w_t + (1 - \rho) E_t \beta_{t+1} \tau_{t+1}, \tag{10}$$

$$\kappa \nu_t^{\psi - 1} = (1 - \rho) q(\theta_t) E_t \beta_{t+1} \tau_{t+1}, \tag{11}$$

where $\beta_{t+1} = \beta \frac{\lambda_{t+1}}{\lambda_t}$ is a stochastic discount factor and τ_t is the Lagrange multiplier associated with employment constraint. The first condition represents current-period marginal value of a job. The second condition is a link between the cost of vacancy and the expected benefit of a vacancy in terms of the marginal value of a worker (adjusted by the job creation rate, $q(\theta_t)$).

Wage bargaining Wages are determined as the outcome of a bilateral bargaining process between workers and firms. Both sides of the bargaining maximize the joint surplus from employment relationship:

$$S_t \equiv \left(\frac{1}{\lambda_t}\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t}\right)^{\eta}\left(\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}\right)^{1-\eta}, \tag{12}$$

where $\eta \in [0,1]$ is the bargaining power of workers, $\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t}$ is the marginal value of a worker to the household's welfare and $\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}$ is the marginal value of a worker to the firm. The term $\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t} = \tau_t$ is given by the first-order condition (10). Recursive representation for $\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t}$ is derived as

$$\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t} = \lambda_t w_t - \lambda_t b - \chi_t + \beta E_t \frac{\partial \mathcal{W}_{t+1}(n_{t+1})}{\partial n_t t+1}\frac{\partial n_{t+1}}{\partial n_t}. \tag{13}$$

Using employment equation (6), it holds $\frac{\partial n_{t+1}}{\partial n_t} = (1-\rho)[1-\theta_t q(\theta_t)]$. All real payments are valued at the marginal utility λ_t. Standard optimality condition for wages may be derived as

$$(1-\eta)\frac{1}{\lambda_t}\frac{\partial \mathcal{W}_t(n_t)}{\partial n_t} = \eta\frac{\partial \mathcal{J}_t(n_t)}{\partial n_t}. \tag{14}$$

Expression for the bargained wage is given after some algebra as

$$w_t = \eta\left[\alpha\frac{y_t}{n_t}\frac{\epsilon_t}{1+\epsilon_t} + \kappa\nu_t^{\psi-1}\theta_t\right] + (1-\eta)\left[b + \chi_t c_t^{\sigma}\right]. \tag{15}$$

Closing the model The model assumes, that unemployment benefits, b, are financed by lump-sum taxes, T, where a condition of balanced budget holds, i.e. $T_t = (1-n_t)b$. Social resource constraint is thus $C_t + \frac{\kappa}{\psi}\nu_t^{\psi} = Y_t$. The technology shock A_t, the labour shock χ_t and the matching shock μ_t are assumed to be independent $AR(1)$ processes (in logs) with coefficients ρ_i, $i \in (A,\xi,\mu)$ and innovations $\epsilon_t^i \sim N(0,\sigma_i^2)$.

Log-linearised model For estimation purposes, I did not use the non-linear form of the model mentioned in the previous section (of course, this form is important to understand the meaning of the key structural model parameters). Instead of that, I use a log-linear version of the model based on my own derivations.[1] In the following equations, the line over a variable means its steady-state value (derived simply from the non-linear equations).[2] The variables with a tilde represent the gaps from their steady-states.

$$\tilde{\lambda}_t = -\delta\tilde{C}_t \qquad \tilde{m}_t = \tilde{\mu}_t + \xi\tilde{u}_t + (1-\xi)\tilde{\nu}_t$$

$$\tilde{q}_t = \tilde{m}_t - \tilde{\nu}_t \qquad \tilde{\theta}_t = \tilde{\nu}_t - \tilde{u}_t$$

$$\tilde{n}_t = -\frac{\overline{u}}{1-\overline{u}}\tilde{u}_t \qquad \tilde{n}_t = \frac{1}{\overline{n}+\overline{vq}}\left[\overline{u}\tilde{n}_{t-1} + \overline{q\nu}(\tilde{\nu}_{t-1}+\tilde{q}_{t-1})\right]$$

$$\tilde{y}_t = (-1-\epsilon)(\tilde{p}_t - \tilde{P}_t) + \tilde{Y}_t \qquad \tilde{y}_t = \tilde{A}_t + \alpha\tilde{n}_t$$

$$\tilde{\tau}_t = \frac{1}{\alpha\frac{\overline{y}}{\overline{n}}\frac{\epsilon}{1+\epsilon}\overline{w} + (1-\rho)\overline{\beta}\overline{\tau}}\left[\alpha\frac{\epsilon}{1+\epsilon}(\tilde{y}_t - \tilde{n}_t) - \overline{w}\tilde{w}_t + (1-\rho)\overline{\tau}\overline{\beta}E_t\left(\tilde{\beta}_{t+1} + \tilde{\tau}_{t+1}\right)\right]$$

$$(\psi-1)\tilde{\nu}_t = \tilde{q}_t + E_t\left(\tilde{\beta}_{t+1} + \tilde{\tau}_{t+1}\right)$$

$$\tilde{\beta}_t = \tilde{\lambda}_t + \tilde{\lambda}_{t-1}$$

$$\tilde{w}_t = \frac{1}{\overline{w}}\left[\eta\left(\alpha\frac{\epsilon}{1+\epsilon}\frac{\overline{y}}{\overline{n}}(\overline{y}_t - \overline{n}_t) + \kappa\overline{\nu}^{\psi-1}\overline{\theta}\left((\psi-1)\tilde{v}_t + \tilde{\theta}_t\right)\right) + (1-\eta)\overline{\chi}\overline{C}^{\sigma}(\tilde{\chi}_t + \sigma\tilde{C}_t)\right]$$

$$\tilde{Y}_t = \frac{1}{\overline{C} + \frac{\chi}{\psi}\overline{\nu}^{\psi}}\left(\overline{C}\tilde{C}_t + \kappa\overline{\nu}^{\psi}\tilde{\nu}_t\right)$$

[1] Log-linear version is not a part of the original contribution of Lubik [4].

[2] Initial steady-state values are calibrated as follows: $\mu^* = A^* = \chi^* = 1$, $\beta^* = 0.99$, $u^* = 0.0763$, $\nu^* = 0.0127$. Remaining steady-states are computed using these values and the prior means of all parameters.

$$\tilde{A}_t = \rho_A \tilde{A}_{t-1} + \epsilon_t^A \quad \tilde{\chi}_t = \rho_\chi \tilde{\chi}_{t-1} + \epsilon_t^\chi \quad \tilde{\mu}_t = \rho_\mu \tilde{\mu}_{t-1} + \epsilon_t^\mu \quad \tilde{Y}_t = \rho_Y \tilde{Y}_{t-1} + \epsilon_t^Y$$

The last equation results from the fact that variable Y is an observed variable. We have thus four shocks (ϵ_t^i for four observed variables – $\tilde{u}$, $\tilde{\nu}$, $\tilde{w}$ and $\tilde{Y}$). The model consists of 17 endogenous variables (variable $(\tilde{p}_t - \tilde{P}_t)$ is a single variable in my application), four shocks and 14 parameters.

3 Data and priors

The model for the Czech economy is estimated using the quarterly data set covering a sample from 1996Q1 to 2009Q4. The observed variables are real output (GDP, in logs), hourly earnings (in logs), unemployment rate and rate of unfilled job vacancies. All data are seasonally adjusted and de-trended (excluding vacancies) using Hodrick-Prescott filter (with the smoothing parameter $\lambda = 1600$). The rate of unfilled job vacancies was demeaned prior estimation. The variables used are expressed as corresponding gaps. The original data are from databases of the OECD and the Czech Statistical Office (CZSO).[3] In the following parts, I will not use the mark ˜ to explicitly express the appropriate gaps.

Description	Parameter	Density	Mean	Std. Dev.
Discount factor	β	Fixed	0.99	—
Labor elasticity	α	Fixed	0.67	—
Demand elasticity	ϵ	Fixed	10	—
Relative risk aversion	σ	Gamma	1.00	0.50
Match elasticity	ξ	Gamma	0.70	0.10
Separation rate	ρ	Gamma	1.00	0.50
Bargaining power of the workers	η	Uniform	0.50	0.3
Unemployment benefits	b	Beta	0.20	0.15
Elasticity of vacancy creation cost	ψ	Gamma	1.00	0.50
Scaling factor on vacancy creation cost	κ	Gamma	0.10	0.05
AR coefficients of shocks	$\rho_{\{\chi,A,\mu,Y\}}$	Beta	0.8	0.2
Standard deviation of shocks	$\sigma_{\{\chi,A,\mu,Y\}}$	Inv. Gamma	0.05	∞
Standard deviation of measurement errors	$\sigma^*_{\{u\}}$	Uniform	0.001	0.0006
Standard deviation of measurement errors	$\sigma^*_{\{w,\nu\}}$	Uniform	0.001	0.0003

Table 1: Parameters description and prior densities

Parameters are estimated using Bayesian techniques combined with Kalman filtering procedures. All computations have been performed using Dynare toolbox [2] for Matlab. Table 1 reports the model parameters and the corresponding prior densities. The priors (and calibrations) are similar to those used by Lubik [4]. On the other hand, the standard deviations are rather uninformative.

4 Estimation results

Table 2 presents the posterior estimates of parameters and 90% highest posterior density intervals. It may be seen (in comparison with the Table 1) that most of the parameters are moved considerably from their prior means. The data seems to be strongly informative. There are some remarkable results which should be emphasized:

- The first surprising estimate is the bargaining power of workers, η. The mean value of this parameter is almost 0.9 with a 90 percent coverage region that is shifted considerably away from the prior density. This implies that the workers can gain the most of their entire surplus. The firms are thus not willing to create vacancies. This result is in sharp contrast to the results of Lubik [4] or Yashiv [6] who aimed to model the U.S. labour market.

[3] GDP at purchaser prices, constant prices 2000, s.a., CZSO, millions of CZK; index of hourly earnings (manufacturing), 2005=100, s.a., OECD; registered unemployment rate, s.a., OECD; unfilled job vacancies, level (transformed to ratio of unfilled vacancies to labour force), s.a., OECD.

	Posterior mean	90% HPDI			Posterior mean	90% HPDI	
σ	1.0399	0.6036	1.4894	ρ_μ	0.8281	0.7298	0.9525
ξ	0.6459	0.5823	0.7099	ρ_Y	0.8695	0.8197	0.9241
ρ	0.0268	0.0086	0.0486	σ_χ	0.0162	0.0108	0.0214
η	0.8694	0.8311	0.9082	σ_A	0.0080	0.0067	0.0093
b	0.2924	0.1318	0.4688	σ_μ	0.0066	0.0059	0.0073
ψ	3.5706	3.0908	3.9944	σ_Y	0.0095	0.0080	0.0108
κ	0.0880	0.0637	0.1199	σ^*_u	0.0005	0.0002	0.0009
ρ_χ	0.7831	0.7208	0.8555	σ^*_w	0.0003	0.0001	0.0004
ρ_A	0.9629	0.9332	0.9996	σ^*_ν	0.0002	0.0001	0.0004

Table 2: Parameter estimates

- The second interesting result is the estimated separation rate, ρ. This parameter is considerably lower than the one estimated by Lubik [4]. Its value supports the view of less flexible Czech labour market with limited ability to destroy old and new matches.
- The third remarkable estimate is the vacancy posting elasticity, ψ. The posterior mean 3.57 is shifted away from the prior mean. The vacancy creation is thus more costly because of increasing marginal posting costs (increasing in the level of vacancies or labour market tightness, θ). Lubik [4] estimated this parameter at the mean value of 2.53. In this case, the high value of ψ may be interpreted as a balancing factor which "restrict" potentially excessive vacancy creation driven by the low bargaining power. In case of the Czech labour market, this higher value provides further evidence of specifically less flexible labour market.
- The estimate of parameter b corresponds to a reasonable value 0.3 which might be in accordance with the real unemployment benefits paid within the Czech social insurance system (30% of average wage).
- The posterior mean of the matching function parameter, ξ, is in accordance with the common values in literature (see Lubik [4] or Christoffel et al. [1]).

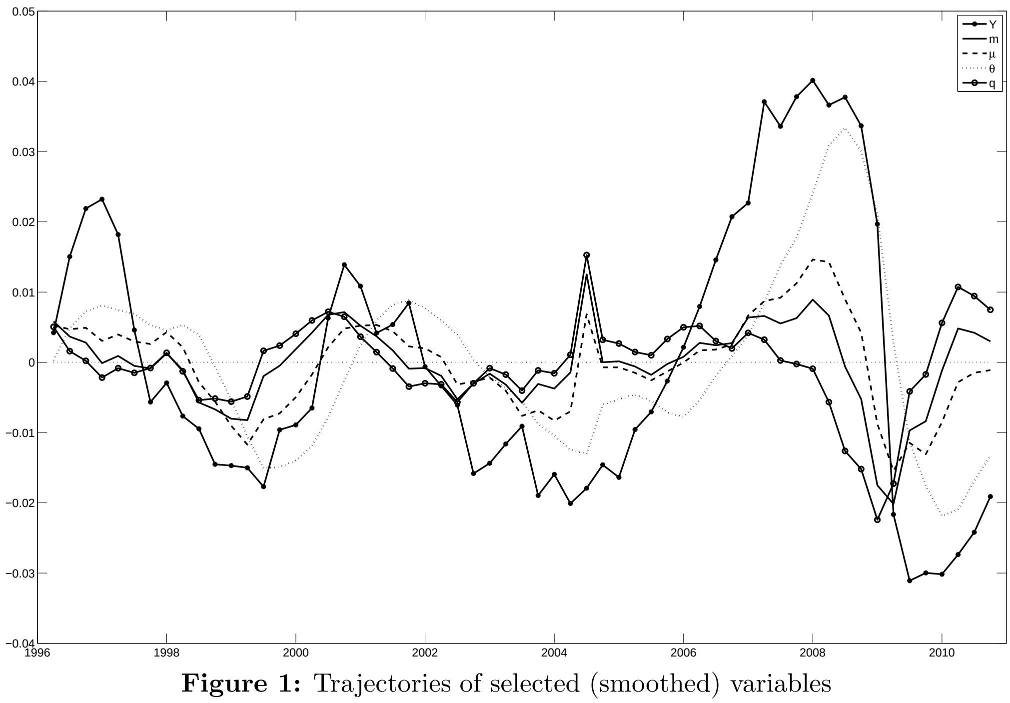


Figure 1: Trajectories of selected (smoothed) variables

Figure 1 presents the trajectories of selected smoothed variables. We can see a relative smooth development of variable q (probability of filling a vacancy) with a sharp decline at the end of the year 2007. This evidence is in favour of conclusion presented by Němec and Vašíček [5] who stressed the role of an obvious lack of employees in the Czech economy. This tendency was reverted as a result of the last global economic slowdown starting at the end of 2008. The efficiency of the matching process is strongly correlated with the output gap (correlation coefficient is 0.8). This indicator is thus seemingly independent of the institutional framework of the Czech labour market. It probably means that the changes in labour market institution have been mostly marginal (with little real impacts). The correlation between output gap and the matching variable m is relative small, 0.4. On the other hand, the correlation between actual value of output gap and the lagged value of m is 0.65. Current value of matching function might be thus an useful indicator of future (one-quarter ahead) changes in the real output (output gap).

5 Model evaluation

In order to see how the model fits the data, sample moments, autocorrelation coefficients and cross-correlations are computed. I computed these statistics from simulation of the estimated models with parameters set at their posterior means. All these statistics correspond to the four observed series (unemployment gap, u, gap of vacancies, ν, gap of the wages, w, and output gap, Y). The results may be found in the Tables 3 and 4.

		Sample moments		Lags for autocorrelation coefficients			
		Mean	Std. dev.	1	2	3	4
u	data	0.00	0.009	0.91	0.71	0.45	0.16
	model	−0.00	0.010	0.88	0.70	0.51	0.35
	90% HPDI	(−0.01, 0.01)	(0.007, 0.014)	(0.79, 0.94)	(0.48, 0.83)	(0.11, 0.72)	(−0.08, 0.62)
ν	data	0.00	0.004	0.91	0.71	0.45	0.17
	model	0.00	0.008	0.72	0.54	0.40	0.29
	90% HPDI	(−0.01, 0.01)	(0.006, 0.011)	(0.55, 0.87)	(0.25, 0.80)	(0.08, 0.73)	(−0.09, 0.67)
w	data	0.00	0.014	0.80	0.53	0.29	0.14
	model	0.00	0.054	0.72	0.52	0.36	0.24
	90% HPDI	(−0.04, 0.04)	(0.041, 0.071)	(0.57, 0.84)	(0.30, 0.72)	(0.06, 0.61)	(−0.09, 0.57)
Y	data	0.00	0.020	0.91	0.74	0.54	0.33
	model	0.00	0.017	0.79	0.62	0.47	0.36
	90% HPDI	(−0.01, 0.01)	(0.012, 0.024)	(0.64, 0.88)	(0.33, 0.77)	(0.09, 0.70)	(0.01, 0.63)

Table 3: Sample moments and autocorrelation coefficients

	Data				Model (90% HPDI)			
	u	ν	w	Y	u	ν	w	Y
u	1.00	−0.90	−0.74	−0.77	1.00	−0.17	−0.12	−0.16
					(1.00, 1.00)	(−0.56, 0.25)	(−0.53, 0.31)	(−0.66, 0.34)
ν	−0.90	1.00	0.80	0.88	−0.17	1.00	0.63	0.72
					(−0.56, 0.25)	(1.00, 1.00)	(0.25, 0.89)	(0.46, 0.87)
w	−0.74	0.80	1.00	0.60	−0.12	0.63	1.00	0.21
					(−0.53, 0.31)	(0.25, 0.83)	(1.00, 1.00)	(−0.26, 0.62)
Y	−0.77	0.88	0.60	1.00	−0.16	0.78	0.21	1.00
					(−0.66, 0.34)	(0.46, 0.87)	(−0.26, 0.62)	(1.00, 1.00)

Table 4: Matrix of correlation

The model is very successful in matching sample moments and autocorrelation coefficients (they are mostly within the appropriate 90% highest posterior density intervals). This ability is not used to be typical for such a small-scale model. But, there is one exception regarding the fit of sample moments. The model predicts higher volatility in wages. This pattern reveals the necessity of enrichment by a new source of wage rigidity (as suggested by Krause and Lubik [3] or Christoffel et al. [1]).

My results are in accordance with the authors arguing that the model with search and matching frictions in the labour market is able to generate negative correlation between vacancies and unemployment (see Krause and Lubik [3]). Unfortunately, the values of cross-correlation coefficients (see the lowest bounds of HPDI in the Table 4) are not sufficient for the correlations of unemployment and the rest of observable variables. The similar experience may be found in the results for U.S. labour market provided by Lubik [4]. Lubik pointed out that this may be due the presence of matching shock, which can act as a residual in employment and wage equations.

6 Conclusion

In my contribution, I investigated structural properties of the Czech labour market using a simple DSGE framework with labour market rigidities. Two sources of rigidities were implemented: wage bargaining mechanism and "‘search and matching"’ process matching workers and firms. Estimated model provides satisfactory description of employment flows in the Czech economy. Parameter estimates provide convincing evidence that wage bargaining process is determined mainly by the power of the unions and that the institutional changes of the Czech labour market in the last fifteen years had only little real impact on the matching effectiveness. Unfortunately, the model predicts higher volatility in wages. This pattern reveals the necessity of enrichment by a new source of wage rigidity as proposed by Krause and Lubik [3] or Christoffel et al. [1].

Acknowledgements

This work was supported by MŠMT project Research centers 1M0524.

References

[1] Christoffel, K., Kuester, K., and Linzert, T.: The Role of Labor Markets for Euro Area Monetary Policy, *European Economic Review* **53** (2009), 908–936.

[2] Juillard, M.: Dynare Toolbox for Matlab, version 4.2.0.

[3] Krause, M., and Lubik, T.: The (ir)relevance of real wage rigidity in the New Keynesian model with search frictions, *Journal of Monetary Policy* **54** (2007), 706–727.

[4] Lubik, T. A.: Estimating a Search and Matching Model of the Aggregate Labor Market, *Economic Quarterly* **95** (2009), 101–120.

[5] Němec, D., and Vašíček, O.: Estimating structural changes of the Czech economy: How convincing are the symptoms of the economic crisis? In: *Proceedings of the Mathematical Methods in Economics 2010* (M. Houda, and J. Fribelov, eds.), České Budějovice, 2010, 469–476.

[6] Yashiv, E. A.: Evaluating the performance of the search and matching model, *European Economic Review* **50** (2006), 909–936.

Application of selected scoring models on corporate credit rating

Martina Novotná[1]

Abstract. The topic of the paper concerns with the area of credit risk measurement possibilities and quantification. Under the credit risk it is possible to consider any risk resulting from the inability or unwillingness of the counterparty to carry out the obligations. Credit rating then reflects the creditworthiness of a particular company, security or obligation. The objective of the paper is to demonstrate an illustrative application of two selected scoring models with respect to the credit rating assessment. By means of multivariate models, we can prove that some financial variables are statistically relevant for rating. Two approaches, linear discriminant analysis and multinomial logistic regression will be used on selected European data. The dataset comprises only nonfinancial corporations from recently acceded EU countries for which is typical the absence of publicly available information concerning external rating. A special attention will be paid to the comparison of models and the interpretation of results.

Keywords: credit rating, model estimation, multinomial logistic regression, multivariate discriminant analysis.

1 Introduction

The goal of the paper is to demonstrate an illustrative application of two selected scoring models that will be applied on credit rating. Two approaches, multivariate linear discriminant analysis and multinomial logistic regression will be used in our analysis. The aim of the analysis is to estimate credit rating models from the sample of European companies. The selected techniques enable us to find the most significant financial variables with respect to the credit rating assessment. First, a brief introduction to the current state of research is mentioned. The next part describes the methodology and a sample used in the analysis. The comparison of two approaches, their results and classification ability of models will be assessed in the last chapter of this paper.

Credit risk models can be used as a guideline when evaluating unrated firms. One of the best known models in this area is applied by E. I. Altman, whose default model is often used as a tool in financial analysis of a company. This model has the ability to identify companies with the possible financial problems and was proposed on the basis of multivariate discriminant analysis. The other research in this area is primarily focused on bond rating and bond rating models. In addition to discriminant analysis, regression analysis became one of the most used methods to estimate rating in the primary research. An approach of the multivariate discriminant analysis was introduced by Pinches and Mingo (1973), Ang and Patel (1978), Belkaoui (1980) and Altman and Katz (1976). Subsequent research was concentrated on comparison of particular statistical methods; e.g. Kaplan and Urwitz (1979) compare ordered probit analysis with ordinary least square regression, Wingler and Watts (1980) compare ordered probit analysis with multiple discriminant analysis. Other studies replicate the process of bond rating model estimation and modify the approaches by considering of new variables, such as the study of Chan and Jegadeesh (2004); or examine the impact of financial variables on credit rating for a given country or region, e.g. Gray, Mirkovic and Ragunathan (2005).

2 Methodology

There is a variety of approaches that can be used for the purposes of credit risk modelling; the paper shows the application of chosen scoring rating models such as discriminant analysis and logistic regression. Both techniques are suitable for this type of problem, where there are more than two categories of dependent variable

[1] Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská třída 33, Ostrava. E-mail: martina.novotna@vsb.cz.

(five rating categories) and each observation is described by several independent variables (financial ratios). Next paragraphs give a brief explanation of the above mentioned techniques.

2.1 Multivariate discriminant analysis

Discriminant analysis is a common statistical method used for separation of groups, and hence a suitable method for credit rating modelling. The analysis can be used for two major objectives: i) description of group separation and ii) prediction or allocation of observations to groups. Discriminant functions are linear combinations of variables that best separate groups, for example the *k* groups of multivariate observations. For the following part of this paragraph, the explanation and definitions were taken from Rencher [4] and Huberty and Olejnik [2].

For k groups with n_i observations in the ith group, we transform each observation vector y_{ij} to obtain $z_{ij} = a'y_{ij}$, i = 1, 2,..., k; j = 1, 2,...,n_i, and find the means $\overline{z}_i = a'\overline{y}_i$, where $\overline{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. We seek the vector **a** that maximally separates $\overline{z}_1, \overline{z}_2, \dots \overline{z}_k$. The separation criterion among $\overline{z_1}, \overline{z_2} \dots, \overline{z_k}$ can be expressed in term of matrices,

$$\lambda = \frac{a'Ha}{a'Ea} \tag{1}$$

where matrix H has a between sum of squares on the diagonal for each of the *p* variables, and matrix E has a within sum of squares for each variable on the diagonal. Another expression of the separation criterion is

$$\lambda = \frac{\text{SSH }(z)}{\text{SSE }(z)} \tag{2}$$

where SSH (*z*) and SSE (*z*) are the between and within sums of squares for z. The main task of the discriminant analysis is to find a set of weights (*a* values) for the outcome variables to determine a linear composite:

$$Z = a_1 Y_1 + a_2 Y_2 + \dots + a_p Y_p \tag{3}$$

so that the ratio (2) is maximized. The discriminant analysis follows by assessing the relative contribution of the $y's$ to separation of several groups and testing the significance of a subset of the discriminant function coefficients. The discriminant criterion (1) is maximized by λ_1, the largest eigenvalue of $E^{-1}H$; the remaining eigenvalues correspond to other discriminant dimensions. The test of significance is usually based on the Wilks' lambda, Λ, the most widely used criterion. The test statistic at the *m*th step is

$$\Lambda_m = \prod_{i=m}^{s} \frac{1}{1+\lambda_i}, \tag{4}$$

which is distributed as $\Lambda_{p-m+1,k-m,N-m+1}$. The statistic,

$$V_m = -\left[N - 1 - \tfrac{1}{2}(p+k)\right] ln\Lambda_m = \left[N - 1 - \tfrac{1}{2}(p+k)\right] \sum_{i=m}^{s} ln(1+\lambda_i), \tag{5}$$

has an approximate χ^2-distribution with *(p-m+1)(k-m)* degrees of freedom.

2.2 Multinomial logistic regression

Logistic regression is another technique which can be used to analyze the relationship of multiple independent variables to a dependent variable. The multinomial logistic regression is a modification of binary logistic, where only two possible outcomes can occur. The definitions and derivations used in this chapter were extracted from Hosmer and Lemeshow [1].

The model for dichotomous outcome variable is based on logistic distribution and we use the quantity $\pi(x) = E(Y|x)$ to represent the conditional mean of Y given x when the logistic distribution is used,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1+e^{\beta_0 + \beta_1 x}}. \tag{6}$$

The central idea of logistic regression is a transformation of $\pi(x)$, so-called logit transformation. This transformation is given by the following equation, showing the case of two dependent (binary) variables.

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x \tag{7}$$

The logit, g(x), is linear in its parameters and may range from $-\infty\ to + \infty$, depending on the range of x. To estimate the logistic regression model, we find the values of parameters β_0 and β_1 which maximize the probability of obtaining the observed set of data. Thus, we must first construct the likelihood function which expresses the probability of the observed data as a function of the unknown parameters. The likelihood function for binary dependent variable can be defined as the log likelihood,

$$L(\beta) = ln[l(\beta)] = \sum_{i=1}^{n}\{y_i ln[\pi(x_i)] + (1 + y_i) ln[1 - \pi(x_i)]\}\ . \tag{8}$$

The multinomial logistic regression is then a modification of the binary alternative. Our analysis of credit rating to five categories requires four logit functions and determination of the so-called reference or baseline category, which is then compared with other logits. A general expression for the conditional probability in the five category model is

$$P(Y = j|\mathbf{x}) = \frac{e^{g_j(x)}}{\sum_{k=1}^{5} e^{g_k(x)}}\ , \tag{9}$$

where $P(Y = j|\mathbf{x})$ = π_j(x) for k = 1, 2, 3, 4, 5 and g_5(x) $= 0$ for the baseline category five.

3 Data description

The sample for analysis must involve financial variables for each company and a rating assessment. For our purposes, companies' credit ratings are extracted from the Amadeus database[2] of public and private European companies. The MORE ratings[3] are comparable across countries; two companies from different countries with the same rating have the same creditworthiness. MORE ratings classify companies similarly as rating agencies, the description of categories is presented in Table 1.

Rating	Description
AAA	The company's capacity to meet its financial commitments is extremely strong.
AA	The company has very strong creditworthiness
A	The company has a high solvency.
BBB	Capital structure and economic equilibrium are considered adequate.
BB	The company is more vulnerable than companies rated 'BBB'.
B	The company presents vulnerable signals with regard to its fundamentals.
CCC	The company has a dangerous disequilibrium on the capital structure and on its economic and financial fundamentals.
CC	The company shows signals of high vulnerability.
C	The company shows considerable pathological situations.
D	The company has not any longer the capacity to meet its financial commitment

Table 1 MORE Rating categories (Bureau van Dijk Electronic Publishing, 2008)

The analysis is focused on eight selected countries from Central and Eastern Europe: the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia and Slovenia. Only very large and large companies[4] from mining, manufacturing and construction industries with MORE rating were selected. The sample contains data from the period 2002 - 2008. The original sample of 4,802 observations was split into two sub-samples: the experimental sample (70 percent of the total sample) for the model estimation and the test sample that will be used for the model verification. Main descriptive statistics were obtained to get an overall picture of the dataset, the experimental sample was adjusted for extreme values (outliers) and assessing normality was carried out.

[2] AMADEUS is a database containing financial information on public and private European companies countries.

[3] The MORE rating is calculated using a unique model that references the company's financial data to create an indication of the company's financial risk level.

[4] Companies in Amadeus database are considered to be very large (or large) when they have operating revenue equal or greater than 100 (10) million euro, or total assets equal or higher than 200 (20) million euro, or number of employees is equal or greater than 1000 (150).

Following suggestions of some authors[5], variables with skewed distribution were transformed by means of logarithm to approach the normal distribution. Finally, the total number of observations is 2,992. Normality of the distributions was accessed by the Kolmogorov-Smirnov test (in our sample, the significance value .000 suggests violation of the assumption of normality for six variables which is common for larger samples like the one in our analysis).

MORE Rating	Number of cases	Marginal percentage	Rating label for analysis
AA	162	5.4%	5
A	721	24.1%	4
BBB	1529	51.1%	3
BB	507	16.9%	2
B	73	2.4%	1
Total	**2992**	**100%**	

Table 2 Rating categories (experimental sample)

The table above shows an uneven representation of rating grades. It is evident that most cases fall within the middle grades A, BBB and BB.

3.1 Selection of independent variables

Previous studies suggest that some variables play an important role in the credit rating process, whereas the others are not relevant for rating. Financial variables in the analysis can be divided into several groups, as is shown in the summary table below. Nine variables entering the analysis were checked for multicolinearity which can substantially affect the resulting models. High level of interdependencies among the variables affects most multivariate analysis (for example, for the purposes of discriminant analysis, it is recommended to use rather lower number of variables than more variables with large interdependencies). Based on the analysis of bivariate correlations between individual variables, two variables should be removed from the experimental sample, ROCE and LogLIQUID. The general description of the final sample including seven independent variables is shown in the following table (Tab. 3).

Category	Economic rationale	Suggested financial variables	Final financial variables
Size	Adequate protection	Total assets (TA)	LogTA
Profitability	Ability to earn a satisfactory returns	Return on total assets (ROA)	ROA
		Return on equity (ROE)	ROE
		Return on capital employed (ROCE)	
Capitalization	Measure of capital structure and leverage	Long term debt to total assets (LTDTA)	LogLTDTA
		Equity to total assets (EQTA)	EQTA
Liquidity	The flow of financial resources	Liquidity ratio	LogCURR
		Current ratio	
Interest coverage	Ability to service the financial charges	Interest cover	LogINTCOV

Table 3 Overview of financial variables

4 Model estimation and interpretation

To obtain the credit rating model, two approaches are applied on the experimental sample, multivariate discriminant analysis (MDA) and multinomial logistic regression (MLR).

4.1 Multivariate discriminant analysis

The discriminant analysis is carried out by means of two methods, the simultaneous method resulting in the model containing all independent variables, and the stepwise method considering only variables with the greatest discriminating ability. Hence, two discriminant models are estimated by two methods.

[5] Pallant, J. SPSS Survival Manual. McGraw-Hill Education. London, 2007.

Tests of equality of group means show that all variables contribute to the model. Based on Wilks' lambda, the variables ROA (.483), LogINTCOV (.524) and EQTA (.557) are those with the greatest ability at discriminating between groups. The assumption of equality of covariances across groups is not validated, however Box's M can be overly sensitive to large data files, which is likely what happened in this case. Since we discriminate between five groups, the analysis results in five discriminant functions, all of them are statistically significant[6].

The stepwise approach gives the model containing five variables with the greatest discriminating ability: ROA, LogINTCOV, EQTA, LogCURR and ROE (the rank of each variable reflects its discriminating ability). The coefficients of classification functions (Fisher's linear discriminant functions) of this model are shown in the Table 4. To classify individual cases, the values of five functions must be calculated and the group corresponding to the function with the highest value is selected.

	1	2	3	4	5
EQTA	.380	.504	.640	.743	.805
ROE	.312	.311	.271	.192	0.046
ROA	-.567	-.374	-0.080	.408	1.246
LogCURR	-11.752	-10.157	-6.906	-1.433	3.330
LogINTCOV	2.279	3.720	6.036	9.879	11.675
Constant	-11.295	-13.962	-21.938	-39.267	65.155

Table 4 Classification function coefficients (MDA)

Classification results show a good ability to correctly classify cases within the five groups. The classification functions are weighted more heavily in favor of classifying group three, because the groups are not equally sized. Results of classification are presented in the table below (Tab. 5) and show overall classification ability of two models.

	Experimental sample	Test sample
Simultaneous method	85.8%	70.6%
Stepwise method	85.7%	70.4%

Table 5 Classification table (MDA)

4.2 Multinomial logistic regression

Multinomial logistic regression allows one to find the coefficients of predictors included in the model by using maximum likelihood method. In direct logistic regression, all predictors enter the equation simultaneously, while the stepwise method allows one to identify the most important variables for classification that should be included in the final model.

In the case of five grouping categories, four logit functions are needed. The category 5 denotes to the highest quality rating category (AA) and is used as a reference value. We form four logits, $g_1(x), g_2(x),\ g_3(x)$, $g_4(x),$ comparing $\mathrm{Y} = 1, \mathrm{Y} = 2, \mathrm{Y} = 3$ and $\mathrm{Y} = 4$ to the reference value. To fit the logistic regression model in equation (9) we estimate the unknown parameters using the maximum likelihood method. Based on the model fitting criteria (-2 Log Likelihood), both models are significantly different with ones with the constant only. The statistical significance of each of the coefficients is evaluated using the Wald test[7]. Two coefficients are not always statistically significant, LogTA in first two equations $g_1(x), g_2(x)$ and EQTA in the next two equations $g_3(x),\ g_4(x)$.

To form the logistic regression model, the conditional probability is expressed for each category in the model. For each case, we calculate probabilities $\pi(1), \pi(2), \pi(3), \pi(4), \pi(5)$ and select the rating category with the

[6] Wilks' lambda is a measure of how well each function separates cases into groups. It is equal to the proportion of the total variance in the discriminant scores not explained by differences among the groups. The associated chi-square statistic tests the hypothesis that the means of the functions listed are equal across groups. The small significance value indicates that the discriminant function does better than chance at separating the groups.
[7] A squared parameter estimate divided by its squared standard error is a Ch-square statistic.

highest probability value. The stepwise approach model includes six following variables, ROA, LogINTCOV, EQTA, LogCURR, ROE and LogTA. Parameter estimates of logit functions are presented in the Table 6.

	1	**2**	**3**	**4**
EQTA	-.377	-.167	.012	.030
ROE	.586	.559	.460	.140
ROA	-2.906	-2.256	-1.619	-.579
LogCURR	-29.815	-21.051	-11.408	-3.575
LogINTCOV	-13.979	-12.616	-9.910	-4.255
LogTA	.152	-.962	-1.425	-1.650
Constant	50.034	48.926	39.171	23.657

Table 6 Parameter estimates (MLR)

Given the observed data (see Tab.2), the “null” model would classify all objects into the modal category, 3, and would be correct in 51.1% of the time. Our models get much better classification ability, as is demonstrated in the Table 7.

	Experimental sample	**Test sample**
Full factors method	88.0%	72.4%
Stepwise method	87.9%	64.5%

Table 7 Classification table (MLR)

5 Conclusion

Two approaches were used to estimate credit rating models in this paper, multivariate discriminant analysis and multinomial logistic regression. The analyses use a large sample of European companies containing financial ratios and MORE rating for each firm. Four credit rating models were estimated, all of them statistically significant with high classification ability. Two models contain seven financial ratios (see Tab. 3), both of them give a good overall classification ability, even when applied on the test sample, 70.6% (MDA) and 72.4% (MLR). Both approaches allow using the stepwise method when variables with low discriminating ability are excluded from the model. By means of this method, simpler models were estimated, as they consider less financial ratios such as EQTA, ROE, ROA, LogCURR, LogINTCOV, or six including an additional variable LogTA in the logistic regression model. Both models provide satisfactory classification ability, 70.4% (MDA) and 64.5% (MLR) on the test sample.

The paper demonstrates that both approaches, MDA and MLR, are suitable for credit rating modelling. Although these methods are based on different methodology, they show very similar results and classification ability. Variables with the greatest ability at discriminating between groups are ROA, LogINTCOV and EQTA. These results confirm the importance of profitability, capitalization and interest coverage for credit rating. In practice, one should focus on these financial variables in particular. This strategy is relevant when it is compared over time for the firm, to the industry (industry and market averages can be used as benchmarks) and economy-wide measures of performance and financial position.

Acknowledgements

The paper is based on research activities sponsored through SGS research project No. SP2011/166. The support is greatly acknowledged.

References

[1] Hosmer, D.W. – Lemeshow, S. *Applied Logistic Regression*. John Wiley & Sons Inc., 2000.
[2] Huberty, C.J. – Olejnik, S. *Applied MANOVA and Discriminant Analysis*. John Wiley & Sons Inc., New York, 2006.
[3] Ong, M. K. *Credit Ratings. Methodologies, Rationale and Default Risk*. Risk Books, 2002.
[4] Rencher, A.C. *Methods of Multivariate Analysis*. John Wiley & Sons Inc., New York, 2002.
[5] Servigny, A. – Renault, O. *Measuring and Managing Credit Risk*. McGraw-Hill Professional, 2004.
[6] Tabachnik, B.G. – Fidell, L.S.: *Using Multivariate Statistics*. Pearson Education, Inc., 2011.

Were Stocks during the Crisis More Jumpy?: A Comparative Study

Jan Novotný[1]

Abstract. I empirically analyze the price jump behavior of the most traded US stocks during the recent financial crisis. I test the hypothesis that the current financial turmoil caused no change in the price jump behavior. I employ data on realized trades for 16 stocks and one ETF from the NYSE TAQ database. These data come at 1-minute frequency and span the period from January 2008 to the end of July 2009. The recent financial crisis is generally understood to start with the plunge of Lehman Brothers shares on September 9, 2008, which lies in the middle of the data set. I employ five model-independent and three model-dependent price jump indicators to robustly assess the price jump behavior. Although the results confirm an increase in overall volatility during the recent financial crisis; the results cannot reject the hypothesis that there was no change in price jump behavior in the data during the financial crisis. This implies that the uncertainty during the crisis was scaled up but the structure of the uncertainty seems to be the same.

Keywords: financial markets, price jumps, extreme price movements, financial crisis.

JEL Classification: G01, P59.
AMS Classification: 62M10.

1 Motivation

Financial markets are uncertain even where there is no crisis. Uncertainty means that when we observe the price process for any financial instrument, we see that the price process follows a stochastic-like path. This path can be with or without a deterministic drift; however, the price process is in any case smeared by noise movements. The noise movements, known as market volatility, make the price unpredictable. However, the unpredictability of the price movements is not a priori a negative feature; it is rather the nature of financial markets since many different interests meet there. Unpredictability, though, can carry important information when the markets are working properly and no one has an inappropriate informative advantage. Thus, it is of great interest to describe the noise movements as accurately as possible [5]. Such a description can then be used both in the financial industry to minimize risk and in theoretical economics, where various models of financial behavior are proposed [3,10,13]. In addition, a deeper empirical understanding of market volatility during the recent financial crisis can shed some light on the crisis itself and thus helps to deal with future crises. In this work, I contribute to this field by studying the behavior of the extreme noise movements of high-frequency stock returns during recent financial crisis.

The goal of this paper is to explicitly answer two questions. First, I ask whether an overall increase in market volatility during the recent financial crisis occurred. Second, I focus on the part corresponding to price jump volatility and ask whether the behavior of price jumps changed during the recent financial crisis. To answer these questions, I employ 16 highly traded stocks and one Exchange Traded Fund (ETF) from the North American exchanges found in the TAQ database. These highly traded stocks represent a significant portion of the traded financial assets. Data from the TAQ database are originally at the tick level; thus, I have to integrate them to a 1-minute frequency. The data set spans from January 2008 to July 2009. It is found that the overall volatility significantly increased in September 2008 when Lehman Brothers filed for Chapter 11 bankruptcy protection.

First of all, the results support the claim that volatility increased during the financial crisis. The volatility soars after the Lehman Brothers problems were announced and the peak lasts until mid-October. Then the volatility decreases but keeps above its pre-crisis level. In the first two months of 2009, the volatility increases again, mainly for the banking industry. The increased levels of volatility are in agreement with general knowledge since they reflect the increase in overall market impatience. The results, however, do not show an increase in price jumps. An overall increase in price jumps would mean a higher rate of market panic and more irrational behavior. A rather stable rate of price jumps, on the other hand, suggests that the proportion of market panic with respect to general impatience remained the same. However, there are some individual cases where the rate of price

[1] CERGE-EI, Charles University and the Academy of Sciences, Prague, Czech Republic; Politickych veznu 7, 111 21 Prague, Czech Republic. jnovotny@cerge-ei.cz.

jumps increased and decreased during the crisis. In addition, it is not possible to draw any industry-dependent conclusions, which is surprising for the banking industry.

Finally, this paper also proves that different price jump indicators measure price jumps very differently. The difference in sensitivity between the indicators, however, is not so easy to describe; this would require a detailed numerical analysis. Such an analysis would be worth, while paying since the exact quantitative connection between the various price jump indicators would enable us to perform a meta-analysis of the results from various papers that use different indicators. The synergy obtained from such a study would draw more complex picture about the market mechanisms governing the spread of information. Such mechanisms play key role when market panic is formed. In addition, this would enable us to better quantitatively describe the irrational behavior of financial markets and thus, hopefully, understand them more deeply.

2 Methodology

In this section, I provide the reader all the ingredients necessary to answer the question about the change in the behavior during the financial crisis. For the sake of the space requirements, I omit formulas, which can be found either in the works I refer to or in the extended version of this manuscript [12].

2.1 Price Jump Indicators

Price jumps are generally understood as abrupt price movements relative to the past behavior. Such a definition is rather vague and thus literature offers us a broad range of explicit price jump indicators. These indicators have different underlying assumptions and approach the art of price jumps from different perspectives. I overcome this problem by choosing a broad range of price jump indicators to obtain robust results. Namely, I consider these indicators:

1. Model independent indicators:
   a) Extreme returns,
   b) Temperature,
   c) p -dependent Realized Volatility,
   d) The Price Jump Index,
   e) The Wavelet Filter,
2. Model dependent indicators:
   a) The Difference between Bi-power Variance and Standard Deviation -- The Differential Approach,
   b) The Difference between Bi-power Variance and Standard Deviation -- The Integral Approach,
   c) Bi-power Test Statistics.

The price jump indicators can be classified into two groups: Model-independent indicators and model-dependent indicators. Model-independent indicators do not assume any price generating process and rather focus on the intuitive properties of price jumps and identify them based on that. In general, there are two types of indicators: one type of indicators is identifying jumps one by one, while the other type of indicators estimates the jumpiness of the time series over a certain period, or, the propensity of a time series to be jumpy.

The model-independent indicators are: the indicator based on extreme returns is advocated in [7] and indicators are defined as those returns, which exceeds a given centile over a certain period. Then, I use temperature based indicators to assess the jumpiness of the time series based on [9]. *p*-dependent Realized Volatility as advocated for example by [4] is another indicator estimating the jumpiness of the data over a certain period of time. This indicator uses different L-measures of a realized volatility and employs its ratio to estimate the relative contribution of extreme price movements. The price jump index [8] represents fourth indicator, which assess jumpiness of the data based on the relative ratio of recent price movements and prevailing realized volatility. Finally, I employ wavelet filter [6] to filter out the fast price changes and estimate their relative contribution to the overall price process.

The model-dependent indicators, as the name suggests, assume some underlying model for the price generating process. Having in hand such an underlying process, one has to estimate the parameters of the underlying model. First indicator from this class is the one based on the difference between the bi-power variance and standard deviation, as advocated by [14], where I employ it in an integral way. It means that I calculate a price jump measure as a cumulative measure of the ratio of price jumps to non-price jump price movements. Second indicator represents a modification of the previous one, where I employ it in a differential way. Such an indicator was used by [2]. Finally, I use indicator of individual price jumps based on the bi-power statistics as was developed by [1,11].

In the recent paper [7], the authors performed a detailed simulation study, where they compare 14 different price jump indicators with respect to two criteria: false positive and false negative misidentification. The results show that the two best performing indicators are Extreme returns (1a) and Bi-power Test Statistics (2c), therefore my analysis is robust with respect to both possible misspecifications.

2.2 Definition of the Financial Crisis

The main scope of this paper is to study, whether the current financial turmoil caused any change in the price jump behavior in the financial markets using high-frequency data. The period of the financial crisis is understood rather intuitively; however, I need explicit definition to carry out the statistical tests. Therefore, I consider the plunge of shares of Lehman Brothers on September 9, 2008 as a main event, which triggered panic all over the financial world. Based on this event, I define financial crisis as a structural break in the economy, which separates different periods. In addition, I consider two options, permanent break and temporary break:

• The Permanent Break (PB): The crisis started on September 9, 2008 and lasted until the end of the sample.

• The Temporary Break (TB): The crisis started on September 9, 2008 and lasted 30 trading days or for one-and-a-half months.

The first scheme is intuitive and is based on the fact that the crisis started with the problems of Lehman Brothers, and the effect of the crisis was permanently present on financial markets at least until the end of July 2009. The second scheme, however, focuses solely on the most problematic days following the plunge of shares. The period of 30 working days was chosen based on the news and the behavior of financial markets. The two schemes thus provide different pictures: The first scheme answers the question about permanent change in the behavior of financial markets, while the second scheme rather focuses on the immediate panic, which spread through the financial markets and affected the trading habits of market participants.

2.3 Hypotheses to Test

The indicators, employed in this work, measure the jumpiness of the financial markets or number of price jumps on a daily basis. The indicators can be divided into two groups according to the way how the daily measure is achieved. First group are the indicators, which by construction estimate one number for every day. Second group of indicators gives an estimate of jumpiness for every tick. Then, the measure of the jumpiness per a given day is obtained based on these tick estimates. These two different groups of indicators also imply the different hypotheses to test with different meaning.

Thus, I form the four different hypotheses, two for each of the two groups of price jump indicators:

Group I: One Number per Trading Day – A

The first group of indicators gives exactly one number per trading day. I divide the sample of trading days into two sub-samples. These two sub-samples correspond to periods without the crisis and period with the financial crisis, where the period of the financial crisis is defined above.

First, I employ the two-sample Wilcoxon test and test whether the estimated price jump measures for the two sub-samples come from the same distribution. The null hypothesis of this test says that the two sub-samples come from the same distribution. The main scope of this test is compare whether the estimated price jump measures changed during the crisis.

Group I: One Number per Trading Day – B

Second, I employ the standard F-test and compare whether the variance of the estimated price jump measures changed during the crisis. The F-test is defined in a standard way with standard deviation corresponding to the two sub-samples. The null hypothesis states, in this case, that the variance of the two sub-samples, i.e., during and outside the crisis, is the same. This test says whether the trading days in either of the two sub-samples were on average more heterogeneous. In other words, this procedure tests the heterogeneity of the trading days between the sub-samples.

Group II: One Number per Tick – A

The second group of price jump indicators gives one number per every tick, in my case one number per minute. Having in hand these numbers, I calculate mean and variance of these numbers per every trading day. Analogously to the previous case, I divide the sample into two sub-samples.

First, I employ again the two-sample Wilcoxon test and test whether the daily means of the estimated price jump measures for the two sub-samples come from the same distribution. The null hypothesis of this test says that the means of the two sub-samples come from the same distribution. The main scope of this test is to compare whether the estimated price jump measures changed during the crisis.

Group II: One Number per Tick – B

Second, I employ again the two-sample Wilcoxon test and test whether the daily variances of the estimated price jump measures for the two sub-samples come from the same distribution. The null hypothesis of this test says that the means of the two sub-samples come from the same distribution. The main scope of this test is to question whether the heterogeneity inside the trading days changed during the financial crisis.

3 Data

I employ a set of 16 stocks and one ETF from the Trade and Quote Database (TAQ) established by NYSE (NYSE, n.d.). Data in my work ranges from Jan-01-2008 to Jul-31-2009. The selected time span covers the critical period of the financial crisis, whose peak was on September 2008. Namely, I use: Apple Inc., Bank of America Corp., Citigroup, Inc. , Chevron Corp., General Electric Co., Google Inc., Hewlett-Packard, Intl. Business Machines Corp., Johnson & Johnson, Coca Cola Company, Microsoft Corp., Pfizer, Procter & Gamble, S&P 500 ETF, AT&T Inc., Wells Fargo & Co. and Exxon Mobil Corp.

All these companies are heavily traded and thus allow me to get enough statistics for every minute. In addition to heavily traded companies, I employ one ETF asset, which mimics the S&P 500 index. Such an asset represents the performance of the entire market and jumps at this particular vehicle correspond to situation, when the entire economy undergoes shocks.

4 Results

The full version of results can be found in [12]. I present results for two price jump indicators found in [7]. First, I define price jump indicator based on the absolute returns. The indicator defines a jump as those returns, for which their absolute value exceeds the given centile. Namely, I calculate 95- and 99-centiles for all absolute returns over the entire sample and use these values as thresholds. The number of returns is depicted in Figure 1.

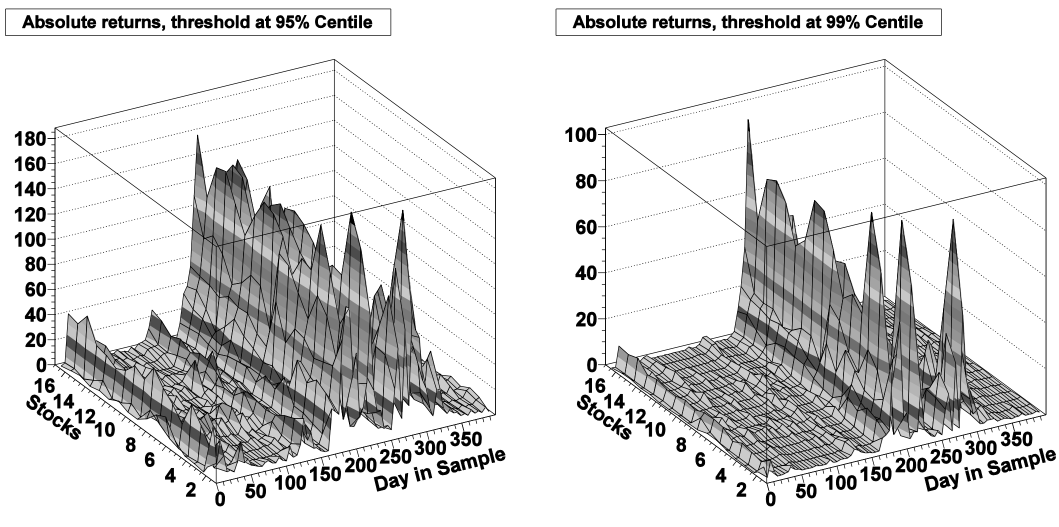


Figure 1 Number of price jumps per day using price jump index based on the extreme returns.

Results suggest that during the period of financial crisis, there were more price jumps defined in this way. However, such a conclusion is too premature since financial crisis was generally characterized by the overall increase in volatility and this can result in the detection of more price jumps compared to out-of-the crisis period.

Then, I define price jump indicator based on the bi-power statistics. The statistics compares the current price movement with the bi-power variance calculated over a moving window of length *n.* I use *n=120* time steps back, or, 2-hours moving window. The advantage of bi-power variance lies in the fact that it is not sensitive to abrupt price movements since it does not consider square of returns but rather a product of two consecutive returns in absolute value. Having in hand such a statistics, we can test for every time step the null hypothesis that there is no price jump. Namely, we have to apply the Group II hypotheses. Results of the two Hypotheses are in Table 1 and Table 2.

ID	Permanent	Temporary	ID	Permanent	Temporary
APPL	-0.141	-0.156	KO	-1.869*	1.948*
BAC	5.416***	0.298	MSFT	-0.405	-0.702
C	3.312***	-1.446	PFE	1.668*	0.280
CVX	0.004	-0.045	PG	-2.674***	-2.289**
GE	6.082***	0.567	SPY	-0.128	-0.781
GOOG	-2.644***	2.016**	T	0.240	-2.186**
HPQ	2.143**	-1.991**	WFC	4.786***	0.032
IBM	1.656*	-1.418	XOM	2.308**	-0.848
JNJ	-1.337	-0.566			

Table 1 The two-sided Wilcoxon statistics. The additional stars denotes, whether we can reject null hypothesis of the two samples to match and the corresponding confidence level: 90% (*), 95% (**) and 99% (***). The overall positive/negative value of the z -statistics suggests that the median of the means is lower/higher during the financial crisis.

ID	Permanent	Temporary	ID	Permanent	Temporary
APPL	1.281*	1.855**	KO	0.975	1.506*
BAC	0.997	2.245***	MSFT	0.832	1.091
C	0.890	1.619**	PFE	0.905	1.231
CVX	0.765*	0.698	PG	1.036	1.103
GE	1.054	1.213	SPY	0.946	1.514*
GOOG	0.933	0.901	T	0.759*	1.107
HPQ	1.083	1.557*	WFC	1.034	1.169
IBM	1.048	1.510*	XOM	1.112	0.735
JNJ	1.028	1.049			

Table 2 Results of the two-sided F-test for the variance. The null hypothesis says that the variances during and outside the crisis match. Stars denote, at what confidence level we can reject the null hypothesis: 90% (*), 95% (**) or 99% (***). In addition, the value of F-statistics higher/lower than one means that variance of the characteristic coefficient during the crisis was higher/lower when compared to the period outside the crisis.

In the case of the Permanent Break, there are several cases where the number of price jumps differs during the crisis. All the banks, GE, and Exxon Mobile are characterized by positive z-values and thus by the lower number of price jumps during the crisis. On the other hand, Google and Procter and Gamble show a higher number of price jumps. In addition, Procter and Gamble is the only one which shows the same change of price jumps also for the Temporary Break. This suggests that the short period immediately after Lehman Brothers' problems was dominated by a huge increase in price jumps. In the case of the remaining stocks, there are no agreements between the different number of price jumps using the Permanent Break and the different number of price jumps using the Temporary Break. This means that the main change in the number of price jumps occurred in the long-time horizon.

In the case of the Permanent Break, the variance in the number of price jumps is not present. On the other hand in the case of the Temporary Break, the difference in the variance is present, namely for Bank of America where the F -statistics are higher than one. This suggests that the variance was higher during the crisis, i.e., the days were very different during the crisis than they were outside the crisis.

5 Conclusion

I performed a detailed technical analysis of price jumps using the high-frequency market data (at a 1-minute frequency) covering 16 major traded stocks and one ETF traded on the main North American stock exchanges. The data spans the period from the beginning of 2009 until the end of July 2009. I answered the main question of this paper whether the behavior of price jumps, understood as extreme and irregular price movements different by their nature from the regular Gaussian noise, changed during the recent financial crisis.

I have employed a broad range of both model-dependent and model-independent indicators to assess the jumpiness of the data day by day. Then, I have defined a financial crisis as a structural break: First, as a permanent break starting at the day when Lehman Brothers' shares plunged, and, second, as a temporary break starting at the same day as above and lasting 30 trading days. The results supports the general understanding that volatility significantly increased after the Lehman Brothers problems and increased uncertainty remained on the market for next month. Then, situation calmed down but it did not return to the pre-critical levels. The results, however, do not confirm the change in the price jump behavior for all the stocks during the crisis. A naïve intuition, which suggests that price were more jumpy, or irrational, does not to be correct. Results rather suggest that the entire price process scaled up; however, the rate of extreme price movements remained the same. Finally, the study also suggests that different price jump indicators estimates price jump differently.

Acknowledgements

I would like to thank to Jan Hanousek, Evzen Kocenda, Randall Filer and Jan Kmenta for their numerous comments and suggestions. This study was supported by GACR Grant 402/08/1376 and by GAUK Grant 271111.

References

[1] Andersen, T. G., Bollerslev, T. and Dobrev, D., 2007. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and i.i.d. noise: theory and testable distributional implications. *NBER Working Paper* No 12963.

[2] Barndorff-Nielsen, O. E. and Shephard, N., 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, Vol. 4, pp. 1– 48.

[3] Bouchaud, J-P., Kockelkoren, J. and Potters, M., 2004. Random walks, liquidity molasses and critical response in financial markets. *Science & Finance (CFM) working paper archive* 500063, Science & Finance, Capital Fund Management.

[4] Dacorogna, Michel M., 2001. *An introduction to high-frequency finance*. Academic Press, San Diego.

[5] Gatheral, J., 2006. *Volatility surface: A practitioner's guide*. John Wiley & Sons, Inc., New Jersey.

[6] Gencay, R., Selcuk, F. and Whitcher, B., 2002. *An introduction to wavelets and other filtering methods in finance and economics*. Elsevier, San Diego.

[7] Hanousek, J., Kocenda, E., and Novotny, J., 2011. The Identification of Price Jumps, *CERGE-EI Working Paper Series*, 2011, No. 434, 48 pages.

[8] Joulin, A., Lefevre, A., Grunberg, D. and Bouchaud, J.-P., 2008. Stock price jumps: news and volume play a minor role. *http://arxiv.org/abs/0803.1769*.

[9] Kleinert, H., 2009. *Path integrals in quantum mechanics, statistics, polymer physics, and financial markets*, 5th Edition. World Scientific, Berlin.

[10] Lahaye, J., Laurent, S. and Neely, C. J., 2009. Jumps, cojumps and macro announcements. Working *Paper of Federal Reserve Bank of St. Louis*, No 2007-032, Revised Version.

[11] Lee, S. S. and Mykland, P. A., 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies,* Vol. 21, No. 6, pp. 2535 – 2563.

[12] Novotny, J., 2010. Were Stocks during the Crisis More Jumpy?: A Comparative Study. *CERGE-EI Working Paper Series*, No. 416, 57 pages.

[13] Merton, R. C., 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, Vol. 3, No. 1-2, pp. 125 – 144.

[14] Pirino, D., 2009. Jump detection and long range dependence. *Physica A*, Vol. 388, pp. 1150 – 1156.

A Note to Vehicle Scheduling with Several Bus Types

Stanislav Palúch[1]

Abstract. The paper studies one type of vehicle scheduling problem with several bus types. Suppose we are given a set of bus trips and a set of buses – bus fleet. The given bus fleet contains several types of vehicles. A subset of appropriate bus types is given for every trip. This paper studies the problem how to find a bus schedule with minimum number of used vehicles.

Keywords: vehicle scheduling, heterogeneous bus fleet, mathematical programming, graph theory, optimization

JEL classification: C61
AMS classification: 68R10, 05C85

1 Introduction

A **bus trip** (sometimes called a journey) is the fundamental notion of vehicle scheduling theory. Bus trip is one move of a bus from a starting bus stop to a finishing bus stop.

Bus trip is determined by geografic location and time position. Geografic location of a trip is a route with starting and finishin bus stop. Several additional bus stops can occur along trip route, however these stops are not important for bus scheduling purposes.

A trip j is defined by four data:

- $t_d(j)$ – departure time of the trip j
- $t_a(j)$ – arrival time of the trip j
- $p_d(j)$ – departure bus stop of the trip j
- $p_a(j)$ – arrival bus stop of the trip j

Therefore we will consider trip j as a quadruple $(t_d(j), t_a(j), p_d(j), p_a(j))$.

Every bus transports takes place in a transportation network. Transportation networt is detemined by a set of nodes and by a set of edges. The nodes are models of street intersections, bus stops and other important places in considered region, the edges represent street segments between adjacent nodes. Two lengths are assigned to every edge: one – distance between adjacent nodes and the second – travel time along corresponding edge. Just mentioned data make it possible to compute space distance matrix and time distance matrix for all nodes of considered transportation network. Namely time distance matrix is important for bus scheduling purposes – we will denote time distance between two nodes u, v by $travel\_time(u, v)$.

Let us define a **precedence relation** $\prec$ on a set of trips

$$j_1 \prec j_2 \iff travel\_time(p_a(j_1), p_a(j_2)) \leq t_d(j_2) - t_a(j_1). \tag{1}$$

If $j_1 \prec j_2$ we will say that the trip j_1 **precedes** the trip j_2. It holds $j_1 \prec j_2$ if there exists enough time for a bus to transfer from arrival place of the trip j_1 to the departure place of the trip j_2 so that it arrives to $p_d(j_2)$ sufficiently early so that it can make the trip j_2.

[1]University of Žlina/Department of Mathematical Methods, Univerzitná 8215/1 01026 Žlina, Slovakia, e-mail: stanislav.paluch@fri.uniza.sk

A **running board** of a bus is a sequence of trips $j_1, j_2, \ldots, j_k$ such that $j_1 \prec j_2 \prec \cdots \prec j_k$. The linkage $j_i \prec j_{i+1}$ is penalized by a cost $c(j_i, j_{i+1})$ which can express dead mileage expenses, line change penalty, waiting time penalty etc. A **bus schedule** is a set of running boards.

Given a set of trips $\mathcal{S}$, we can formulate two fundamental vehicle scheduling problems:

VSP1: To arrange all trips from $\mathcal{S}$ into minimum number of running boards.

VSP2: To arrange all trips from $\mathcal{S}$ into minimum number of running boards with minimum total cost of all linkages.

Standard vehicle scheduling assumes that all buses are the same. Practical experiences show that bus providers use several types of buses with different size and capacity. In this case a set $F(j)$ of available vehicles is assigned to every trip j according to traffic demand (e.g. number of passengers requiring this trip), according to parameters of trip route (e.g. narrow street segments eliminate use of large vehicles) and according to possibility and/or necessity to provide trips with certain bus type (e.g. trips for disabled people).

Scheduling taking into account these additional constraints will be called a **Vehicle Scheduling with Heterogeneous Bus Fleet** – VSHBF problem. Two bus type instance of VSHBF is studied in [1].

Standard vehicle scheduling problem can be transformed to an assigning problem and therefore we have a polynomial complexity algorithm for it. However, additional constraints make VSHBF problem hard.

There are many additional constraints imposed on running boards e.g. every running board has to contain a safety break and meal break, driver working time is limited from abow and sometimes from below etc. These constraints depend on legislation of corresponding country, on the way of driver duties scheduling, regional traditions and can even vary from bus provider to bus provider. There are even more general scheduling problems as scheduling with flexible trips – see [5] or [6]

We have developed a sophisticated method (see [3], [4]) which starts with a bus schedule which is optimal with respect to the number of vehicles but need not minimize total cost not to be feasible with respect to additional constraints. This neighborhood search procedure outputs a bus schedule which keeps the number of used vehicles and which is suboptimal with total complex cost expressing all desired objectives and constraints. Unfortunately, this method is not able to lower the number of used vehicles. Therefore it is necessary that it starts with a solution having, if possible, minimum number of used vehicles.

2 Bus number minimization with homogeneous bus fleet

2.1 Graph formulation and algorithm for VSP1

Let S be a set of trips. **Trip digraph** $G_S(V, E)$ of S is a digraph with the vertex set $V = S$ and with the edge set $E = \{(i, j) | i, j \in S,\ i \prec j\}$. The set E contains all ordered pairs (i, j) of trips such that trip i precedes trip j.

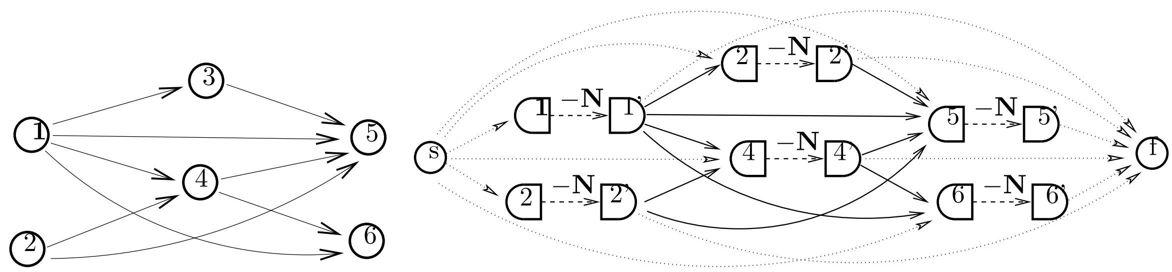


Figure 1: Trip digraph and extended trip digraph.

Extended trip digraph of the set S is a digraph $G_S^A = (V^A, E^A)$, which has originated from trip digraph G_S by

1. splitting every vertex i (i. e. every trip) i into two parts – departure part i and arrival part i' and adding directed arc (i, i') from departure to arrival part with large negative edge weight $-N$ (dashed arcs in fig. 1).
2. adding to fictive vertices s, f and all edges of the type (s, i) and all edges of the type (i', f) with zero ege weight (dotted lines in fig. 1).
3. attaching zero weight to all arcs of the type (i', j), $i, j \in V$.

Both digraphs G_S and G_S^A are acyclic. Every path in G_S is a feasible running board. The same holds for every s-f path in G_S^A.

Hence the problem VSP1 – to arrange all trips from S into minimum number of running boards – can be solved in corresponding trip digraph G_S as to cover all vertices of G_S with minimum number of disjoint paths, what is equivalent as to cover all vertices of G_S^A with minimum number of disjoint s-f paths.

A **semipath** in a digraph is a path in which edges can be used in right and in reverse direction. The **length of a semipath** is the sum of costs of edges used in right direction minus the sum of costs of edges used in reverse direction.

Algorithm 1. Covering all vertices of G_S^A with minimum number of disjoint s-f paths

Step1: Find a shortest (s, f) - path in G_S^A. Mark the arcs of that path as used, all other arcs as unused.

Step2: While the set contains an unused arc of the type (i', i) do:
Find a shortest (s, f) - semipathpath in G_S^A.
Mark arcs with right direction of that path as used.
Mark arcs with reverse direction of that path as unused.

Step3: Used arcs of the type (i', j) define trip linkages from what corresponding running boards can be easily constructed.

The resulting bus schedule is optimal with respect to the number of used vehicles and can be used as a starting solution for further cost optimizing process.

2.2 Vehicle minimization as an attachment problem

Let x_{ij} be a binary decision variable, $x_{ij} = 1$ if trip j is immediately linked behind trip i in a running board of a bus, $x_{ij} = 0$ otherwise. Let c_{ij} be a constant, $c_{ij} = 1$ if $i \prec j$, $c_{ij} = 0$ otherwise.
VSP1 is equivalent to the following problem:

$$\text{Maximize} \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij} \tag{2}$$

$$\text{subject to} \quad \sum_{i=1}^{n} x_{ij} \leq 1 \qquad j = 1, 2, \dots n \tag{3}$$

$$\sum_{i=j}^{n} x_{ij} \leq 1 \qquad i = 1, 2, \dots n \tag{4}$$

$$x_{ij} \in \{0, 1\} \tag{5}$$

2.3 Graph coloring formulation and algorithm for VSP1

Let S be a set of trips. **Collision graph** $G_C = (V, H)$ of S is a graph, with the vertex set $V = S$ and with the edge set

$$H = \{\{i, j\} | i, j \in V, i \not\prec j \text{ and } j \not\prec i\}$$

The set H is the set of such unordered pairs of trips which cannot be serviced by one bus simultanously. Such trips are called **incompatible trips**. Every independent set in G_C represents a feasible running board and vice versa. Hence to find an optimum solution of VSP1 means to find an optimum vertex coloring of G_C.

Let x_{ij} be a binary decision variable, $x_{ij} = 1$ if and only if the trip i is in the running board j, otherwise $x_{ij} = 0$. Denote by n the number of trips in S, let m be an upper bound of number of buses. Then VSP1 can be formulated as follows:

$$\text{Minimize} \sum_{i=1}^{n} \sum_{j=1}^{m} j x_{ij} \tag{6}$$

$$\text{subject to} \quad \sum_{j=1}^{m} x_{ij} = 1 \quad \text{for all } i \in V \tag{7}$$

$$x_{ik} + x_{jk} \leq 1 \quad \text{for all } i,\ j \text{ such that } \{i, j\} \in H \text{ and for } k = 1, 2, \dots m \tag{8}$$

$$x_{ij} \in \{0, 1\} \tag{9}$$

Just formulated model contains $k.|H|$ conditions (8). These constraints can be replaced by conditions

$$n x_{ik} + \sum_{j \in V(i)} x_{jk} \leq n \quad \text{for all } i \in V \text{ and for } k = 1, 2, \dots m \tag{10}$$

Let us note that the objective function (6) could be replaced by $\sum \sum x_{ij}$, however item $j x_{ij}$ in (6) will guarrantee that the resulting set of running boards is in the form $\{1, 2, \dots, w\}$ – i.e running boards with lowest numbers. Constraints (7) say that every trip is exactly in one bus running board. Constraints (8) say that no pair of incompatible trips can be in the same running board. Every constraint of the type (10) says that trip i can not be in the same bus running board with any other incompatible trip.

This formulation converts polynomial problem VSP1 into NP-hard graph coloring problem therefore it has no practical meaning. Hovewer, it has theoretical importance to Vehicle Scheduling with Heterogeneous Bus Fleet - VSHBF problem which is by our conjecture a hard problem. For more information about graph coloring formulation see [2].

3 Bus number minimization with heterogeneous bus fleet

3.1 Graph coloring formulation for heterogeneous fleet case

Denote by $F(i)$ the set of vehicles which can provide the trip i. Let $V(i)$ be the set of all trips incompatible with trip i. Mathematical model for VSHBF can be obtained by replacing constraints (7) by constraints (12) which say that every trip i has to be serviced only by a feasible bus type.

$$\text{Minimize} \sum_{i=1}^{n} \sum_{j=1}^{m} j x_{ij} \tag{11}$$

$$\text{subject to} \quad \sum_{j \in F(i)} x_{ij} = 1 \quad \text{for all } i \in V \tag{12}$$

$$n x_{ik} + \sum_{j \in V(i)} x_{jk} \leq n \quad \text{for all } i \in V \text{ and for } k = 1, 2, \dots m \tag{13}$$

$$x_{ij} \in \{0, 1\} \tag{14}$$

3.2 Two bus type problem

In the two bus type problem we have two types of trips and two types of buses. The trips of the first type are crowded trips requiring service by high capacity buses of the first type like hinged buses (we will call them maxibuses). The rest of trips are ordinary trips of the second type requiring ordinary buses. Ordinary trip can be serviced by maxibus too, but this is not a desirable instance and should occur only if necessary.

The simplest attitude to this problem is to decompose it into two independent scheduling problems - one for crowded trips and maxibuses and one for ordinary trips and ordinary buses. However, this attitude need not to be optimal since maxibuses can service several ordinary trips what can decrease the number of ordinary buses. Nevertheless just mentioned decomposition gives us the exact number of necessary maxibuses and a upper bound of ordinary buses.

Let us partition the set of trips S into two subsets - the first is the set of so called must-trips and the second is the set of so called may-trips. Algorithm 1 can be modified in order to give a bus schedule with minimum number of vehicles containing all must-trips and maximum possible number of may-trips.

Here is the following modification (L is again a large number):

Algorithm 2. Modification of Algorithm 1 for must- and may- trips.

1. For all arcs of the type (i, i') (dashed lines in fig.1) set edge weight to $-L^2$ if i is a must-trip. For may-trips set edge widht equal to $-L$.
2. Modify stop condition in Step2 of Algorithm 1 as follows: While the set E contains an unused edge of the type (i, i') where i is a must-trip do:

Several may-trips remain not scheduled after finishing Algorithm 2.

Now we are prepared to formulate an algorithm for exact minimization of maxibuses and suboptimal minimization of ordinary buses.

Algorithm 3. Suboptimal algorithm for two bus type problem

Let S be the set of all trips, denote by $\mathcal{C}$ the set of all crowded trips and denote by $\mathcal{O}$ the set of all ordinary trips. Clearly $S = \mathcal{C} \cup \mathcal{O}$.

Step1: Declare all crowded trips from the set as must-trips and all other trips as may-trips.
Run Algorithm 2.

Commentary: The result is the minimum cardinality set of running boards containing all crowded trips from $\mathcal{C}$ and as much as possible ordinary trips from $\mathcal{O}$ which do not increase the number of maxibuses.
Many ordinary trips remian still unscheduled. Denote by $\mathcal{O}_1$ the set of scheduled ordinary trips, let $\mathcal{O}_2$ be the set of unscheduled ordinary trips in this Step1.

Step2: Declare all unscheduled trips from the set $\mathcal{O}_2$ as must-trips and all trips from the set $\mathcal{O}_1$ as may-trips.
Run Algorithm 2.
Commentary: The result is the minimum cardinality set of running boards for all ordinary trips unscheduled in Step1 and as much as possible ordinary trips scheduled in Step1. Denote by $\mathcal{O}_3$ the set of scheduled ordinary trips in this step and denote by $\mathcal{O}_4$ the set of unscheduled ordinary trips. Clearly $\mathcal{O}_4 \subseteq \mathcal{O}_1$ and therefore trips from $\mathcal{O}_4$ do not increase the number of maxibuses.

Step3: Run Algorithm 1 for all crowded trips from the set $\mathcal{C}$ and for all unscheduled ordinary trips from the Step2 - i.e. for all trips from the set $\mathcal{O}_4$.
Commentary: The result is the set of running boards for maxibuses containing all crowded trips and all trips unscheduled in Step2.

Step4: The union of running boards from Step3 and Step4 is the resulting suboptimal bus schedule for the given set of trips S.

This algorithm was used for many real world computations with great success. However, the result of this algorithm is only the first step of optimization procedure for practical use. It is used as the starting solution for the second optimization step with more complex objective function and is imbedded into vehicle and crew optimizing system KASTOR. System KASTOR was successfully used in municipal and regional bus transport of following towns: Tachov, Uherské Hradiště, Most and Litvínov, Prachatice, Strakonice, Nymburk, Lysá n/L., Milovice, Prievidza, Martin - Vrútky, Piešťany, Považská Bystrica, Trenčín and others. Optimization savings ranged from 5% to 20% of dead mileage and number of used vehicles.

Unfortunately, several cases occurred when Algorithm 3 gave more ordinary buses than the exact minimum. Therefore we proposed the following procedure:

- Run Algorithm 1 for all crowded trips from S.
  The result is the exact minimum number of maxibuses n_M.
- Run Algorithm 1 for all trips from S regardless of the trip and bus type.
  The result is a lower bound $LB(n_{ALL})$ of all buses.

  Since n_M is exact minimum of maxibuses, we have a lower bound $LB(n_{\mathcal{O}}) = LB(n_{ALL}) - n_M$ of ordinary buses.
- Run Algorithm 1 for all ordinary trips.
  The result is a upper bound $UB(n_{\mathcal{O}})$ of ordinary buses in two bus type schedule.
  Upper bound of all buses is $UB(n_{ALL}) = UB(n_{\mathcal{O}}) + n_m$.

The values n_M, $UB(n_{\mathcal{O}})$ can be used to reduce the size of graph coloring model (11) – (14) by reducing the sizes of sets F(i).

The lower bound can be used in the following way. If the degree of an ordinary trip i (i.e. the number of incompatible trips with the trip i) is less than $LB(n_{ALL})$ it can be colored by one of colors from the set $\{1, 2, \ldots, LB(n_{ALL})\}$ regardless of coloring of its neighbors. Therefore such trip can be removed from the graph G_C. A sequence of such removals can reduce the problem size significantly.

Similarly, a crowded trip can be removed from the graph G_C if its degree is less than n_M. In practice $n_M << LB(n_{ALL})$ that's why such reduction will be probably negligible.

Acknowledgements

This research was supported by grant VEGA 1/0374/11 – Modelling and optimization of mobility and infrastructure in logistic networks.

References

[1] Czimmermann, P., Peško, Š.: Autobusové rozvrhy s dvoma typmi vozidiel – Bus Schedules with Two Types of Vehicles, *3-rd International Conference APLIMAT 2004*, pp. 321-326

[2] Palúch, S.: Bus Scheduling as a Graph Coloring Problem, *Komunikcie/Communications*, vol.**4**, 2003

[3] Palúch, S.: Graph Theory Approach to Bus Scheduling Problem, *Studies of the faculty of management science and informatics*, vol.**9**, October, 2001, pp. 53-57

[4] Palúch, S.: Simultaneous Bus and Crew Scheduling. *International Scientific Conference on Mathematics, Žilina, June 30. - July 3.*, 1998

[5] Peško, Š.: Flexible Bus Scheduling with Overlapping Trips, In: *Quantitative Methods in Economics (Multiple Criteria Decision Making)*, June 5-7, 2008

[6] Peško, Š.: The Minimum Fleet Size Problem for Flexible Bus Scheduling, Studies of the faculty of management science and informatics, Vol. 9, October, 2001, pp. 59-65

[7] Plesník, J.: Grafové algoritmy - Graph Algorithms, VEDA, Bratislava, 1983

Q-model of investment

Václava Pánková[1]

Abstract. Most part of econometric models of investment is based on neoclassical theory of factor demand. *Q*-model of investment is derived using first order optimality conditions of the optimizing problem to maximize value of the firm subject to usual capital movement equation. Marginal *Q* defined as a ratio of shadow value to purchase cost is a theoretical concept which in practice is not observable and also it is hardly to be computed on the base of the definition. As a result of computations, an average *Q* can be found using a VAR model and panel data of firms. The logic of construction of the model which reflects the fact that only a small part of firms is quoted on stock markets, is described and an application to a sub-sample of firms selected from the official industry segment "Steam and Hot Water Supply Industry" of the Czech Republic is given. The results show that only a small part of them exhibit a chance to profit from further investment.

Keywords: optimal investment, Tobin´s *Q*, implementation of *Q*, panel data, VAR model

JEL Classification: C44, G31, C13
AMS Classification: 91B28, 91G70

1 Introduction

Alternative theories of investment differ in the way how is the desired value of capital modeled and how are the cost adjustments treated. A wide survey is given e.g. in [5]. None of the alternative approaches has proved its superiority over the others; each describes and accents only a part of a complex reality. In [4], an equivalency of the neoclassical and the Tobin´s theories is explained, including necessary assumptions. According to the Tobin's theory, investment depends on the ratio *Q* of the market value of business capital assets to their replacement value. To appreciate non-quoted firms, a construction based on forecasts from a VAR model can be useful as it shown e.g. in [2] or [7]. VAR models are usually used to describe long-run processes (e.g.[4]) but here a panel structure with dominating number of units will be estimated and then necessary magnitudes forecasted. An application concludes the text.

2 Theoretical concept

Most part of econometric models of investment is based on neoclassical theory of factor demand. The firm´s objective is to maximize the value of the equity owned by its shareholders. The optimization problem can be formulated as follows.

$$V_t(K_{t-1}) = max_{I_t,L_t,M_t}\{\Pi_t(K_t, L_t, M_t, I_t) + \beta_{t+1}E_t(V_{t+1}(K_t))\}$$

subject to

$$K_t^i = (1-\delta^i)K_{t-1}^i + I_t^i \quad \text{for} \quad i = 1, \dots, N$$

where V_t is the maximized value of the firm in period *t*, $\Pi_t(K_t, L_t, M_t, I_t)$ is the firm´s net revenue function, $K_t = (K_t^1, \dots, K_t^N)$ is a vector of *N* types of capital inputs, $L_t = (L_t^1, \dots, L_t^R)$ is a vector of *R* types of labor inputs, $M_t = (M_t^1, \dots, M_t^S)$ is a vector of *S* types of current inputs, $I_t = (I_t^1, \dots, I_t^N)$ is a vector of gross investments in each type of capital, $\beta_{t+1} = (1 + \rho_{t+1})^{-1}$ is the firm´s discount factor with ρ_{t+1} being the risk-free rate of interest between periods *t* and *t*+1 and $E_t(V_{t+1}(K_t))$ denotes the expected value. δ^i is the rate of depreciation of capital of type *i*, assumed to be exogenous and fixed.

[1] University of Economics, Prague, Dept. of Econometrics, pankova@vse.cz

The solution of the optimization problem can be characterized by the first order conditions

$$-\left(\frac{\partial \Pi_t}{\partial I_t^i}\right) = \lambda_t^i \quad \text{for} \quad i = 1, \ldots, N \tag{1}$$

$$\lambda_t^i = \left(\frac{\partial \Pi_t}{\partial K_t^i}\right) + (1-\delta^i)\beta_{t+1}E_t(\lambda_{t+1}^i) \quad \text{for} \quad i = 1, \ldots, N \tag{2}$$

$$\left(\frac{\partial \Pi_t}{\partial L_t^i}\right) = 0 \quad \text{for} \quad i = 1, \ldots, R$$

$$\left(\frac{\partial \Pi_t}{\partial ML_t^i}\right) = 0 \quad \text{for} \quad i = 1, \ldots, S$$

where $\lambda_t^i = \frac{1}{1-\delta^i}\left(\frac{\partial V_t}{\partial K_{t-1}^i}\right)$ is the shadow value of inheriting one additional unit of capital of type *i* in period *t*. Without any adjustment costs

$$\Pi_t(K_t, L_t, M_t, I_t) = p_t F(K_t, L_t, M_t) - p_t^K I_t - w_t L_t - p_t^M M_t$$

with $F(K_t, L_t, M_t)$ being the production function, p_t the price of firm´s output, $p_t^K = (p_t^{K1}, \ldots, p_t^{KN})$ are prices of capital goods, $w_t = (w_t^1, \ldots, wL_t^R)$ wage rates for each type of labor, p_t^M=$(p_t^{M1}, \ldots, p_t^{MS})$ are prices of inputs. Including adjustment costs we have

$$\Pi_t(K_t, L_t, M_t, I_t) = p_t F(K_t, L_t, M_t) - G(K_t, L_t, H_t, I_t) - p_t^K I_t - w_t L_t - p_t^M M_t$$

with $H_t = (H_t^1, \ldots, H_t^R)$ being vector of hiring of each type of labor. We then obtain

$$\frac{\partial \Pi_t}{\partial I_t} = -p_t\left(\frac{\partial G}{\partial I_t}\right) - p_t^K$$

and

$$\frac{\partial G}{\partial I_t} = \left(\frac{\lambda_t}{p_t^K} - 1\right)\frac{p_t^K}{p_t} = (Q_t - 1)\frac{p_t^K}{p_t} \tag{3}$$

after substituting into (1).

The optimal capital stock is characterized by $Q_t = \frac{\lambda_t}{p_t^K}$, what is ratio of shadow value to purchase cost; such a Q_t is known as marginal one (for further inference see e.g. [5]).

Marginal *Q* is a theoretical concept which in practice is not observable and also it is hardly to be computed on the base of the definition. As a result of computations, an average *Q* can be found. Both *Q*´s are equivalent only under certain assumption precised e.g. by Hayashi, . [6].

3 Practical approach

The calculation of *Q* with a panel data set is proposed e.g. in [2] and deals with the formula

$$Q_{it} = \frac{V_{it} + D_{it} - N_{it}}{K_{it}} \quad . \tag{4}$$

Here, the *Q* is computed for firm *i* at period *t* using the market value of equity market V_{it} , value of outstanding debt D_{it} , remaining assets besides the capital stock N_{it} and replacement value of capital stock K_{it} .

As for the V_{it} , the firms quoted on the stock-markets are easy to be evaluated but the following approach enables to process also non-quoted firms. Taking the values of pre-tax profits *PTP*, sales *S* and cash flow *CF*, a VAR model with one lag can be formulated as

$$\begin{aligned} PTP_{it} &= \alpha_{10} + \alpha_{11}PTP_{it-1} + \alpha_{12}S_{it-1} + \alpha_{13}CF_{it-1} + u_{1it} \\ S_{it} &= \alpha_{20} + \alpha_{21}PTP_{it-1} + \alpha_{22}S_{it-1} + \alpha_{23}CF_{it-1} + u_{2it} \\ CF_{it} &= \alpha_{30} + \alpha_{31}PTP_{it-1} + \alpha_{32}S_{it-1} + \alpha_{33}CF_{it-1} + u_{3it} \end{aligned} \tag{5}$$

Analyzing a group of firms, we are using panel data comprising short time series of a representative number of firms. As for the estimate of the VAR model proposed, the dominant factor is the panel data structure. If there are the firms of different size comprised in the data set, the deviations from firms´ individual means can be used instead of levels. So, the equations do not contain firm specific effects. The VAR structure is very convenient to help to produce forecasts of

$$\hat{V}_{it} = \sum_{\tau=1}^{\infty} P\hat{T}P_{it+\tau}\delta_{i\tau}^{\tau} \quad \text{with} \quad \delta_{i\tau}^{\tau} = \frac{1}{(1+r_{t\tau})^{\tau}}, \; r \text{ being a market interest rate.} \tag{6}$$

Looking for a firm´s Q then the formula

$$Q_{it} = \frac{\hat{V}_{it} + D_{it} - N_{it}}{K_{it}} \tag{7}$$

instead of (4) is used. In practice, the question of the number of summands in (6) arises. A common practice is to deal according to an empirically verified rule that it is sufficient to consider four years as a relevant time horizon.

4 Application

The methodology described above will be now illustrated using a group of companies from an energetic industry.

In general, industry of energetic comprises firms of large and very large size as well as that of medium and small size.

The formers are first of all mines and power stations. In a country with ten millions inhabitants, what the Czech Republic is, there are only few such enterprises. They are quoted on stock markets and besides “too big to fail” in a political context of the economy. That is why the technique described above is not appropriate to be applied to them.

The letters are mostly firms of a regional importance. As their managements are not obliged to make economic information public, it is rather difficult to find relevant data. In the database Amadeus32 [1], seventeen medium size Czech firms from the segment “Steam and Hot Water Supply Industry” are comprised which have used to form the data set describing the period from 1997 to 2006.

As for the firms, the VAR model of type (5) was estimated as

$$\underset{}{P\hat{T}P_{it}} = \underset{(824)}{-814.4} + \underset{(0.114)}{0.425PTP_{it-1}} + \underset{(0.012)}{0.021S_{it-1}} + \underset{(0.087)}{0.096CF_{it-1}}$$

$$\hat{S}_{it} = \underset{(2377)}{5110.45} - \underset{(0.330)}{0.533PTP_{it-1}} + \underset{(0.034)}{0.888S_{it-1}} + \underset{(0.250)}{0.439CF_{it-1}}$$

$$C\hat{F}_{it} = \underset{(889.7)}{-213.7} - \underset{(0.123)}{0.141PTP_{it-1}} + \underset{(0.013)}{0.028S_{it-1}} + \underset{(0.093)}{0.802CF_{it-1}}$$

The Q´s were computed according to the equation (4). The results are summarized by the help of Table 1.

unit	1	2	3	4	5	6	7	8	9
Q	0.12	0.30	0.27	-0.32	0.00	5.30	-0.09	4.67	0.05
unit	10	11	12	13	14	15	16	17	
Q	0.10	0.08	0.02	0.23	0.17	1.80	0.21	0.74	

Table 1 Computed Q´s

An interpretation (see [3]) is that $Q > 1$ motivates to invest, $Q = 1$ is an equilibrium state when there is no incentive for the firm to invest, $Q < 1$ relates to an unprofitable environment.

It is evident, that the results are not very optimistic. Among the 17 units, only two exhibit their expected profit of a unit of capital to be higher than a unit of additional cost and hence, a motivation to a further investment.

5 Conclusions

Optimal investment strategy of 17 companies belonging to the industry segment "Steam and Hot Water Supply Industry" of the Czech Republic was studied. Following the concept of Tobin´s Q, only 3 companies among the 17 cases exhibit a chance to profit from further investment. The rest of firms refer to the case $Q < 1$, hence an unprofitable environment. A rising of prices of steam and hot water could be an evident and straightforward way how to change the situation. But the problem, of course has also a social dimension because among their customers there are not only firms but also the households. There is a common consensus that private ownership should be supported (private monopoly, in fact), with some price regulation and an explicit funding rule. If such a private monopoly goes bankrupt, one could auction its network and assets and never even interrupt service - ideally.

Acknowledgement

Financial support of GACR 402/09/0273 is gratefully acknowledged by the author.

References

[1] Amadeus32, database, Copyright © 2008 Bureau van Dijk Electronic Publishing

[2] Behr, A., Bellgardt, E.: Dynamic Q-investment functions for Germany using panel balance sheet data. Discussion paper 23/02, economic Research Centre of the Deutsche Bundesbank, 2002

[3] Berndt, E.: The Practice of Econometrics, Addison-Wesley, 1991

[4] Hančlová, J., Lukáčik, M., Szomolanyi, K.: A VEC model of Czech and Slovac economies. In: ed. Houda, Friebelová, Mathematical Methods in Economics, 2010, p. 208 – 213, ISBN: 978-80-7394-218-2

[5] *Handbook of Econometrics*, Volume 6, Part 1, 2007, Pages 4417-4498, Chapter 65: Microeconometric Models of Investment and Employment (Stephen Bond and John Van Reenen)

[6] Hayashi, F.: Tobin´s marginal Q and Average Q: A Neoclassical Interpretation, Econometrica, Vol. 50, No 1, 1982

[7] Pánková, V.: Tobinovo Q – teorie a aplikace. *Politická ekonomie*, 2005, roč. LIII, č. 5, p. 601–608. ISSN 0032-3233.

Fuzzy approach to modification by enlargement of supply chain based on logistic indicators dimensions

Martin Pech[1], Jaroslava Smolová[2]

Abstract. As globalization took place throughout the twentieth century supply chains became longer and more complex, a number of firms realized the potentials of supply chain in day-to-day operations management. The supply chain integrates, coordinates and controls the movement of goods, materials, and information form a supplier through a series of intermediate customers to the final consumer. The supply chain links all companies, which became members of this process. However, many relationships of the underlying system and the complexity of interactions within prevent from gather comprehensive and complete information's. These conditions under vagueness claim to fuzzy approach application that proposes a more robust and computational capable way of measurement in supply chains.
The main aim of the paper is modification model of supply chain by adding new companies with capability of bind to chain. The purpose of verification is to analyzing measures changes, which are obtained by evaluation of logistic indicators dimensions and non-probabilistic entropy. Joined companies that are potentially able to connect to hubs are chosen by cluster analysis results. Comparing is oriented to examination of differences between original supply chain model and modification. Based on conditions for selecting companies from each cluster, original supply chain dimensions measures should be changed according to orientation and characteristics of different clusters.

Keywords: fuzzy approach, supply chain, logistic indicators, entropy.

JEL Classification: C21, L14, L60
AMS Classification: 28D20, 03B52

1 Introduction

With the increasing complexity of supply chains, it becomes even more important to make the right decision about establishing a physical network infrastructure and optimize its performance [4]. The supply chain links all companies, which became members of this process. They integrate, coordinate and control the movement of goods, materials, and information form a supplier through a series of intermediate customers to the final consumer [2]. It becomes important that supply chain partners understand their individual strengths and sources of value creation. This enables a clear positioning of supply chain partners on global, regional and local playing fields [4].

Over the past years, most of existing supply chain management models are solely optimized by only one objective for a single firm, and lack the consideration of strategic partnership [8]. Evaluating strategic behavior of supply chain partners with a growing number of firms connected to network is going important. Strategic partnership lie in common supply chain strategy design, information sharing and balanced objectives and business functions set. The value of information sharing increases as the service level at the supplier, supplier-holding costs, demand variability and offset time increase, and as the length or the order cycle decreases [6]. The paper concerns with the uncompleted indicators information sharing problem.

2 Methodology

The methodology work on the assumption that soft measures have better capability to quantify imprecision and vagueness arising in the assessment respondents' opinions about relevance of some characteristic and hidden features in supply chains, specifically indicators information sharing. These conditions under vagueness claim to fuzzy approach application that proposes a more robust and computational capable way of measurement of information incompleteness gained by questionnaire research.

[1] University of South Bohemia in České Budějovice/Faculty of Economics, department of management, Studentská 13, 370 05 České Budějovice, mpechac@ef.jcu.cz.
[2] University of South Bohemia in České Budějovice/Faculty of Economics, department of management, Studentská 13, 370 05 České Budějovice, smoloj@ef.jcu.cz.

The main aim of the paper is modification of real supply chain by adding new companies with capability of bind to chain. Fuzzy approach to enlargement of supply chain can bring some relevant information in building “optimal networks”. Modification is analyzed by measures changes, which are obtained by evaluation of logistic indicators dimensions and its non-probabilistic entropy. Fuzzy approach is also proposed here as an alternative approach to handle effectively the impreciseness and uncertainty.

Special fuzzy decision methodology develops Wang & Shu [15], who provides an alternative framework to handle supply chain uncertainties and to determine supply chain inventory strategies. A fuzzy model based on possibility theory allows decision makers to express their risk attitudes and to analyze the trade-off between customer service level and inventory investment in the supply chain. Green supply chain (GSC) strategy ranking model is used by Chen, Ma [5] to integrate the fuzzy attribute values by the HWA operator that is applied to transform the fuzzy decision matrices regarding attribute values into a complex decision matrix.

Fuzzy approach is adopted from the research of Soyer, Kabak, Asan [12], elaborated and applied to logistic domain in our papers [9], [11] (see for more) and is defined and modified as follows:

Fuzzification is originated from scales used in questionnaire and maintains the scale’s basic characteristics. Fuzzy set *A* on a universal set *U* is uniquely determined by its membership function: $\mu_A: U \rightarrow [0,1]$. In this case are fuzzy sets, by means of fuzzy membership functions, used to represent successfully the vagueness inherent in indicators assessment. The construction of the fuzzy set membership of indicators is undertaken (linear membership functions are used). The parameters associated with the membership function are provided by expert judgments. The linear membership function of indicators: a) is continuous; b) it maps an interval [*a*, *e*] to [0, 1]; c) it is monotonically increasing; d) it isn’t restricted; e) is defined by the value *x* as average of all respondent judgments for a given indicator and parameters *a*, *b*, *c*, *d*, *e*.

$$\mu(x,a,b,c,d,e) = \begin{cases} 0, & a \le x \le b \\ \frac{1}{2}\cdot\left(\frac{x-b}{c-b}\right), & b < x < c \\ 0{,}5, & x = c \\ 1-\frac{1}{2}\cdot\left(\frac{d-x}{d-c}\right), & c < x < d \\ 1, & d \ge x \ge e \end{cases} \tag{1}$$

where parameter *a* is represent as minimum, *e* as maximum of relevant scale and *b*, *c*, *d* as facultative parameters. In case study is selected parameter *b* defined as *a* + 5% of scale length and *d* as *e* – 5 % of scale length and *c* = (*a* + *e*) / 2

Fuzzy evaluation of dimensions is based on the presence (measured by gradual membership to relevant dimension) of indicators in companies of researched supply chain. Each dimension (denoted by CS) is consisted of number indicators. While the presence of indicators is expressed by using of membership function, the fuzzy approach enables according to paper [12] to interpret the results in compliance with different decision levels (medium, high, very high) indicating gradual membership to dimensions.

$$\mu_{C-L}(x) = \sup_{r_1,\ldots,r_n \in CS} \left\{ \bigwedge_{k=1}^{n} \mu_v(x_{rk}) \right\} \tag{2}$$

where C-L = {medium, high, very high} denotes the dimension decision levels and CS the dimension set (we use five dimensions New supplier selection - N, Evaluation of suppliers - E, Storage - S, Customers – C, Transport - T); *n* represents the number of indicators that should exist for a given dimension and dimension decision level, and *x* is a vector of all average values of indicators. Results are given for all of 5 supply chain dimensions of companies connected in network.

Deffuzification (COA method is used) is used for consolidate fragmented decision levels for all decision levels that provides rank of measures supply chain dimension.

Non-probabilistic entropy, the common measure of fuzziness, is selected for quantified the vagueness of the dimension sets. Non-probabilistic entropy is based on De Luca, Termini special case in special circumstances [1], which is viewed as a degree of uncertainty inherent in each dimension set that we are exposed to in any judgment about these dimension set. De Luca, Termini are used Shannon´s function, and they defined a measure that became largest at the grade of membership of 1/2 [14]. This type of uncertainty differs from probabilistic uncertainty (randomness) and non-specificity in that it deals with situations where the boundaries of the sets under consideration are not sharply defined [12]. Measure of non-probabilistic entropy used in our study is based on the intersection and union of the fuzzy set and its complement set. Shang & Jiang [10] defined it follows:

$$E(A)=\frac{1}{n}\sum_{i=1}^{n}\frac{\mu_{A\cap\bar{A}}(x_i)}{\mu_{A\cup\bar{A}}(x_i)} \tag{3}$$

where n denotes the number of indicators in a related dimension; x_i is the average of all respondents' judgments for a given indicator i and $\mu_A(x_i)$ denotes the degree of belongingness of indicator i to the dimension set A. $\bar{A}$ is the complement set of A (Soyer, Kabak, & Asan 2007). High level of entropy (value close to one) in the supply chain has the effect of impeding flow by building obstacles, the bigger this obstacle is, the longer lead times and less predictable operations, making more uncertain the state of the system, and as a consequence, a bigger amount of information is required to monitor and manage that system [7].

3 Results

The data from questionnaire research we performed in 2008-2010. Involved companies are focused on different production groups and 18 of them are related to supply chain (we have discovered and elaborated mutual interconnection of 18 companies; obvious proofs of other connections are not provided). The main aim of the research was finding the most frequented logistic indicators which are used in asked Czech companies [9]. Used questionnaire was divided into five dimensions, where in every dimension several indicators marked by respondents according to frequency of use are given: New supplier selection (N), Evaluation of suppliers (E), Storage (S), Customers (C) and Transport (T). The questionnaire does not deal with indicators, which are used for performance of production.

3.1 Fuzzy evaluation of original supply chain

Supply chain consists of 18 companies mainly concerned on production of automobiles. Base for verification of linkages was information on web sites of these companies and structured interviews. Results of dimensions evaluation are adopted from paper [11]. Membership degrees of a predefined linear membership function (equation 1) are obtained by average responses of collected data from questionnaire. Degrees of membership shows to what extent the supply chain evaluates an indicator represented by a particular fuzzy set. From indicators measures are gained corresponding aggregated dimensions fuzzy evaluations. Three different decision levels ("average", "high", "very high") for each dimension evaluation are calculated by equation 2. They and represent different gradual membership to dimensions and strictness of decisions to evaluation results.

For example dimension Evaluation of supplier (E) have degree of membership 0,81 at level "average", 0,53 at level "high" and degree 0,47 at level "very high". Even, if the decision level is increased to "very high" or "high" with stricter decision level (and lower membership degrees), dimension could have higher overall rank (and importance to supply chain) in contrast to different dimension. All fuzzy measures of dimensions according to different decision levels are not included in the paper (for more see [11]). For simplify are only corresponding deffuzified values of fuzzy measures are exemplified.

Fragmented decision levels are consolidated by defuzzification method centre of area (COA) for all dimensions (these results provide rank of measures supply chain dimension). Results of original supply chain after deffuzifying are in the centre of figure 1 (with evaluations E = 0,60 – Evaluation of supplier; N = 0,44 - New supplier selection; C = 0,48 - Customers; S = 0,17 - Storage; T = 0,14 - Transport). Highest values indices that companies connected in supply chain have high potential of information sharing in supplier-customers relationships. According to results dimensions and results of indicators, depicted supply chain indicates attributes of cooperative supply chain type [3]. Indicators dimensions in original supply chain are ranked as follows: E > C > N > S > T.

In the supply chain, the most uncertain value with fuzziness (non-probabilistic entropy, E_f) of 0,56 has dimension customers (C). With the intention of simplifying the analysis, and based on the experience and common sense, we assume that higher value concludes high uncertainty and thus poor judgment about dimension evaluation. Contrariwise dimension set with a fuzziness of 0,33 indices poor level of uncertainty, which is attended by high level of information. Because of the non-probabilistic entropy of all dimensions is not greater than 0,60, calculated evaluations have relatively high capability of information notice [11].

3.2 Cluster analysis

Results of cluster analysis are used for presented modification of supply chain. For clustering (agglomerative hierarchical, k-means) are used all companies examined in questionnaire research. In AHC clustering we are tested five aggregation criterions in conjunction with several dissimilarity methods, that were obtained by using Euclidean, Bhattacharya, Chebychev, Mahalanobis, Manhattan distance and single, strong linkage and Ward´s algorithm as aggregation method.

According to AHC results (and Dendrogram) four clusters are used for next phase of k-means clustering. The k-means method is carried out to divide the observations into homogeneous clusters, based on their description by a set of supply chain dimensions. There are four clusters determined which have different features and characteristics expressed by dimensions. We have identified one cluster with very strong linkages to all dimensions evaluations (cluster 3), cluster with weak linkages (cluster 4) and two clusters that have partial strong dependence on dimensions evaluations. According to those characteristics are clusters named and described:

Cluster 1 associates companies focused on relationship (up and down stream cooperation in supply chain). This cluster contains 62 companies (36 % of all these companies are focused on engineering it is 50 % of all asked engineering companies). Seven of these companies fulfill EU terms for big companies and the rest are small and middle sized companies.

Cluster 2 consists of companies focused on downstream cooperation. Producers of consumers' goods are dominant for this cluster. On the second place by the number of companies stand building companies and engineering industry. Only 5 of these 51 companies fulfill EU terms for big companies. Suppliers are very important especially for SME's, these companies have only week bargain power to their suppliers, and on the other hand these companies are very sensitive on delivery price.

Cluster 3 represents companies focused on reporting by indicators in each of five dimensions. Totally account of this group companies is 49. More than half of these companies are represented by food producers, second important group are engineering companies. In this cluster there are 41% of asked big companies, they are mainly from food industry and build industry. There are two main groups of companies: First of them concentrated big companies with conducted evaluation system of process and performance indicators. Second group of companies are new firms, which try to create new information system for performance or process evaluation.

Cluster 4 associates companies, which do not use so many indicators as the clusters before. Come up to expectation, this cluster is created by small companies across the branches. Companies monitored only a few indicators, which are connected with accounting.

3.3 Modification of supply chain

The main purpose of this phase is to analyzing measures changes, which are obtained by adding new set of companies into supply chain (based on cluster analysis results). Enlargement of this network is only hypothetical, but new joined companies have capability of linkage to supply chain and they are potentially able to connect to hubs. Figure 1 draw four main modification of original supply chain: four ellipses across the original network with new linkages grouped by certain clusters. From each of four clusters with different features and characteristics are selected 2-5 companies, which are nearest to clusters centers (centroids) and have potential to become supply chain member. We make an effort to select companies, which represent related clusters, but their characteristics and features that influences connection to network are important too. Different number of companies, which are fulfilled these conditions are also obtained.

As we are modified original supply chain with selected companies from each cluster, changes based on orientation and characteristics of different clusters are expected. As depict figure 1 (computed values for supply chain enlarged by certain cluster), some results of dimensions measures are not influenced by clusters characteristics.

For example companies in Cluster 1 are mostly focused on relationships (up and down stream cooperation in supply chain), but some measures of dimensions Evaluation of Supplier (E), New supplier selection (N) and Customers (C) are decreased. However, focusing of Cluster 2 on downstream cooperation leads to Evaluation of Supplier (E) and New supplier selection (N) dimensions measures increase. With connection to this case it is important to note that some features and characteristics of clusters should not selected companies from each clusters represented (because of having different distance from each clusters centers).

The highest measures after deffuzification (COA method) for each dimension modification (and supply chain enlargement) are: Evaluation of supplier (E = 0,64) is the highest after adding companies from Cluster 3, New supplier selection (N = 0,49) after adding companies from Cluster 2, Customers (C = 0,55) measures after adding companies from Cluster 1 and measures of Storage (S = 0,17) and Transport (T = 0,19) after adding companies from Cluster 4. So, managing of supply chain could be focused on emphasize some of dimensions, which are increased by adding companies from relevant cluster [11].

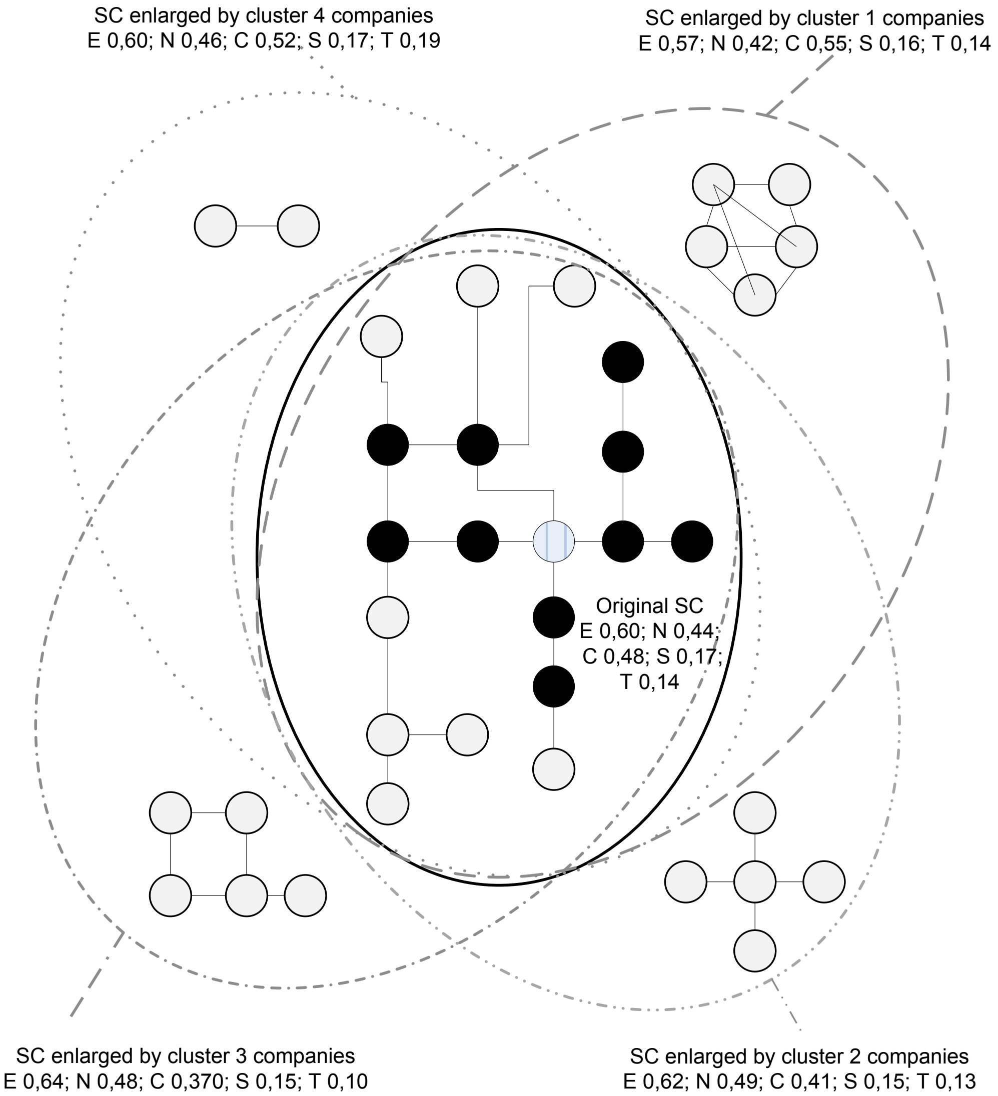


Figure 1 Supply chain fuzzy evaluation (original data set and modifications)

Results of modifying supply chain by adding new companies from different clusters are arranged in Table 1, Figure 2 where only evaluation measures and non-probabilistic entropy changes are included. Highlighted are changes values (in bold changes higher than 0,05), which are indicated very high influences of new added companies to current supply chain.

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	E_f	COA	E_f	COA	E_f	COA	E_f	COA
Evaluation of supplier (E)	-0,03	-0,03	**-0,05**	+0,02	-0,04	+0,04	-0,04	±0,00
New supplier selection (N)	**-0,06**	-0,02	+0,01	**+0,05**	+0,02	+0,04	+0,01	+0,02
Customers (C)	+0,02	**+0,07**	**+0,16**	**-0,07**	**+0,17**	**-0,11**	-0,02	+0,04
Storage (S)	-0,03	-0,01	-0,04	-0,02	-0,03	-0,02	+0,02	±0,00
Transport (T)	-0,01	±0,00	-0,03	-0,01	**-0,07**	-0,04	**+0,07**	**+0,05**

Table 1 Supply chain evaluation and entropy changes [11]

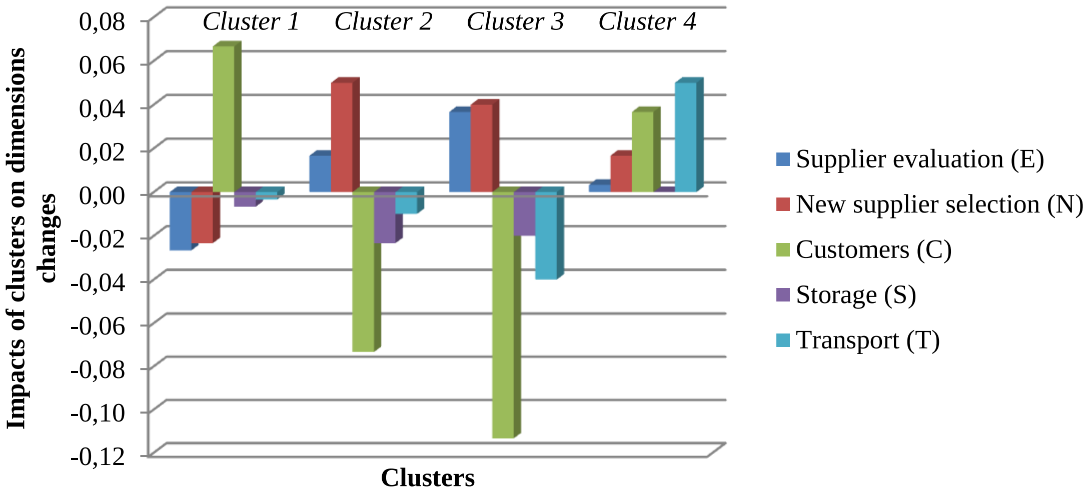


Figure 2 Supply chain evaluation after modification [11]

Following three types of changes are determined:

1. Selected companies from cluster 1 have high positive impact on dimension Customers (C = + 0,07) and some negative impact on other dimensions. Selected companies from cluster 1 should be focused on up and down stream cooperation, but in this case is group of SME's chosen. This group of companies is very often close to customers and their production consists mainly of job-order manufacturing.
2. Selected companies from cluster 2 and cluster 3 have positive impact on supplier evaluation dimensions - Evaluation of supplier (E = +0,02 and E = +0,04) and New supplier selection (N = +0,05 and N = +0,04), the rest of dimensions evaluations are decreased, particularly dimension Customers (C = -0,07 and C = -0,11). Because of companies focused on downstream cooperation in cluster 2, this modification meets our expectation. Companies in cluster 2 are oriented on quality of deliveries (for example buying materials and semi-finished products), customers and good downstream cooperation. To the contrary companies selected to the cluster 3 have not positive impacts to all dimensions as is expected. These companies are focused on reporting by indicators in each of five dimensions with conducted evaluation system of process and performance. Although this cluster has big potential to sharing information in supply chain, it is not able to come up to expectation. Creating good information system brings time period, when companies monitored big amount of information and will precise their information system in future.
3. Selected companies from cluster 4 have surprisingly positive impact on all dimensions evaluations (two of these have neutral impact to measures), particularly on Transport (T = +0,05). These companies monitored only a few indicators, which are connected with accounting. Reasons for not using many indicators of these five dimensions are very specific and narrow production portfolio (only 2 of 3 different products), short-run production system, and orientation on providing services (especially transport companies).

All changes in dimensions evaluations are interconnect with non-probabilistic entropy (E_f), where greater values of entropy reveals the higher degree of uncertainty inherent in each dimension set that we are exposed to in any judgment about these dimension set [11]. For example increasing values of non-probabilistic entropy after adding companies from cluster 2 and cluster 3 are determined evaluations of dimension Customers (C = +0,16 and C = +0,17). These results also have thanks to high degree of non-probabilistic entropy small reliability as state Spartalis, Iliadis, Maris [13].

4 Conclusions

The paper is focused on evaluation of logistic dimensions in supply chain, where the uncertainty surrounding supply chain performance measurement arises from the vagueness and deals with application of fuzzy approach that provides a formal method for modeling imprecise or incomplete relationships inherent in supply chains. Results of fuzzy evaluations shows strong linkage with suppliers and this implies orientation on indicators of quality of buying materials and semi finished products. According to results dimensions evaluation and results of indicators, depicted supply chain indicates attributes of cooperative supply chain type, where companies need to learn to trust their business partners and to build enduring partnership [3].

Enlargement of original supply chain is only hypothetical, but new joined companies have capability of linkage to supply chain and they are potentially able to connect to hubs. Our findings show three types of changes influenced by selected companies from: a) cluster 1 with high positive impact to dimension Customers (C) and some negative impact to other dimensions; b) cluster 2 and cluster 3 with positive impact to supplier evaluation dimensions (E, N), the rest of dimensions evaluations are decreased, particularly dimension Customers (C); c) cluster 4 with surprisingly positive impact to all dimensions evaluations, particularly to Transport (T). All changes in dimensions evaluations are interconnect with non-probabilistic entropy, where greater values of entropy reveals the higher degree of uncertainty inherent in each dimension set. Calculated evaluations have relatively high capability of information notice. New approach has potential for comparing information sharing in different supply chains or networks.

Acknowledgements

Paper is one of the solutions of grant research MSMT Contact MEB061008 “Cross border enterprise – B2B marketing and distribution in Austria and Czech Republic”.

References

[1] De Luca, A., Termini, S.: Definition of nonprobabilistic entropy in setting of fuzzy sets theory. *Information and Control* 4 (20), 1972, 301-312. ISSN 0019-9958.

[2] Emmett, S., Crocker, B.: *The relationship-driven supply chain : creating a culture of collaboration throughout the chain*. Gower, Aldershot, 2006. 187 p. ISBN 978-0-566-08684-7.

[3] Gattorna, J.: *Dynamic Supply Chain Alignment*. Gower, Farnham, 2009. 440 p. ISBN 978-0-566-08822-3.

[4] Gattorna, J. L.: *Gower Handbook of Supply Chain Management*. 5 ed. Gower Publishing, Aldershot, 2003. 692 p. ISBN 0566 08581 X.

[5] Chen, H. J., Ma, F. Model for Ranking Green Supply Chain Strategies Based on MDEA. In Qi, E. S., Cheng, G., Shen, J. A., Dou, R. L. *2009 Ieee 16th International Conference on Industrial Engineering and Engineering Management, Vols 1 and 2, Proceedings*. New York: Ieee, 2009. ISBN 978-1-4244-3670-5. s. 1530-1533.

[6] Karaesmen, F., Buzacott, J. A., Dallery, Y.: Integrating advance order information in make-to stock production systems. *IIE Transactions* 8 (35), 2002, 649-662. ISSN 0740-817X.

[7] Martinez-Olvera, C.: Entropy as an assessment tool of supply chain information sharing. *European Journal of Operational Research* 1 (185), 2008, 405-417. ISSN 0377-2217.

[8] Min, H., Zhou, G.: Supply chain modeling: past present and future. *Computers & Industrial Engineering* 1 (53), 2002, 231-249. ISSN 0360-8352.

[9] Pech, M., Smolova, J. Using of fuzzy entropy as a supportive method for managing the real supply chain: a case study. In: *28th International Conference on Mathematical Methods in Economics 2010*. University of South Bohemia, Faculty of Economics, Ceske Budejovice, 2010. 505-510. ISBN 978-80-7394-218-2.

[10] Shang, X. G., Jiang, W. S.: A note on fuzzy information measures. *Pattern Recognition Letters* 5 (18), 1997, 425-432. ISSN 0167-8655.

[11] Smolová, J., Pech, M.: Fuzzy approach to supply chain management. *Economics Working Papers* 1 (1), 2011, 7-56. ISSN 1804-5618.

[12] Soyer, A., Kabak, Ö., Asan, U.: A fuzzy approach to value and culture assessment and an application. *International Journal of Approximate Reasoning* 2 (44), 2007, 182-196. ISSN 0888-613X.

[13] Spartalis, S., Iliadis, L., Maris, F.: An innovative risk evaluation system estimating its own fuzzy entropy. *Mathematical and Computer Modelling* 1-2 (46), 2007, 260-267. ISSN 0895-7177.

[14] Suzuki, T., Kodama, T., Furuhashi, T., et al.: Fuzzy modeling using genetic algorithms with fuzzy entropy as conciseness measure. *Information Sciences* 1-4 (136), 2001, 53-67. ISSN 0020-0255.

[15] Wang, J. T., Shu, Y. F.: Fuzzy decision modeling for supply chain management. *Fuzzy Sets and Systems* 1 (150), 2005, 107-127. ISSN 0165-0114.

On the reducibility of the pickup and delivery problem

Jan Pelikán[1]

Abstract. The paper deals with the pickup and delivery problem with transfers and split delivery. The reducibility of the vehicle routing problem is studied in literature. Only for a special case of the vehicle routing problem the reducibility is proved and the non reducibility is proved for many cases of the problem. The reducibility of the problem allows a remarkable simplification of the problem. The reducibility for the pickup and delivery problem is defined and the non reducibility is proved for the case triangle inequality of the distance matrix and a special case of the transport demand matrix. The reducibility is valid only for the symmetric demand matrix with only full capacity of vehicle demand and symmetric nonnegative cost matrix with triangle inequality.

Keywords: pickup and delivery problem, case study, integer programming, heuristics.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction – the description of reducibility of the vehicle routing problem

The reducibility feature of the problem was studied for the vehicle routing problem with split demand (SDVRP). The reducibility of the problem allows a remarkable simplification of the problem [1]. The instance of the vehicle routing problem with split demand is reducible if there is an optimal solution in which each node is served by as many full load depot – node routes as possible.

The reducibility was proved for SDVRP with integer demand in nodes and the capacity of vehicles V=2 and symmetric costs matrix satisfying the triangle inequality. The problem with V=2 is called as the skip deliver problem (SDP).

For the non symmetric costs matrix it was shown that the reducibility of the SDP is not valid even if .the costs matrix satisfies the triangle inequality.

No validity of the reducibility is proved for V>2 even if the costs are symmetric with the triangle inequality [1].

2 Pickup and delivery problem with transfers and transport demand split (SDPDPT)

Let us have a distribution network with a set of n nodes and the cost matrix C of distances between all pair of nodes, where c_{ij} is the costs - distance between nodes i and j. Let us denote q_{kl} the amount of goods that has to be transported from node k to node l. Vehicles with capacity V are used for pickup and delivery and they can start in any node. All routes have to be cyclical, each vehicle has to come back to the node it starts from. The objective is to minimize the length of all the routes. The optimal solution is a set of cyclical routes, for each of them it is specified the depot, which cover all demands for pickup and delivery.

The problem is solved in two phases: the first phase finds the optimal number of vehicles going through each arc, i.e. from each node to each node to cover all pickup and delivery demand given by matrix Q. The second phase utilizes the optimal solution from the first phase and creates the cyclical routes and their depots which minimize a number of transfers.

---

[1] Jan Pelikán, University of Economics Prague, Faculty of Informatics and Statistics, Department of Econometrics, Prague 3, W. Churchill sq.4, pelikan@ vse.cz

3 The transportation costs optimization model

Let Q is a matrix of transport demands, C a matrix of transportation costs. The unique capacity vehicles are available with capacity V. We define two artificial nodes: the node 0, the source of all goods and the node n+1 as the sink. All transport demands will origin in the node 0 and finish in to node n+1. The following variables are defined in the mathematical model:

$y_{ij} \geq 0$ is integer – number of vehicles going through the arc (i, j) in the direction from i to j, $(i,j=0,1,...,n+1, i\neq j)$,

$x^{kl}_{ij} \geq 0$ amount of goods (part of the total amount q_{kl}) transported from node i to node j, $(i,j=0,1,...,n+1, i\neq j, k,l=1,2,...,n, k\neq l)$.

Mathematical model:

$$z = \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} y_{ij} \to \min \tag{1}$$

$$\sum_{i=1}^{n} y_{ij} = \sum_{i=1}^{n} y_{ji},\ j = 1,2,...,n, \tag{2}$$

$$\sum_{i=0}^{n} x_{ij}^{kl} = \sum_{i=1}^{n+1} x_{ji}^{kl},\ j = 1,2,...,n; k,l = 1,2,...n, k \neq l, \tag{3}$$

$$\sum_{k,l} x_{ij}^{kl} \leq V y_{ij}, i, j = 1,2,...,n, i \neq j, \tag{4}$$

$$\begin{aligned} &x_{0,k}^{kl} = q_{kl}\ , k,l = 1,2,...,n, k \neq l; \quad x_{0,j}^{kl} = 0, k,l = 1,2,...,n, k \neq l; j = 1,2,...,n+1, j \neq k, \\ &x_{l,n+1}^{kl} = q_{kl}\ ,\ k,l = 1,2,...,n, k \neq l; \quad x_{j,n+1}^{kl} = 0, k,l = 1,2,...,n, k \neq l; j = 0,1,...,n, j \neq l, \end{aligned} \tag{5}$$

$$\begin{aligned} &x_{ij}^{kl} \geq 0, i, j = 0,1,...,n+1,\ i \neq j,\ k,l = 1,2,...,n,\ k \neq l, \\ &y_{kl} \geq 0, \text{integers},\ k,l = 1,2,...,n,\ k \neq l. \end{aligned} \tag{6}$$

The objective (1) corresponds to the sum of the evaluations of all the arcs in the solution, i.e. the total length of all routes. Equations (2) assure the vehicle will leave the location that it will visit. With respect to equations (3), amount of goods being transported from k to l entering node j leaves this node. Inequalities (4) disable exceeding the capacity of the vehicle transporting goods between nodes i and j. Equations (5) assure that both the total flow from the source 0 to node k and the total flow from node l to the sink n + 1 are equal to the total requirement q_{kl}. All other flows from the source and to the sink are set to 0.

The second phase consists in the proposal of cyclical routes and their depots on the base of the results from the first phase. In the model described above, the optimal value of y_{ij} determines a number of vehicles going through arc (i, j) from node i to node j. Respecting equations (2), a number of vehicles entering each node equals to a number of vehicles leaving it. For generation of cyclical routes in the form of path in the network $(i_1, i_2,..., i_t)$, the following general algorithm can be used:

Algorithm for the route generation:

Step 1. If $y_{ij} = 0$ for all arcs $(i, j) \in E$, it is not possible to generate any route, otherwise select any arc $(i_1, i_2) \in E$, $y_{i_1,i_2} > 0$. Set $y_{i_1,i_2} = (y_{i_1,i_2} - 1)$ and $t = 2$.

Step 2. Repeat while $i_1 \neq i_t$: Select any arc $(i_t, i_{t+1}) \in E$, $y_{i_t,i_{t+1}} > 0$. Set $y_{i_t,i_{t+1}} = (y_{i_t,i_{t+1}} - 1)$ and $t = t + 1$.

Proposition 1.

If $\sum_{i=1}^{n} y_{ij} = \sum_{i=1}^{n} y_{ji}$ for j = 1,2,…,n, then the previous algorithm will generate cycles.

Proof.

If, in the step 2, the path (i_1, i_2, ...,i_t) is not closed, i.e. $y_{it,\,j}$= 0 for all nodes j, then $1 < \sum_{i=1}^{n} y_{ij} \neq \sum_{i=1}^{n} y_{ji} = 0$ for $j=i_t$ that contradicts the assumption.

4 Reducibility of SDPDPT

First question is how to define the reducibility of SDPDPT. Let us define the reducibility in the following way.

Definition.

The instance of SDPDPT is reducible if the optimal solution exists such that for all arc (i, j) with λ= min $\{\lfloor q_{ij}/V \rfloor, \lfloor q_{ji}/V \rfloor\}$ hold $y_{ij} \geq \lambda$ and $y_{ji} \geq \lambda$.

Comment. In the previous definition there exists simple cyclic route i-j-i for λ vehicles which ensure λV transport demand of q_{ij} and q_{ji}.

Proposition 2.
Let us suppose the triangle inequality for the matrix C and any arc (i, j) with $0 < q_{ij} < (\lceil q_{ij}/V \rceil)V$. Then the instance PDP is not reducible.

Proof.
The proof is based on the following example.
Lets have PDP with 4 nodes and the demand matrix Q and the costs matrix C:

$$Q = \begin{pmatrix} 0,10,5,5 \\ 10,\ 0,\ 5,5 \\ 5,\ 5,0,5 \\ 5,\ 5,5,0 \end{pmatrix}, \quad C = \begin{pmatrix} 0,1.5,1,1 \\ 1.5,0,1,1 \\ 1,1,0,2 \\ 1,1,2,0 \end{pmatrix}$$ and vehicle capacity V=10.

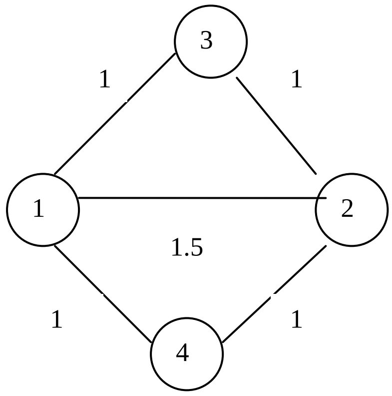
We will demonstrate for this instance that the solution with reduced demand is not optimal. Now we reduce the demands and next we find the optimal solution for the problem with reduced demands. The direct route is 1-2-1 of the length is 3, this route covers transport demands $q_{12}=q_{21}=10$.
The rest of demands is covered by the route: 1-3-2-4-2-3-1 of the length 6.
Total routes costs are 3 + 6 = 9.

node	pickup	delivery	Load
1	q_{13}=5, q_{14}=5		10
3	q_{32}=5	q_{13}=5	10
2	q_{21}=5, q_{24}=5	q_{32}=5, q_{12}=5	10
4	q_{41}=5, q_{42}=5	q_{24}=5, q_{14}=5	10
2	q_{23}=5	q_{42}=5	10
3	q_{31}=5	q_{23}=5	10
1		q_{41}=5, q_{31}=5	10

Table 1. Route 1-3-2-4-2-3-1

The optimal solution of the instance without reduced demands represents two routes:
route A: 1-3-2-4-1 of length 4,
route B: 1-4-2-3-1 of length 4, total costs are 8, i.e. lower costs are achieved than the costs of the solution obtained with reduction of the demands.

Figure 1. The not reducible instance PDP



Route A

Node	Pickup	delivery	Load
1	q_{13}=5, q_{12}=5		10
3	q_{32}=5	q_{13}=5	10
2	q_{21}=5, q_{24}=5	q_{32}=5, q_{12}=5	10
4	q_{41}=5	q_{24}=5	10
1		q_{21}=5, q_{41}=5	

Route B

Node	Pickup	delivery	Load
1	q_{14}=5, q_{12}=5		10
4	q_{42}=5	q_{14}=5	10
2	q_{21}=5, q_{23}=5	q_{42}=5, q_{21}=5	10
3	q_{31}=5	q_{23}=5	10
1		q_{21}=5, q_{31}=5	

Table 2. Non reducible solution

Lemma: If the non negative costs matrix satisfies the triangular inequality then the length of the arc (i,j) c_{ij} is less than or equal to the length of any path from node i to node j.
Proof.
Let us have the path $(k_1, k_2, \ldots, k_s)$, where $k_1=i$ and $k_s=j$. Then
$c_{ij} \leq c_{i,k2}+c_{k2,j} \leq c_{k1,k2}+c_{k2,ks} \leq c_{k1,k2}+c_{k2,k3}+c_{k3,ks} \leq \ldots \leq c_{k1,k2}+c_{k2,k3}+\ldots+c_{ks-1,ks}$.

Proposition 3.
If C is nonnegative symmetric matrix with triangular inequality and Q is a symmetric matrix with $q_{ij}=k_{ij}V$, where k_{ij} is integer. Then the instance of PDP is reducible and the optimal solution is $y_{ij}=k_{ij}$ for all i,j .

Proof.
The optimal number of vehicles going through arc (i,j) is k_{ij}. If a number of vehicles providing the transport of the demand q_{ij} is $y_{ij}=k_{ij}-1$, so the path of the vehicle in place of going through the arc (i,j) with costs c_{ij} has to be replaced by a path from node i to node j. But the length of path, according to the lemma, is greater than or equal to c_{ij}.
Comment 1.
If the matrix C satisfies triangular inequality and the matrix Q is not symmetric and it holds
$q_{ij} = k_{ij} V$, where k_{ij} is integer for all $i,j=1,2,\ldots,n$, $i \neq j$, and $\sum_{i=1}^{n} k_{ij} = \sum_{i=1}^{n} k_{ji},\ j=1,2,\ldots,n$, then the optimal solution is $y_{ij}=k_{ij}$, for all $i,j=1,2,\ldots,n$, $i \neq j$.

Proof.
The optimality follows from the Lemma. The existence of cyclic paths results from
$\sum_{i=1}^{n} k_{ij} = \sum_{i=1}^{n} k_{ji}, j = 1,2,\ldots,n$ (see Proposition 1).

Comment 2.
If the matrix C satisfies triangular inequality and the matrix Q is not symmetric holding $q_{ij} = k_{ij} V$, where k_{ij} is integer for all $i,j=1,2,...,n$, $i \neq j$, then the optimal solution is $y_{ij}=k_{ij}+x_{ij}$, $i,j=1,2,...,n$, where $X=\{ x_{ij} \}$is optimal solution of the transportation problem with n sources and n destinations, cost matrix C, capacities of sources are $a_j = \max\left\{\sum_{i=1}^{n} k_{ji} - \sum_{i=1}^{n} k_{ij}, 0\right\} j = 1,2,\ldots,n$, and demands $b_j = \max\left\{\sum_{i=1}^{n} k_{ij} - \sum_{i=1}^{n} k_{ji}, 0\right\} j = 1,2,\ldots,n$ of destinations.

Proof. It follows from Comment 1and Proposition 1.

5 Conclusions

The paper describes a new kind of the pickup and delivery problem and it's mathematical model. It is shown that this problem is not reducible for wide range of instances of the problem.

Acknowledgements

This research was supported by GAČR grant No. GACR 402/09/0041 and the project FIS 18/2010 founded by the Internal Grant Agency of the University of Economics Prague.

References

[1] Fábry J.: Optimization of routes in pickup and delivery problem. České Budějovice 08.09.2010 – 10.09.2010. In: *Mathematical Methods in Economics.* České Budějovice : University of South Bohemia, 2010, s. 128-131 ISBN 978-80-7394-218-2.

[2] Xu, H., Chen, Z., Rajagopal, S., Arunapuram, S.: Solving a practical pickup and delivery problem., Transportation science 2003, vol. 37, No. 3. pp. 347-364, ISBN 978-80-7394-218-2.

[3] Lagova M, Kalcevova J.: *Automatic Testing Framework for Linear Optimization Problems*. Liberec 17.09.2008 – 19.09.2008. In: *Mathematical Methods in Economics 2008 [CD-ROM].* Liberec : TU, 2008, s. 317-321. ISBN 978-80-7372-387-3.

[4] Mašátová, J., KALČEVOVÁ, J., KUNCOVÁ, M., Bokšteflová, B., Smrčka, P.: Comparison of the Universities from the Operational Research Point of View. České Budějovice 08.09.2010 – 10.09.2010. In: *Mathematical Methods in Economics [CD-ROM].* České Budějovice : University of South Bohemia, 2010, s. 423–428. ISBN 978-80-7394-218-2.

[5] Kuncová, M., Kalčevová, J., Novotný, J., Vojáčková, H.: Multicriteria Decision in the Analysis of the Market Environment for the Notification Process of the Broadband Network Development. České Budějovice 08.09.2010 – 10.09.2010. In: *Mathematical Methods in Economics [CD-ROM].* České Budějovice : University of South Bohemia, 2010, s. 386–391. ISBN 978-80-7394-218-2.

[6] Savelsbergh, M., W., P., Sol, M.: The general pickup and delivery problem. Transportation Research, no. 29, pp.17-29, 1995.

[7] Smrcka P, Kalcevova J: *Implementation of single-criteria discrete methods*. České Budějovice 08.09.2010 – 10.09.2010. In: *Mathematical Methods in Economics.* České Budějovice : University of South Bohemia, 2010, s. 572-577 ISBN 978-80-7394-218-2.

Self-adaptive Genetic Algorithm for Solving Travelling Purchaser Problem

Radomír Perzina[1]

Abstract. The aim of this paper is presentation of encoding for self-adaptation of genetic algorithms which is suitable for travelling purchaser problem. Comparing to previous approaches here is designed the encoding for self-adaptation not only one parameter or several ones but for all possible parameters of genetic algorithms at the same time. The proposed self-adaptive genetic algorithm is then applied for solving various instances of travelling purchaser problem and compared with a standard genetic algorithm. The main advantage of this approach is that it makes possible to solve wide range of combinatorial problems without setting parameters for each type of problem in advance.

Keywords: Genetic Algorithms, Travelling Purchaser Problem, Travelling Salesman Problem, Self-adaptation, Combinatorial Optimization.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Genetic algorithms are search algorithms based on the idea of natural selection and natural genetics. It is well known that efficiency of genetic algorithms strongly depends on their parameters. These parameters are usually set up according to vaguely formulated recommendations of experts or by the so-called two-level genetic algorithm where at the first-level genetic algorithm optimizes parameters of the second-level. A self-adaptation seems to be a promising way of genetic algorithms where the parameters of the genetic algorithm are optimized during the same evolution cycle as the problem itself. The aim of this paper is to present encoding and genetic operators for self-adaptation of genetic algorithms that is suitable for solving combinatorial problems, particularly the travelling purchaser problem. Comparing to previous approaches (e.g. [1], [11], [20]) we designed the encoding for self-adaptation not only one parameter but for all or nearly all possible parameters of genetic algorithms at the same evolution process. Moreover, the parameters are encoded separately for each element of a chromosome.

The proposed self-adaptive genetic algorithm was already successfully used for solving real timetabling problem [14], [15], [16] and travelling salesman problem [17]. Here we apply it for solving another combinatorial optimization problem – Travelling Purchaser Problem (TPP) which is generalization of travelling salesman problem. The problem is known to be NP-complete and therefore it is not known any algorithm for solving it in polynomial time. A large number of various methods have been already proposed in the literature for solving the travelling purchaser problem. The simplest method is enumeration of all possibilities which is suitable only for problems with small number of markets and products. Methods for real life problems with high number of markets and products are usually based on some heuristics and can be divided into two main categories – constructive and local optimization methods. Constructive methods build the path from the beginning using various rules [6]. Local optimization methods start from some initial solution, usually generated randomly or by another algorithm, which is then locally optimized [19]. Other methods are inspired by nature, e.g. tabu search, simulated annealing, evolution algorithms or ant colony optimization [21].

2 Encoding

Encoding is a major element of every genetic algorithm. The structure of our self-adaptive genetic algorithm's encoding is depicted on Figure 1. The idea of the proposed encoding consists in redundancy of information through hierarchical evaluation of individuals

[1] School of Business Administration, Silesian University, Department of Mathematical Methods in Economy, University Square 1934/3, Karviná, Czech Republic, perzina@opf.slu.cz.

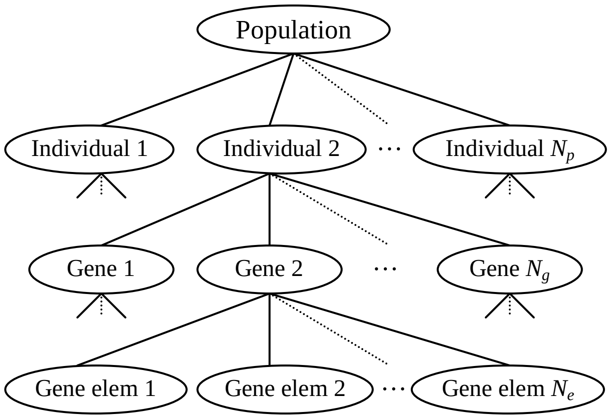


Figure 1 The structure of a population

As we can see, in the population each individual is composed of N_g genes where each gene is linked to exactly one optimized variable. Each gene is composed of N_e gene elements. The number of gene element varies for each gene and it is continuously updated throughout evolution. Each gene element contains low-level parameters, which encode optimized variables and parameters of evolution. All parameters are listed in Table 1.

Name	Description	Range
x^E	Optimized variable	<0;1>
q_m^E	Parameter of mutation	<-1;1>
r_m^E	Radius of mutation	<0;0.5>
p_c^E	Probability of crossover	<0;1>
r_c^E	Ratio of crossover	<0;1>
q_d^E	Parameter of deletion	<-0.1;0.1>
q_u^E	Parameter of duplication	<-0.1;0.1>
s_m^E	Identifier of myself for mating	<0;1>
s_w^E	Wanted partner for mating	<0;1>
r_r^E	Ratio of replacement	<0;1>
r_t^E	Ratio of population for selection	<0;1>
r_p^E	Ratio of population for 2nd partner selection	<0;1>
c_d^E	Coefficient of death	<0;1>
N_p^E	Wanted size of population	<0;1>

Table 1 The structure of a gene element

The upper index "^{E}" denotes, that it is a gene element value of the parameter. As the encoding is hierarchical, there are several levels of the parameters so gene values of parameters are marked by the upper index "^{G}", individual values by "^{I}" and population values by "^{P}". Since genetic operators are applied only to the low level values of parameters (gene element), the upper level values of parameters cannot be updated directly through evolution process, but only indirectly by evaluation mechanism from low level values.

2.1 Gene Parameters Evaluation

The encoding is polyploiditial, so each gene is composed of N_e gene elements. The number of gene elements is variable and undergoes evolution. Gene values of gene elements are evaluated by formula (1)

$$X^G = \frac{1}{N_e}\sum_{i=1}^{N_e} X_i^E \qquad (1)$$

where X stands for parameters that must be evaluated, i.e. x, s_m, s_w, r_r, r_t, r_p, c_d, N_p, i_t.

2.2 Individual Parameters Evaluation

Parameters concerning the whole individual, such as s_m^I, s_w^I, r_r^I, r_t^I, r_p^I, c_d^I, N_p^I are evaluated by formula (2)

$$X^I = \frac{1}{N_g}\sum_{i=1}^{N_g} X_i^G \tag{2}$$

The number of genes N_g is not variable, because one gene contains exactly one optimized variable.

2.3 Population Parameters Evaluation

Parameters concerning the whole population, such as r_r^P, r_t^P, c_d^P, N_p^P are evaluated as weighted average with weights according to their relative fitness w_f, defined as

$$w_f = \frac{N_p^P - i + 1}{\frac{\left(1 + N_p^P\right)N_p^P}{2}} \tag{3}$$

where i is index of i^{th} individual in population sorted by fitness in descending order, i.e. the individual with the highest value of the fitness function has the value of i equal to 1, the individual with the second highest value of the fitness function has the value of i equal to 2 etc.

3 Genetic Operators

As the proposed encoding is specific, the genetic operators must be adjusted to fit the encoding. There are used not only common genetic operators such as selection, crossover or mutation, but also some specific ones, as described in following paragraphs.

3.1 Selection

In genetic algorithms the selection of both mating parents is usually based on their fitness, but this is not true in nature. In nature a winner of a tournament selects his partner according to his individual preferences. Important is that he cannot take into account his genotype, i.e. directly the values of his genes nor his fitness, but only his phenotype, i.e. only expression of the genes to the outside. In a similar way we try to imitate nature by using parameters s_m^I and s_w^I. The parameter s_w^I represents individual preferences for mating and the parameter s_m^I represents individual's phenotype for mating. So the first parent is selected by a tournament selection method with variable ratio of population r_t^P from which the fittest individual is selected. The second parent is selected according to individual's preferences represented by the parameter s_w^I, i.e. the first parent chose an individual with the minimal value of expression $\left|s_w^I - s_m^I\right|$, but selection is made from only limited ratio of population r_p^I.

3.2 Crossover

The crossover operator is applied to every gene element of the first parent with the probability p_c^E. The crossover itself proceeds only between gene elements of mating parents according to formula

$$X_3^E = X_1^E + \left(X_2^E - X_1^E\right)\cdot r_c^E \tag{4}$$

where X stands for all parameters of a gene element (see Table 1), r_c^E is a ratio of crossover of the first parent defined in this gene element, the lower index "$_1$" denotes the gene element of the first parent, the index "$_2$" the second parent and the index "$_3$" denotes the child of both parents. The gene element of the second parent is selected randomly, but it is of the same gene as the gene element of the first parent.

3.3 Mutation

The mutation operator is applied to every gene element with probability $p_m^E = \left| q_m^E \right|$. Notice that probability of mutation is calculated as the absolute value of the parameter of mutation, because the mean value of p_m^E should be zero. Moreover, every gene element has its own probability of mutation. The mutation formula is defined as

$$X_{new}^E = X_{old}^E + \left(X_{\max}^E - X_{\min}^E \right) \cdot U\left(-r_m^E, r_m^E \right) \tag{5}$$

where X stands for all parameters of the gene element, $U(a,b)$ is a random variable with uniform probability distribution in the interval $\langle a;b \rangle$, X_{new}^E is the value of the parameter after mutation, X_{old}^E is the original value of the parameter, $X_{\max}^E$ ($X_{\min}^E$) is the maximal (minimal) allowed bit element value of the parameter in Table 1.

3.4 Duplication

The duplication operator is applied to every gene element with probability $p_u^E = \left| q_u^E \right|$. The gene element is duplicated (copied) with the same value of all parameters with the only exception that the values of parameter q_u^E of both gene elements are divided by a coefficient 2, in order to inhibit exponential growth of bit elements.

3.5 Deletion

The deletion operator is applied to every gene element with probability $p_d^E = \left| q_d^E \right|$. It means that the gene element is removed from the gene. By deletion and duplication operators the degree of polyploidity is controlled.

3.6 Replacement of Individuals

For every individual the parameter of a life strength – L is defined. When the individual is created, its life strength L is set to one and in every generation it is multiplied by the coefficient c_L defined as

$$c_L = 1 - c_d^P \left(1 - w_f \right) \tag{6}$$

Evidently, through evolution, a less fitter individual causes the greater decrease in L. In every generation all X^P parameters are evaluated and by using the above listed genetic operators $N_p^P \cdot r_r^P$ new individuals are created. Then a randomly selected individual is killed with probability (1-L). This process of individuals eliminating is repeated until only N_p^P individuals survive in the population.

4 Travelling Purchaser Problem

Travelling Purchaser Problem (TPP) is an interesting generalization of the well known Travelling Salesman Problem (TSP). This problem was firstly introduced by Ramesh [18] which can be defined as follows. Consider a domicile denoted by 0, a set of markets denoted by $M = \{1, 2, \ldots, m\}$, travel costs c_{ij} on each edge (i, j) linking two markets, and a set of products $K = \{1, 2, \ldots, \mathrm{n}\}$. Denote by M_k the set of markets selling product k and by p_{ki} the price of product k at market i. The objective of the TPP is to construct a tour through a subset of the m markets and the domicile and to purchase each of the n products at one of these markets such as to minimize the sum of the traveling and purchasing costs. In this paper it is assumed that if a product is available at a given market, its quantity is sufficient to satisfy demand. Such version of the problem is called the Uncapacitated Traveling Purchaser Problem. The most common TPP applications are in vehicle routing and warehousing [19].

The model of TPP used in this paper is defined as follows:

$$z = \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{k=1}^{n} \sum_{i=1}^{m} p_{ki} y_{ki} \to \min \tag{7}$$

s.t.

$$x_{ij} = \begin{cases} 1 & \text{if edge } (i,j) \text{ belongs to the solution} \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } i > j, i \in \{1,2,\ldots,m\} \text{and } j \in \{1,2,\ldots,m\}$$

$$y_{ki} = \begin{cases} 1 & \text{if product } k \text{ is bought in market } i \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } i \in \{1,2,\ldots,m\} \text{ and } k \in \{1,2,\ldots,n\}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{m} x_{ij} \leq n \ \text{ for } i > j$$

$$\left|\sum_{i=1}^{m} x_{ij} - 1\right| = 1 \text{ for } i \in \{1,2,\ldots,m\}$$

$$\sum_{k=1}^{n}\sum_{i=1}^{m} y_{ki} = n$$

The proposed self-adaptive genetic algorithm was used to solve the above problem. Each gene of the chromosome represents one real variable within the interval <0;1> which corresponds to exactly one pair (i, k), $i \in \{1, 2 ,..., m\}$, $k \in \{1, 2 ,..., n\}$, so there are ($m \cdot n$) genes in the chromosome. Genes are then sorted in ascending order and this order represents order in which product are bought and first occurrence of variable i indicates order in which markets are visited. Once all products are bought the reminder of genes in the chromosome is ignored. The fitness function f for the genetic algorithm is the negative value of z in (7), i.e. $f = - z$.
To make the idea behind decoding the chromosome clearer, a simple example will be provided. Let's suppose we have 4 markets (m_1, m_2, m_3, m_4) and 3 products (k_1, k_2, k_3), so the chromosome will have 12 genes. Let's say after evaluation, the gene values of parameters are:

Pair	**(1,1)**	**(1,2)**	**(1,3)**	**(2,1)**	**(2,2)**	**(2,3)**	**(3,1)**	**(3,2)**	**(3,3)**	**(4,1)**	**(4,2)**	**(4,3)**
Value	0.93	0.02	0.82	0.21	0.32	0.59	0.45	0.14	0.05	0.48	0.79	0.25

Table 2 Gene values

Then the order of pairs after sorting according gene values of parameters is listed in the table 3.

Pair	**(1,2)**	**(3,3)**	**(3,2)**	**(2,1)**	**(4,3)**	**(2,2)**	**(3,1)**	**(4,1)**	**(2,3)**	**(4,2)**	**(1,3)**	**(1,1)**
Value	0.02	0.05	0.14	0.21	0.25	0.32	0.45	0.48	0.59	0.79	0.82	0.93

Table 3 Sorted gene values

Therefore markets will be visited in order: m_1, m_3, m_2. Market m_4 will not be visited because after using 4th gene (2,1) in sorted chromosome we have bought all products. In the table 3 markets where we have bought products are underscored, i.e. $m_1 - k_2$, $m_3 - k_3$ and $m_2 - k_1$.

5 Numerical Experiments

To test performance of the proposed self-adaptive genetic algorithm (SAGA) we have randomly generated 5 instances of TPP with various numbers of markets (20, 40, 60, 80 and 100 markets) and products (5, 10, 15, 20 and 25 products). All instances contain markets distributed on the plane 100 x 100 units. Product prices are randomly generated within the range <1; 20> units. The results were then compared with the simple genetic algorithm (SGA) on these instances. The simple genetic algorithm used a binary encoding, the size of population was 30 individuals, probability of mutation 0.003 and elitism was used. We executed both algorithms 10 times and measured minimal, maximal and average value of the fitness function. Each algorithm was stopped after the fitness function was evaluated $N_f = 10000$ times. The results of both algorithms are shown in the table 2. For easier readability the values in the table are rounded and positive, i.e. they represent total cost of the tour.

	SAGA			**SGA**		
Instance	**Min**	**Avg**	**Max**	**Min**	**Avg**	**Max**
TPP20/5	241	278	317	241	267	326
TPP40/10	394	416	487	397	410	452
TPP60/15	503	522	554	514	525	568
TPP80/20	617	649	686	633	664	704
TPP100/25	720	763	801	742	791	813

Table 4 SAGA and SGA comparison

The table 4 shows that for instances with smaller number of markets and products SGA found better average solution than SAGA which found better average solution for more complex problems. It is also worth to compare minimum values of the objective function which was never worse for SAGA than for SGA in all instances.

6 Conclusion

In this paper we have presented a self-adaptive genetic algorithm with self-adaptation of all its parameters. It was shown that the self-adaptive genetic algorithm was able to effectively solve not just timetabling problem for which the algorithm was originally developed but also other combinatorial problems such as the travelling purchaser problem. The parameters of the SAGA converge during process of evolution so that it makes possible to solve wide range of combinatorial problems without setting parameters for each type of problem in advance.

The aim of this paper was not to present perfect optimization algorithm that can outperform any other one. There are many optimization algorithms that can find near optimal solution in much short time for complex travelling purchaser problems, but they achieve such good results just for the problem they were designed. Unlike these algorithms, the self-adaptive genetic algorithm can effectively handle various problems as it is able to optimize its parameters during evolution.

Acknowledgements

This research was supported by the grant project of GACR No. 402090405.

References

[1] Bäck, T.: Self-Adaptation in Genetic Algorithms. In: *Proceedings of the First European Conference on Artificial Life*. MIT Press, Cambridge, 1992.

[2] Boctor, F, Laporte, G., Renaud, J.: Heuristics for the traveling purchaser problem. *Computers & Operations Research* 30. Pergamon, 2003.

[3] Brailsford, S. C. et al.: Constraint Satisfaction Problems: Algorithms and Applications. *European Journal of Operational Research*, vol 119, 1999, pp. 557-581.

[4] Buzacott, J.A., Dutta, S.K.: Sequencing many jobs on a multipurpose facility. *Naval Research Logistics Quarterly* 1971.

[5] Dorigo, M., Stutzle, T.: *Ant Colony Optimization*. MIT Press, Cambridge, 2004.

[6] Golden, B.L., Levy, L., Dahl, R.: Two generalizations of the traveling salesman problem. *Omega*, 9, 1981.

[7] Laarhoven P.J.M., Aarts, E.H.L.: *Simulated Annealing: Theory and Applications*. Kluwer Academic Publishers, Norwell, 1987.

[8] Johnson, D.S., McGeoch, L.A.: The Traveling Salesman Problem: A Case Study. In: *Local Search in Combinatorial Optimization*.Wiley and Sons, New York, 1997.

[9] Lawer, E.L. et al.: *The Traveling Salesman Problem*. John Wiley, Chichestar, 1985.

[10] Lin, S., Keringhan B.W.: An Effective Heuristic Algorithm for the Traveling Salesman Problem. In *Operations Research*, 21. 1973.

[11] Marsili, S. L., Alba, P. A.: *Adaptive Mutation in Genetic Algorithms*. Soft Computing, vol. 4, no. 2. Springer – Verlag, New York, 2000.

[12] Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd ed. Springer – Verlag, New York, 1996.

[13] Pearn, W.L., Chien, R.C.: Improved solutions for the traveling purchaser problem. *Computers & Operations Research*, 25, 1998.

[14] Perzina, R.: A Self-adapting Genetic Algorithm for Solving the University Timetabling Problem. In: *Proceedings of the 8th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2004)*. IIIS, Orlando, 2004.

[15] Perzina, R.: Solving the University Timetabling Problem with Optimized Enrollment of Students by a Parallel Self-adaptive Genetic Algorithm. In: *Proceedings of the 6th International Conference on the Practice and Theory of Automated Timetabling*. Masarykova univerzita, Brno, 2006.

[16] Perzina, R.: Solving Multicriteria University Timetabling Problem by a Self-adaptive Genetic Algorithm with Minimal Perturbation. In *Proceedings of the 2007 IEEE International Conference on Information Reuse and Integration*. IEEE Publishing, Las Vegas, 2007.

[17] Perzina, R.: Self-Adaptive Genetic Algorithm and Travelling Salesman Problem. In: *Mendel 2010. 16th International Conference on Soft Computing*. VUT Brno, Brno, 2010.

[18] Ramesh, T.: *Traveling purchaser problem*. Opsearch 1981.

[19] Singh, K.N., Van Oudheusden, D.L.: A branch and bound algorithm for the traveling purchaser problem. *European Journal of Operational Research* 1997.

[20] Stefano, C. D., Tettamanazi, A. G. B.: An Evolutionary Algorithm for Solving the Schoool Timetabling Problem. In: *Proceedings of the Evo Workshops 2001*. Springer, Como, 2001.

[21] Vos, S.: Dynamic tabu search strategies for the traveling purchaser problem. *Annals of Operations Research*, 63, 1996.

Routing Problem with Many Starting Points and Cost Dependent on Starting Point

Štefan Peško[1]

Abstract. This paper studies a modification of vehicle routing problem formulated as follows: We are given n customers. These customers are to be served by several servicemen who depart from home, travel from customer to customer and return home. The cost of every such route depends on starting point and travelled distance. The goal is to minimize the total cost of all such routes covering all customers. We present a novel MILP formulation of this problem based on the Makhorin formulation of the traveling salesman problem for the serviceman route.

Keywords: routing problem, starting point dependence, traveling salesman problem, MILP.

JEL classification: C61
AMS classification: 90C08, 90C11

1 Motivation

This paper is motivated by chat with collega Kaukič [5]. He formulated the traveling salesman problem with starting dependent cost given by cash discount for salesman – pensioner. The traveling salesman problem (TSP) is the search for the shortest tour that visits each node exactly once in the given set of nodes. This problem is one of the well-known NP-hard combinatorial optimization problems [1]. The aim of Kaukič's modification (KTSP) is to find the starting city (home city of a pensioner) and the cheapest route that visits all points in a given set of cities and return to the starting city. Then cost of optimum pensioner's tour is equal to minimum cost of the TSP tour multiplied by pensioner's cash discount.

We study a generalisation of the KTSP as a starting points dependent routing problem formulated as follows (SPDRP): We are given n customers. These customers are to be served by several servicemen who depart from home, travel from customer to customer and return home. The cost of every such route depends on starting point and traveled distance. The goal is to minimize the total cost of all such routes covering all customers. It is easy see that the SPDRP for only one pensioner is the KTSP.

Our mixed integer linear programming (MILP) formulation of the SPDRP is based on Makhorin formulation [2] for the classical TSP.

2 Makhorin formulation of the TSP

In the case of the TSP for n cities, the mathematical description can be a digraph $G = (V, H, c)$ with finite set of verticies $V = \{1, 2, \dots, n\}$, where each city is denoted by vertex $i \in V$ and arcs $(i, j) \in H$ connecting two cities $i, j \in V, i \neq j$ and cost c_{ij} is assigned for every arc $(i, j) \in H$.

The goal is to find a Hamiltonian cycle $1 = \pi_1 \to \pi_2 \to, \dots, \to \pi_n \to 1$ in digraph G that has minimum total cost

$$c(\pi) = \sum_{i=1}^{n-1} c_{\pi_i,\pi_{i+1}} + c_{\pi_n,1}.$$

Note that for the complete digraf G the Hamiltonian cycle exists always. The symmetric case of the TSP can be modelled via symmetric costs $c_{ij} = c_{ji}$.

[1]University of Žilina, Faculty of Management Science and Informatics, Department of Mathematical Methods, Univerzitná 8215/1, 01 026 Žilina, Slovak Republic , e-mail: stefan.pesko@fri.uniza.sk

Let $x_{ij} \in \{0,1\}$ be decision variable for $(i,j) \in H$ and $x_{ij} = 1$ means that the arc (i,j) is an element of the Hamiltonian cycle i.e. $\pi = (\ldots, i, j, \ldots)$. Let $y_{ij} \geq 0$ be flow variables for $(i,j) \in H$.

The Makhorin formulates and solves effectively small TSP as a following mixed integer linear programming problem (MTSP):

$$\begin{aligned}
\sum_{(i,j)\in H} c_{ij} \cdot x_{ij} &\rightarrow \min && && (1)\\
\sum_{j\in V:(i,j)\in H} x_{ij} &= 1 && \forall i \in V, && (2)\\
\sum_{i\in V:(i,j)\in H} x_{ij} &= 1 && \forall j \in V, && (3)\\
y_{ij} &\leq (n-1)\cdot x_{ij} && \forall (i,j) \in H, && (4)\\
\sum_{(1,k)\in H} y_{1k} &= n-1 && && (5)\\
\sum_{k\in V:(k,i)\in H} y_{ki} &= \sum_{k\in V:(i,k)\in H} y_{ik} + 1 && \forall i \in V - \{1\}, && (6)\\
x_{ij} &\in \{0,1\} && \forall (i,j) \in H, && (7)\\
y_{ij} &\geq 0 && \forall (i,j) \in H. && (8)
\end{aligned}$$

The objective function (1) minimizes the cost of the Hamiltonian cycle. Constraints (2), (3) and (7) define a feasible solution of the assignment problem in digraf G. Constraints (5), (6) and (8) define a spanning tree in the digraph G with root 1. The coordination constraint (4) requires that all arcs (i,j) of spanning tree ($y_{ij} > 0$) will be $n-1$ selected arcs in the desired Hamiltonian cycle ($x_{ij} = 1$).

The formulation MTSP eliminates a subtour via requirement that a spanning tree contained in assigment of vertices. This idea is used in our MILP model.

3 Formulation of the SPDRP

The mathematical description of the SPDRP is modelled in a weighted digraph $\vec{\mathcal{G}} = (V, H, c, w)$ with finite set of vertices $V = \{0, 1, 2, \ldots, n\}$, where vertex 0 is fictive and real vertices represent points of customers. Vertices of customers $i \in V - \{0\}$ are given weight (positive real number) w_i that is a weight of serviceman from home point i. The set of arcs H contains a real connection between customers and fictive connections $(0,i)$ and loops (i,i) for $i \in V - \{0\}$. The real arcs (i,j) are given weight by a distance from i to j which is positive real number d_{ij}.

The goal is to find a permutation π on the set of real vertices formed from cycles $C_{k_1}, C_{k_2}, \ldots, C_{k_m}$ and home points $\{k_1, k_2, \ldots, k_m\} \subset V$ of servicemen with minimum total cost

$$c(\pi) = \sum_{C_k \in \pi} w_k \cdot d(C_k),$$

where

- k is exactly one starting vertex (serviceman home point) in the cycle C_k,
- $d(C_k)$ is the length of cycle C_k counted as the sum of its arc distances,
- m is a number of cycles of permutation π that is a positive integer variable.

Let $x_{ijk} \in \{0,1\}$ be decision variable for $(i,j) \in H, k \in V$. The value $x_{ijk} = 1$ means that the arc $(i,j) \in H, i \neq 0$ is an element of the cycle $C_k = k \rightarrow \cdots \rightarrow i \rightarrow j \rightarrow \cdots \rightarrow k$. The value $x_{iik} = 1$ means that vertex i is not a vertex of the cycle C_k. The value $x_{0kk} = 1$ means the k is the only starting point in the cycle C_k. At last let $y_{ij} \geq 0$ be flow variables for arcs $(i,j) \in H$ interpreted as in the Makhorin model.

We define $W = V - \{0\}$ as the set of real vertices for easier identification. Now we can formulate for the SPDRP as a following mixed integer linear programming problem (RP):

$$\sum_{(i,j)\in H} \sum_{k\in W} c_{ij} \cdot w_k \cdot x_{ijk} \quad \to \quad \min \tag{9}$$

$$\sum_{j\in W:(i,j)\in H} x_{ijk} \quad = \quad 1 \qquad \forall i,k \in W \tag{10}$$

$$\sum_{i\in W:(i,j)\in H} x_{ijk} \quad = \quad 1 \qquad \forall j,k \in W, \tag{11}$$

$$\sum_{k\in W} x_{iik} \quad = \quad n-1 \qquad \forall (i,i) \in H, \tag{12}$$

$$\sum_{j\in W:(r,j)\in H} x_{rjr} \quad = \quad 1 \qquad \forall r \in W, \tag{13}$$

$$\sum_{(i,i)\in H} x_{iir} \quad \geq \quad n \cdot x_{rrr} \qquad \forall r \in W, \tag{14}$$

$$n \cdot \sum_{k\in W} x_{ijk} \quad \geq \quad y_{ij} \qquad \forall (i,j) \in H, i \neq j, \tag{15}$$

$$y_{0,r} + n \cdot x_{rrr} \quad \leq \quad n \qquad \forall r \in W, \tag{16}$$

$$\sum_{i\in V:(i,r))\in H} y_{ir} - \sum_{j\in V:(r,j)\in H} y_{rj} \quad = \quad 1 \qquad \forall r \in W, \tag{17}$$

$$\sum_{i\in V:(i,r)\in H} y_{ir} + \sum_{j\in V:(r,j)\in H} y_{rj} \quad \geq \quad 1 \qquad \forall r \in W, \tag{18}$$

$$\sum_{(0,j)\in H} y_{0,j} \quad = \quad n, \tag{19}$$

$$x_{ijk} \quad \in \quad \{0,1\} \qquad \forall (i,j) \in H, k \in V, \tag{20}$$

$$y_{ij} \quad \geq \quad 0 \qquad \forall (i,j) \in H. \tag{21}$$

The objective function (9) minimizes the total weighted cost of all cycles of a solution. Constraints (10), (11), (12) and (20) define a feasible solution of the three index assignment problem in subgraph of digraf $\vec{\mathcal{G}}$ in the set of verticies W. Constraints (17), (18), (19) and (21) define a spanning tree in the digraph $\vec{\mathcal{G}}$ with root 0. Costraints (13) and (14) ensure selection of exactly one home point r for every feasible cycle C_r. Condition (15) requires that arc $(i,j) \in H, i \in W, j \in W$ of spanning tree with $y_{ij} > 0$ is a selected arc of cycle C_k where $x_{ijk} = 1$.

4 Two examples

Considering RP model is relatively complicated and so we will ilustrate the central idea of the model on a small instance for the MTSP.

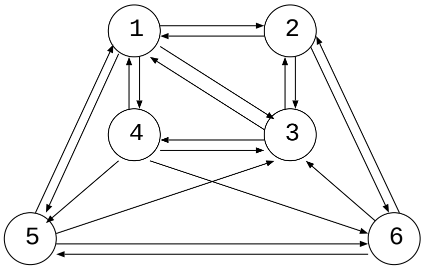


Figure 1: Digraph $\vec{G} = (V, H, c)$.

Example 1. (MTSP)
Let a diagram of digraph $\vec{G} = (V, H, c)$ be given on figure 1. For all $(i,j) \in H$ we define $c(i,j) = 1$ if

$i < j$ else $c(i,j) = 2$. After optimization [2] is the Makhorin flow given by five nonzero wariables:

$$y_{12} = 5,\ y_{26} = 4,\ y_{63} = 3,\ y_{34} = 2,\ y_{45} = 1.$$

The corresponding minimum Hamiltonian cycle $\pi = 1 \to 2 \to 6 \to 3 \to 4 \to 5 \to 1$ with cost $c(\pi) = 8$ as depicted in figure 2.

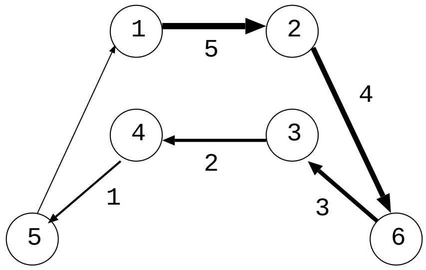


Figure 2: Hamiltonian cycle and Makhorin rooted spanning tree (with intensive arcs).

Now we can ilustrate the frame of the RP model based on the digraph $\vec{G}$ from Example 1.

Example 2. (SPDRP)
Let us add to the digraph $\vec{G}$ the fictive vertex 0 and fictive arcs (dashed lines) from the fictive vertex to the real vertices from $W = \{1, \ldots, 5\}$. We got the digraph $\vec{\mathcal{G}} = (V, H, d, w)$ with $d(i,j) = c(i,j)$ for $(i,j) \in H \cap W \times W$ depiced on the left part of the figure 3. Let $w_1 = w_6 = 20$ and $w_2 = w_3 = w_4 = w_5 = 10$ are weights of real vertices. After optimization via the RP formulation in [2] we have the Makhorin flow

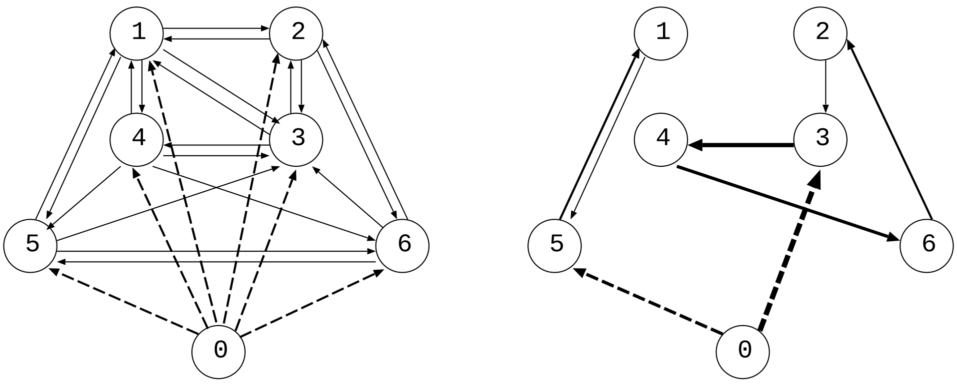


Figure 3: Digraph $\vec{\mathcal{G}} = (V, H, c, w)$ with two cycles C_5 and C_3.

given by six nonzero variables:

$$y_{03} = 4,\ y_{34} = 3,\ y_{46} = 2,\ y_{62} = 2,\ y_{05} = 2\ y_{51} = 1.$$

And so one servicemen will start from home vertex 3 of the cycle $C_3 = 3 \to 4 \to 6 \to 2 \to 3$ and the other servicemen will start from vertex 3 of the cycle $C_5 = 5 \to 1 \to 5$. This two cycles of the optimum solution $\pi = (C_3, C_5)$ with cost $c(\pi) = 10 \cdot d(C_3) + 10 \cdot d(C_5) = 80$ as depicted on the right part of the figure 3 where we can see how home vertices 3, 5 are assigned by a value $y_{03} > 0, y_{05} > 0$.

5 Future work

Computation experiments with small instances $6 \leq |V| \leq 30, 30 \leq |H| \leq 120$ of the RP formulation shows that the SPDRP is possible to solve by MILP solver of GNU Linear Programming Kit [2], from 10 seconds into the 110 minutes approximately. It is posible to show that the RP is NP-hard problem. We hope that our experience with the differencial evolution for small TSPs [3] and novel bacterial evolution algorithm for cost variable TSP [4] can be used for solving the real SPDRP problems with acceptable results.

Acknowledgements

The research of author is supported by Slovak Scientific Grant Agency under grant VEGA 1/0374/11

References

[1] Garey, M. R. and Johnson, D. S.: Computers and Intractability *A Guide to the Theory of NP-Completeness.*, A Series of Books in the Mathematical Sciences. W. H. Freeman and Co.. pp. x+338. ISBN 0-7167-1045-5, 1979.

[2] Makhorin, A.: GNU Linear Programming Kit, *Reference Manual*, Verson 4.29, Moscow, pp.154, 2008, Available at http://www.gnu.org/software/glpk/

[3] Peško, Š.: Differential Evolution for Small TSPs with Constraints, In: *Proceedings of the fourth Internattional Scientific Conference: Challenges in Transport and Communications*, Part III, September 14 - 15, Pardubice, Czech Republic, ISBN 80-7194-880-2, (2006), 989–994.

[4] Foldesi, P, Botzheim, J.: Solution for Modified Traveling Salesman Problem with Variable Cost Matrix Using Bacterial Evolutionar Algorithm, *Acta Technica Jaurinensis serial Logistics* Vol.**18**, No.**2** (2008), 159–171.

[5] Kaukič, M.: *private communication.*

The Dynamics of the World Agricultural Production and Consumption

Vítězslav Píša[1], Jan Brůha[2]

Abstract. The purpose of the paper is to analyze the dynamics of the world agriculture production and consumption using the index decomposition analysis (IDA). The changes in both variables are split to shifts in chosen types of level, intensity and structural indicators. The world agricultural production is explained by changes in total agricultural employment (scale effect) and regional composition of land (structural effect), productivity of labor and capital (intensity effects). In addition, evolution of world food consumption can be described by changes in total population (scale effect) and its composition over the world (structural effect), changes in GDP per capita and structure of consumption (intensity effects). Structural effects are relatively small in comparison with intensity and scale effects.

Keywords: Decomposition analysis, Index numbers, World Agricultural Production and Consumption

JEL classification: C43, N50, Q11
AMS classification: 49M27

1 Introduction

The dynamics of the agricultural production and food consumption interests researchers, policy-makers, or planners around the world for a variety of reasons. The main reason is the assessment of the world's ability to feed itself (Islam, 1995), but other reasons like impacts of agricultural production on composition of land or regional relationships between agricultural supply and demand are important too. The goal of this paper is to contribute to this research agenda by analysis of regional and intensity shifts in agricultural production and consumption around the world.

To achieve the goal of the paper, the index decomposition analysis (hereafter, IDA) is used. The IDA is a commonly adopted tool for determination of the impact of indicators such as population, economic activity and its structure, technology and possibly other factors on a chosen indicator (Ang [2]). Nevertheless, alternative approaches, such as econometric decomposition analysis (Stern [9]) or structural decomposition analysis (Hoekstra & van der Bergh [6]) have been proposed too. Each approach has different data requirements and provides different pieces of information but there can also been found several common features[1]. This paper uses the IDA to isolate the impacts of changes in labor productivity, capital formation, technology and constitution of land on world food production and impacts of shifts in GDP, its structure and population on world agricultural consumption. Namely, logarithmic mean Divisia index method II (hereafter LMDI II) proposed by Ang and Choi [1] or Ang [4] is chosen as appropriate tool for the analysis.

The rest of the paper is organized as follows: Section 2 describes the main stylized facts for the food production and consumption in past years. Section 3 describes the methodology behind the IDA and section 4 shows the results of agricultural production and consumption decomposition analysis. Section 5 concludes.

[1]IREAS, Institute for Structural Policy, Štěpánská 45, 110 00, Prague 1, Czech Republic. Jan Evangelista Purkyně University in Ústí nad Labem, e-mail: vitezslav.pisa@centrum.cz

[2]Jan Evangelista Purkyně University in Ústí nad Labem, e-mail: jan_bruha@yahoo.co.uk

[1]For example, structural decomposition analysis is often cited together with IDA (eg. Hoekstra & van der Bergh [6]). In both methods, Laspeyeres (weighting by base year), Paasche (weihgting by target year) and Marshall-Edgeworth (weighting by the mean of base and target year) indices can be applied. However, only in IDA approach, different types of Divisia indices can further be employed.

2 Data and stylized facts

The data for the analysis come mainly from the World Bank website.[2] The world has been divided among seven aggregated regions: Europe, non-European developed countries, Middle East, rest of the former USSR, the rest of Asia, Africa, and Latin America. Grouping of the region respect preferentially economic structures of particular states and their location.

For the purpose of the analysis, we use the value added in agricultural sector measured in 2000 US dollars as variable expressing regional agricultural production[3]. The real agricultural value added increased about 42 % between the year 1990 and 2007. The growth can be split into two periods: the years (1990 - 1995) witnessed moderate increase oscillating around 1 % per year. Moreover a very negligible drops in real value added were recorded in the years 1991 and 1995. Since the year 1996 the total real value added in agriculture has permanently been increasing with annual growths oscillating around 2 % and ranging from 1,6 % to 4,8 %.

Food consumption in the selected regions is computed as the real value added created in the agriculture plus the real value of net agricultural exports. The total indicator of food consumption in selected regions grew annually about 2 % on average. Real agricultural consumption increased the most in Asian, African and Middle East countries (around 3 % a year on average). A relatively high annual growth in real consumption was also indicated in the case of non European developed countries (1.47 %). Looking at regional composition of overall food consumption, Asian countries increased their share from 30 % to 37 % in the analyzed period while African countries raised their share from 2.7 % to 3.1 %. The share of European countries on overall food consumption declined by approximatively 4 % from 20.6 % to 16.4 %.

3 The index decomposition analysis

3.1 General theory

The goal of the IDA is to understand historical changes in a social, economic, environmental, or agricultural indicator, and to gauge the driving forces or determinants that underlie these changes. The application of the IDA to agricultural indicators has been used especially in assessing the influence of the population size, the amount of arable land, sectoral shifts, capital formation, or technology changes.

Let us consider the indicator Φ, which is given as:

$$\Phi_t = \Upsilon_t \sum_i \phi_{1it} \dots \phi_{Mit}, \tag{1}$$

where Υ_t is the scale measure[4], and the summation runs over countries, commodities, or another interesting dimension. The goal is to decompose the change in the indicator into a number of determinants.

If observations were available in continuous time, the decomposition would be straightforward: the percentage change in the indicator $\dot{\Phi}_t/\Phi_t$ could be written as follows:

$$\frac{\dot{\Phi}_t}{\Phi_t} = \frac{\dot{\Upsilon}_t}{\Upsilon_t} + \frac{\sum_i \frac{\dot{\phi}_{1it}}{\phi_{1it}} \phi_{1it} \dots \phi_{Mit}}{\sum_i \phi_{1it} \dots \phi_{Mit}} + \dots + \frac{\sum_i \frac{\dot{\phi}_{Mit}}{\phi_{Mit}} \phi_{1it} \dots \phi_{Mit}}{\sum_i \phi_{1it} \dots \phi_{Mit}} \tag{2}$$

where $\frac{\dot{\Upsilon}_t}{\Upsilon_t}$ is the growth in the scale measure, and the expression $\frac{\sum_i \frac{\dot{\phi}_{mit}}{\phi_{mit}} \phi_{1it} \dots \phi_{Mit}}{\sum_i \phi_{1it} \dots \phi_{Mit}}$ could be interpreted as the weighted percentage change in the factors ϕ_{mit}. The problem is that observations are not available in continuous time, and therefore discrete-time approximations should be used.

[2] World Development Indicators & Global Development Finance downloadable at: *http://databank.worldbank.org.*

[3] Authors are aware that selected indicator need not necessarily correspond with nutritional values that particular states produce, i.e. the question of ability to feed the world population need not exactly be addressed.

[4] Such as the total population if the aggregate food consumption is investigated, or the total agricultural land if the production is investigated, but it may represent also the real GDP if the relation between wealth and food production/consumption is investigated

A discrete-time decomposition approximation can adopt an additive or a multiplicative mathematical form. The additive form decomposes the difference in the indicator Φ between times t_1 and t_2 into the sum of determinants D_i and a residual term $\tilde{R}$:

$$\Phi_{t_2} - \Phi_{t_1} = D_1 + D_2 + \ldots + D_N + \tilde{R}. \tag{3}$$

The multiplicative form decomposes the relative growth of the indicator into the product of determinant effects:

$$\frac{\Phi_{t_2}}{\Phi_{t_1}} = D_1 \times D_2 \times \ldots \times D_N \times \tilde{R} \tag{4}$$

A number of mathematical forms for the additive as well as multiplicative decomposition forms has been proposed. Ang[2], [4] provide useful overviews of mathematical forms and their useful properties. The following four properties are particularly relevant to the index decomposition analysis:

Exactness: an exact decomposition has no residual; in the additive case this means that the residual equals 0, while it equals 1 in the multiplicative case.

Time reversal: the decomposition satisfies this property whenever the decomposition yields the reciprocal results after the reversal of the time periods.

Factor reversal: concerns the invariance with respect to the permutation of determinants.

Robustness: a decomposition is robust if it does not fail when it comes across zero (or even negative) values in the dataset.

3.2 LMDI II

This paper applies the LMDI II suggested by Ang and Choi [1] or Ang [4] as the preferred method under a wide range of circumstances: the LMDI II satisfies the four requirements mentioned above and has no residual (i.e. $\tilde{R} = 0$ in the additive case and $\tilde{R} = 1$ in the multiplicative case.[5]) While LMDI II has both a multiplicative and an additive form the multiplicative form will be applied for subsequent analysis.

The multiplicative LMDI II is defined as follows:

$$D_j^{t_2,t_1} \equiv \exp\left(\sum_i \frac{\mathcal{L}(\Phi_{it_2}, \Phi_{it_1})}{\mathcal{L}(\Phi_{t_2}, \Phi_{t_1})} \log\left(\frac{\phi_{jit_2}}{\phi_{jit_1}}\right)\right), \tag{5}$$

where $\Phi_{it} \equiv \prod_{j=1}^{m} \phi_{jit}$ and $\mathcal{L}$ is so-called logarithmic average:

$$\mathcal{L}(x_1, x_2) \equiv \begin{cases} \frac{x_1 - x_2}{\log x_1 - \log x_2} & \text{if } x_1 \neq x_2 \\ x_1 & \text{otherwise.} \end{cases}$$

The residual term satisfies $R = 1$, since the *LMDI II* is an exact approach.

The *intensity effect* is than given as:

$$D_a^{t_2,t_1} = \exp\left(\sum_i \frac{\mathcal{L}(a_{it_2}s_{it_2}, a_{it_1}s_{it_1})}{\mathcal{L}(\sum_j a_{jt_2}s_{jt_2}, \sum_j a_{jt_1}s_{jt_1})} \log\left(\frac{a_{it_2}}{a_{it_1}}\right)\right),$$

and the *structure* effect is given as follows:

$$D_s^{t_2,t_1} = \exp\left(\sum_i \frac{\mathcal{L}(a_{it_2}s_{it_2}, a_{it_1}s_{it_1})}{\mathcal{L}(\sum_j a_{jt_2}s_{jt_2}, \sum_j a_{jt_1}s_{jt_1})} \log\left(\frac{s_{it_2}}{s_{it_1}}\right)\right).$$

[5]This is advantage since decomposition based on Laspeyres indices could suffer from large unexplained residuals in some cases.

4 Empirical Results

The decomposition is provided both for agricultural production and agricultural consumption for seventeen year period 1991-2007. The explanations of changes in both indicators are, however, different.

4.1 The Agricultural Production

In the case of agricultural production, the *scale effect* explains the changes in real production in agriculture as a result of changes in total agricultural employment $\mathcal{L}$. Further, the *structural effect* D_s is the result of changes in composition of agricultural land according to selected regions over the world. The *structural effect* is assumed to be negative since the agricultural production gradually moves to less productive world regions as the best land tend to be used first; this is so-called Ricardo effect. The *intensity effect* is in the case of agricultural production consequently decomposed into following sub-effects:

Intensity effect of employment D_a: reflects changes in number of employees per land in agriculture, the assumption is that growing productivity in agricultural sector will lead rather to decline in this indicator because there is pressure on reduction of employees in agricultural sector and simultaneous increase in value added in agriculture,

Intensity effect of capital utilization D_b: are changes in the capital intensity per employee (the capital intensity is measured as value added per one tractor). This effect is expected to be positive as higher capital intensity should lead to an increase in the agriculture output per worker.

Equations (1), (4), and from the common known approximation $\log(\frac{X_2}{X_1}) \cong \frac{X_2}{X_1} - 1)$ imply:

$$\frac{A^s_{t2} - A^s_{t1}}{A^s_{t1}} = \frac{\mathcal{L}_{t2} - \mathcal{L}_{t1}}{\mathcal{L}_{t1}} + \log D_a + \log D_b + \log D_s, \tag{6}$$

where A^s stands for real agricultural value added, $\mathcal{A}$ is total area of agricultural land and whole fraction addresses scale effect, logarithms of D_a, D_b are particular intensity effects, and the logarithm of D_s shows structure effect.

The lower subfigure of Figure (1) shows the decomposition of changes in the real value added (production). The increases in value added are mainly driven by the scale effect and capital equipment of labor. The employment intensity effect adversely affects changes in the value added during the whole analyzed period. The structural effect rather shows that the agricultural production is indeed moved to poorer countries. The structural effect is weakened by the fact, that the value added in agriculture naturally increases also in developed countries as a result of production of 'more luxurious' agricultural products.

A deeper analysis of the regions shows that the value added per square kilometer of agricultural land favorably developed mainly in Asian countries, which supported growths in the total agricultural value added. Growing productivity was also witnessed in the European region up to the year 2003, then after a relatively high improvement in the year 2004, continuous drops were observed in the years 2005, 2006 and 2007. In Europe, the main driver for the highest agricultural production increase in observed period was a relatively large improvement in value added per agricultural land and a large decrease in employees.

4.2 The Real Agricultural Consumption

In the case of agricultural consumption, the *scale effect* explains the changes in real consumption of agricultural goods as a result of changes in total population L with an expectation of positive relationship since more people are assumed to consume more food. Furthermore, the *structural effect* reflects changes in composition people in the world. The idea of this effect is that negative effect shows growing number of people in poorer regions with smaller demand on food. The *intensity effect* is in the case of real agricultural consumption split into two sub-effects:

Intensity effect of GDP D_a: is measured by changes in ratio of GDP per capita, it is assumed a positive impact on consumption since higher income usually lead to higher consumption,

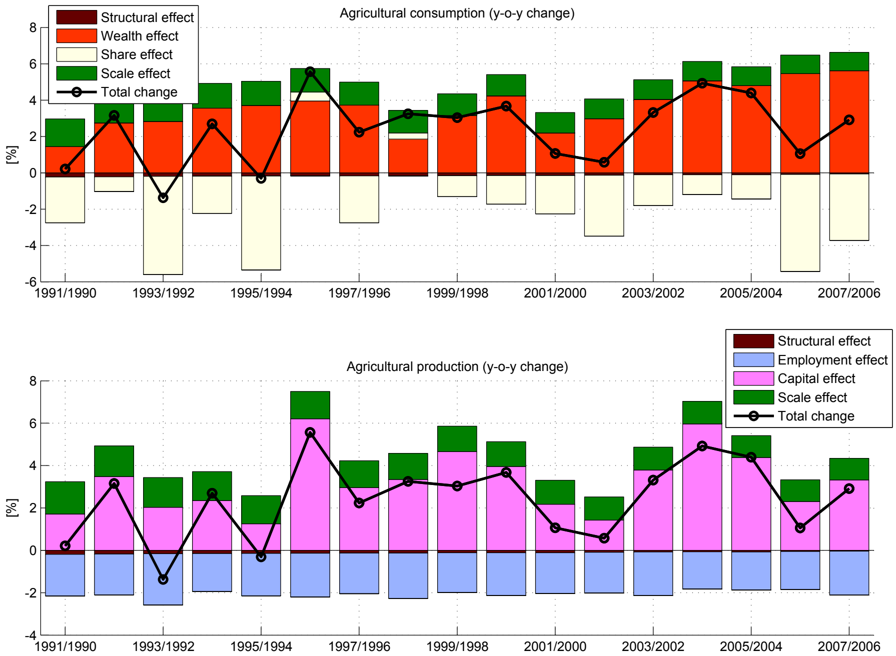


Figure 1: Results

Intensity effect of agricultural consumption share D_b: reflects changes in share of consumption of food in GDP, the effect should be negative since food is a basic commodity and the share of expenditure on food in total income declines as the society becomes richer.

Again, using the same approximation we get following formula:

$$\frac{A_{t2}^d - A_{t1}^d}{A_{t1}^d} = \frac{L_{t2} - L_{t1}}{L_{t1}} + \log D_a + \log D_b + \log D_s, \tag{7}$$

where A^d stands for agricultural demand/consumption, L expresses population, D_a intensity effect of GDP, D_b intensity effect of agricultural consumption share and D_s structural effect.

The upper subfigure of Figure (1) shows the decomposition of changes in real agricultural consumption. It can be seen that annual changes in world agricultural consumption range from approximately - 2 % to 4.7 %. The growths are mainly driven by intensity effect of GDP. In addition, the growing population also significantly contributed to increases in agricultural demand. On the contrary, the share of consumption in GDP mainly hampered the growth (the only exception are the years 1996 and 1998). The structural effect is relatively small but its negative direction occurred as expected.

Inspecting the data more deeply, it can be seen that the share of people with lesser food demand (i.e. in the rest of Africa and Latin America) was gradually increasing causing the structural effect to be negative. Intensity effect of GDP was mainly driven by Asian, non European developed and European countries.

5 Conclusion

This paper attempts at explaining the changes in the world agricultural demand and supply by changes of other relevant socioeconomic and agricultural indicators. For the analysis, IDA was chosen as appropriate tool with LMDI II as flagship to this approach. Generally, we witnessed moderate increase in both, world agricultural consumption and world agricultural production (the only exception is the former USSR where the level of value added is in the year 2007 approximately the same as in the year 1990). The changes in both variables were described by changes in level, intensity and structural effects. The world agricultural production is explained by changes in total agricultural employment (scale effect), regional composition of land (structural effect), productivity of labor and capital utilization (intensity effects). In addition, evolution of world food consumption can be described by changes of total population (scale effect) and its composition over the world (structural effect), changes in GDP per capita and structure of consumption (intensity effects). The structural effects are relatively negligible compared to scale and intensity effects.

Acknowledgements

The research on this paper has been supported by the grant of the Grant Agency of the Czech Republic **No. 402/11/2500**. The support is gratefully acknowledged.

References

[1] Ang, B.W., Choi, K.H. (1997). Decomposition of aggregate energy and gas emission intensities for industry: a refined Divisia index method. *The Energy Journal 18(3)*, pp. 59-73.

[2] Ang, B.W. (1999). Decomposition methodology in energy demand and environmental analysis. In van den Bergh J.C.J.M. (Ed.) *Handbook of environmental and resource economics*, pp. 1165-1179, Edward Elgar.

[3] Ang, B.W., Liu F. L. (2001). A new energy decomposition method: perfect in decomposition and consistent in aggregation *Energy 26 (6)*, pp. 537-548.

[4] Ang, B.W. (2004). Decomposition analysis for policymaking in energy: which is the preferred method? *Energy Policy 32*, pp. 1131-1139.

[5] Ang, B.W., Liu F.L., Chung, H.S. (2004). A generalized Fisher index approach to energy decomposition analysis. *Energy Economics 26*, pp. 757– 763.

[6] Hoekstra, R., van der Bergh, J.C.J.M. (2003). Comparing structural and index decompostion analysis. *Energy Economics 25*, pp. 39-64.

[7] Islam, N. (1995). Population and Food in the Early Twenty-First Century: Meeting Future Food Demand of an Increasing Population. International Food Policy Research Institute, Washington DC.

[8] Seigel, I.H. (1945). The 'ideal' index-number formula. *Journal of the American Statistical Association 40*, pp. 520–523.

[9] Stern, D. (2002). Explaining Changes in Global Sulfur Emissions: An Econometric Decomposition Approach. *Ecological Economics 42*, pp. 201-220.

The optimization process of the stock control management by means of the random method of storing

Miroslav Plevný[1], Tereza Kubišová[2]

Abstract. Supplying firms offering their customers parts or components for final assembly are forced, due to strong competition and pressure to reduce prices, to meet the uncompromising requirements of the customers for accuracy and timeliness of their supplies and their quality. It is therefore necessary to cope with the big logistic flows in the field of internal logistics. Suppliers in the car industry may be a good example of such firms.
The article presents one of the possible methods how to estimate the benefits related to the change of the stock control management aimed at choosing a suitable strategy of placing items in the rack section of the store. A procedure of calculating the estimated savings in the storage area after the transition from the fixed position storage to the random (chaotic) method of storage supported by the appropriate software programme is the main focus of the presented paper. In the conclusion other benefits for a company are outlined that are also related to such a change.

Keywords: stock management, random storage, minimization of the storage area.

JEL Classification: C15, O14
AMS Classification: 90B05, 90B90

1 Strategy of a store arrangement

Storing and keeping at least a certain amount of stocks influences the competitiveness of companies. That is why reducing stocks and limiting storage area becomes a focal point for the management. A high level of stocks presents a high financial burden for a company in the form of unused capital. Tidying the company´s own storage area by means of a transition from storing in the fixed positions to random storing assisted by the appropriate software might be one of the suitable solutions. This way it is possible to achieve a more effective use of storage areas and thanks to faultless records of stocks it is even possible to reduce the amount of stocks in the store. [2] It is possible to show [1] [8] that even by using very simple exact methods it is possible to achieve significant savings of the company expense.

In professional literature arranging the store is sometimes called a placing strategy. The placing strategy determines the allocation of stockkeeping units between the storage zone and storage positions. The main aim of the strategy of the storage arrangement is to satisfy the demand by means of as few storage locations as possible. Safety, accessibility, short storage cycle and a minimum number of transfers are secondary aims. In reality the following basic ways of store arrangement are the most common:

- **Fixed arrangement or storing in the reserved place**
  In this way the fixed storing locations are reserved for the maximum stock level that is expected in connection with a particular item and blocked for all the others.
  This system consists in the fact that the same products are always stored in the same position and therefore a sufficient method of labelling the individual storage positions and the staff´s good knowledge of the store are important here. The material in this case is stored according to numbers and the speed of turnover or according to other criteria, e.g.: compatibility, complementarity and popularity of products.

- **Random (chaotic) storing**
  Vacant storage locations may be used for storing any random material item. Random storing is currently managed by a software application specifically designed for storing thanks to which the company is well informed about the individual components and by this for example the FIFO monitoring (or any other monitoring system) is enabled. [6]

- **Zone arrangement**
  It is a compromise between the random and fixed store arrangements. This method enables increasing the capacity of the demand resulting from higher reserves as these are necessary for some items in the separated zones. This way of storing is often used for storing the material requiring special treatment. It is typical

[1] University of West Bohemia/Faculty of Economics, Husova 11, Pilsen, CZ, plevny@kem.zcu.cz.
[2] University of West Bohemia/Faculty of Economics, Husova 11, Pilsen, CZ, kubisova@students.zcu.cz.

namely for chemical industry in which the individual chemicals have to be stored under exactly stipulated conditions.

In this article we will show a possible example of the optimization process of the stock control management from the point of view of the change of the storing system. The aim is, apart from the savings, to make logistic processes more effective, to reduce the number of errors caused by human factor thanks to the FIFO mode monitoring. This way the customer audit gets better results and the overall image of the company is likely to improve.

The placing strategy has been chosen as a solution, i.e. the allocation of stockeeping units among the zones and storing positions by means of the lowest number of the storage locations. This means especially the transition from storing <u>in the fixed positions</u> to <u>the random, computer assisted way of storing</u>. In the fixed storage system the fixed storing locations are reserved for the maximum expected stock level of a particular item and blocked for all the others. In the random storage system the storage locations may be used for storing any random material item and the store is managed by a software application thanks to which the company is well informed about the individual components, which enables monitoring such processes as for example stocking out in the FIFO mode. Thanks to this transition a more effective use of the storage area is to be achieved and thanks to the faultless records of stocks the amount of stocks in the store is to be reduced. [4][6][10]

2 The original method of storing

This method uses rack and block systems for storing several types of pallets. [3] The material and documents entry is done by means of the company main standard information system. It is the company information system that is linked to the financial module of the SAP system. According to the delivery notes the identification cards are printed with the information about the individual items which are then fixed to the packaging of the individual components. The storage itself is carried out according to these cards, the components are always placed in the allotted position on the rack or block with the same labelling.

The extraction from the store is based on the FIFO mode, but its observance depends on the human factor and any possible nonobservance may only be revealed by looking into the system. The material is transported to a transfer point where another employee takes it over. This employee carries out a system transfer of the material by means of the above mentioned information system.

3 Introducing the random way of storing

The random way of storing may be introduced in such a part of the store that is equipped by a palette rack system in which there are 2,800 palette positions in a monitored store. For the random store to function the wireless technology and bar codes has been chosen and a new information logistic system has been installed as well. This system has also a few useful functions such as records of transactions, reporting, a stocktaking mode etc.

Safety is another advantage as thanks to the system of user accounts the individual employees operating in logistics have access to some functions only. Some new hardware is necessary for these purposes, namely the following devices:

- bar code scanners,
- bar code label printer,
- wi-fi routers.

Labelling the individual palette positions in the store is another necessary prerequisite.

The new system based on random storing works as follows. The entry of material must be entered in the system and this enables the identification cards to be printed using the new feature of bar codes. These cards are then affixed to the individual packages of the material. The labelled material that is unloaded from the means of transport is registered by the system as being in the REC (receiving) position. From this position the material is then stored into a random location in the store which is also labelled by a bar code. Other operations in the store are run in a similar way, as for example transfers, quality control and the like.

The extraction of the material into the production takes place on the basis of an order created by the operators in the production by scanning the code of the needed material. The storeman scans the code of the order and the location appears on the display of the reader where the required material is located together with the information about the oldest date of receipt, and this way the FIFO regulations are complied with. The material is then transported to the transfer point. Here it is taken over by the company production operators. The workplace in which the system transfer used to be carried out can be cancelled as transfers are carried out automatically. When the above procedure is completed another order may be entered into the system. The system, apart from the above described process, also enables other functions, such as: information in the package, as the scanner reader shows the number of the component, number of pieces in the package and the location where the component is located;

quality control; dispatch of material and stocktaking, in other words mapping if the existing state of the stored palettes in the individual locations corresponds to the state in the system.

4 The estimate of the savings in the storage locations

The introduction of the random method of storage assumes **a better use of the storage locations**. In reality this means savings in the palette positions. The part of the store formed by these newly created vacant palette locations can be removed, or, as the case may be, used for other purposes.

4.1 Measuring a store utilization

In the below analysis of a store utilisation the vacant palette positions on the racks were physically monitored and counted within a period of a two months at the beginning of the year 2011. 33 measurements were taken in total. The measured values shown in Table 1 are stated in the order in which they were taken.

Measurement No.	Number of vacant positions	Measurement No.	Number of vacant positions	Measurement No.	Number of vacant positions
1	**944**	12	**1021**	23	**905**
2	**1067**	13	**966**	24	**969**
3	**923**	14	**917**	25	**1043**
4	**990**	15	**986**	26	**921**
5	**851**	16	**1016**	27	**958**
6	**932**	17	**970**	28	**877**
7	**1097**	18	**951**	29	**1011**
8	**945**	19	**947**	30	**981**
9	**919**	20	**1009**	31	**966**
10	**953**	21	**973**	32	**923**
11	**891**	22	**930**	33	**953**
average: $\bar{x}$ = 961					

Source: Own

Table 1: Vacant palette positions in the store

As Table 1 shows, it was found that, on average, out of the total K = 2,800 rack palette positions in the store **$\bar{x}$ = 961 of them are vacant, which represents 34.3%** of the total capacity of the store.

4.2 Statistical analysis of the measured figures

The measured values of the vacant palette locations in the given store show, as expected, a random character. From the character of the measured value it is possible to assume that it is a random value with a normal distribution. Let us first test the assumption by the Kolmogorov-Smirnov test (hereinafter only K-S test) which belongs to the non-parametric tests of the statistical hypotheses. Unlike the test χ^2 it is possible to use it even in the case of a small size sample, and it is based on the originally observed values and not the figures categorized in groups.

The test may be used to verify the hypothesis that the acquired sample arise out of the distribution with the continuous distribution function $F_0(x)$. In place of the theoretical parameters (which fully determine the given function) we have to use their estimates in practice often. Let us formulate the null hypothesis as the equivalence of the empirical distribution function $F_n(x)$ and the theoretic distribution function $F(x)$, i.e.

$$H_0: F_n(x) = F_0(x), \quad -\infty < x < \infty \qquad (1)$$
$$H_1: F_n(x) \neq F_0(x), \quad -\infty < x < \infty$$

Following statistics will be used as the test criterion

$$D_n = sup\ \left| F_n(x) - F_0(x) \right| = max\ (D_1^*, D_2^*, \ldots, D_n^*), \qquad (2)$$

where

$$D_i^* = \max\left\{ \left| F_0(x_i) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F_0(x_i) \right| \right\} \qquad \text{pro } i = 1, 2, \ldots, n. \qquad (3)$$

The hypothesis H_1 has at the level of significance α a critical domain the set of values $D \geq D_n,(1-\alpha)$, where $D_n,(1-\alpha)$ are $100(1-\alpha)$% quantiles of K-S statistic if the hypothesis H_0 is valid. If the value of the test criterion

D is smaller than the quantile D_n,(1– α), we do not reject the hypothesis H_0 i.e. the equivalence of both distribution functions. Hereafter it is possible the examined sample treat as a sample from the set with the distribution function $F_0(x)$. If testing normality we usually take as $F_0(x)$ the distribution function of the normal distribution, where for the mean value μ and variance σ^2 we use their estimates, e.g. arithmetic average and selection variance. [5]

The following parameters related to the sample were determined by a standard procedure:

- arithmetic average $\overline{x}$ = 960,76
- sample variance s^2 = 2757,75
- standard deviation s = 52,51

The final value of the test criterion was calculated by the above procedure:

$$D_{33}^* = \quad 0{,}10480166 \tag{4}$$

The value of the criterion (4) is compared with the quantiles for the K-S test stated in the following table:

	Quantiles of distribution D_n,(1- α) for various levels of significance				
(N)	0.20	0.15	0.10	0.05	0.01
30	0.190	0.200	0.220	0.240	0.290
35	0.180	0.190	0.210	0.230	0.270

Source: Statistica, 2011

Table 2: Quantiles of distribution D_n,(1- α) when H_0 is valid for the K-S test for one sample

For the number of observances being 33 the values of quantiles for the individual levels of significance are located within the range of the values of the last two rows. As is obvious from the table, the test criterion is smaller than the quantiles for all the given levels of significance α = 20%, α = 15%, α = 10%, α = 5% and α = 1%. We do not reject the null hypothesis and hereafter we will assume that the sample originate from the normal distribution.

Now it is possible to calculate the quantiles for the individual probabilities. The values of the quantiles of the normalized normal distribution for p=0,05 and 0,01 can be found for example in the tables [5]: $u_{0,05}$ = –1,6449; $u_{0,01}$ = –2,3263. By substitution into the formula $u_p = \frac{x_p - \mu}{\sigma}$, after conversion $x_p = u_p \cdot \sigma + \mu$ or for example by means of the function NORMINV in MS Excel it is possible to get the values of the quantiles of the value X, i.e. x_p [5]: $x_{0,05}$ = 874,38 ; $x_{0,01}$ = 838,59 (we used the estimates for the values μ and σ). From Table 1 it is obvious that the least measured value in the store was 851 vacant positions, which corresponds to the quantile in the range of 1 – 5%, and is nearest to the 2% quantile. An approximately 2% probability exists that in the store there will be fewer than 851 vacant palette positions. We will work with this value further.

4.3 The quantification of the safety stock level

For calculating the savings of the palette positions it is necessary to know the size of the overall safety stock level in the store, as the resulting savings theoretically correspond to one half of the overall capacity of the store reduced just by the overall size of the safety stock level.

As a Just-in-Time supplier the firm is forced to keep a relatively high safety stock level. For further proceeding let us assume that the total safety stock level remains unchanged. This article does not deal with determining the optimum safety stock level. The safety stock level was in the past set subjectively when defining the size of the reserved area for each individual item. At present there are no records of its size in the company (only the overall stock level of a given item is monitored). The company only gave the estimated extent of the size of the safety stock level in the volume of approximately 30-50% of all the stock. To be able to carry out further calculations it will be necessary to estimate the total size of the safety stock level.

For further calculations let us consider two different inventory models. The first one assumes an approximately constant speed of drawing stocks from the store for each item, when, at the same time, the length of the replenishing cycle is constant. It is the well-known EOQ model, e.g. [7].

As is obvious from Figure 1, in this case the average size of the unused positions in the store is equal to one half of the size of one consignment. From this it is easily possible to express an estimate of the existing size of the safety stock level w from the known size of the store capacity K and the average volume of the unused positions $\bar{x}$ as follows:

$$w = K - 2\,\bar{x} \tag{5}$$

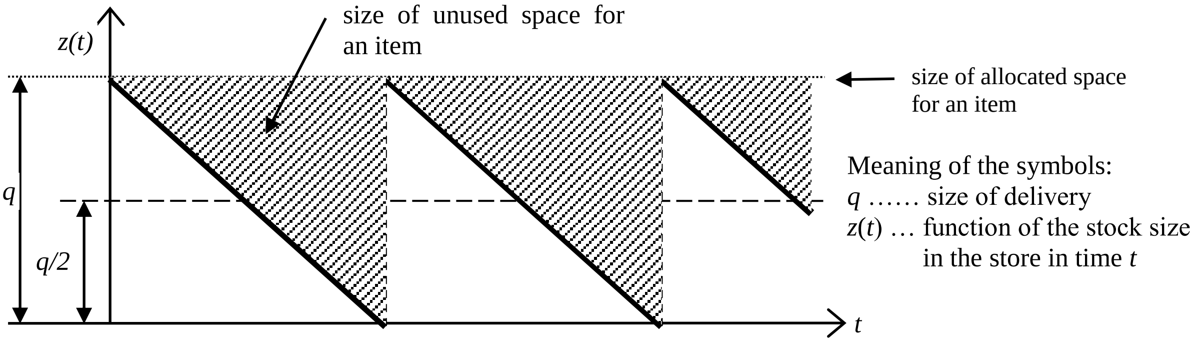


Figure 1: Model 1 – a typical course of a stock level function z(*t*)

From Fig. 1, on the basis of the similarity of the triangles, it is also obvious that the below relationship is valid:

$$\bar{x} = q/2 \tag{6}$$

For the second considered inventory model let us consider a more complicated but in reality a more probable course of the development of stock levels. Similarly as in the first case let us consider constant deliveries of stocks *q* by means of the signal level of stocks *s*, but the speed of drawing stocks from the store will be different as well as the length of the individual cycles. A typical course of the function of the stock size is illustrated in Figure 2 where the following symbols are used:

w .... size of the safety stock level,
s .... level of the signal level of stock,
w_i …size of the drawn (–) or undrawn (+) safety stock level within the cycle *i*,
t_i …. length of the cycle *i*.

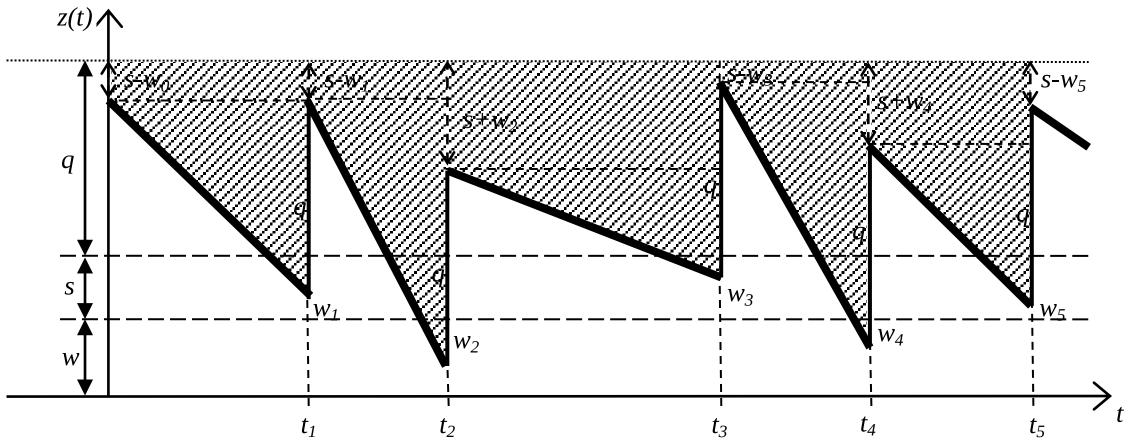


Figure 2: Model 2 – an example of the course of a stock level function z(t)

Let us further assume that the size of the allocated space is determined in such a way that it can hold a delivery even in case there is only minimum drawing within the replenishment time of the *i*-th delivery, i.e. it is given by the sum of the values $q + s + w$ (see Figure 2).

As is obvious from Figure 2, in this case the determination of the average size of the unused positions in store is much more complicated. Using the above stated symbols it is possible to write down the estimate of the average volume of the unused positions $\bar{x}_i$ in the *i*-th cycle in the length t_i as follows:

$$\bar{x}_i = \left[\frac{(q + w_{i-1} + w_i)\, t_i}{2} + (s - w_i)t_i\right] / \, t_i = \left[\frac{(q + w_{i-1} + w_i)}{2} + (s - w_i)\right] \tag{7}$$

For all the monitored period it is then possible to estimate the average volume of the unused positions $\bar{x}$ according to the following relation:

$$\bar{x} = \sum_{i=1}^{n} \left[\frac{(q + w_{i-1} + w_i)}{2} + (s - w_i)\right] / \, n \tag{8}$$

From the relation (8), after modifications, it is possible to get:

$$\bar{x} = q/2 + s + \sum_{i=1}^{n} \left[\frac{(w_{i-1} - w_i)}{2}\right] / \, n \tag{9}$$

Using the proven normality of the number of the vacant palette positions it is reasonable to assume that the value of the sum in the relation (9) is close to null and therefore the following relation is valid:

$$\bar{x} = q/2 + s \tag{10}$$

From the above it is possible to express an estimate of the size of the safety stock level *w* from the known size of the capacity of the store *K* and:

$$w = K - q/2 - \bar{x} = K - q - s \tag{11}$$

From the above it is obvious, when the assumptions for model 2 are valid, that the number of the vacant palette positions depends on the level of the signal level of the stock level because of the various volumes of drawing stocks in the period between the order and the delivery. If we express the level of the signal stock relatively as a proportion of the size of an order $s = k * q$, where $k \in \langle 0; 1\rangle$, then the relation (11) can be written as follows:

$$w = K - \bar{x}\,\frac{k+1}{1/2+k} \tag{12}$$

If we now summarize the findings resulting from models 1 and 2, then in case of model 1 the number of the vacant palette positions forms 50% of the current stock. If there are on average 961 vacant palette positions it is possible to estimate the size of the stock-in-trade as a double of this value, i.e. 2 x 961 = 1922 palette positions. The number that remains for the full capacity of the store to be reached should corresponds to the total size of the safety stock level, i.e. 2800 – 1922 = 878 of the palette positions, which is 31.4% of the overall capacity of the store. This figure corresponds to the bottom limit of the safety stock level of the given company.

In case of model 2 the estimate of the safety stock level size (11) depends on the level of the signal stock level. If we, for example, consider a relatively high level of the signal stock level in the amount of 50% of the delivery size, than, according to the relation (12) we gain, in our case, the estimate of the overall sum of the safety stock level in the following way: 2800 – 961 (1.5 / 1) = 1358.5 palette positions, i.e. 48.5% of the total capacity of the store. This figure corresponds well to the top limit of the safety stock level of the given company.

4.4 Calculating the savings of the palette positions

Let us summarize the known or found values. The size of the store is 2,800 palette positions; on average there are 961 vacant palette positions and there is an approximately 2% probability that there will be fewer than 851 vacant palette positions; the overall size of the safety stock level was determined in range from 878 to 1359 palette positions. It is obvious that the size of the safety stock level does not depend on the method of storing. Let us then assume that by the introduction of the random method of storing the safety stock level remains at the same level.

In reality, the savings that can be expected as a result of the transfer from the fixed to the random method of storing can be calculated as follows. On average there are 961 vacant palette positions in the store and there is a 98% probability that this value will not decrease lower than 851 palette positions. In reality, in the case of the random store, it is possible to reduce the storage area by the above mentioned circa 850 palette positions, which amounts to approximately **30.3% of the total capacity of the store** with the rack system.

Note: In case that after the above reduction there will be a temporary need of a larger storage area (approx. a 2% probability), it is possible to solve the problem in the given case by temporary placing the necessary number of the palettes in the part of the store meant for the block storage, which is the part of the store not dealt with in this study.

The calculated savings may be expressed as savings of a few rows of the rear racks or as savings of approx. 680 m^2 of the storage area. The company can use the free space in the store for other projects, or, as the case may be, they may consider renting the given storage area. With an assumed average price of the storage rentals of EUR 4.-/ m^2 per month the company might theoretically earn EUR 2.720.-, which is approx. CZK 65,280.- per month (assuming the exchange rate is CZK 24.-/EUR). [9]

5 Conclusion

The total quantifiable earnings of the introduction of the random method of storing as seen from the point of view of the storage area can be estimated in the amount of CZK 65,280.- per month. The introduction of the random store can, at the same time, help remove the operation of the system transfer on the transfer line, by which as many as three employees can be made redundant. This way the company may save another sum of approx. CZK 120,000.- per month (assuming the calculated total costs per employee are CZK 40,000.-/month).

It can be stated there are no doubts that introducing the random method of storing seems beneficiary for the company. Apart from the above stated quantifiable savings there are other benefits as well. The main ones are as follows: reducing the time necessary for searching the component in the store thanks to a clear and better quality material identification; accelerating the system of receiving and extracting the material; 100% compliance with the FIFO mode as it no more depends on the human factor but is monitored by the system; making the process of stocktaking more effective as it is done by means of bar code scanners; increasing the level of safety as a result of applying the user accounts; monitoring the work of the individual employees and possible detecting errors thanks to the function of the transaction records; observing the flow of material in all the stages of the logistic process and the like.

References

[1] DANĚK, J.; PLEVNÝ, M.; TEICHMANN, D. The rational operation of the urban transport line network by minimisation of the needed vehicles. *E & M Ekonomie a management*, 2010, Vol. XIII, No. 2, p. 53-61. ISSN 1212-3609.

[2] DORŇÁK, D., SCHWOB, R. *Snižujte zásoby! Materiálový sklad a spolupráce s dodavateli* [online] Journal *IT Systems*, 2008, Vol. 9, No. 5, p. 14, ISSN 1802-615X [cited 2011-02-21] Available at www: <http://www.systemonline.cz/aps-scm/materialovy-sklad-a-spoluprace-s-dodavateli.htm>

[3] *Ewals Cargo Care* [online] Česká Lípa: Ewals Cargo Care s.r.o. [cited 2011-01-12] Available at www: <http://www.ewals.cz/cz/home.html>

[4] GUDEHUS, T., KOTZAB, H. *Comprehensive Logistic*. London: Springer, 2011, p. 892, ISBN 3540307222

[5] HINDLS, R., HRONOVÁ, S., SEGER, J., FISCHNER, J. *Statistika pro ekonomy.* 7th ed. Praha: Professional Publishing, 2006, p. 422, ISBN 80-86946-16-9

[6] LAMBERT, D.M., STOCK, J.R, ELLRAM, L.M. Logistika. 2nd ed. Brno: CP Books, a.s., 2006, ISBN 80-251-0573-3

[7] LUKÁŠ, L. Pravděpodobnostní modely. Pilsen: University of West Bohemia, 2005, p. 210, ISBN 80-7043.

[8] PLEVNÝ, M. Application of Quantitative Optimization Techniques to Small Businesses. In *Mathematical Methods in Economics 2004*. Brno: Masaryk University Brno, 2004, pp. 256-260, ISBN 80-210-3496-3.

[9] *Realitní kancelář Svoboda a Wiliams* [online] Praha, Svoboda&Williams, [cited 2011-05-04] Available at www: <http://www.svoboda-williams.com/nemovitosti-v-praze/informace-ze-sveta-realit/pronajem-kancelarskych-obchodnich-skladovych-prosotr/>

[10] SCHENK, M., WIRTH, S., MULLER, E. *Factory Planning Manual.* London: Springer, 2010, 410 s., ISBN 978-3-642-03634-7

Construction of Monge matrices in max-min algebra

Alena Pozdílková[1]

Abstract. Max-min algebra and its various aspects have been intensively studied by many authors because of its applicability to various areas, such as fuzzy systems, knowledge management and others. Binary operations of addition and multiplication are replaced in max-min algebra by operations of maximum and minimum. The article is focused on the special type of matrices – Monge matrices. These matrices have been studied in max-plus algebra, in which they have various applications. Monge matrices in max-min algebra will be described in the article, first a special type – binary Monge matrices and then general Monge matrices. Monge matrices in max-min algebra can also be useful in simplifying some problems or algorithms. The basic properties of Monge matrices will be described. We study various methods of creating complex Monge matrices from the given ones. Construction of any Monge matrix will be shown. Minimal elements in a Monge matrix are investigated and their position in the matrix is described for many minimal elements, also operations with rows and columns of minimal elements will be described.

Keywords: Monge matrix, binary Monge matrix, max-min algebra.

AMS Classification: 90C49, 90C70, 90B80

1 What is max-min algebra

Max–min algebra $(\overline{R}, \oplus, \otimes)$ is an algebraic structure with two binary operations ⊕, ⊗ on the set $\overline{R} = R \cup \{-\infty, \infty\}$, which is an extension of the set of all real numbers. The operations minimum and maximum in max-min algebra are derived from the linear ordering in the set of real numbers.

The operations in max-min algebra are defined as follows:

$$\text{for } x, y \in \overline{R}: \ x \oplus y = \max(x, y),\ x \otimes y = \min(x, y).$$

For matrices A, B over $\overline{R}$ we define operations ⊕, ⊗ analogously as in linear algebra over R with addition and multiplication. We assume matrices A, B of suitable types.

The product of two vectors is illustrated on following example.

Example 1.

$$\begin{pmatrix} 3 & 7 & 2 \end{pmatrix} \otimes \begin{pmatrix} 5 \\ 4 \\ 9 \end{pmatrix} = (3 \otimes 5) \oplus (7 \otimes 4) \oplus (2 \otimes 9) = 3 \oplus 4 \oplus 2 = 4$$

2 Monge matrices in max-min algebra

Monge matrices in max-min algebra are special matrices which meet following condition: for all elements $a \in A$ and all $i, j, k, l \in \mathbb{N}$, where A is a matrix of type (m, n), $m \geq 2, n \geq 2,$ i and k are row indexes, where $i < k$ and j and l are column indexes, where $j < l$:

$$a_{i,j} \otimes a_{k,l} \leq a_{i,l} \otimes a_{j,k} \tag{1}$$

This condition has to be fulfilled by all quadruples of such elements from given matrix, which are placed in rectangle.

Following example illustrates positions of elements of (1) in Monge matrix.

[1] University of Hradec Králové, Faculty of Informatics, Dept. of information technology, Rokitanského 62, Hradec Králové, 500 03, Czech Republic, e-mail: alena.pozdilkova@uhk.cz.

Example 2.

$$\begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \boldsymbol{a_{i,j}} & \cdots & \boldsymbol{a_{i,l}} & \cdots \\ \cdots & \vdots & \ddots & \vdots & \cdots \\ \cdots & \boldsymbol{a_{j,k}} & \cdots & \boldsymbol{a_{k,l}} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

We could also formulate condition (1) in a following equivalent way:

$$a_{i,j} \otimes a_{k,l} \leq a_{i,j} \otimes a_{k,l} \otimes a_{i,l} \otimes a_{j,k} \tag{2}$$

From this formulation is obvious, that the minimum of the quadruple is $a_{i,j}$ or $a_{k,l}$. Consequently the minimum of each quadruple is the element in the upper left corner or in the bottom right corner. This formulation will be useful in description of minimal elements in Monge matrix.

Theorem 1. Elementary Monge matrix of type (2, 2) is a Monge matrix, if the minimal element is placed in the upper left corner or in the bottom right corner of this matrix - position (1, 1) or (2, 2).

Proof. This theorem follows from the Monge property (2).

Because direct method of verification of Monge property is very hard and for large matrices is necessary to verify many quadruples, it is useful to use following Theorem.

Theorem 2. Any matrix fulfils the Monge property, if all its sub-matrices of type (2, 2), consisting of two consecutive rows and columns, fulfill the Monge property.

Proof. First will be proved joining of two elementary Monge matrices.

Let us have matrices $M_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, M_2 = \begin{pmatrix} b & e \\ d & f \end{pmatrix}$ and the joined matrix $M = \begin{pmatrix} a & b & e \\ c & d & f \end{pmatrix}$. We will show that M is a Monge matrix. It means that for all quadruples must be fulfilled Monge property. For quadruples $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} b & e \\ d & f \end{pmatrix}$ it is clear, because $M_1\ and\ M_2$ are Monge matrices. We must show it only for quadruple $\begin{pmatrix} a & e \\ c & f \end{pmatrix}, let\ us\ denote\ it\ M_3$. There are four possible cases:

1. $\min M_1 = a, \min M_2 = b$: then $a < b, e, d, f$ and $\min M_3 = a$,
2. $\min M_1 = a, \min M_2 = f$: then $a < c, f < e$, so the minimal element in M_3 is placed in the upper left corner or the bottom right corner and M_3 is a Monge matrix,
3. $\min M_1 = d, \min M_2 = b$: this situation is not possible, because b, d are in both matrices M_1, M_2 ,
4. $\min M_1 = d, \min M_2 = f$: then $f < b, e, d$ and $d < a, c$ what implies $\min M_3 = f$.

In the same way we can construct a matrix of type (2, n) and then extend this matrix to the type (m, n), which is composed of Monge quadruples.

Example 3 -- Monge matrix of type (5, 5) with elements 0, 1, … 9.

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{2} & \mathbf{5} & \mathbf{7} \\ \mathbf{1} & \mathbf{3} & \mathbf{5} & \mathbf{5} & \mathbf{7} \\ \mathbf{1} & \mathbf{6} & \mathbf{7} & \mathbf{8} & \mathbf{3} \\ \mathbf{1} & \mathbf{9} & \mathbf{7} & \mathbf{4} & \mathbf{2} \\ \mathbf{2} & \mathbf{6} & \mathbf{3} & \mathbf{3} & \mathbf{0} \end{pmatrix}$$

The simplest type of Monge matrices are binary Monge matrices. All elements in binary Monge matrices may acquire only two values – 0, 1.

Example 4 – binary Monge matrix of type (5, 5).

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

3 Minimal elements in Monge matrices and their positions

Theorem 3. If the minimal element in Monge matrix of type (m, n) is unique, $\boldsymbol{m \geq 2, n \geq 2}$, then this element is placed in the upper left corner or in the bottom right corner of this matrix - position (1, 1) or (m, n).

Proof. This theorem follows from the Monge property (2) (minimum of all quadruples is always in the upper left corner or in the bottom right corner of this quadruple), what is clearly fulfilled for matrices of type (2, 2). If we apply this property for more elements, if the minimal element is not placed in the upper left corner or in the bottom right corner, the matrix is not Monge. This element must be pushed into one of corners on the basis of the Monge property. Proof by contradiction will be used. It is illustrated by matrix, which was Monge and minimal element is added to any position (except of the upper left or bottom right corners).

$$\begin{pmatrix} 3 & 3 & 8 & 8 \\ 5 & 6 & 9 & 6 \\ 5 & 6 & 4 & 4 \\ 9 & 5 & 4 & 1 \end{pmatrix} \xrightarrow{minimal\ element\ 0\ added} \begin{pmatrix} 3 & 3 & 8 & 8 \\ 5 & \mathbf{0} & 9 & 6 \\ 5 & 6 & 4 & 4 \\ 9 & 5 & 4 & 1 \end{pmatrix}$$

Now we want to convert this matrix to Monge matrix. We have more possibilities – one of them is to move the minimal element to the (1, 2) position. Because Monge property is not fulfilled for the quadruple $\begin{pmatrix} 3 & 8 \\ 0 & 9 \end{pmatrix}$, the minimal element must be moved for example to the upper left corner. All adjacent quadruple must be checked for Monge property. Therefore 0 in previous position can be replaced by 5. These steps are illustrated by this matrix, which is still not the Monge matrix.

$$\begin{pmatrix} 3 & \mathbf{0} & 8 & 8 \\ 5 & 5 & 9 & 6 \\ 5 & 6 & 4 & 4 \\ 9 & 5 & 4 & 1 \end{pmatrix}$$

The Monge property is not fulfilled for example for quadruple $\begin{pmatrix} 3 & 0 \\ 5 & 5 \end{pmatrix}$. One of possibilities is to move the minimal element 0 to the (1, 1) position and place for example element 3 instead of element 0.

Now all adjacent quadruple as well as the whole matrix fulfill Monge property, so we have the new Monge matrix with the minimal element in the in the upper left corner.

$$\begin{pmatrix} 0 & 3 & 8 & 8 \\ 5 & 5 & 9 & 6 \\ 5 & 6 & 4 & 4 \\ 9 & 5 & 4 & 1 \end{pmatrix}$$

There are many possibilities in this construction, but it is clear to prove, that all of them leads to the Monge matrix, in which is the minimal element 0 in the upper left corner or in the bottom right corner. ■

Theorem 4. If more minimal elements in Monge matrix of type (m, n) exist, then these elements are placed in this matrix convexly around the upper left corner or convexly around the bottom right corner. It means that minimal elements form a convex set around the upper left corner and the bottom right corner. These minimal elements can be in both these areas or only in one of them. Moreover in Monge matrix may exist rows or columns, in which are only minimal elements. These rows or columns can be placed anywhere in the Monge matrix.

Proof. The first part of proof is analogous as in the previous Theorem. Now will be proved, that Monge matrix can contain row or column with only minimal elements, and this row or column can be placed anywhere. Let us have any Monge matrix and we will prove that after addition of row of minimal elements this matrix will be still Monge. The Monge matrix after addition a row of minimal elements looks like this:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ min & min & min \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

For all quadruples, which does not contain elements from added minimal row, the Monge condition is fulfilled. If a quadruple contains elements from added minimal row, two elements will be from minimal row, two will be from origin matrix. Elements from minimal row form the minimal row of quadruple represented as a matrix of type (2, 2), which means that Monge property is fulfilled. Analogously we can prove addition of column of minimal elements. ■

Theorem 5. All sub-matrices obtained from Monge matrix are Monge matrices too.

Proof. This property clears from Monge property, because all quadruple, contained in sub-matrix, are contained in origin Monge matrix too and therefore the Monge property is fulfilled. ■

Theorem 6. If row or column of minimal elements or more rows or columns of minimal elements are added into Monge matrix, the matrix fulfills the Monge condition.

Proof. This property was proved in Theorem 4, the proof for more rows or columns is analogous. ■

Theorem 7. Row or column of minimal elements or more rows or columns of minimal elements can be deleted from Monge matrix of type (m, n), if $\boldsymbol{m} > \mathbf{2}$, respectively $\boldsymbol{n} > \mathbf{2}$.

Proof. This property clears from Theorem 5. ■

Theorem 8. Intersection of Monge matrices is a Monge matrix too.

Proof. This property clears from Theorem 3, because intersection of Monge matrices is a sub-matrix of any of these Monge matrices. ■

4 Creating more complex Monge matrices

In this chapter will be shown two ways, how to join Monge matrices. These ways can be used more times or combined to get more complex Monge matrices.

Theorem 7 – joining Monge matrices by using one common column.

Let us have two Monge matrices denoted M_1, M_2 (M_1 of type (m, n), M_2 of type (m, p)) which have one common column (the $n-th$ column of matrix $M_1$is the same as the first column of matrix M_2). Now we join these two Monge matrices into matrix M of type (m, n + p – 1) by using one common column. Then the matrix M is a Monge matrix.

Proof. This Theorem clears from Theorem 2 and its proof. ■

Remark 1. Previous construction shows transitivity in Monge matrices and can be clearly extended to matrices of type (m, n).

Remark 2. Monge matrices can be analogously joined by using a common row.

Theorem 8. – Joining Monge matrices without common column

Let us have two Monge matrices denoted M_1, M_2 (M_1 of type (m, n), M_2 of type (m, p)). Now we join these two Monge matrices side by side into matrix M of type (m, n + p). When Monge property is fulfilled for all quadruples, which are placed at $n-th$ and $n+1-th$ column and two adjoining rows, then M is a Monge matrix.

Proof. The proof follows from transitivity, which was proved in previous Theorem and Remark 1. ■

Remark 3. Quadruples need not necessarily be placed in two adjoining rows, by using transitivity can be proved, that quadruples can be in any rows.

Remark 4. The Theorem 8 can be analogously formulated for two Monge matrices above themselves, only by changing rows and columns.

5 Construction of Monge matrices

In this chapter construction of any Monge matrix will be shown. At first binary Monge matrix will be constructed and then this knowledge will be extended for Monge matrix with 3 different elements (0, 1, and 2), four different elements and so on.

5.1 Construction of binary Monge matrices

Binary Monge matrices can be constructed according to Theorem 2 by placing elements 0 convexly around the upper left corner or convexly around the bottom right corner. Other elements are 1. Row or column (or more rows or more columns) of elements 0 can be added according to Theorem 4.

$$\begin{pmatrix}0&0&.&.\\0&.&.&.\\0&.&.&0\\.&.&.&0\end{pmatrix}\xrightarrow{adding\ 1}\begin{pmatrix}0&0&1&1\\0&1&1&1\\0&1&1&0\\1&1&1&0\end{pmatrix}\xrightarrow{adding\ rows\ or\ columns\ of\ 0}\begin{pmatrix}0&0&1&0&1\\0&1&1&0&1\\0&0&0&0&0\\0&1&1&0&0\\1&1&1&0&0\end{pmatrix}$$

5.2 Construction of (0, 1, 2) Monge matrices

Monge matrices with 3 different elements (for example 0, 1 and 2) can be constructed in a similar way as binary Monge matrices. At first elements 0 will be placed convexly around the upper left corner or convexly around the bottom right corner. Remaining elements are 1 and 2 and will cover the rest of the matrix. The rest of the matrix is divided to more rectangular matrices with non-empty intersections, which cover the origin matrix, expect of elements 0. The rectangular matrices will be filled now. The first rectangular matrix (the first from the upper side of the matrix for this example) will be filled with elements 1 and 2 according to previous chapter (Construction of binary Monge matrices). Now we have filled an intersection of the first and the second matrix and it can be completed the second matrix (like a binary matrix with few pre-filled elements). Now we have filled an intersection of the second and third matrix and we continue until the whole matrix is filled. Eventually row or column (or more rows or more columns) of elements 0 can be added according to Theorem 4.

In following example the rest of matrix is covered with three binary rectangular matrices with non-empty intersections, the elements of intersection are highlighted in each step.

Example 5.

$$\begin{pmatrix}0&0&.&.&.\\0&.&.&.&.\\0&.&.&.&.\\.&.&.&.&0\\.&.&.&0&0\end{pmatrix}\xrightarrow{filling\ the\ first\ binary\ sub\ matrix}\begin{pmatrix}0&0&1&1&2\\0&.&\mathbf{1}&\mathbf{2}&2\\0&.&\mathbf{2}&\mathbf{2}&1\\.&.&.&.&0\\.&.&.&0&0\end{pmatrix}$$

$$\begin{pmatrix}0&0&1&1&2\\0&.&\mathbf{1}&\mathbf{2}&2\\0&.&\mathbf{2}&\mathbf{2}&1\\.&.&.&.&0\\.&.&.&0&0\end{pmatrix}\xrightarrow{filling\ the\ second\ binary\ sub\ matrix}\begin{pmatrix}0&0&1&1&2\\0&1&1&2&2\\0&1&2&2&1\\.&\mathbf{2}&\mathbf{2}&1&0\\.&.&.&0&0\end{pmatrix}$$

$$\begin{pmatrix}0&0&1&1&2\\0&1&1&2&2\\0&1&2&2&1\\.&\mathbf{2}&\mathbf{2}&1&0\\.&.&.&0&0\end{pmatrix}\xrightarrow{filling\ the\ third\ binary\ sub\ matrix}\begin{pmatrix}0&0&1&1&2\\0&1&1&2&2\\0&1&2&2&1\\1&2&2&1&0\\2&1&1&0&0\end{pmatrix}$$

$$\begin{pmatrix}0&0&1&1&2\\0&1&1&2&2\\0&1&2&2&1\\1&2&2&1&0\\2&1&1&0&0\end{pmatrix}\xrightarrow{adding\ rows\ or\ columns\ of\ 0}\begin{pmatrix}0&0&0&1&1&0&2\\0&1&0&1&2&0&2\\0&1&0&2&2&0&1\\1&2&0&2&1&0&0\\2&1&0&1&0&0&0\end{pmatrix}$$

5.3 Construction of Monge matrices in general way

Construction of general Monge matrices is only extension of construction of Monge matrices with 3 different elements.

1. At the first step elements 0 (minimal elements in general way) will be placed convexly around the upper left corner or convexly around the bottom right corner of the whole matrix.
2. At the second step the rest of matrix will be covered with rectangular sub-matrices with non-empty intersections. In these matrices will be elements 1 placed convexly around the upper left corner or convexly around the bottom right corner of these sub-matrices, sometimes the pre-filled elements in intersections must be preserved.

3. This algorithm will be repeated until left only two elements, and then the sub-matrices will be filled like binary Monge matrices.

4. Eventually row or column (or more rows or more columns) of elements 0 can be added according to Theorem 4.

Acknowledgements

The support of the Czech Science Foundation project # 402/09/0405 and the project INKOV #CZ.1.07/2.3.00/20.0001, financed from EU and Czech Republic funds, is kindly acknowledged.

References

[1] Butkovič, P., Zimmermann K.: *A strongly polynomial algorithm for solving two-sided system of (max, plus)-linear equations*. Discrete Applied Math 154 (2006), 437-446.

[2] Cunninghame-Green, R. A.: *Minimax Algebra.* Lecture Notes in Economics and Mathematical systems 166, Springer-Verlag, Berlin, 1979.

[3] Krbálek, P., Pozdílková, A., *Maximal solutions of two-sided linear systems in max-min algebra,* Kybernetika 46 (2010).

[4] Gavalec, M., Zimmermann K., *Solving systems of two sided (max,min)-linear equations*, Kybernetika 46 (2010), 405-414.

[5] Burkard, R., E.: *Monge properties, discrete convexity and applications*, European J. Oper. Res. 176 (2007), 1-14.

[6] Burkard, R., E., Klinz, B., Rudolf, R.: *Perspectives of Monge properties in optimization*, Discrete Applied Math. 70 (1996), 95-161.

[7] Gavalec, M., Plavka, J.: *An* $O(n^2)$ *algorithm for maximum cycle mean of Monge matrices in max-algebra*, Discrete Applied Math. 127 (2003), 651-656.

[8] Gavalec, M., Plavka, J.: *Structure of the eigenspace of a Monge matrix in max-plus algebra*, Discrete Applied Math. 156 (2008), 596-606.

[9] Gavalec, M., Plavka, J.: *Structure and dimension of the eigenspace of a concave Monge matrix,* Discrete Applied Math. 157 (2009), 768-773.

On Necessary Transversality Condition for Infinite Horizon Optimal Control Problems

Pavel Pražák[1]

Abstract. The present paper is intended as a brief overview that involves infinite horizon optimal control problems in continuous time. Such problems can be frequently found in economic dynamics models and especially in the literature dealing with the economic growth. First, the analogy of necessary conditions as in Pontryagin maximum principle will be presented. Then some remarks on transversality conditions follow. The paper contains a few solved and illustrative problems as well with the purpose to better understand the abstract concepts.

Keywords: optimal control, dynamic economic models, ordinary differential equation, Pontryagin maximum principle, discounted autonomous infinite horizon problems, transversality condition.

JEL Classification: C61, O20, O40
AMS Classification: 49K15

1 Introduction

During the second half of the 20th century the theory of optimal control proved to be very useful in modelling of some specific economic phenomena, see [4], [9], [12] or [13]. Many of economics models are formulated as finite time optimal control problems. Nevertheless there exist other models, for which it is problematic to find a reason why to use a finite time horizon. The infinite time horizon can be used to model a very long time period conveniently, see [1]. Typical infinite horizon problems include almost all growth models, see e.g. [3], that affirms that infinite horizon control optimal models are frequently used in economics modelling. Therefore it is worthwhile to know some mathematical issues concerning this problem. As valuable references dealing with infinite horizon optimal control and its applications to economics can be noticed [12], [5] or [1].

2 Infinite Horizon Control Problems

Before we start to deal with the infinite horizon problem we introduce basic notations. Let m, n and r be positive integers and let $I = [0,\infty)$ be the right-unbounded time interval called the infinite **planning horizon**.

2.1 State variable and control variable

Let us suppose that the state of an economic system at time $t \in I$ can be described by an n dimensional real column vector $\mathbf{x}(t) \in R^n$, $t \in I$, that is called the **state variable**. Suppose further that the state variable can be influenced by choosing an r dimensional real column vector $\mathbf{u}(t) \in U$, $t \in I$, that is called the **control variable** and where U is a closed subset of R^r called the **control set** or **control region.** It is assumed that the evolution of the state can be described by an ordinary differential equation

$$\dot{\mathbf{x}}(t)=g(t, \mathbf{x}(t), \mathbf{u}(t))$$

called **state equation** and that **initial conditions** and **terminal conditions** are given. Suppose that the function g: $I \times R^n \times U \to R^n$ is a continuous one and its first derivatives with respect to n arguments from R^n exist and are continuous over the set $I \times R^n \times U$. Now it is possible to define a „correct process" for optimal control.

2.2 Admissible Pair

The pair of functions $(\mathbf{x}, \mathbf{u})$: $I \to R^n \times R^r$ is called **admissible** for infinite horizon optimal control problem if $\mathbf{x}$ is continuous and piecewise smooth on the interval I, it means that its first derivative $\dot{\mathbf{x}}(t)$ with respect to the variable t is a piecewise continuous function on the interval I, $\mathbf{u}$ is piecewise continuous function on the set I and the state variable $\mathbf{x}$ satisfies **initial condition**, **state equation** and **terminal conditions**. The corresponding control variable $\mathbf{u}$ will be called the **admissible control**.

2.3 Objective Functional and Optimal Solution

The aim of the decision-maker is to choose an admissible control $\mathbf{u}$ in an optimal way. For this purpose, we suppose that there is an **objective functional** J: $R^n \times U \to R$ which is given by the improper integral

[1] University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, Hradec Králové, pavel.prazak@uhk.cz

$$J(\mathbf{x},\mathbf{u}) = \int_0^\infty f(t,\mathbf{x}(t),\mathbf{u}(t))dt .$$

The functional can be either convergent or divergent. It is the reason why it is necessary to use a slightly different approach to define the optimal process. There are several criteria of optimality, see e.g. [5] or [12].

3 The Maximum Principle for Infinite Horizon Control Problem

The proof that the maximum principle holds true for the optimal control problem on an infinite domain was originally given in the paper [8]. Here we present a result given in [12] Chapter 3. To present necessary conditions for optimal control process we use a function H: $I \times R^n \times U \times R \times R^n \to R$ that is called **Hamiltonian**. This function is defined as follows:

$$H(t, \mathbf{x}, \mathbf{u}, p_0, \mathbf{p}) = p_0\, f(t, \mathbf{x}, \mathbf{u}) + \langle \mathbf{p}, g(t, \mathbf{x}, \mathbf{u})\rangle,$$

where the symbol $\langle\cdot, \cdot\rangle$ means the scalar product of the **adjoint function p** and $g(t, \mathbf{x}, \mathbf{u})$. Now we can state the following version of maximum principle for infinite horizon problems. If $(\hat{\mathbf{x}},\hat{\mathbf{u}})$ is optimal for the infinite optimal control problem then there exist a constant p_0, with $p_0 = 0$ or $p_0 = 1$, and a continuous piecewise smooth function $\mathbf{p}$: $I \to R^n$ such that

(i) $(p_0, \mathbf{p}(t)) \neq 0 \in R^{n+1}$ for all $t \in I$;

(ii) the control $\hat{\mathbf{u}}$ maximizes the Hamiltonian $H(t, \mathbf{x}, \mathbf{u}, p_0, \mathbf{p}(t))$ for $\mathbf{u} \in U$, i.e.

$$H(t, \hat{\mathbf{x}}(t), \mathbf{u}(t), p_0, \mathbf{p}(t)) \leq H(t, \hat{\mathbf{x}}(t), \hat{\mathbf{u}}(t), p_0, \mathbf{p}(t))$$

for all $t \in I$ and all $\mathbf{u} \in U$;

(iii) wherever $\hat{\mathbf{x}}$ is continuous, the adjoint function $\mathbf{p}$ satisfies

$$\dot{p}_i = -\frac{\partial H(t,\hat{\mathbf{x}}(t),\hat{\mathbf{u}}(t), p_0,\mathbf{p}(t))}{\partial x_i},$$

where $i = 1, ..., n$.

3.1 Difficulties with Transversality Condition for Infinite Horizon Problems

The formulated version of maximum principle is significantly weaker than the maximum principle for the finite time horizon since there is missing the formulation of an analogue to the **transversality conditions**. The fact that transversality conditions are missing means that maximum principle does not give enough information to find candidates for optimality. Therefore it would be useful to find some versions of transversality conditions at least for a less general optimal control problem on the infinite domain. Before we deal with this problem in more details we shortly describe doubts connected with an analogue of transversality condition similar to finite horizon optimal control problems. It is well known that in the case of a finite horizon $T < \infty$ the optimal problem with $x(T)$ free has the transversality condition $p(T)$=0, see e.g. [12] Chapter 2. If we think of an infinite horizon problem with no terminal condition it would be natural to consider that the transversality condition is $p(t) \to 0$ as $t \to \infty$ but it seems that generally this is not true. First it was shown in [7] that the transversality condition, similar to the finite horizon problem, need not to be valid. Instead of this known example we introduce an example given in [2]. Another popular transversality condition is $p(t)\hat{x}(t) \to 0$ as $t \to \infty$. As we will observe at the second counterexample we have to be very careful with application of this transversality condition too.

Counterexample 1

Let us consider the following optimal control problem

$$\max_{(x,u)} \int_0^\infty e^{-t} \ln\frac{1}{x(t)}dt,$$

subject to

$$\dot{x}(t) = u(t) - x(t), x(0) = \frac{1}{2}$$

and $u\in[0, 1]$. First note that the optimal process can be found without using Pontryagin maximum principle. The solution to the state equation is

$$x(t) = \frac{1}{2}e^{-t} + \int_0^t u(s)e^{s-t}ds.$$

Since $u\in[0, 1]$ and the definite integral is monotonic we obtain

$$0 \leq \int_0^t u(s)e^{s-t}ds \leq 1 - e^{-t}.$$

Thus $1/2e^{-t} \leq x(t) \leq 1\text{-}1/2e^{-t} < 1$, $t \geq 0$. If we use this observation we can find that

$$0 < \ln\frac{1}{x(t)} \le \ln 2e^{t}.$$

The lower limit is valid for $u(t)$=1 and the upper limit is valid for $u(t)$=0. Hence the objective functional reaches its maximum for $\hat{u}(t) = 0, t \ge 0,$ and $\hat{x}(t) = 1/2e^{-t}, t \ge 0,$ which can be considered as the unique optimal pair.

Now we use Pontryagin maximum principle to formulate the necessary condition for optimal process. The Hamiltonian is given by

$$H(t, x, u, p, p_0) = -p_0 e^{-t} \ln x + p(u-x)$$

If we apply the maximum condition for the Hamiltonian we obtain

$$p(t)u(t) \le p(t)\hat{u}(t),$$

for all $u \in [0,1]$ and all $t \in [0,\infty)$. Since we know that $\hat{u}(t) = 0, t \ge 0,$ from the latter condition we gain that $p(t) \le 0$ for all $t \ge 0$. The adjoint equation is

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(t, \hat{x}(t), \hat{u}(t), p(t)) = p_0 e^{-t}\frac{1}{\hat{x}(t)} + p(t) = 2p_0 + p(t),$$

since $\hat{x}(t) = 1/2e^{-t}, t \ge 0.$

- Let us consider the first possibility p_0=0. Then $p(t)=p(0)e^{t} < 0$ for all $t\ge 0$ and $p(0)<0$, therefore $p(t) \to -\infty$ as $t\to\infty$.
- Now let us consider the second possibility p_0=1. Then $p(t)=(p(0)+2)e^{t} - 2$. Since we know that $p(t)\le 0$, $t\ge 0$, we can consider that either $p(0)=-2$ or $p(0)<-2$. In the first case we obtain $p(t)=2$, $t\ge 0$, and thus $p(t)\to 2$ as $t\to\infty$. In the second case we again obtain $p(t) \to -\infty$ as $t\to\infty$.

These results mean that, at each case, the limit transversality condition $p(t) \to 0$ as $t \to \infty$ is not valid.

Counterexample 2

Consider the problem

$$\max_{(x,u)} \int_0^{\infty} -u^2(t)e^{-\rho t}\,dt,$$

subject to $\dot{x}(t) = u(t)e^{-at}, x(0) = 0,$ with the terminal condition $\lim_{t\to\infty} x(t) \ge K$ and $u\in R$. The constants ρ, a and K are considered to be positive and $\rho - 2a < 0.$

If we consider that p_0 =1, the Hamiltonian is

$$H(t, x, u, p) = -u^2 e^{-\rho t} + pue^{-at},$$

which is concave in the variable u. Since

$$\frac{\partial H}{\partial u}(t, \hat{x}, \hat{u}, p) = -2\hat{u}e^{-\rho t} + pe^{-at} = 0$$

it follows that

$$\hat{u}(t) = \frac{1}{2}p(t)e^{(\rho - a)t}.$$

The adjoint equation is

$$\dot{p}(t) = -\frac{\partial H}{\partial x}(t, \hat{x}(t), \hat{u}(t), p(t)) = 0.$$

with the solution $p(t) = A$, where A is a constant. Thus

$$\hat{u}(t) = \frac{1}{2}Ae^{(\rho - a)t}.$$

The state equation for x becomes

$$\dot{\hat{x}}(t) = ue^{-at} = \frac{1}{2}Ae^{(\rho - 2a)t}, \hat{x}(0) = 0.$$

By integrating of this equation we obtain its solution

$$\hat{x}(t) = \frac{A}{2(2a - \rho)}(1 - e^{(\rho - 2a)t}).$$

Thus

$$\lim_{t\to\infty}\hat{x}(t) = \frac{A}{2(2a-\rho)}$$

and admissibility requires that

$$\frac{A}{2(2a-\rho)} \geq K,$$

which means that $A \geq 2K(2a\text{-}\rho)$. It is possible to show that if $A = 2K(2a\text{-}\rho)$ is chosen the optimal solution can be obtained. This quite tedious analysis is omitted here, for details see [12]. At any case, since $K > 0$ we obtain $A > 0$. This particularly means that

$$\lim_{t\to\infty} p(t) = A > 0.$$

Moreover the present example also shows that

$$\lim_{t\to\infty} p(t)\hat{x}(t) = \frac{A^2}{2(2a-\rho)} > 0$$

thus the popular transversality condition $p(t)\hat{x}(t) \to 0$ as $t \to \infty$ is not valid in general.

4 Discounted Autonomous Infinite Horizon Problems

Many economical infinite time horizon optimal control problems are formulated as autonomous problems with the discount factor and therefore we will consider these problems in more details now. Let the state variable $\mathbf{x} \in R^n$, the control variable $\mathbf{u} \in R^r$ and let $\rho > 0$ be a **discount factor**; then a typical infinite horizon optimal control problem in economics takes the following form:

$$\max_{(x,u)} \int_0^\infty e^{-\rho t} f(\mathbf{x}(t), \mathbf{u}(t))dt\ ,$$

subject to

$$\dot{\mathbf{x}}(t) = g(\mathbf{x}(t), \mathbf{u}(t)),\ \mathbf{x}(0) = \mathbf{x}_0$$

and

$$\mathbf{u}(t) \in U \text{ for all } t \in [0, \infty).$$

where f and g are considered to be continuously differentiable functions with respect to $\mathbf{x}$ and $\mathbf{u}$ and $U \subseteq R^r$. For this special kind of the optimal control problem it is possible to formulate a necessary transversality condition. We shall refer to these specific problems shortly as DAP (discounted autonomous problems).

4.1 Michel's Theorem

Since the discount factor $e^{-\rho t}$ was used in the objective functional it is convenient to introduce the current value formulation of the maximum principle. This formulation uses the current value Hamiltonian H^c that can be obtained if the ordinary Hamiltonian H is multiplied by the factor $e^{\rho t}$, i.e.

$$H^c(t, \mathbf{x}, \mathbf{u}, \mathbf{p}) = e^{\rho t}H(\mathbf{x}, \mathbf{u}, \mathbf{p}) = p_0 f(\mathbf{x}, \mathbf{u}) + e^{\rho t} \langle \mathbf{p}, g(\mathbf{x}, \mathbf{u})\rangle.$$

Introducing the adjoint function $\boldsymbol{\lambda}(t) = e^{\rho t}\mathbf{p}(t)$ and putting $\lambda_0 = p_0$ as the current value of the shadow price for the given problem, it is possible to write the current value Hamiltonian in the form

$$H^c(t, \mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) = \lambda_0 f(\mathbf{x}, \mathbf{u}) + \langle \boldsymbol{\lambda}, g(\mathbf{x}, \mathbf{u}) \rangle.$$

The maximum principle with a transversality condition for the given DAP was first given in [11] and can be formulated as follows, cf. e.g. [6]. Let admissible pair $(\hat{\mathbf{x}}, \hat{\mathbf{u}})$ be an optimal solution of DAP and the objective functional is convergent for all $(\hat{\mathbf{x}}, \hat{\mathbf{u}})$. Then there exists a continuous and piecewise continuously differentiable function $\boldsymbol{\lambda}(t)$, $\boldsymbol{\lambda}(t) \in R^n$ and a constant λ_0, $\lambda_0 \geq 0$ such that for all t, $t \geq 0$, the following conditions hold:

(i) $(\lambda_0, \boldsymbol{\lambda}(t)) \neq 0 \in R^{n+1}$ for all $t \in I$;

(ii) the control $\hat{\mathbf{u}}$ maximizes the current value Hamiltonian $H^c(t, \mathbf{x}, \mathbf{u}, \lambda_0, \boldsymbol{\lambda})$ for $\mathbf{u} \in U$, i.e.

$$H^c(t, \hat{\mathbf{x}}(t), \mathbf{u}(t), \lambda_0, \boldsymbol{\lambda}(t)) \leq H^c(t, \hat{\mathbf{x}}(t), \hat{\mathbf{u}}(t), \lambda_0, \boldsymbol{\lambda}(t))$$

for all $t \in I$ and all $\mathbf{u} \in U$;

(iii) wherever $\hat{\mathbf{u}}$ is continuous, the adjoint function $\boldsymbol{\lambda}$ satisfies

$$\dot{\lambda}_i - \rho\lambda_i = -\frac{\partial H^c(t, \hat{\mathbf{x}}(t), \hat{\mathbf{u}}(t), \lambda_0, \boldsymbol{\lambda}(t))}{\partial x_i}\ ,$$

where $i = 1, ..., n$,

(iv) moreover

$$e^{-\rho t}H^c(t, \hat{\mathbf{x}}(t), \hat{\mathbf{u}}(t), \lambda_0, \boldsymbol{\lambda}(t)) = \rho\lambda_0 \int_t^\infty e^{-\rho s} f(\hat{\mathbf{x}}(s), \hat{\mathbf{u}}(s))ds.$$

As a consequence of the last relation it is possible to write the following Michel's transversality condition as

$$\lim_{t\to\infty} e^{-\rho t} H^c(\hat{\mathbf{x}}(t), \hat{\mathbf{u}}(t), \lambda_0, \boldsymbol{\lambda}(t)) = 0.$$

4.2 Example: Optimal Investment Rate

To illustrate how to use Michel's transversality condition we present a model formulated in [9]. Let t, $t \geq 0$, denotes the time variable. Let P, $P=P(x)$, be the profit rate that can be earned with a stock of productive capital x, $x=x(t)$. It is assumed that P is a concave function and $P'(0) > 0$. The capital stock decays at a constant proportionate rate $b > 0$. The investment cost C, $C=C(u)$, is an increasing convex function of the gross investment rate u, $u=u(t) \geq 0$, with $C''(0) = 0$. We will seek the investment rate u that maximizes the present value of the profit stream over the infinite time horizon $[0, \infty)$. The problem can be shortly written as

$$\max_{(x,u)} \int_0^\infty e^{-\rho t}(P(x(t)) - C(u(t)))dt,$$

where ρ, $\rho \in (0, 1)$, is an average interest rate, subject to the state equation

$$\dot{x}(t) = u(t) - bx(t),$$

with the initial condition $x(0) = x_0 \geq 0$ and $u \in [0, \infty)$. To be able to find an explicit solution of the given problem we will assume that

$$P(x) = ax - \frac{x^2}{2}, C(u) = cu^2,$$

where a and c are positive real constants.

The current value Hamiltonian of the given problem is

$$H^c(t,x,u,\lambda_0,\lambda) = \lambda_0\left(ax - \frac{x^2}{2} - cu^2\right) + \lambda(u - bx)$$

First let us assume that $\lambda_0 = 0$ and apply maximum principle (ii). This principle gives a condition that can be rewritten as $\hat{\lambda}(\hat{u} - u) \geq 0$ for all $u \in [0,\infty)$. From this condition it is possible to deduce that $\hat{\lambda} = \hat{\lambda}(t) = 0, t \in [0,\infty)$, which is in contradiction with (i) in maximum principle. This observation allow us to concentrate only on the regular case when λ_0=1. The current value Hamiltonian is a concave function of the variable u so it is possible to concentrate only on the positive values of optimal investment rate. Application of the maximum principle (ii) leads to the relation

$$\frac{\partial H^c}{\partial u}(t, \hat{x}(t), \hat{u}(t), \lambda(t)) = -2c\hat{u}(t) + \lambda(t) = 0$$

which means that

$$\lambda(t) = 2c\hat{u}(t).$$

By using the necessary condition (iii) in Michel's theorem it is possible to write the adjoint equation as

$$\dot{\lambda}(t) - \rho\lambda(t) = -\frac{\partial H^c}{\partial x}(t, \hat{x}(t), \hat{u}(t), \lambda(t)) = -[a - \hat{x}(t) - b\lambda(t)]$$

If we combine the last two equations we obtain a differential equation for optimal rate of investment in the form

$$\dot{\hat{u}}(t) = (\rho + b)\hat{u}(t) + 2c\hat{x}(t) - 2a.$$

This equation together with the state equation for optimal capital stock that can be rewritten in the form

$$\dot{\hat{x}}(t) = \hat{u}(t) - b\hat{x}(t),$$

which forms a linear system of two ordinary differential equations. Note that the stationary solution to this linear system is the unique point

$$\begin{pmatrix} x^\circ \\ u^\circ \end{pmatrix} = \frac{2a}{b^2 + \rho b + 2c} \cdot \begin{pmatrix} 1 \\ b \end{pmatrix}.$$

All parameters a, b, c, and ρ are positive constants, thus $x^\circ > 0$ and $u^\circ > 0$. Omitting details the general solution to the given system of ordinary differential equations can be written as

$$\hat{x}(t) = A_1 e^{\alpha_1 t} + A_2 e^{\alpha_2 t} + x^\circ \text{ and } \hat{u}(t) = B_1 e^{\alpha_1 t} + B_2 e^{\alpha_2 t} + u^\circ,$$

respectively, where

$$\alpha_1 = \frac{\rho + \sqrt{(\rho + 2b)^2 + 8c}}{2}, \alpha_2 = \frac{\rho - \sqrt{(\rho + 2b)^2 + 8c}}{2},$$

and A_1, A_2, B_1 and B_2 are real constants. It is evident that $\alpha_1 > \rho > 0$ and $\alpha_2 < 0$. Now it is possible to make the account of the limit transversality condition (iv). This condition requires that

$$\lim_{t\to\infty} e^{-\rho t}[\left(a\hat{x}(t) - \frac{\hat{x}(t)^2}{2} - c\hat{u}(t)^2 \right) + 2c\hat{u}(t)(\hat{u}(t) - b\hat{x}(t))] = 0.$$

Since $\alpha_1>0$ it is necessary that $A_1 = B_1 = 0$. These observations enable us to find unknown values of the remaining constants A_2 and B_2. Since $\hat{x}(0) = A_2 + x^\circ$ and $\hat{u}(0) = B_2 + u^\circ$, we can use the first relation to find A_2 and the second relation to find B_2. The details are omitted here but we can summarize that the only candidate for optimal pair is

$$(\hat{x}(t), \hat{u}(t)) = ((x_0 - x^\circ)e^{\alpha_2 t} + x^\circ, (\alpha_2 + b)(x_0 - x^\circ)e^{\alpha_2 t} + u^\circ).$$

If we remind that $\alpha_2<0$ then it is obvious that

$$(\hat{x}(t), \hat{u}(t)) \to (x^\circ, u^\circ)$$

as $t\to\infty$ for all initial conditions $x_0 \geq 0$. This feature of convergence of the optimal process to the stationary solution of the given state equation and the adjoint equation respectively is very important for infinite time horizon optimal control problems. More details and further analysis can be found in [6].

5 Summary

The recent literature on economics growth deals with infinite time horizon optimal control models frequently. Such problems offer two mathematical difficulties that have to be solved. On one hand, there is an objective functional expressed as an improper integral. Thus the concept of optimality has to be formulated more precisely. In the rest of the paper it was supposed that the objective functional is convergent for all admissible pairs and the concept of strong optimality was involved. The problem of convergence of the objective functional was not solved in this paper. On the other hand, Pontryagin maximum principle for infinite time horizon models can be formulated without transversality condition. This complication means that there is not enough conditions to characterize optimal solution. The paper also provided a notice that the natural transversality condition is not necessary valid for infinite time horizon problems. A less known counterexample than the counterexample of Halkin, cf. [7], was introduced here. The less general optimal control problem has to be formulated to find at least some transversality condition. Therefore the discounted autonomous problem for infinite time problem was introduced. For this problem Michel's transversality condition can be stated. In the paper the use of transversality condition of this type was introduced as well. It was shown that this condition can be used to decide whether the optimal control problem is regular, i.e. λ_0=1, or irregular, i.e. λ_0=0, and that it can be also used to specify candidates for an optimal solution, although this could be quite complicated.

Acknowledgements

The support of the Czech Science Foundation #402/09/0405 is kindly acknowledged.

References

[1] Acemoglu D., *Introduction to Modern Economic Growth*, Princeton, Princeton University Press, 2009.

[2] Aseev S. M., Kryazhyimskiy A. V., The Pontryagin Maximum Principle and transversality Condition for a Class of Optimal Control Problems with Infinite Time Horizons, *SIAM Journal on Control and Optimization*, 43(3), 2004, pp. 1094–1119.

[3] Barro R., Sala–i–Martin X., *Economic Growth*. Cambridge, Massachusetts, MIT Press, 2004.

[4] Caputo M. R., *Foundations of Dynamic Economic Analysis, Optimal Control Theory and Applications*, Cambridge, Cambridge University Press, 2005.

[5] Carlson D. A., Haurie A., *Infinite Horizon Optimal Control, Theory and Applications*, Berlin, Springer-Verlag, 1987.

[6] Grass D. et al., *Optimal Control of Nonlinear Processes*, Berlin, Springer-Verlag, 2008.

[7] Halkin H., Necessary conditions for optimal control problems with infinite horizons. *Econometrica*, *42* (2), 1974, pp. 267–272.

[8] Chiang A.C., *Elements of Dynamic Optimization*, Illinois, Waweland Press, 1992.

[9] Kamien M. I., Schwartz N. L., *Dynamic optimization: The calculus of variations and optimal control in economics and management* (2nd ed.), Amsterdam, North-Holland, 1991.

[10] Nevima J., Ramík, J. Application of multicriteria decision making for evaluation of regional competitiveness. In *Mathematical Methods in Economics 2009*. Kostelec nad Černými lesy 9.-11.9.2009. Czech University of Life Sciences Prague: Faculty of Economics and Management, 2009.

[11] Michel, P., On the transversality condition in infinite horizon optimal problems. *Econometrica, 50* (4), 1982, pp. 975–985.

[12] Seierstad A., Sydsæter K., *Optimal control theory with economic applications*. Amsterdam, North-Holland, 1986.

[13] Sethi S. P., Thompson G. L., *Optimal control theory applications to management science and economics* (2nd ed.). Boston, Kluwer Academic, 2000.

Simulation Model of the Logistic Flows between Warehouse and Production

Peter Princ[1], Roman Černohous[2], Martina Kuncová[3]

Abstract. Simulation belongs to methods that help managers to solve their problems – especially when they are looking for problematic places in production and distribution or when they want to have better knowledge of all the processes. The aim of the simulation model is to describe the situation and simulate the real processes and the situations that may happen. In our study we try to solve the real problem of the logistic flows between the warehouse and the production by using simulation model. The problem is not only in distribution and transportation of the final products and supplying parts (from the production place to the warehouse or the other way round) but also with handling of the empty containers back to the warehouse. The aim of the simulation model is to find out the main "bottlenecks" of the system, which are the places or the resources with limited capacity. Another goal is to find the optimal number of MHE – material handling equipment (trucks, trains, forklifts) that are used for transportation. The model that is presented in this article was created using SIMUL8 software.

Keywords: simulation model, logistic flows, SIMUL8

JEL Classification: C63
AMS Classification: 68U20

1 Introduction

Amid economic turmoil demand for cars plummeted at the turn of years 2008 and 2009. Most markets are now recovering and the industry begins to operate as normal. The same situation is also in the business of automotive suppliers. In our study we focused on a factory of automotive supplier company in which pumps for passenger cars are made. The headquarters decided to start a new production line and therefore there were requirements for analysing the behaviour of the production processes under new conditions.

The company chose to solve their unanswered questions by using simulation model because of the difficulty of analytical solving related to the introduction of new production line to the original production zone of their factory. The basic idea consists of finding new solutions for the whole range of processes in the factory, including logistic flows that we modelled using simulation models.

Banks [2] defines simulation as the imitation of the operation of a real-world process or system over time. Simulation involves generation of an artificial history of the system and the observation of that artificial history to draw inferences concerning the operating characteristics of the real system that is represented. Simulation is an indispensable problem-solving methodology for the solution of many real-world problems. It is used to describe and analyze the behaviour of a system, ask "what if" questions about the real system, and aid in the design of real systems.

Model is a representation of a system or process. Simulation model is a representation that incorporates time and changes that occur over time. Discrete model is one that changes only at discrete points in time, not continuously. Model may incorporate logical, mathematical and structural aspects of the system or process. Time is a critical component [3].

Our model is based on discrete-event simulation method. Dlouhý [4] analyzed different types of simulations and management methods and classified them by various criteria. According to this classification we place our model to the class of Methods of Dynamic Process Analysis.

[1] University of Economics, Prague, Dept. Of Econometrics, nám. W. Churchilla 4, Praha 3, peter.princ@vse.cz
[2] University of Economics, Prague, Dept. Of Econometrics, nám. W. Churchilla 4, Praha 3, xcerr22@vse.cz
[3] University of Economics, Prague, Dept. Of Econometrics, nám. W. Churchilla 4, Praha 3, kuncovam@vse.cz

Nowadays, many problems grouped to the complicated area of production systems, such as logistic flows, are solved using simulation software Simul8. Several case studies with similar approach to problem solving were published [1].

2 Model description

The main idea of the project was to create a model for simulating logistic flows between warehouse and production area. The model contains simulation of manipulation equipment, complex material flow from the reception area of material to the final testing and the expedition of final production.

Simulation concept was based on production plans for years 2010 and 2011 with weekly shift patterns. In this way, a model for warehouse and logistics with other material flows and handling parameters was set. Simulation time was set for one week, goal date for simulation corresponds with the production plan for December 2011, and also a sublevel of simulation for new product introduction in November 2011was defined. The simulation was based on data from SAP software, information from foremen and internal information systems used in the enterprise. The model was created in the program Simul8 Professional and the data were imported to the model from MS Excel which allows changes in input data depending on simulated scenario. We define three basic scenarios for November 2010, November 2011 and November 2011 High which is based on high utilization.

2.1 Simul8

SIMUL8 is a software package designed for Discrete Event Simulation (www.simul8.com). It allows user to create a visual model of the system under investigation by drawing objects directly on the screen of his computer. Typical objects may be queues or service points. The characteristics of the objects are defined in terms of capacity, speed, etc. Once the system is modelled a simulation can be undertaken. The flow of work items around the system is shown by animation on the screen, and for that reason the appropriateness of the model can be easily assessed. When the structure of the model has been confirmed, a number of trials can be run under different conditions and the performance of the system can be described statistically. Statistics of interest may be average waiting times or utilization of Work Centres or Resources. [5]

2.2 Simulation goals

We defined three main areas of simulation goals:

1. Specification of requirements for logistics
   a) Material flow intensities to washing machines, transport of final products to the warehouse
   b) Reaction time – delivery time limits for transfer of parts to washing machine area
   c) Frequency of pallet transport from testing zone
2. Specification of required logistic capacities
3. Simulation of scenarios and variants for material handling – trains, forklifts, reaches trucks, VNA trucks and irregular vehicle assembling waste

2.3 Limitations in the model

The following special circumstances had to be accounted for by this model:

- door clearance – each vehicle has different size
- the corridor between racks can be served only by the VNA truck with exemptions at the outer racks which can be served by the reach truck
- lifted weights by warehouse operators have to respect the legislative norms
- warehouse area consists of 12 long racks and 4 short racks ordered to section in compliance with their type
- circulated process of KLT containers
- inadequate ramps for vans – vans are served at the fifth ramp which is not located in the central reception and expedition area, and was not planned for this purpose
- reception process was simulated as 60 minutes delay because of required administration and sample testing
- velocity of each vehicle – forklifts and reach trucks have velocity 8-10 km/h and VNA trucks have velocity 9 km/h

2.4 Logistic flows in the model

In our simulation model we used these logistic flows:

- Transport of supplying parts to washing machines and their collection to racks in production area - material for production of pumps, which is stored in the KLT containers on pallets, is transported from racks in the warehouse by the VNA truck or the reach truck to the store area which is located in the north end of each rack in the warehouse. Then it is handled by the forklift pallet truck in the case of a request for the whole pallet and by the train if there is not a request for the whole pallet of material from the production area. In the case of transportation by train, the driver of reach truck chooses only the exact amount of KLT containers that are loaded onto the train cars. Special care in our model is dedicated to heavy parts for production because of requirements for lifted weight by warehouse operators. Seventy percent of all material from warehouse is transported to the production area by train.
- Transport of empty containers and pallets back to the warehouse - highly used KLT containers are assembled to pallets directly in the area of washing machines and completed pallets are transported to packaging process that is followed by storage in the warehouse. Others KLT containers are transported to the warehouse where pallets for each type of empty KLT containers are located. The pallets are usually completed in 3 to 4 days and then packed and stored in the warehouse.
- Handling of empty containers for final products at the end of assembly line
- Transport of final products to the warehouse
- Flows related to other operations - reception and expedition (Figure 1) consist of 4 ramps for trucks and 1 ramp for vans which is situated in the warehouse. This location of the van ramp caused problems with cleanliness. In our model we monitor volume of traffic at each ramp. This area is open for 12 hours per day with 4-hour peak in traffic from 8 am to 12 am, and the percentage proportion between trucks and vans is 50% to 50%. Each reception process includes unloading pallets from the truck by the forklift and their transportation to the reception area where administration and controlling process are operated. Afterwards, the forklift takes the pallets to the warehouse where they are transported by the reach truck to the racks. The expedition runs in the same way. The length of the loading and unloading process of the pallets from/onto the vehicle depends on the amount of pallets for each type of vehicle; for trucks we count with time for 20 pallets and for vans with time for 4 pallets. The flow of material in this area is simulated in one way and this has no influence on behaviour of the system. It is simulated as another reception process with unique characteristics.

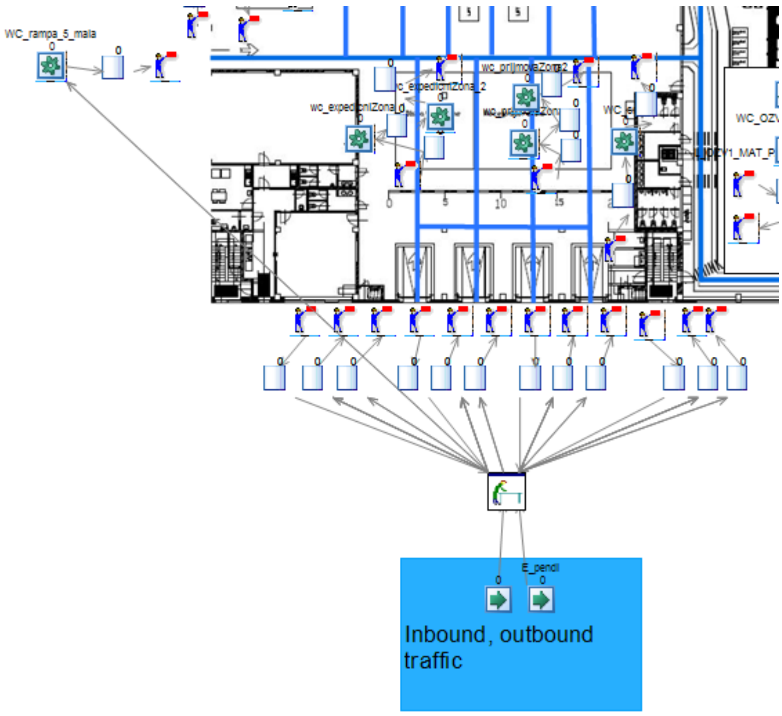


Figure 1 Reception and Expedition simulated in one way

- Another material flow is for production of cylinders on the CNC machines. The products from the CNC machines are transported to another manufacturer and after their finalization they are transported back through reception area to the production zone by a specialized vehicle.
- The waste from CNC machines, which is approximately 50% of material for one cylinder and the waste from final products, which is about 5% of total production, is transported by an irregular vehicle to the waste zone located outside the factory. Additionally, final products addressed for testing zone are transported by an irregular vehicle.

All the logistic flows are shown in the Figure 2.

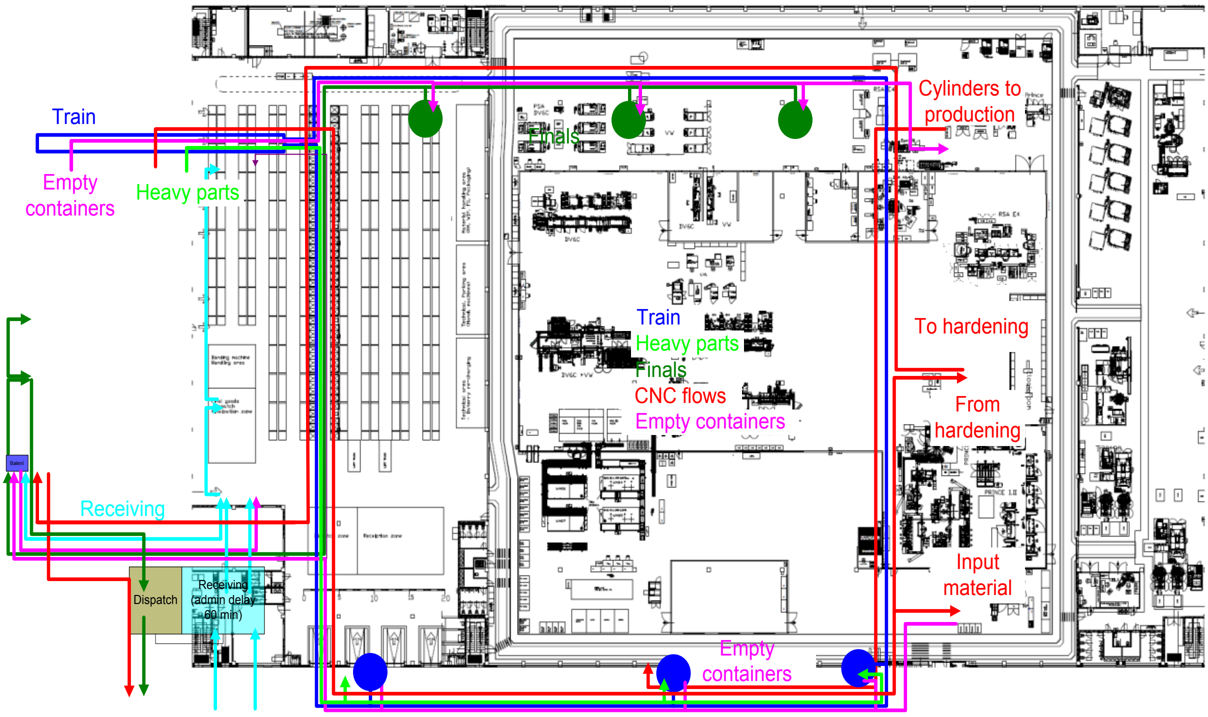


Figure 2 Logistic flows in the model

3 Simulation results

The most important bottleneck in our simulation is the long handover from logistics to washing machine racks. Original length of handover was 2 minutes and according to our results from simulation, this time is unacceptably long for full production capacity. This was caused by checking and counting of all parts in each KLT container before loading to the rack. Maximal feasible handover time is 1.3 minutes proved by simulation, desirable time would be 0.5 minutes. As possible solutions we see change of organization in sharing responsibility for checking and counting parts in KLT or design of an IT solution for full traceability and identification in this process.

Another bottleneck, the OZV pallet position, resulted from the fact, that the production area was not designed for such amount of production. Therefore we suggest shortening the time of handover, enlarging the area for pallet position, and consequently modifying the number of racks in the washing machine zone. The most frequently used materials should be supplied on pallets according to available space and pallet position and not in the KLT containers. This would also cause more workload and lifted weight for washing machine operators and this change has to comply with the legislative norms.

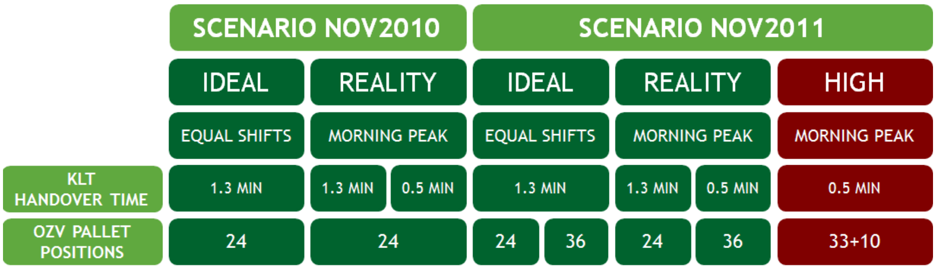


Figure 3 Bottlenecks in different scenarios

Maximum reaction time for logistics for transport of material from the warehouse to the production area was examined for all scenarios and their variants based on various breakdowns, such as failure of vehicles, and this time was verified to be 75 minutes in the worst case. The maximum reaction time was accepted by the warehouse and production management.

Number of operators in logistics was again examined for different scenarios and various variants, based on material flows and their peaks, handover times, and OZV pallet positions. For instance, in scenario for November 2010, ideal material flow without peaks, handover time 1,3 minutes and 24 pallet positions in the washing machine zone we acquired the following results (Table 1).

Personnel	Count needed	Utilization (%)	Lifted weight
VNA drivers	2	56,3	1356 kg/shift
WH operators	6	83,4	4701 kg/shift

Table 1 Results for VNA drivers and washing machine operators - scenario November 2010

Utilization and count of needed manipulation techniques based on the same characteristics as in the example above are shown in the Table 2.

Vehicle	Count needed	Utilization (%)
Trains	2	44,0
Forklifts	4	73,8
VNA trucks	2	56,3

Table 2 Results for handling equipment – scenario November 2010

Warehouse area is divided into the manipulation zone and the expedition zone, and for this reason we tested various proportions of areas with different scenarios. We propose rebuilding one unloading bay to handling area for vans which could be done during full operation according to our scenarios.

As we compare actual situation in the factory, the size of area used by logistics, and the size of area used by production zone to the others manufacturers in this field of business with benchmarking method, we come to the conclusion that there is inappropriate proportion between logistics (22%) and production zone (78%). For instance, automotive supplier in Czech Republic has the proportion 60% to 40% favouring the production zone and the automotive producer in Czech Republic has the proportion 50% to 50%. Therefore, we see an obstacle in future expansion of business in this company’s production area.

4 Conclusion

Nowadays the simulation models are used for analysing enormous range of business problems when it is difficult to find solutions through classic analytical approach. In our case study, we examine using simulation model for logistic flows between warehouse and production. We contemplated different variants based on various characteristics as number of operators, number of vehicles, number of pallet positions, delivery time peaks etc. We found two bottlenecks in the original system of logistic processes; the first one is the length of handover in the washing machine area between logistics and operators of washing machines, and the second one is the inappropriate size of area for OZV pallet positions. According to our study, the factory has unsuitable size of warehouse area to the production area.

Our proposal for solving the bottlenecks consists of expediting the process of handover by sharing responsibility in organization structure or by designing appropriate IT solution, and of enlarging the area of pallet positions and using pallets for the most frequently used parts. Additionally, we suggested enlarging the warehouse capacity.

Acknowledgements

The research project was supported by Grant No. IGA F4/16/2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Aguirre, A., and Méndez, C. A.: Applying a Simulation-based Tool to Productivity Management in an Automotive-parts Industry. In: *Proceedings of the 2008 Winter Simulation Conference* (Mason, S. J., Hill, R. R., Mönch, L., Rose, O., Jefferson, T., and Fowler, J. W., eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2005, 1838-1846.

[2] Banks, J., Carson II, J. S., Nelson, B. L., and Nicol, D. M..: *Discrete-Event System Simulation*, 3rd Ed. Upper Saddle River, New Jersey: Prentice-Hall, 2000.

[3] Carson II, J. S.: Introduction to Modeling and Simulation. In: *Proceedings of the 2005 Winter Simulation Conference* (Kuhl, M. E., Steiger, N. M., Armstrong, F. B., and Joines, J. A., eds.). Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, 2005, 16-23.

[4] Dlouhy, M.: A Classification of Simulation and Management Methods. In: *Proceedings of 28th International Conference on Mathematical Methods in Economics 2010* (Houda, M., and Friebelova, J., eds.). University of South Bohemia, Faculty of Economics, Ceske Budejovice, Czech Republic, 2010, 111-116.

[5] Shalliker, J. and Ricketts C.: *An Introduction to SIMUL8*, Release nine. School of Mathematics and Statistics, University of Plymouth, 2002.

Pairwise comparison method in multi-criteria optimization - an alternative approach

Jaroslav Ramík[1]

Abstract. In this paper an alternative approach to pair-wise comparison method in multi-criteria optimization is proposed and discussed. The aim is to study the unification of multiplicative preference relation as a preference representation structure and additive preference relation, called also fuzzy preference relation, both based on pair-wise comparisons in multipurpose decision-making problems. Under different preference representation structures, the multiplicative preference relations and additive preference relations are incorporated in the decision problem by means of a transformation function between multiplicative and additive preference relations. Some theoretical results on relationships between multiplicative and additive preference relations are presented and comparative study to the classical pair-wise comparison in AHP is discussed. Illustrative numerical examples are supplemented.

Keywords: multi-criteria optimization, AHP, pair-wise comparison matrix, reciprocity, consistency.

JEL Classification: C44
AMS Classification: 90B50, 90C29, 91B08

1 Introduction

Decision making in situations with multiple criteria and/or persons is a prominent area of research in normative decision theory. This topic has been widely studied e.g. in [2 - 8]. Here, we do not distinguish between “criteria” and “persons”, and interpret the decision process in the framework of multipurpose decision making (MPDM), see e.g. [4]. In an MPDM problem, we have a set of alternatives to be analyzed according to different purposes in order to select the best one(s). For each purpose a set of evaluations about the alternatives is known. The classical pair-wise comparison method requires the decision-maker (DM) to express his/her preferences in the form of a pair-wise comparison matrix. The pair-wise comparison method is a powerful inference tool that can be also used as a knowledge acquisition technique for knowledge-based systems. It is useful for assessing the relative importance of several objects, when this cannot be done by direct rating. In fact, this perspective has been recently used for measuring the importance of a web site [1]. As it is known, most decision processes are based on preference relations, in the sense that processes are linked to some degree of preference of any alternative over another. The use of preference relations is usual in decision making [3, 8]. Therefore, to establish properties to be verified by such preference relations is very important for designing good decision making models. One of these properties we investigate in this paper is the so called consistency property. The lack of consistency in decision making can lead to inconsistent conclusions. That is why it is so important, if not crucial, to study conditions under which consistency is satisfied [2, 4 - 8]. On the other hand, perfect consistency is difficult to obtain in practice, particularly when measuring preferences on a set with a large number of alternatives. Here, we deal with two types of consistency: multiplicative, which is traditional, see e.g. [7], and additive, which is a new one, see also [3, 4].

2 Multiplicative and additive relations

Formally, the problem can be formulated as follows. Let $X = \{x_1, x_2,...,x_n\}$ be a finite set of alternatives. These alternatives have to be classified from best to worst, using the information given by a DM. However, it can often be difficult for the DM to express exact estimates of the ratios of importance of alternatives. The DM has his/her own criteria, ideas, attitudes, motivations and personality, it is quite natural to think that different DMs will give their preferences in a different way. This leads us to assume that the DMs´ preferences over the set of alternatives, X, may be represented at least in the following two ways: multiplicative and additive preferences.

2.1 Multiplicative preferences

Let us assume that the DM´s preferences on X are described by a preference relation, $A \subset X\times X$ given by a positive nxn matrix $A = \{a_{ij}\}$, where a_{ij} indicates a ratio of preference intensity for alternative x_i to that of x_j , i.e., it is

[1] Silesian University/School of Business Administration, Department of Mathematical Methods in Economics, University Square 1934/3, 733 40 Karviná, ramik@opf.slu.cz.

interpreted as x_i is a_{ij} times as good as x_j. According to [7], T. Saaty suggests measuring a_{ij} using a ratio scale, and precisely the 1 to 9 scale: a_{ij} = 1 indicates indifference between x_i and x_j , a_{ij} = 9 indicates that x_i is absolutely preferred to x_j , and $a_{ij} \in \{2, 3, ... ,8\}$ indicates intermediate evaluations. The elements of $A = \{a_{ij}\}$ should satisfy the following condition [7].

A positive nxn matrix $A = \{ a_{ij} \}$ is *multiplicative reciprocal* (*m-reciprocal*), if

$$a_{ij}. a_{ji} = 1 \text{ for all } i, j. \tag{1}$$

If, e.g., x_i is 3 times as good as x_j, then the goodness of x_j is $^1/_3$ with respect to the goodness of x_i.

A positive nxn matrix $A = \{a_{ij}\}$ is *multiplicative consistent* (or, *m-consistent*), see [3, 7], if

$$a_{ij} = a_{ik}. a_{kj} \text{ for all } i, j, k. \tag{2}$$

If, e.g., x_i is 3 times as good as x_k and x_k is 2 times as good as x_j, then x_i is 3·2 = 6 times as good as x_j.

Notice that a_{ii} = 1 for all i, and also (2) implies (1), i.e. an m-consistent positive matrix is m-reciprocal (however, not vice-versa). Then, (2) can be rewritten equivalently as

$$a_{ik}\, a_{kj}\, a_{ji} = 1 \text{ for all } i, j, k. \tag{2´}$$

2.2 Additive preferences

The above mentioned interpretation of preferences on X described by a positive multiplicative preference relation matrix A is, however, not always appropriate for a DM. Evaluating the preference of two elements of a pair, say, x_i and x_j with respect to e.g. "design" property of a product might cause a problem. Here, saying e.g. that x_i is 3 times as good as x_j is peculiar. A more natural procedure is the following: divide 100% of the property into two parts and then assign the first part of it to the first element and the rest to the second one. In other words, when comparing x_i with x_j the DM assigns the value a_{ij} to x_i and a_{ji} to x_j, whereas $a_{ij} + a_{ji} = 1$ (i.e. 100%). With this interpretation, a DM´s preferences on X can be represented also by a *fuzzy preference relation*, $R \subset X \times X$, with membership function [13], $\mu_R : X \times X \to [0;1]$, where $\mu_R (x_i, x_j) = a_{ij}$ denotes the preference degree or intensity of the alternative x_i over x_j , see also [3, 7]. In what follows, we will not use fuzzy preference nomenclature for the above mentioned preference relations. Instead, we will use *additive preference nxn matrices*, $A = \{a_{ij}\}$, with $0 \le a_{ij} \le 1$ for all i, j. Here, a_{ij} = 0,5 indicates indifference between x_i and x_j, a_{ij} = 1 indicates that x_i is absolutely preferred to x_j, a_{ij} = 0 indicates that x_j is absolutely preferred to x_i, and, a_{ij} > 0,5 indicates that x_i is preferred to x_j. The same interpretation is appropriate also for ordinal preference relations. We have the following definition:

An nxn matrix $A = \{a_{ij}\}$ with $0 \le a_{ij} \le 1$ for all i and j is *additive reciprocal* (*a-reciprocal*), see [4, 8], if

$$a_{ij} + a_{ji} = 1 \text{ for all } i, j. \tag{3}$$

Evidently, if (3) holds for all i and j, then a_{ii} = 0,5 for all i.

An nxn matrix $A = \{a_{ij}\}$ with $0 \le a_{ij} \le 1$ for all i and j is *additive consistent* (or, *a-consistent*), if it is a-reciprocal and

$$a_{ij} a_{jk} a_{ki} = a_{ik} a_{kj} a_{ji} \text{ for all } i, j, k. \tag{4}$$

If $a_{ij} > 0$ for all i and j, then (4) can be presented as

$$\frac{a_{ij}}{a_{ji}} \frac{a_{jk}}{a_{kj}} = \frac{a_{ki}}{a_{ik}} \text{ for all } i, j, k. \tag{5}$$

3 Multiplicative and additive consistent matrices

In this section we shall investigate some relationships between m-consistent and a-consistent pair-wise comparison matrices. We start with the result following from (5). The proofs of the following propositions will be given elsewhere.

Proposition 1. *Let* $A = \{a_{ij}\}$ *be a positive nxn matrix with* $0 < a_{ij} < 1$ *for all i and j.* $A = \{a_{ij}\}$ *is a-consistent iff* $B = \{b_{ij}\} = \{\frac{a_{ij}}{1-a_{ij}}\}$ *is m-consistent.*

Proposition 2. *Let* $A = \{a_{ij}\}$ *be a positive nxn matrix.* $A = \{a_{ij}\}$ *is m-consistent iff* $B = \{b_{ij}\} = \{\frac{a_{ij}}{1+a_{ij}}\}$ *is a-consistent.*

We have also the following result, see [7], characterizing an m-consistent matrix by a vector of weights.

Proposition 3. *Let* $A = \{a_{ij}\}$ *be a positive nxn matrix. A is m-consistent iff there exists a vector* $w=(w_1, w_2,..., w_n)$ *with* $w_i > 0$ *for all* $i=1,2,...,n$, *and* $\sum_{j=1}^{n} w_j = 1$ *such that*

$$a_{ij} = \frac{w_i}{w_j} \text{ for all } i,j=1,2,...,n. \tag{6}$$

In the following proposition we derive a similar characterization of a-consistent matrices.

Proposition 4. *Let* $A = \{a_{ij}\}$ *be a positive nxn matrix with* $0 < a_{ij} < 1$ *for all i and j. A is a-consistent iff there exists a vector* $v=(v_1, v_2,..., v_n)$ *with* $v_i > 0$ *for all* $i=1,2,...,n$, *and* $\sum_{j=1}^{n} v_j = 1$ *such that*

$$a_{ij} = \frac{v_i}{v_i + v_j} \text{ for all } i,j=1,2,...,n. \tag{7}$$

In practice, perfect consistency is difficult to obtain, particularly when measuring preferences on a set with a large number of alternatives. In the following section we deal with problem of inconsistency of pair-wise comparison matrices.

4 Inconsistency of pair-wise comparison matrices

If for some $i, j, k=1,2,...,n$, (2) does not hold, or, on the other hand, if for some $i, j, k=1,2,...,n$, (4) does not hold, than A is said to be *inconsistent*. In order to measure the grade of inconsistency several consistency indices have been proposed in the literature. In Analytic hierarchy process (AHP), see [7], in order to measure consistency T. Saaty proposed the consistency ratio (CR). For the prioritization procedure based on geometric mean, the geometric consistency ratio was proposed in [7], with an interpretation analogous to that considered for *CR*. Here, we shall deal with two classes of methods for measuring inconsistency of pair-wise comparison matrix: the first class is based on the principal eigenvalue of the matrix, or, the corresponding principal eigenvector, the second one is based on minimizing the distance to a special ratio matrix.

4.1 Consistency based on the principal eigenvalue

In AHP the m-consistency of a positive nxn matrix $A = \{a_{ij}\}$ is measured by the *m-consistency index* CI_n as

$$CI_n = \frac{\lambda_{\max} - n}{n-1} \tag{8}$$

where $\lambda_{\max}$ is the *principal eigenvalue* of A. It holds $CI_n \geq 0$, see [7]. The "importance" of the alternatives in X is determined by the vector of weights $w = (w_1, w_2,..., w_n)$, with $w_i > 0$, for all $i = 1,2,...,n$, and $\sum_{j=1}^{n} w_j = 1$, the normalized *principal eigenvector of A*, i.e.

$$Aw = \lambda_{\max} w. \tag{9}$$

As the weight w_i is interpreted as the relative importance of the alternative x_i, the alternatives $x_1, x_2,...,x_n$ in X are ranked according to the corresponding weights.

If $A = \{a_{ij}\}$ is an $n\times n$ positive m-reciprocal (PmR) matrix, then A is m-consistent iff $CI_n = 0$, see [7]. To provide a consistency measure independent of the dimension of the matrix, n, T. Saaty proposed the *m-consistency ratio* (CR). (T. Saaty himself called it *consistency ratio*, in order to distinguish it here from the other consistency measures, we call it *m-consistency ratio CR*.) This is obtained by taking the ratio between CI_n to its expected value over a large number of positive reciprocal matrices of dimension n, whose entries are randomly chosen in the set of values. For this consistency measure, T. Saaty proposed a 10% threshold for the CR to accept the estimation. In practical decision situations inconsistency is "acceptable" if $CR < 0.1$, see e.g. [7].

4.2 Consistency based on minimizing the distance

Here, we propose a new definition of consistency of a positive square matrix based on the distance to a special matrix, see also [5]. For this purpose, we propose the following definitions.

Definition 1. *Let* M^n *be the set of all* $n\times n$ *matrices with positive elements, and let* d *be a real function defined on* $\mathsf{M}^n\times\mathsf{M}^n$, *i.e.,* $d : \mathsf{M}^n\times\mathsf{M}^n \to \mathbf{R}$ *satisfying*:

(i) $d(A,B) \geq 0$ *for all* $A,B \in \mathsf{M}^n$.
(ii) $d(A,B) = 0$ *iff* $A = B$.
(iii) $d(A,B) + d(B,C) \geq d(A,C)$ *for all* A, B, $C\in \mathsf{M}^n$.

The function d is called the metric function on M^n.

Definition 2. *Let* $A = \{a_{ij}\}$ *be a positive nxn matrix, d be a metric function on* M^n. $I_d^m(A)$ *defined as*

$$I_d^m(A) = \inf\{ d(A,W) \mid w_i > 0, i = 1,2,\ldots,n, \sum_{j=1}^{n} w_j = 1, W = \{\frac{w_i}{w_j}\} \} \tag{10}$$

is said to be the distance multiplicative-consistency index (dm-consistency index) of *A*. *If* $I_d^m(A) > 0$, *then A is called* distance multiplicative-inconsistent (dm-inconsistent).

An optimal solution $w^* = (w_1^*,\ldots,w_n^*)$ *of the minimization problem*

$$\begin{aligned} & d(A,W) \to \min; \\ & s.t. \\ & w_i > 0, i = 1,2,\ldots,n, \sum_{j=1}^{n} w_j = 1, W = \{\frac{w_i}{w_j}\} \end{aligned} \tag{11}$$

is called the vector of weights corresponding to dm-inconsistent matrix *A*.

The dm-consistency index of A is defined as the distance of A and $W = \{\frac{w_i}{w_j}\}$, where the ratio matrix W is "as close as possible" to the matrix A. Based on Definition 2 (ii), and Proposition 3, for a positive nxn m-consistent matrix A, we obtain $I_d^m(A) = 0$. Below we present two examples of metric functions *on* M^n.

Example 1. (l^p*-distance*)

$$d_1(A,W) = \left(\sum_{i,j} \left| a_{ij} - \frac{w_i}{w_j} \right|^p \right)^{\frac{1}{p}}, p > 0. \tag{12}$$

Example 2. (*Logarithmic* l^p*-distance*)

$$d_2(A,W) = \left(\sum_{i,j} \left| \ln a_{ij} - \ln \frac{w_i}{w_j} \right|^p \right)^{\frac{1}{p}}, p > 0. \tag{13}$$

Here, $p = 2$ in (12) is the well known "least squares" distance framework, whereas in (13) we obtain the logarithmic least squares distance.

Further, we define another concept of consistency of positive a-reciprocal matrix based on the distance to a special matrix. For this purpose we have the following definition.

Definition 3. *Let* Let $A = \{a_{ij}\}$ *be a positive nxn matrix such that* $0 < a_{ij} < 1$ *for all i and j, and let d be a metric function on* M^n. $I_d^a(A)$ *defined as*

$$I_d^a(A) = \inf\{ d(A,V) \mid v_i > 0, i = 1,2,\ldots,n, \sum_{j=1}^{n} v_j = 1, V = \{\frac{v_i}{v_i + v_j}\} \}. \tag{14}$$

is said to be the distance additive-consistency index (da-consistency index) of *A*. *If* $I_d^a(A) > 0$, *then A is called* distance additive-inconsistent (da-inconsistent).

An optimal solution $v^* = (v_1^*,\ldots,v_n^*)$ *of the minimization problem*

$$\begin{aligned} & d(A,V) \to \min; \\ & s.t. \\ & v_i > 0, i = 1,2,\ldots,n, \sum_{j=1}^{n} v_j = 1, V = \{\frac{v_i}{v_i + v_j}\} \end{aligned} \tag{15}$$

is called the vector of weights corresponding to da-inconsistent matrix *A*.

The da-consistency index of *A* is defined as the minimal possible distance of *A* and $V = \{\frac{v_i}{v_i + v_j}\}$, where matrix *V* is "as close as possible" to the matrix *A*. Again, based on Definition 2 (ii), and Proposition 4, if *A* is a positive *nxn* a-consistent matrix, then $I_d^a(A) = 0$. Below we present more two examples of metric functions *on* M^n.

Examples 3.

$$d_3(A,V) = \left(\sum_{i,j} \left| a_{ij} - \frac{v_i}{v_i + v_j} \right|^p \right)^{\frac{1}{p}}, \ p > 0. \tag{16}$$

Example 4.

$$d_4(A,V) = \left(\sum_{i,j} \left| \ln a_{ij} - \ln \frac{v_i}{v_i + v_j} \right|^p \right)^{\frac{1}{p}}, \ p > 0. \tag{17}$$

5 Illustrative examples

In this section, we present several numerical examples of pair-wise comparison matrix in order to illustrate the results from Section 3 and 4.

Example 5.

$$A = \begin{pmatrix} 1 & 1,\overline{6} & 3,\overline{3} & 10 \\ 0,6 & 1 & 2 & 6 \\ 0,3 & 0,5 & 1 & 3 \\ 0,1 & 0,1\overline{6} & 0,\overline{3} & 1 \end{pmatrix}$$

Here, as it can be easily verified by (2), $A = \{a_{ij}\}$ is m-consistent. Also, $w = (0,5; 0,3; 0,15; 0,05)$ is the corresponding vector of weights from Proposition 3 with $A = \{w_i/w_j\}$. Let $B = \{b_{ij}\} = \{ \frac{a_{ij}}{1 + a_{ij}} \}$, i.e.

$$B = \begin{pmatrix} 0,5 & 0,63 & 0,77 & 0,91 \\ 0,38 & 0,5 & 0,67 & 0,86 \\ 0,23 & 0,33 & 0,5 & 0,75 \\ 0,09 & 0,14 & 0,25 & 0,5 \end{pmatrix}$$

By Proposition 2, B is a-consistent, as it can be easily verified by (4). The above vector w verifies Proposition 4.

Example 6.

$$C = \begin{pmatrix} 1 & 2 & 3 & 9 \\ 0,5 & 1 & 2 & 6 \\ 0,\overline{3} & 0,5 & 1 & 3 \\ 0,\overline{1} & 0,1\overline{6} & 0,\overline{3} & 1 \end{pmatrix}$$

Here, as it can be easily verified by (2), $C = \{c_{ij}\}$ is m-inconsistent as $a_{12}\, a_{23} = 2.2 = 4 \neq 3 = a_{13}$.

The vector of weights $w = (w_1, w_2, w_3, w_4)$, with $w_i > 0$, for all $i = 1,2,3,4$, and $\sum_{j=1}^{4} w_j = 1$, satisfying

$$Cw = \lambda_{\max} w$$

is calculated as follows: $w = (0{,}503 , 0{,}290 , 0{,}155 , 0{,}052)$ with $\lambda_{\max} = 4{,}01$ and consistency ratio CR = 0,00384.

By solving (15) the vector of weights $v = (v_1, v_2, v_3, v_4)$ is calculated as follows: $v = (0{,}437, 0{,}270, 0{,}181, 0{,}112)$ with dm-consistency index $I_d^m(C) = 0{,}0497$ with l^1-distance function $d_1(A,W) = \sum_{i,j} \left| a_{ij} - \frac{w_i}{w_j} \right|$.

6 Conclusion

In this paper we defined two types of pair-wise comparison relations as well as the concepts of reciprocity and consistency. We investigated mutual relations between two types of consistency: multiplicative and additive consistency, and derived some necessary and sufficient conditions for their existence. Moreover, we dealt with the problem of inconsistency defining two inconsistency indices. Finally, some illustrating examples were presented and discussed.

Acknowledgements

This research has been supported by GACR project No. 402090405.

References

[1] Brin, S., Page. L., The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* 30 (1998), 107-117, ISSN 1389-1286.

[2] Chiclana, F. Herrera, F., Herrera-Viedma, E., Alonso, S., Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations. *European Journal of Operat. Res*. 182 (2007) 383-399, ISSN 0377-2217.

[3] Fodor, J., Roubens, M., *Fuzzy Preference Modelling and Multicriteria Decision Support*. Kluwer, Dordrecht, 1994.

[4] Herrera-Viedma, E., Herrera, F., Chiclana, F., Luque, M., Some issues on consistency of fuzzy preference relations. *European Journal of Operat. Res*. 154 (2004) 98-109. ISSN 0377-2217.

[5] Ramík, J., Korviny, P., Inconsistency of pair-wise comparison matrix with fuzzy elements based on geometric mean. *Fuzzy Sets and Systems* 161 (2010), 1604–1613. ISSN 0165-0114.

[6] Ramík, J., Vlach, M., *Generalized concavity in optimization and decision making*. Kluwer Publ. Comp., Boston-Dordrecht-London, 2001, 305 p. ISBN 0-7923-7494-9.

[7] Saaty, T. L., *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991, ISBN 0-9620317-2-0.

[8] Tanino, T., Fuzzy Preference Orderings in Group Decision Making. *Fuzzy Sets and Systems* 12 (1984) 117-131, ISSN 0165-0114.

Monotone eigenspace structure of a max-Łukasiewicz fuzzy matrix

Imran Rashid[1], Martin Gavalec [2], Richard Cimler [3]

Abstract.

The structure of the set of all monotone eigenvectors (monotone eigenspace) is studied of a given fuzzy (max,$Ł$)-matrix with the Łukasiewicz fuzzy triangular norm. Necessary and sufficient conditions are presented under which the monotone eigenspace of a given matrix is non-empty. The structure of the eigenspace is then described. The work is a continuation of previous works of the authors, in which analogous problems were solved for fuzzy (max,t)-matrices with the drastic and the product fuzzy triangular norms.

Keywords: Łukasiewicz triangular norm, max-t fuzzy algebra, eigenproblem, monotone eigenvector
AMS classification: 08A72, 90B35, 90C47

1 Introduction

The eigenproblem for a fuzzy matrix corresponds to finding a stable state of a complex discrete-events system, described by a given transition matrix and fuzzy state vectors. Hence, investigation of the eigenspace structure in fuzzy algebras is important for applications, see [2], [3], [8]. The eigenproblem has been studied by many authors in the case of max-min fuzzy algebra, see [1], [4], [10]. Complete structure of the eigenspace in max-min fuzzy algebra was described in [5] and for fuzzy max-t algebras for the drastic t-norm and for the product t-norm in [6] and [7] respectively.

In this paper the structure of the eigenspace for matrices in max-Łukasiewicz algebra is considered. For simplicity, the eigenproblem is studied for three-dimensional fuzzy matrices and strictly increasing eigenvectors. Necessary and sufficient conditions are shown under which the eigenspace restricted to strictly increasing eigenvectors of a given matrix is non-empty, and the structure of the strictly increasing eigenspace is described. Then, using simultaneous row and column permutations of the matrix, complete characterization is given of the general eigenspace structure of an arbitrary three-dimensional fuzzy matrix.

2 Eigenvectors in max-t algebra

Let us denote the real unit interval as $\mathcal{I} = \langle 0,\, 1\rangle$, let t be one of the triangular norms used in fuzzy theory. By max-t algebra we understand a triple $(\mathcal{I}, \oplus, \otimes)$ with $\oplus = \max$ and $\otimes = t$, binary operations on $\mathcal{I}$. For a given natural n, we denote $N = \{1, 2, \dots, n\}$. Further, the notation $\mathcal{I}(n, n)$ ($\mathcal{I}(n)$) denotes the set of all square matrices (all vectors) of a given dimension n over $\mathcal{I}$. Operations $\oplus$, $\otimes$ are analogously extended to matrices and vectors in a formal way.

The eigenproblem for a given matrix $A \in \mathcal{I}(n, n)$ in max-t algebra consists of finding an eigenvector $x \in \mathcal{I}(n)$ for which $A \otimes x = x$ holds true. Eigenspace of a matrix $A \in \mathcal{I}(n, n)$ is denoted by $\mathcal{F}(A) := \{\, x \in \mathcal{I}(n);\, A \otimes x = x \,\}$.

[1]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: imran.rashid@uhk.cz
[2]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: martin.gavalec@uhk.cz
[3]University of Hradec Králové, Faculty of Informatics and Management, Rokitanského 62, 50003 Hradec Králové, Czech Republic, e-mail: richard.cimler@uhk.cz

Investigation of the eigenspace structure can be simplified by permuting any vector $x \in \mathcal{I}(n)$ to an increasing form. For given permutations φ, $\psi \in P_n$ we denote by $A_{\varphi\psi}$ the matrix with rows permuted by φ and columns permuted by ψ, and we denote by x_φ the vector permuted by φ, see [5], [6].

Theorem 1. *[5] Let $A \in \mathcal{I}(n,n)$, $x \in \mathcal{I}(n)$ and $\varphi \in P_n$. Then $x \in \mathcal{F}(A)$ if and only if $x_\varphi \in \mathcal{F}(A_{\varphi\varphi})$.*

We define the *increasing* eigenspace of a matrix $A \in \mathcal{I}(n,n)$ as

$$\mathcal{F}^{\leq}(A) := \left\{ x \in \mathcal{I}(n);\ A \otimes x = x,\ (\forall i,j)[\, i \leq j \ \Rightarrow\ x_i \leq x_j \,] \right\} ,$$

and the *strictly increasing* eigenspace as

$$\mathcal{F}^{<}(A) := \left\{ x \in \mathcal{I}(n);\ A \otimes x = x,\ (\forall i,j)[\, i < j \ \Rightarrow\ x_i < x_j \,] \right\} .$$

Similar notation $\mathcal{I}^{\leq}(n)$ and $\mathcal{I}^{<}(n)$ will be used without the condition $A \otimes x = x$.

In fuzzy sets theory, various triangular norms are being used. The most frequent of them are

Gödel norm	$G(x,y) = \min(x,y)$
product norm	$\mathrm{prod}(x,y) = (x \cdot y)$
drastic norm	$\mathrm{drast}(x,y) = \begin{cases} \min(x,y) & \text{if } \max(x,y) = 1 \\ 0 & \text{if } \max(x,y) < 1 \end{cases}$
Łukasiewicz norm	$Ł(x,y) = \max(x+y-1, 0)$

The max-t fuzzy algebra with the Gödel norm is a special case of the max-min algebra.The eigenspace for this case is described in [5]. The eigenspace for the max-drast fuzzy algebra and max-prod fuzzy algebra with the drastic norm and product norm was discussed in [6] and [7] respectively.

3 Eigenvectors in max-Łukasiewicz algebra

In the rest of this paper we shall work with the max-Łukasiewicz fuzzy algebra $(\mathcal{I}, \oplus, \otimes_l)$ with binary operation $\oplus = \max$ and $\otimes_l = Ł$. Hence, for vectors $x, y \in \mathcal{I}(n)$ we have

$$(x \oplus y)_i = \max(x_i, y_i)$$
$$(x \otimes_l y)_i = \begin{cases} x_i + y_i - 1 & \text{if } \min(x_i + y_i - 1, 0) = 0 \\ 0 & \text{otherwise} \end{cases}$$

for every $i \in N$.

The proofs of the propositions and theorems in this section can be found in [9].

Proposition 2. *Let $A \in \mathcal{I}(n,n)$, $x \in \mathcal{I}^{<}(n)$. Then $x \in \mathcal{F}^{<}(A)$ if and only if for every $i \in N$ the following hold*

$$a_{ij} \otimes_l x_j \ \leq\ x_i \quad \text{for every} \quad j \in N, \quad j \geq i \tag{1}$$
$$a_{ij} \otimes_l x_j \ =\ x_i \quad \text{for some} \quad j \in N, \quad j \geq i \tag{2}$$

Theorem 3. *Let $A \in \mathcal{I}(n,n)$ and $x \in \mathcal{I}^{<}(n)$. Then $x \in \mathcal{F}^{<}(A)$ if and only if for every $i \in N$ the following hold*

(i) $a_{ij} \leq 1 + x_i - x_j$ *for every* $j \in N, j \geq i$

(ii) *if* $i = 1$, *then* $x_1 = 0$ *or* $a_{1j} = 1 + x_1 - x_j$ *for some* $j \in N$

(iii) $a_{ij} = 1 + x_i - x_j$ *for some* $j \in N, j \geq i > 1$

The following theorem describes necessary conditions under which a square matrix can have a strictly increasing eigenvector.

Theorem 4. *Let $A \in \mathcal{I}(n,n)$. If $\mathcal{F}^{<}(A) \neq \emptyset$, then the following conditions are satisfied*

(i) $\quad a_{ij} < 1$ *for all* $i, j \in N, i < j$,

(ii) $\quad a_{nn} = 1$.

Remark 1. It can be easily seen that in the case $n = 2$ the necessary conditions in Theorem (4) are also sufficient. Namely, if $n = 2$ then we have two conditions: $a_{12} < 1$ and $a_{22} = 1$. Then, arbitrary vector $x_1 = 0$ and $0 < x_2 \leq 1 - a_{12}$ fulfills the conditions of Theorem (3). Hence the vector $x = (x_1, x_2)$ with $x_1 = 0$ and $0 < x_2 \leq 1 - a_{12}$ is a strictly increasing eigenvector of A. In the particular case when $a_{11} = 1$, x_1 can even take arbitrary values from the interval $\langle 0, 1)$. This result is explicitly formulated in the following theorem.

Theorem 5. *Let $A \in \mathcal{I}(2,2)$. Then $\mathcal{F}^{<}(A) \neq \emptyset$ if and only if $a_{12} < 1$ and $a_{22} = 1$. In this case,*

$$\begin{aligned}
&\textit{if } a_{11} = 1, \textit{ then} \\
&\qquad \mathcal{F}^{<}(A) = \big\{(x_1, x_2) \in \mathcal{I}(2) : x_1 \in \langle 0, 1),\ x_1 < x_2 \leq \min(1, 1 + x_1 - a_{12})\big\}, \qquad (3) \\
&\textit{if } a_{11} < 1, \textit{ then} \\
&\qquad \mathcal{F}^{<}(A) = \big\{(x_1, x_2) \in \mathcal{I}(2) : 0 = x_1 < x_2 \leq 1 - a_{12}\big\} \qquad (4) \\
&\quad \textit{or} \quad \mathcal{F}^{<}(A) = \big\{(x_1, x_2) \in \mathcal{I}(2) : 0 < x_1 \leq a_{12}, x_2 = 1 + x_1 - a_{12}\big\}.
\end{aligned}$$

In the next theorem, a necessary and sufficient condition for the existence of a non-zero constant eigenvector is presented. The set of all constant eigenvectors of a matrix A is denoted by $\mathcal{F}^{=}(A)$.

Theorem 6. *Let $A \in \mathcal{I}(n,n)$, then there is a non-zero constant eigenvector $x \in \mathcal{F}^{=}(A)$ if and only if*

(i) $\quad \max\{\, a_{ij};\, j \in N\,\} = 1$ *for every* $i \in N$.

Theorem 7. *Let $A \in \mathcal{I}(n,n)$. If the condition (i) of Theorem (6) is satisfied, then*

$$\mathcal{F}^{=}(A) = \big\{(c, c, , \ldots, c);\ c \in \mathcal{I}\big\} \ .$$

If A does not satisfy the condition (i), then $\mathcal{F}^{=}(A) = \big\{(0, 0, , \ldots, 0)\big\}$.

4 Eigenvectors in the three-dimensional case

In this section we consider the three-dimensional eigenproblem in the max-Łukasiewicz fuzzy algebra. In other words, we assume $n = 3$, hence we work with matrices in $\mathcal{I}(3,3)$ and vectors in $\mathcal{I}(3)$. The results from the previous sections will be extended and a complete description of the eigenspace will be given.

The following theorem describes the necessary and sufficient conditions under which a three-dimensional fuzzy matrix has a strictly increasing eigenvector.

Theorem 8. *Let $A \in \mathcal{I}(3,3)$. Then $\mathcal{F}^{<}(A) \neq \emptyset$ if and only if the following conditions are satisfied*

(i) $\quad a_{12} < 1$, $\quad a_{13} < 1$, $\quad a_{23} < 1$,
(ii) $\quad a_{22} = 1 \quad$ *or* $\quad a_{13} < a_{23}$,
(iii) $\quad a_{33} = 1$.

Proof. Let $\mathcal{F}^{<}(A) \neq \emptyset$, i.e. there exists $x \in \mathcal{F}^{<}(A)$. The conditions (i) and (iii) follow directly from Theorem 3. To prove the condition (ii), let us assume that $a_{22} < 1$. Then by (iii) of Theorem (3)we get $a_{22} = 1 + x_2 - x_2$ or $a_{23} = 1 + x_2 - x_3$. The first equation implies $a_{22} = 1$, which is a contradiction. Therefore we must have $a_{23} = 1 + x_2 - x_3 < 1$. By (i) of Theorem (3), we have $a_{13} \leq 1 + x_1 - x_3 = 1 + (x_1 - x_2) + (x_2 - x_3) = (1 + x_2 - x_3) + (x_1 - x_2) < 1 + x_2 - x_3 = a_{23}$. This implies $a_{13} < a_{23}$.

Conversely, suppose that conditions (i), (ii) and (iii) are satisfied. To show that $\mathcal{F}^{<}(A) \neq \emptyset$, we have two cases:

Case 1. If $a_{22} < 1$, then put $x_1 = 0$ and choose $x_2 \leq \min(1 - a_{12}, a_{23} - a_{13})$ and $0 < x_2 < 1 - a_{13}$, and put $x_3 = x_2 + (1 - a_{23})$. By our assumption, $1 - a_{12} > 0$, $a_{23} - a_{13} > 0$ and $1 - a_{13} > 0$. Therefore the choice of x_2 fulfills the conditions. That is $x_2 \leq \min(1 - a_{12}, a_{23} - a_{13})$ and $0 < x_2 < 1 - a_{13}$ is always possible. Also by assumption, $a_{23} < 1$ implies that $1 - a_{23} > 0$. Therefore $x_3 = x_2 + (1 - a_{23}) > x_2$. Moreover, $x_2 \leq \min(1 - a_{12}, a_{23} - a_{13}) \leq a_{23} - a_{13} \leq a_{23}$, i.e. $x_2 \leq a_{23}$. From this we have $x_3 = x_2 + (1 - a_{23}) \leq 1$, i.e. $x_3 \leq 1$. This shows that $x \in \mathcal{I}^<(3)$. To show that $x \in \mathcal{F}^<(A)$, consider $x_2 \leq \min(1 - a_{12}, a_{23} - a_{13})$ and $0 < x_2 < 1 - a_{13}$. This implies that $x_2 \leq 1 - a_{12}$ and $0 < x_2 < 1 - a_{13}$ or $a_{12} \leq 1 - x_2$ and $a_{13} < 1 - x_2$ satisfying condition (i) of Theorem (3). Choice of $x_1 = 0$ satisfies the condition (ii) of Theorem (3). Also, $x_3 = x_2 + (1 - a_{23})$ or $a_{23} = 1 + x_2 - x_3$ satisfies condition (iii) of Theorem (3). Hence $\mathcal{F}^<(A) \neq \emptyset$.

Case 2. If $a_{22} = 1$, then put $x_1 = 0$, choose $0 < x_2 < \min(1 - a_{12}, 1 - a_{13})$ and choose x_3 such that $x_2 < x_3 \leq \min(1 - a_{13}, x_2 + (1 - a_{23}))$. The choice $0 < x_2 < \min(1 - a_{12}, 1 - a_{13})$ is always possible because by our assumption $1 - a_{12} > 0$ and $1 - a_{13} > 0$. Also, $1 - a_{23} > 0$ by the same argument and therefore $x \in \mathcal{I}^<(3)$. Consider $x_2 < \min(1 - a_{12}, 1 - a_{13}) \leq 1 - a_{12}$. Then $x_2 < 1 - a_{12}$ implies that $a_{12} < 1 - x_2$. Similarly $a_{13} < 1 - x_2$ and by $x_3 \leq \min(1 - a_{13}, x_2 + (1 - a_{23}))$ we have $a_{23} \leq 1 + x_2 - x_3$, showing that condition (i) of Theorem (3)is satisfied. Conditions (ii) and (iii) of Theorem (3)are satisfied by the choice of $x_1 = 0$ and assumption $a_{22} = 1$ respectively. Hence $\mathcal{F}^<(A) \neq \emptyset$. $\square$

Theorem 9. *Let $A \in \mathcal{I}(3,3)$ satisfies the conditions (i), (ii) and (iii) of Theorem 8. Then $\mathcal{F}^<(A)$ is the union of two disjoint sets $\mathcal{F}_0^<(A)$ and $\mathcal{F}_1^<(A)$.*

$\mathcal{F}_0^<(A)$ consists exactly of all vectors $x = (x_1, x_2, x_3) \in \mathcal{I}^<(3)$ with $x_1 = 0$ and satisfying conditions

$$\begin{aligned}&\textit{if} \quad a_{22} < 1, \ \textit{then}\\ &\quad 0 < x_2 \leq \min(1 - a_{12}, a_{23} - a_{13}), \ x_2 < 1 - a_{13} \ \textit{and} x_3 = x_2 + (1 - a_{23}) \ ,\end{aligned} \tag{5}$$

$$\begin{aligned}&\textit{if} \quad a_{22} = 1, \ \textit{then}\\ &\quad 0 < x_2 < \min(1 - a_{12}, 1 - a_{13}) \ \textit{and} x_2 < x_3 \leq \min(1 - a_{13}, x_2 + (1 - a_{23})) \ .\end{aligned} \tag{6}$$

$\mathcal{F}_1^<(A)$ consists exactly of all vectors $x = (x_1, x_2, x_3) \in \mathcal{I}^<(3)$ with $x_1 > 0$ and satisfying conditions

$$\begin{aligned}&\textit{if} \quad a_{11} = 1, a_{22} = 1, \ \textit{then}\\ &\quad x_1 < 1, x_1 < x_2 < 1 \ \textit{and} x_2 \leq x_1 + (1 - a_{12}), x_3 \leq \min(x_1 + (1 - a_{13}), x_2 + (1 - a_{23}), 1) \ ,\end{aligned} \tag{7}$$

$$\begin{aligned}&\textit{if} \quad a_{11} = 1, a_{22} < 1, \ \textit{then}\\ &\quad x_1 < a_{23}, x_2 \leq \min(a_{23}, x_1 + (1 - a_{12}), x_1 + (a_{23} - a_{13})), x_3 = x_2 + (1 - a_{23}) \ ,\end{aligned} \tag{8}$$

$$\begin{aligned}&\textit{if} \quad a_{11} < 1, a_{22} = 1, \ \textit{then}\\ &\quad x_1 < a_{12}, x_2 = x_1 + (1 - a_{12}), x_3 \leq \min(1, x_1 + 2 - (a_{12} + a_{23}))\\ \textit{or} &\quad x_1 \leq a_{13}, x_1 + (a_{23} - a_{13}) \leq x_2 < x_1 + (1 - a_{13}), x_3 = x_1 + (1 - a_{13}) \ ,\end{aligned} \tag{9}$$

$$\begin{aligned}&\textit{if} \quad a_{11} < 1, a_{22} < 1, \ \textit{then}\\ &\quad x_1 \leq a_{13}, x_2 = x_1 + (a_{23} - a_{13}), x_3 = x_1 + (1 - a_{13})\\ \textit{or} &\quad x_1 \leq a_{12} + a_{23} - 1, x_2 = x_1 + (1 - a_{12}), x_3 = x_1 + 2 - (a_{12} + a_{23}) \ .\end{aligned} \tag{10}$$

Proof. The proof of the theorem can be found in [9].

$\square$

Remark 2. The inequalities $x_1 < a_{12}$, $x_1 \leq a_{13}$, $x_1 \leq a_{12} + a_{23} - 1$ in conditions (9) and (10) can only be satisfied if $a_{12} > 0$, $a_{13} > 0$, $a_{12} + a_{23} - 1 > 0$ respectively. E.g., in the condition (10), when $a_{12} + a_{23} - 1 < 0$ then there are no eigenvectors of the corresponding type. If $a_{13} = 0$ or $a_{12} + a_{23} - 1 = 0$, then we have a solution $x_1 = 0$, $x_2 = 1 - a_{12} = a_{23}$ and $x_3 = 1$ described in condition (6).

5 Algorithm and examples

Based on Theorems 5, 7, 8, 9, algorithms for computation of the eigenspace of a given max-Łukasiewicz matrix A have been prepared and encoded as computer program in Java language. In Figure 1, an output of the program is shown.

Figure 1: Output of the program

Example 1. Considerations from this section will be demonstrated by computing the eigenspace $\mathcal{F}(A)$ of a two-dimensional matrix $A \in \mathcal{I}(2,2)$. As any two-dimensional vector $x = (x_1, x_2)$ is either strictly increasing, or strictly decreasing, or constant, the eigenspace can be written in the form $\mathcal{F}(A) = \mathcal{F}^{<}(A) \cup \mathcal{F}^{>}(A) \cup \mathcal{F}^{=}(A)$. The eigenspace $\mathcal{F}^{>}(A)$ is computed using the permuted matrix $A_{\varphi\varphi}$ with $\varphi = (2 \quad 1)$, according to Theorem 1.

$$A = \begin{pmatrix} 0.7 & 0.3 \\ 0.2 & 1 \end{pmatrix} , \qquad A_{\varphi\varphi} = \begin{pmatrix} 1 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$$

By Theorem 5 and Theorem 7 we immediately get that the eigenspace $\mathcal{F}(A)$ consists of

- strictly increasing eigenvectors: $(0 = x_1 < x_2 \leq 0.7)$ or $(0 < x_1 \leq 0.3,\ x_2 = x_1 + 0.7)$,
- no strictly decreasing eigenvectors
- exactly one constant eigenvector: $x = (0,0)$

Example 2. In this example we change the input $a_{11} = 0.7$ to $b_{11} = 1$, and leave the other inputs unchanged. Similarly as above, we compute the eigenspace $\mathcal{F}(B) = \mathcal{F}^{<}(B) \cup \mathcal{F}^{>}(B) \cup \mathcal{F}^{=}(B)$ of matrix B using the permuted matrix $B_{\varphi\varphi}$.

$$B = \begin{pmatrix} 1 & 0.3 \\ 0.2 & 1 \end{pmatrix} , \qquad B_{\varphi\varphi} = \begin{pmatrix} 1 & 0.2 \\ 0.3 & 1 \end{pmatrix}$$

The eigenspace $\mathcal{F}(B)$ consists of

- strictly increasing eigenvectors: $0 \leq x_1 < 1$, $x_1 < x_2 \leq \min(1, x_1 + 0.7)$
- strictly decreasing eigenvectors: $0 \leq x_1 < 1$, $x_1 < x_2 \leq \min(1, x_1 + 0.8)$
- constant eigenvectors $x = (c,\ c)$ with $0 \leq c \leq 1$

Example 3. Let us consider the matrix

$$A = \begin{pmatrix} 0.6 & 0.8 & 0.3 \\ 0.5 & 0.9 & 0.4 \\ 0.3 & 0.7 & 1 \end{pmatrix}$$

Matrix A satisfies the conditions (i), (ii) and (iii) of Theorem 8, hence $\mathcal{F}^{<}(A) \neq \emptyset$. By Theorem 9, the strictly increasing eigenspace of A is $\mathcal{F}^{<}(A) = \mathcal{F}^{<}_0(A) \cup \mathcal{F}^{<}_1(A)$. Since $a_{22} < 1$, then by condition (5) of Theorem 9

$$\mathcal{F}_0^<(A) = \{(x_1, x_2, x_3) \in \mathcal{I}(3) : 0 = x_1 < x_2 \leq 0.1, x_2 < 0.7,\ x_3 = x_2 + 0.6\}$$

Also, since $a_{11} < 1$, $a_{22} < 1$, then by condition (10) of Theorem 9 we have $\mathcal{F}_1^<(A) = \mathcal{F}_{1a}^<(A) \cup \mathcal{F}_{1b}^<(A)$. with

$$\mathcal{F}_{1a}^<(A) = \{(x_1, x_2, x_3) \in \mathcal{I}(3) : 0 < x_1 \leq 0.3, x_2 = x_1 + 0.1, x_3 \leq x_1 + 0.7\}$$
$$\mathcal{F}_{1b}^<(A) = \{(x_1, x_2, x_3) \in \mathcal{I}(3) : 0 < x_1 < 0.2, x_2 = x_1 + 0.2, x_3 = x_1 + 0.8\}$$

By Theorem 7 we get that A has only one constant eigenvector $(0, 0, 0)$.

Acknowledgements

The support of the project INKOV # CZ.1.07/2.3.00/20.0001, financed from EU and Czech Republic funds, and the Czech Science Foundation project # 402/09/0405 is kindly acknowledged.

References

[1] K. Cechlárová, *Efficient computation of the greatest eigenvector in fuzzy algebra*, Tatra Mt. Math. Publ. **12** (1997), 73-79.

[2] G. Cohen, D. Dubois, J. P. Quadrat, M. Viot, *A linerar-system-theoretic view of discrete event processes and its use for performance evaluation in manufacturing*, IEE Transactions on Automatic Control, AC-30 (1985), 210-220.

[3] R.A. Cuninghame-Green: *Minimax Algebra*, Lecture Notes in Economics and Mathematical Systems **166**, Springer–Verlag, Berlin, 1979.

[4] R.A. Cuninghame-Green: *Minimax Algebra and Application*, in: Advances in Imaging and Electron Physics **90**, (P. W. Hawkes, Ed.). Academic Press, New York, 1995.

[5] M. Gavalec: *Monotone eigenspace structure in max-min algebra*, Lin. Algebra Appl. **345** (2002), 149–167.

[6] M. Gavalec, I. Rashid: *Monotone eigenspace structure of a max-drast fuzzy matrix*, Proc. of the 28th Int. Conf. Mathematical Methods in Economics, University of South Bohemia, České Budějovice (2010), 162–167.

[7] M. Gavalec, I. Rashid: *Monotone eigenspace structure of a max-prod fuzzy matrix*, submitted to Fuzzy Sets and Systems, February 2010.

[8] G. Olsder, *Eigenvalues of dynamic max-min systems*, in: Discrete Events Dynamic Systems 1, Kluwer Academic Publishers, 1991, 177-201.

[9] I. Rashid: *Eigenspace structure of a max-t fuzzy matrix*, Doctoral Thesis, University of Hradec Králové, May 2011.

[10] E. Sanchez, *Resolution of eigen fuzzy sets equations*, Fuzzy Sets and Systems **1** (1978), 69-74.

Optimal Strategy in Sports Betting

Jiří Rozkovec[1]

Abstract. The aim of this paper is to find the optimal strategy for a bettor who bets on soccer matches. We will consider only single odds, i.e. either a win for the home team, a draw, or a win for the away team. No double chances or multiple odds etc. The result of a particular match is a random variable and the probability of its result is inverse in proportion to the odd given by a bookmaker. Further, the odds of the home team and the odds of the away team are not independent variables – they are also inversely proportional. Thus the odds of a draw are only a transformed probability, which is a complement of two other probabilities to get one as their sum. Therefore the bettor chooses the matches and the events only according to the home odds and does not take into account the position on the table, the latest results etc. The optimal strategy should identify the odds and the result which the bettor should bet on. For this optimization we use historical data. For every match we have three odds (home win, draw, away win). We keep only the odds which correspond to the right result. The other ones we substitute with zeros. The modified triples of odds are sorted by the odds for home wins and then they are smoothed by moving averages. If we find a contiguous segment of smoothed values greater than one, it means that there are intervals of odds (home win) where the odds for home wins bring a profit. We also use the same method for the odds for draws and also for away wins. The data was evaluated on an annul basis from 2001 till 2010. Specifically, we analysed the top european soccer leagues, the English Premier League, German 1st Bundesliga, Italian Serie A etc.

Keywords: odds, sports betting, moving averages, profitable intervals.

JEL Classification: C13
AMS Classification: 62G05

1 Introduction

Every week milions of bettors bet on hundreds of sporting matches. Some of them do this only for fun, but some of them are professionals, who earn by betting a lot of money. Sport betting, when compared to lotteries, needs expertise and experience concerning the selected branch of sport. We will try to find a strategy which brings a profit regardless of circumstances connected with the chosen match. The following calculations result from the data of the English Premier League from August 2008 till May 2009.

1.1 Assumptions

We consider the result of any match as a random variable. The probabilities of all the possible outcomes (home win, draw, away win) are transformed by bookmakers into the odds. This is the only information which we have and use in our strategy. Secondly, there is a negative correlation between home odds and away odds - the greater the home odds (i.e. the lower probability of the home win), the lesser the away odds (the greater probability of the away win) and vice versa - see Fig. 1. When we tried to find any regression function for the mentioned relation between the home odds (denoted as X) and the away ones (Y), we estimated the parameters[2] of a hyperbola:

$$Y = 2{,}66/(X - 1) + 1$$

[1] Technical University of Liberec, Faculty of Economics, Department of Economic Statistics, Studentska 2, 46117 Liberec, Czech Republic, jiri.rozkovec@tul.cz

[2] At confidence level 95%.

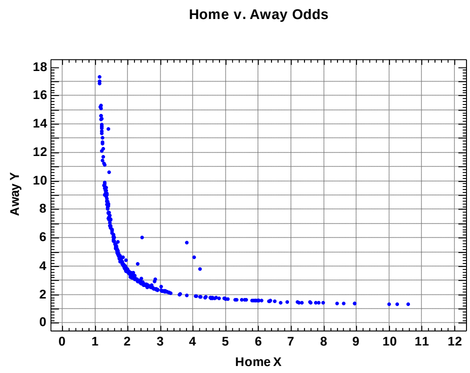


Figure 1 Odds – Home win v. Away win

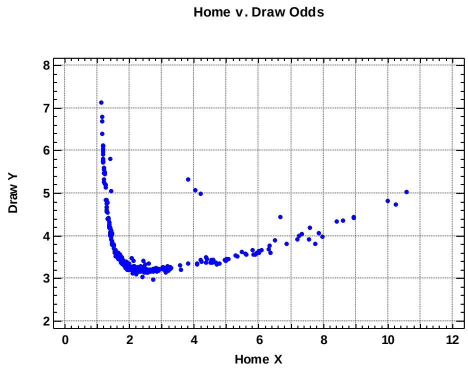


Figure 2 Odds – Home win v. Draw

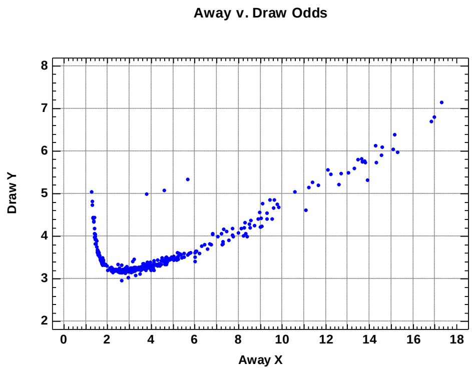


Figure 3 Odds – Away win v. Draw

The third property of the odds is that there are only two degrees of freedom[3] when a bookmaker sets the odds. When he creates two odds, the last one is just a complementary probability to get one summarizing all three probabilities. For illustration we present Fig. 2 and Fig. 3 with the other possible pairs of the odds.

[3] Thereinafter we deal only with soccer matches which have three possible outcomes.

2 Data Arrangement

To find the strategy we deal with historical data - the one year history of the odds of a particular soccer league. We know all the starting odds[4] (home, draw, away) of all the matches and their results. We denote the number of matches as n, the home win as '1', the draw as '0' and away win as '2'. Then j-th match is represented by four numbers: r_{j1} - home win odds, r_{j0} - draw odds, r_{j2} - away win odds and the result $v_j \in \{1,0,2\}$. The basic question for the bettor is whether he bets on the favorite team or not. Though the favorites win very often, but there are low odds. The odds for the opponents are much higher, but the frequency of their success is naturally lower (see also [1]).

With respect to the patterns mentioned above we sort the matches (the triples of odds corresponding to the matches) ascending from the home odds (follows as indicator) and get a matrix $\boldsymbol{R}(n\times 3)= r_{jk}$, $j=1,...,n$, $k=1,0,2$. Further we denote the first column of matrix $\boldsymbol{R}$ (home odds) as a vector $\boldsymbol{r}_1$, the second column as $\boldsymbol{r}_0$ (draw odds) and the third one as $\boldsymbol{r}_2$ (away odds).

Now we transform $\boldsymbol{r}_1$ to $\boldsymbol{c}_1$ this way:

$$c_{j1} = \begin{cases} r_{j1} & v_j = 1 \\ 0 & v_j \neq 1 \end{cases} \tag{1}$$

for $j=1,...,n$. In other words, if the homes won, then $c_{j1}=r_{j1}$. Otherwise $c_{j1}=0$, if the home team did not win the particular match. If the bettor had bet a unit of currency for the home win on every match, then c_{j1} would have equaled to the bettor's income. Analogously we get vectors $\boldsymbol{c}_0$ and $\boldsymbol{c}_2$ respectively, where:

$$c_{j0} = \begin{cases} r_{j0} & v_j = 0 \\ 0 & v_j \neq 0 \end{cases} \tag{2}$$

$$c_{j2} = \begin{cases} r_{j2} & v_j = 2 \\ 0 & v_j \neq 2 \end{cases} \tag{3}$$

So we transform matrix $\boldsymbol{R}=(\boldsymbol{r}_1,\boldsymbol{r}_0,\boldsymbol{r}_2)$ to matrix $\boldsymbol{C}=(\boldsymbol{c}_1,\boldsymbol{c}_0,\boldsymbol{c}_2)$. There are some selected values of $\boldsymbol{R}$ and $\boldsymbol{C}$ in Tab. 2[5]. It was said that when the bettor bets a currency unit on a particular outcome (e.g. a draw) of every match, such a strategy will be called a *unit strategy*. This means the total money staked is n units. If we sum components of vector $\boldsymbol{c}_1$, we will get the total income

$$P_1 = \sum_{j=1}^{n} c_{j1}$$

which the bettor would have got, if he had bet on a home win for all the matches of the selected league during one season. Instead of P_1 we can use a mean P_1/n. If $P_1>n$, $(P_1/n)>1$ respectively, then this strategy generates a profit from the home wins (analogously for the draw and away win).

P_1	P_0	P_2
312,54	340,55	355,29

Table 1 The total incomes

[4] We use so called *european format* of the rate, a real number greater than 1.
[5] There are 380 matches per year, so $\boldsymbol{R}$ and $\boldsymbol{C}$ have 380 rows.

In Tab. 1 there are values of P_1, P_0, P_2 and all of them are less than 380. This means the unit strategy generates a loss in general for all the outcomes. Naturally, betting agencies construct the odds just to prevent such simple “money machine” strategies like the unit strategy is.

Played from 16.8.2008 to 24.5.2009				$\boldsymbol{R}=(\boldsymbol{r}_1,\boldsymbol{r}_0,\boldsymbol{r}_2)$			$\boldsymbol{C}=(\boldsymbol{c}_1,\boldsymbol{c}_0,\boldsymbol{c}_2)$		
j	**Home team**	**Away team**	**Result** $\boldsymbol{v}_j$	$\boldsymbol{r}_{j1}$	$\boldsymbol{r}_{j0}$	$\boldsymbol{r}_{j2}$	$\boldsymbol{c}_{j1}$	$\boldsymbol{c}_{j0}$	$\boldsymbol{c}_{j2}$
1	Manchester U.	Sunderland	1	1,16	6,5	17	1,16	0	0
2	Chelsea	W. Bromwich	1	1,18	6	17	1,18	0	0
3	Manchester U.	Stoke City	1	1,18	6	17	1,18	0	0
4	Manchester U.	Hull City	1	1,18	6	17	1,18	0	0
5	Arsenal	Hull City	2	1,20	5,75	15	0	0	15
6	Manchester U.	W. Bromwich	1	1,20	5,75	15	1,20	0	0
…	…	…	…	…	…	…	…	…	…
194	Aston Villa	Everton	0	2,20	3,25	3,20	0	3,25	0
195	Blackburn	Newcastle	1	2,20	3,25	3,20	2,20	0	0
196	Blackburn	West Ham	0	2,20	3,25	3,20	0	3,25	0
197	Bolton	Wigan	2	2,20	3,25	3,20	0	0	3,20
…	…	…	…	…	…	…	…	…	…
378	W. Bromwich	Chelsea	2	9	4,50	1,35	0	0	1,35
379	Wigan	Manchester U.	2	9	4,75	1,33	0	0	1,33
380	Stoke City	Manchester U.	2	9,50	5	1,30	0	0	1,30

Table 2 Odds and Profits (English Premier League 2008/2009)

3 Profitable intervals

However, if we further analyse the columns of matrix $\boldsymbol{C}=(\boldsymbol{c}_1,\boldsymbol{c}_0,\boldsymbol{c}_2)$, we find that there are subsequences with a dense occurance of positive numbers (odds), i.e. with a rare frequency of zeros. It is clear that the vector $\boldsymbol{c}_1$ has these positive values c_{j1} however for smaller values of j, for $\boldsymbol{c}_2$ it is reversed. There are positive c_{j2} most frequently at the end of this vector. See Fig. 4 and Fig. 5.

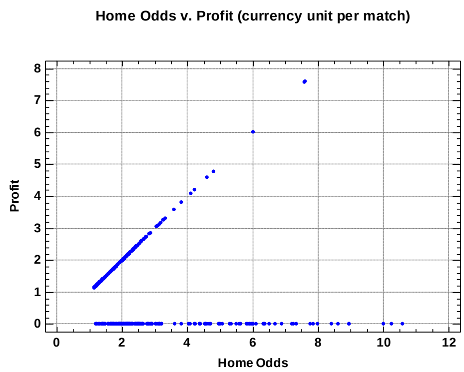


Figure 4 Vector $\boldsymbol{c}_1$

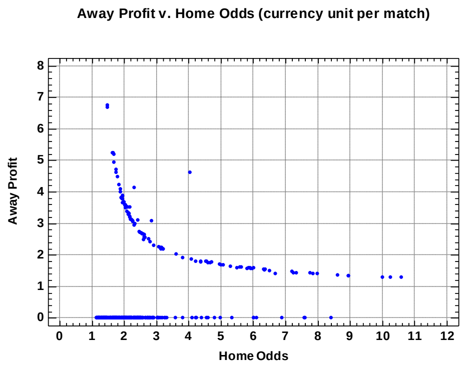


Figure 5 Vector $\boldsymbol{c}_2$

Based on these facts, the goal is to identify these (profitable) segments within each vector $\boldsymbol{c}_k$, where the unit strategy should on average bring a profit. The simplest tools which we can use for this are *simple moving averages*. Their order *m* we decided to set approximately to 10% of observations, i. e. *m*=39. Thus we smooth the values of the vector $\boldsymbol{c}_k$ using the formula:

$$\hat{c}_{jk} = \frac{1}{m}\sum_{i} c_{j+i,k} \quad i = -(m-1)/2, ..., (m-1)/2, j = (m+1)/2, ..., n-(m-1)/2. \quad (4)$$

and get the vector $\boldsymbol{c}_k$. For example see Fig. 6, the moving averages are denoted as MA(39).

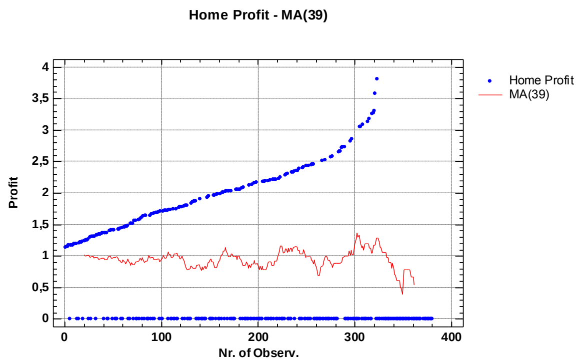


Figure 6 Moving Averages for Home Win

The profitable intervals are indicated by the values of the moving averages (from (4)) which are greater than one. From Fig. 6 we get intervals [2,18; 2,52] (65 obs.) and [2,58; 8,82] (76 obs.). It is necessary to say that these intervals always come from the values of the home odds apart from the outcome {1, 0, 2} which we analyse, because from these odds flow the other ones (see paragraph 1). To sum the profitable intervals for all the outcomes there is Tab. 3.

Home Win	**Draw**	**Away win**
[2,18; 2,52]	[1,18; 1,68]	[1,76; 2,21]
[2,58; 5,82]	[1,73; 1,95]	[2,87; 10,58]
	[2,07; 2,72]	

Table 3 Profitable intervals

As stated this summary is from 2008/2009 of the English Premier League. We have processed the history from 2001 and put the found profitable intervals together - see Fig. 7-9.

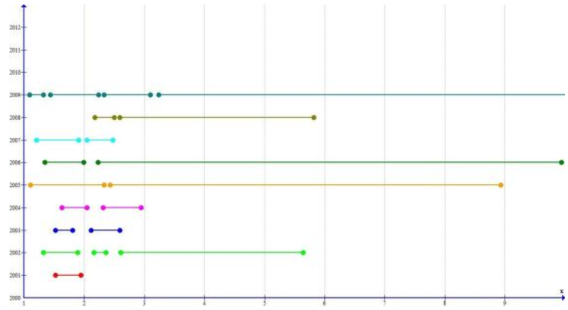


Figure 7 Profitable Intervals for Home Win

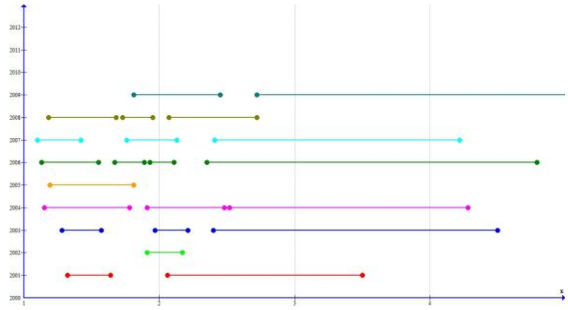


Figure 8 Profitable Intervals for Draw

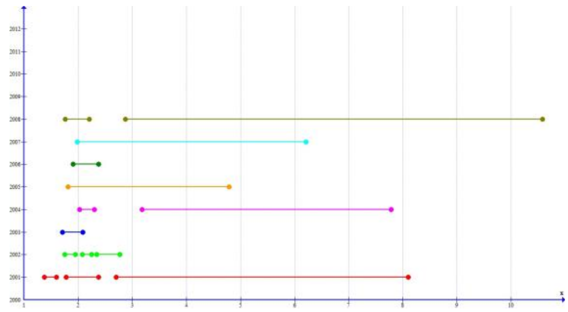


Figure 9 Profitable Intervals for Away Win

4 Conclusion

We tried to analyse a behaviour of the outcomes as a random variable at soccer matches in the top european soccer leagues. It was found that the odds are mutually dependent. Further, from the history we built the intervals of the home odds where the unit strategy would bring a profit, when the bettor bets on the particular outcome, whereas this outcome is determined only by the home odds. The profitable intervals shows irregularity, although we could find any intersection of them across the years of the researched period (Fig. 7 - Fig. 9). The reasons above all are: changes in teams, bookmakers' errors etc. For the next research we could consider any up-to-date precising of valuation to build like a combination of the historical data and the results of the current season.

Acknowledgements

Special thanks to Ing.Ondřej Dus and students Martina Slabinová and Daniela Halková for various support during writing of this paper.

References

[1] DUS, O. Optimalizace v kurzovém sázení. [Diplomová práce]. Liberec: Technická univerzita v Liberci, Ekonomická fakulta, 2010.

[2] Výsledky evropských fotbalových lig. [online]. Available from: http://www.liga.cz

Quantification of Framing Effect using ANP and AHP

Jan Rydval[1]

Abstract. Decision-making is influenced by the quality of the information, and by the size of the framing effect. In the decision-making process the framing effect arises as a set of opinions and expectations of involved subjects. Each decision-maker has his own preferences and expectations that create his own view (frame). These frames may negatively influence the information sharing in the decision-making processes and therefore it is important to analyse the framing effect. This paper compares two methods serving to the determination and quantification of the framing effect namely the Analytic Network Process and the Analytic Hierarchy Process.

The result of this approach will be the set of weights of all viewpoints included into the frames. Because of the different nature of the inside structure of both methods different results are expected. Values of these weights can serve to analysis of the framing effect, because all subjects in the decision-making process will receive the information about the importance of different viewpoints and then they can rearrange their opinions and expectations. It will have positive impact on the overall outcome of the decision-making process too.

Keywords: Framing Effect, Analytic Hierarchy Process, Analytic Network Process, Pairwise Comparison

JEL Classification: C44
AMS Classification: 90C29

1 Introduction

The information receiving process is the integral part of any decision-making processes. As mentioned by Fagley, Coleman and Simon [3] decision-making is influenced by the quality of the information and by the effect of information distortion (framing effect). And as Tversky and Kahneman [12] mentioned the framing effect included in some information may influence decisions significantly.

There are various views on a particular issue in decision process. This various views (or in other words frames) create several issues. As Bishop [1] suggests firstly if we do not sort information according to its relevance because we cannot properly decide whose view is the most important, we may face the problem of overloading with too much information resulting in either poor information acquisition or the whole process is very time consuming and thus very ineffective. On the contrary to this situation preferring the certain point of view we may omit the information needed for successful decision making.

To limit these negative frames we need firstly to define and understand them, secondly, as Druckman [2] points out, we need to evaluate them, and thirdly to use the appropriate method to reduce them. For this evaluation as Fagley, Coleman and Simon [3] write we need to know the importance of various frames and included viewpoints. These frames can be analysed using multiple attribute decision-making methods.

The Analytic Hierarchy Process is a structured technique for dealing with complex decisions. It helps decision makers find the decision that suits the best their objectives and their understanding of the issue. (Saaty [7])

The Analytic Network Process belongs to the multiple criteria decision making methods. It decomposes decision problems into a network of sub-problems that can be more easily analysed and evaluated. It is specific for this method that the human judgment is included. (Saaty [8])

These two methods of pairwise comparison differ in the view on the dependency of evaluated criteria; therefore their results also are different. While AHP looks at the criteria as independent, ANP considers them to be interdependent and able to influence each other. (Saaty, [10])

The main goal of this paper is to define and quantify the framing effect, its main set of points of view and their components, and their evaluation using the Analytic Hierarchy Process and the Analytic Network Process and the comparison of their results.

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

2 Materials and Methods

2.1 Framing Effect

Druckman [2] pointed out that individual decisions are influenced by the presented information and by the formulation of problems. People often accept and use information in a form in which it is obtained, without thinking about its form. As Rydval and Hornická [5] mentioned we can therefore define the framing effect as a set of preferences and expectations of involved subjects belonging to a particular decision-making problem. Each subject has his own preferences and expectations that create his own view (frame). The framing effect is made up of these frames. As Tversky and Kahneman [12] mentioned the framing effect influence the way of information interpretation or misinterpretation, so it may influence decision-making significantly.

Is a bottle half full or half empty? Companies prefer to raise prices instead of reporting their reductions. In employment contracts there is stated base salary plus bonuses, and allowances, rather than minus fines and penalties.

This example shows that the framing effect represents actually a distortion of information received and processed in the various decision-making processes. This then leads to information "wrapped" by redundant components such as misleading information, disinformation and various incomplete and inaccurate information. That is a way framing effect arises.

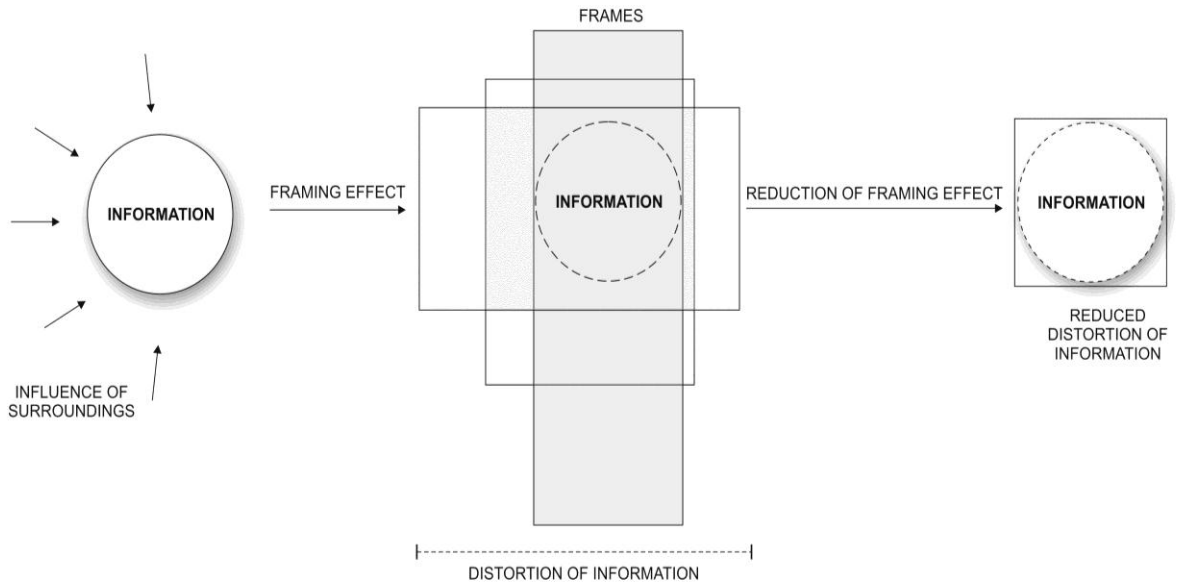


Figure 1 Framing effect (Rydval [6])

As shown in figure 1 each information or problem is affected by its surroundings which consist of various subjects. These subjects influence the problem with their opinions or expectations and thus create the particular frames of the whole framing effect. The influence of the mentioned frames is minimized by the reduction of the framing effect. Quantification of frame's importance and its components is necessary to reduce their consequences. This quantification can be provided by the Analytic Hierarchy Process or the Analytic Network Process.

2.2 Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) is a decision making approach structuring alternatives into a weighted multiple choice criteria hierarchy. AHP allows a better and more efficient identification of selection criteria, their weighting and analysis. This way AHP reduces significantly the decision cycle. (Saaty [7])

The basic elements of the AHP method are following:

- Decomposing is the first process and its aim is to structure the problem into humanly-manageable sub-problems. It splits the problem, which has no structure at this point, into sub-modules that will become sub-hierarchies.
- Weighting means assigning a relative weight to each criterion, based on its importance within the node to which it belongs.

- Evaluating value alternatives and compare each one to others. Using AHP, a relative value for each alternative is assigned to each leaf within the hierarchy, then to the branch the leaf belongs to, and so on, up to the top of the hierarchy, where an overall value is calculated.
- Selecting the alternative that best fits the requirements. (Saaty [7])

2.3 Analytic Network Process

The Analytic Network Process (ANP) is a generalization of the Analytic Hierarchy Process (AHP), by considering the dependence between the elements of the hierarchy. Many decision problems cannot be structured hierarchically because they involve the interaction and dependence of higher-level elements in a hierarchy on lower level elements. Therefore, ANP is represented by a network, rather than a hierarchy. (Saaty [8])

The basic elements of the ANP method are following:

- The first step of ANP is based on the creation of a control network which describes dependency among decision elements. The ANP allows
  - inner dependence within a set (clusters) of elements, and
  - outer dependence among different sets (clusters).
- In the second step pairwise comparisons of the elements within the clusters and among the clusters are performed according to their influence on each element in another cluster or elements in their own cluster. So the ANP prioritizes not only decision elements but also their groups or clusters as it is often the case in the real world. The consistency of these comparisons has to be controlled.
- The third step consists of the supermatrix construction. The priorities derived from the pairwise comparisons are entered into the appropriate position in this supermatrix. This supermatrix has to be normalized using clusters weights.
- In the fourth step the limiting supermatrix is computed and global preferences of decision elements are obtained. These preferences serve as the best decision selection or for the purpose of analysis of preferences of decision-making elements. (Saaty [9], [10])

2.4 SuperDecisions software

This method is carried out by the SuperDecisions software system (SuperDecisions [11]). The SuperDecisions software implements the ANP and also AHP developed by Dr. Thomas Saaty. The program was written by the Creative Decisions Foundation.

3 Results and Discussion

3.1 Case study – Framing effect in education

This case study was published at ERIE International Conference (Rydval, Brožová [4]). The published version contained only the analysis of the framing effect using AHP but in this paper the analysis of the framing affect using ANP has been introduced. In education process the framing effect can be defined as shown in figure 2. The issue of the framing in education is affected by many various subjects. These are for example teacher's personality, student's personality, labour market demand, natural conditions etc.

To simplify this framing model in our case study we will deal only with the following subjects: Student, Teacher and Labour Market. Each of these subjects has its own opinions and expectations, which follow:

Student's frame

- Difficulty of subject (How difficult it is for student to understand this subject)
- Time consumed by preparation
- Teaching method
- Practical application of acquired knowledge (and its application on job market)
- Student's ability to learn

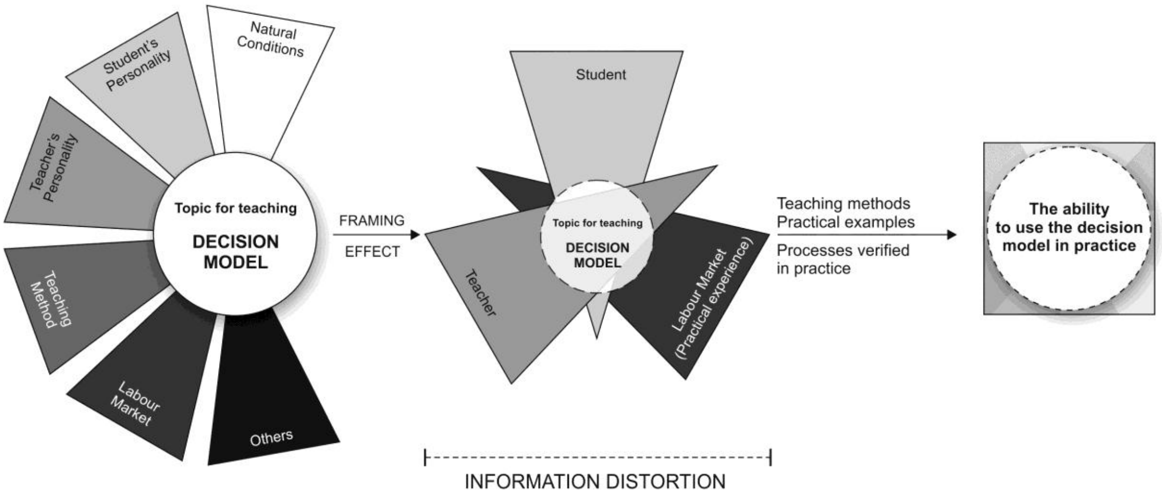


Figure 2 Framing effect in education

Teacher's frame

- Prescribed curriculum
- Previous training (teacher's formal training, other courses and seminars)
- Teaching method
- Teacher's ability to pass information (to teach)
- Subject's relevance (the significance of the subject for labour market)
- Student's ability to self-study

Labour market's frame

- Practical application (of knowledge acquired from the study)
- Skills needed for the particular job (technical, human and conceptual skills)
- Ability to adapt to market changes
- Job performance
- Analytical thinking
- Self-Reliance and initiative

These frames and included points of view have their weights of significance, which have been identified using appropriate questionnaires as follows: pairwise comparisons of frames according to the viewpoint of all subjects, and pairwise comparisons of opinions according to the viewpoint of all frames.

Each subject has his own evaluated preferences and expectations, preferences and expectations of other subjects are not important for others. That means the subject creates its own view (frame) as shown in table 1. Individual frames present the risk of a negative (or some time positive) influence on the decision process, in our case on the education process.

For example teachers generally put too much emphasis on the subject's relevance and they do not take into account the difficulty of taught subject as shown in table 1. It can lead to excessive demands on student who may then fail to learn the subject. This is the risk possessed by frames and the very reason why they have to be reduced.

The influence of the frames can be quantified using the AHP or the ANP. The SuperDecision Software (SuperDecisions [11]) was used to calculate the weights of the frames by synthesizing the preferences from the different frame's perspective.

All necessary comparisons were made and finally the decision alternatives (opinions and expectations) were preferred. Thus a complex quantification of the frames and all opinions and expectations of the education framing model were evaluated as shown in table 1 (column AHP and ANP results). These new views on the situation quantify the influence of the particular frames and their elements (other subjects' views) and also the proportion of their possible negative impact.

Above mentioned example of teachers' opinion which result in the problem of student failing in a particular subject is solved through the complex view. According to the synthesised weights of the individual opinions and expectations teacher should put less emphasis on the subject's relevance and at the same time teacher should take into account other aspects of the whole case study. Teaching method should present the subject to student in a proper form and take into account the practical use of the subject especially on the labour market. On the other hand student should not be concerned only with the difficulty of subject but he should focus mainly on how to improve his innovative thinking, skills needed for the particular job and the ability to self-study.

Set of preferences and expectations	Labour market's frame	Student's frame	Teacher's frame	AHP result	ANP results
Adaptation to market changes	0,19	0	0	0,0647	0,0931
Analytical thinking	0,11	0	0	0,0315	0,0394
Innovations and initiative	0,17	0	0	0,0592	0,1216
Job performance	0,30	0	0	0,1083	0,0211
Skills needed for the particular job	0,23	0	0	0,0627	0,1863
Difficulty of subject	0	0,42	0	0,1198	0,0185
Practical application (S)	0	0,23	0	0,0758	0,0487
Teaching method (S)	0	0,35	0	0,1117	0,0450
Information sharing	0	0	0,12	0,0472	0,0239
Prescribed curriculum	0	0	0,15	0,0598	0,0191
Previous training	0	0	0,14	0,0595	0,0308
Student self-studying	0	0	0,13	0,0483	0,1412
Subject's relevance	0	0	0,28	0,0769	0,0404
Teaching method (T)	0	0	0,18	0,0746	0,1710
Control Sum of weights	1	1	1	1	1

Table 1 Frames in education process

Figure 3 shows how the influence of individual frames is synthesised. Analysis of these synthesised weights can result in the answer how to reduce the risk of making a wrong decision in decision making process or in the words of education framing model the risk of negative influence on the education process. The reduction of impact of individual frames was achieved through synthesising weights from the whole perspective. The new views aggregate the views of all participating subjects.

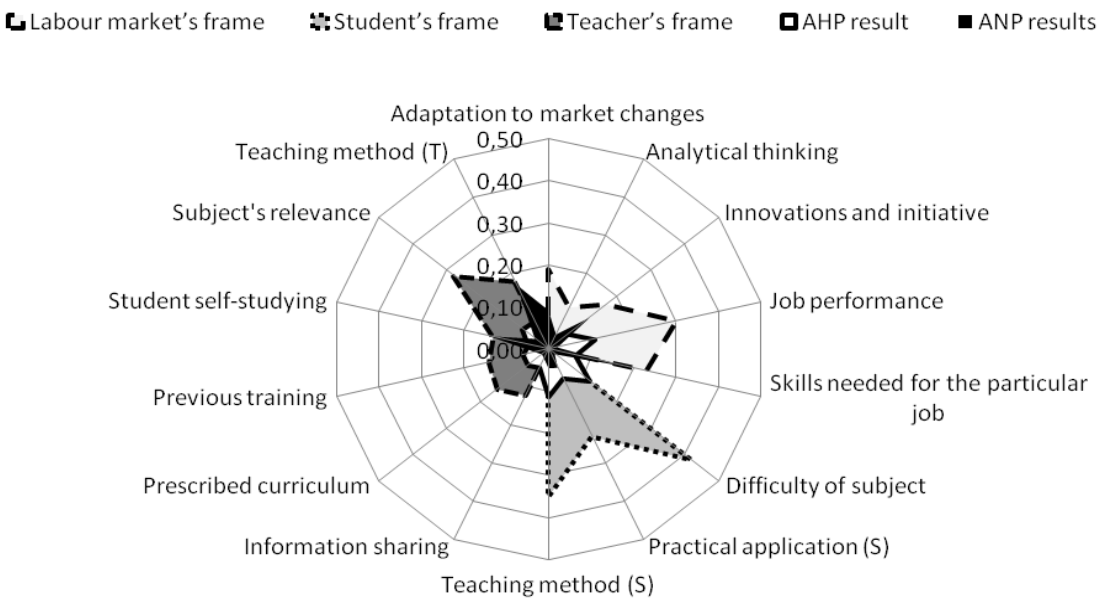


Figure 3 Weights in education framing model

4 Conclusion

This paper deals with the factors affecting our rational thinking, with our ability to make rational decisions, and particularly it serves to explain the framing effect in decision process and its quantification through AHP and ANP. The framing effect influences the ability to make the rational choice mostly in a negative way and thus it makes decision-making processes very difficult and reduces the quality of decisions. It can have fatal consequences and for example it can negatively affect passing the information in the education process.

This paper shows in its case study the emergence of the three dominant frames that influence the outcome of the education process. Using AHP and ANP new views (frames) from the whole perspective was created. In these new frames all previous dominant frames are aggregated and their influence is reduced. It can help the decision maker to move closer to a suitable rational decision and positively influence the overall outcome of the decision process.

ANP is a more general form of the AHP used in multi-criteria decision analysis. AHP structures a decision problem into a hierarchy with a goal, decision criteria, and alternatives, while the ANP structures it as a network. Both then use a system of pairwise comparisons to measure the weights of the components of the structure, and finally to rank the alternatives in the decision. In the AHP, each element in the hierarchy is considered to be independent of all the others. The decision criteria are considered to be independent of one another, and the alternatives are considered to be independent of the decision criteria and of each other. But in many real-world cases, there is interdependence among the items and the alternatives. ANP does not require independence among elements, so it can be used as an effective tool in these cases.

Therefore the weights determined by AHP for the particular preferences are different from the ones determined by ANP. It can be clearly seen in a case of the preferences "Student self-studying" and "Teaching method". Their newly calculated weights from the whole perspective are for the particular situation more than twice as high. And because of the interdependence among the items and the alternatives is for these situations, where there are preferences and expectations dependent on each other, better to use ANP.

Using ANP the more detailed interpretation of its results can be achieved. The results can be interpreted from the view of individual preferences of participating subjects. The education process can be adjusted according to the results of case study. The future research may focus on this specific advantage of ANP in more details.

References

[1] Bishop, J.: *Chapter 15 Drinking from the Fountain of Knowledge: Student Incentive to Study and Learn – Externalities, Information Problems and Peer Pressure*. Handbook of the Economics of Education, 2006, Volume 2, Pages 909-944, ISBN: 978-0-444-52819-3

[2] Druckman, J. N.: *Evaluating framing effects.* Journal of Economic Psychology, 2001, Volume 22, Issue 1, ISSN 0167-4870.

[3] Fagley, N. S., Coleman, J. G., Simon A. F.: *Effects of framing, perspective taking, and perspective (affective focus) on choice.* Personality and Individual Differences, 2010, Vol: 48, Issue: 3, ISSN 0191-8869.

[4] Rydval, J., Brožová, H.: Quantification of Framing effect in education Process using ANP. Proceedings of Efficiency and Responsibility in Education International Conference 2011, Prague, CULS, 2011, ISBN 978-80-213-2183-0.

[5] Rydval, J., Hornická A.: *Reduction of framing effect in transport logistic using decision support system*. Sborník - Think Together 2011, Praha, ČZU, 2011, ISBN 978-80-213-2169-4

[6] Rydval, J.: *Reducing The Framing Effect in Decision Processes*. Sborník - UCOLIS 2010 - University Conference in Life Sciences – Proceedings, Praha, ČZU, 2010, ISBN 978-80-213-2141-0.

[7] Saaty, T. L.: *Decision Making with Dependence and Feedback: The Analytic Network Process*. 1996, ISBN 0-9620317-9-8, RWS

[8] Saaty, T. L.: *Decision Making with Dependence and Feedback: The Analytic Network Process, The Analytic Hierarchy Process Series*. Pittsburgh: 2001, IX, RWS Publications.

[9] Saaty, T. L.: *How to make a decision: The analytic hierarchy process*. European Journal of Operational Research, 1990, Volume 48, Issue 1,

[10] Saaty, T.L.: *The Analytic Hierarchy Process (AHP) for Decision Making and the Analytic Network Process (ANP) for Decision Making with Dependence and Feedback, Creative Decisions Foundation*. 2003

[11] SuperDecisions Software for Decision-Making. URL <http://www.superdecisions.com/>

[12] Tversky, A., Kahneman, D.: *The framing of decisions and the psychology of choice*. Science 211, 1981

Bayesian estimation of model with financial frictions on Czech data

Jakub Ryšánek[1], Jaromír Tonner[2], Osvald Vašíček[3]

Abstract. As the global economy seems to be recovering from 2009 financial crisis, we find it desirable to look back and analyze Czech economy *ex post*. We work with Swedish New Keynesian model of a small open economy which embeds financial frictions in light of financial accelerator literature. Without explicitly modeling the banking sector, this model serves as a tool to understand how a negative financial shock may spread into real economy and how monetary policy may react. Financial variables turn out to be a significant driver in explaining the business cycle dynamics. We use Bayesian techniques to re-estimate original model parameters to adjust the model structure closer to evidence stemming from Czech data. Our attention focuses on a series of experiments in which we generate *ex post* forecasts of the economy prior to 2009 crisis and illustrate that monetary policy response to upcoming crisis in case of the Czech Republic might have been even more aggressive in terms of policy rate cut.

Keywords: financial frictions, Bayesian methods.

JEL Classification: C53 E32 E37

1 Introduction

Recent financial turmoil has proved that macroeconomic modeling disregarding the existence of banking sector can no longer be treated as a robust framework for policy analysis and macroeconomic forecasting. After Lehman Brothers went bankrupt in late 2008, leakage of collapsing financial sector into real economy was clearly witnessed by many. General uncertainty about future trends led banks to reduce their credit activities and central banks promptly reacted with severe policy rate cutting. Not surprisingly, market interest rates in this situation remained high and banks were not willing to loosen credit conditions since they needed to compensate for increasing default rates at that time.

It turns out that financial accelerator approach, which stems mainly from Bernanke *et al.* [2], comprises a channel through which recent events can be explained. This approach was modified multiple times during past decade – see Dib and Christensen [5] among many others – and has become a standard toolkit in modeling financial frictions. We work with a model of Christiano *et al.* [4] which was originally developed for Swedish economy and which makes use of financial accelerator in explaining business cycle fluctuations. Section 2 describes the model structure, section 3 specifies issues and results regarding Bayesian estimation of the model, section 4 demonstrates an experiment what monetary policy implications are triggered when a model with financial frictions is put to use.

2 Model

Christiano *et al.* [4] depart from what has become standard New Keynesian model in Smets and Wouters [7] and Christiano *et al.* [3] by adding financial frictions block[4] as in Bernanke *et al.* [2]. Their model is easy to implement in a small open economy setting since it contains an exogenous block of foreign variables following Adolfson *et al.* [1].

Banks in this model do not play an important role because they only function as intermediaries. Entrepreneurs are risk taking subjects who borrow from a bank and invest in capital. Upon successful investment, they profit from a positive return on capital net of bank loan and the interest. Entrepreneurs also face shocks to return on capital. This shock can be both positive and negative. For the entrepreneur there exists certain threshold val-

[1] Faculty of Informatics and Statistics, University of Economics, Prague, Department of Econometrics, nám. W. Churchilla 4, 130 67, Praha, jaac@atlas.cz (corresponding author).
[2] Faculty of Economics and Administration, Masaryk University, Lipová 41a, 602 00 Brno and Macroeconomic Forecasting Division, Czech National Bank, Na Příkopě 28, 115 03, Praha 1, Jaromir.Tonner@cnb.cz.
[3] Faculty of Economics and Administration, Masaryk University, Lipová 41a, 602 00 Brno, osvald@econ.muni.cz.
[4] Their model also incorporates a sophisticated labor market block which endogenizes the unemployment rate.

ue, $\omega = \varpi$, of this shock such that return on assets times volume of assets covers the bank loan and the interest, in which case the entrepreneur is left with nothing but has not defaulted, or

$$(1+r)A\omega = (1+i)B, \tag{1}$$

where r is return on assets, A is the volume of assets, ω is an idiosyncratic shock to return on assets distributed lognormally with mean centered at one, i is the interest rate and B is the bank loan. Equation (1) can be rearranged to get ω explicitly:

$$\omega = (1+i)/(1+r)\ B/A.$$

Due to balance sheet constraints, because

$$A = N + B, \tag{2}$$

the ratio B/A can be complemented with a ratio of net worth over assets (N/A) of which inverse ratio is usually referred to as *leverage ratio* in financial frictions literature. Threshold value of the idiosyncratic shock thus depends inversely on the leverage ratio, which serves as a constraint in the model.

Equilibrium loan contract is the one in which entrepreneurs maximize their expected welfare given threshold value ϖ and the value for leverage. Associated problem can be written as

$$\frac{E_t\{(1+r_{t+1})A_t\}}{(1+r_t^d)N_t} - bank\ share,$$

where expected nominal value of assets is taken relative to a guaranteed profit resulting from depositing net worth in the bank with the deposit rate r^d. The total expected profit is free of the amount which goes back to the bank. This bank share is dependent on the cut-off value ϖ (see Christiano *et al.* [4]) and problem maximization is subject to banks' zero profit condition, which is described in next paragraph. Taking derivatives with respect to ϖ and the leverage yield FOCs that can be combined together to rule out the Lagrangian multiplier.

As mentioned above, banks have only passive role in the model and their expected revenue corresponds with risk-free rate of return $(1+i_{\text{risk-free}})B$. The entrepreneurs who survived must pay back $(1+i)B$ and those who went bankrupt lose everything. Banks must, however, pay monitoring cost, μ, in order to reveal true condition of defaulted entrepreneur's assets. Since ω is a random variable whose distribution is known to be lognormal, we can work with its cumulative distribution function, and banks' clearing (zero profit) condition can then be written as

$$(1+i_{risk-free})B_t = [\varpi \times prob\{\omega \geq \varpi_t\} + \omega(1-\mu) \times prob\{\omega < \varpi_t\}](1+r_t)A_t, \tag{4}$$

where ϖ is the threshold value for the idiosyncratic shock that makes entrepreneur break even. This is also an equation entering the model after linearizing.

The last model equation from the financial frictions block determines law of motion for the net worth of entrepreneurs. If an entrepreneur ends up in black numbers at the end of period, his bank loan is paid off and the excess amount can serve as net worth for the next period. If the entrepreneur goes bankrupt and loses everything, an initial transfer is made to him at the beginning of next period so that he is eligible to apply for new bank loan. This occurs with probability $(1-\gamma)$. The underlying equation for evolution of net worth, nw, reads as follows

$$nw_{t+1} = \gamma[(1+r_t)A_t\varpi_t - (1+i_t)B_t] + initial\ transfer.$$

Small open economy setting is captured in the model via exogenous foreign block of variables that evolve according to a vector autoregressive scheme:

$$\begin{pmatrix} y_t \\ \pi_t \\ r_t \\ \mu_{z,t} \\ \mu_{\psi,t} \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ 0 & 0 & 0 & \rho_{\mu_z} & 0 \\ 0 & 0 & 0 & 0 & \rho_{\mu_\psi} \end{bmatrix} \cdot \begin{pmatrix} y_{t-1} \\ \pi_{t-1} \\ r_{t-1} \\ \mu_{z,t-1} \\ \mu_{\psi,t-1} \end{pmatrix} + \begin{bmatrix} \sigma_y & 0 & 0 & c_{14} & c_{15} \\ c_{21} & \sigma_\pi & 0 & c_{24} & c_{25} \\ c_{31} & c_{32} & \sigma_r & c_{34} & c_{35} \\ 0 & 0 & 0 & \sigma_{\mu_z} & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\mu_\psi} \end{bmatrix} \cdot \begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{\pi,t} \\ \varepsilon_{r,t} \\ \varepsilon_{\mu_z,t} \\ \varepsilon_{\mu_\psi,t} \end{pmatrix},$$

where technology processes, μ's, are modeled as mutually uncorrelated AR(1) processes and foreign GDP (y_t), inflation (π_t) and interest rate (r_t) form a Cholesky block in the VAR.

3 Estimation

3.1 Data

To capture financial frictions in real data, we need to introduce two observable variables concerning financial frictions – a measure of the interest rate spread and a measure of entrepreneurial net worth.

We calculate the interest rate spread as follows. Dipping PRIBOR in 2009 crisis was not immediately followed by general decrease of market interest rates. Therefore we take the average interest rate of newly issued credit to non-financial corporate obligors and subtract 3 month PRIBOR. The idea behind this choice is that non-financial corporates serve both as a representative market interest rate and also one that is close enough to entrepreneurial borrowing in model of Christiano et al. [4]. On the other hand 3 month PRIBOR, being strongly correlated with regulated 2 week PRIBOR, is believed to trace closely monetary policy actions. Figure 1 shows quarter-to-quarter percentage changes in the interest rate spread. An obvious upward shift in 2009 coincides with increased risk during the post financial turmoil period.

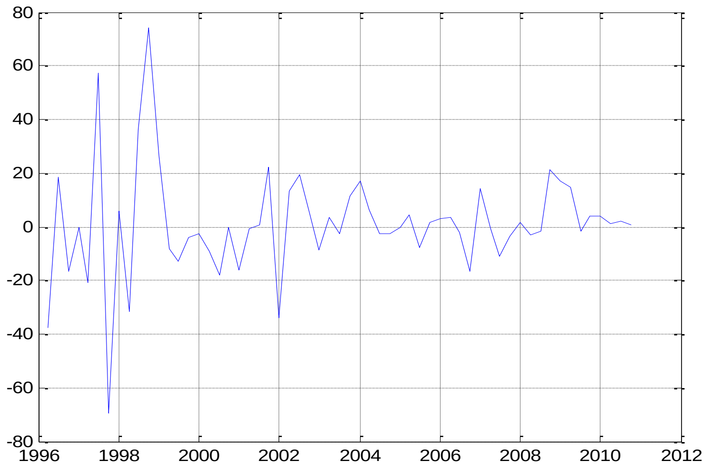


Figure 1 Interest rate spread, % changes

While it is quite intuitive to find a reasonable data counterpart for the interest rate spread, the situation concerning the entrepreneurial net worth is somewhat trickier. We take the approach of Christiano *et al.* [4] and approximate the net worth with aggregate stock market index even though this choice has its drawbacks. Let us suppose for a moment that stock prices are determined entirely by dividend payments. It is a well known fact that dividend payments behave very smoothly over the time and companies try hard to accomplish this. During the Great depression in the USA, firms were even willing to sell their physical capital in order to keep the dividend payments high, as Christiano *et al.* [4] emphasize. Thus, real net worth of companies does not correspond to the value of share prices since, from this viewpoint, these are determined by dividend payments. Real net worth is rather deducible from risk taking behavior of managers and others who possess direct control over a company for which, of course, we do not have publicly known data. On the other hand, no matter how accurate the approximation via stock market index is, this measure is very sensitive to general public opinion and is quite often used as a leading indicator. Figure 2 shows quarter-to-quarter percentage changes in Czech PX index. Double digit falls precede the moment when, in the first quarter of 2009, the real economy was hit most by global crisis.

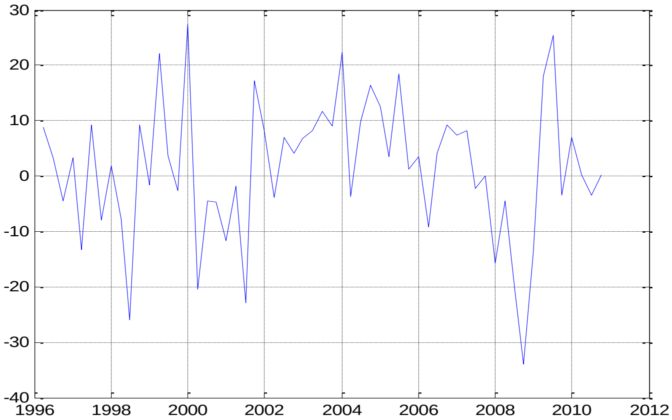


Figure 2 Stock market index (PX), % changes

3.2 Priors and posteriors

In an attempt to reveal financial frictions specific to Czech economy, we process Christiano *et al.* [4] model through a Bayesian estimation routine in Dynare[5], which is a suitable software that is capable of handling rational expectations models. The basic setup of prior assumptions concerning relevant model parameters and respective posterior estimates are summarized in table 1. Note that this table shows only selected parameters that we find interesting.

Parameter	Explanation	Distribution	Prior mean	Posterior mean
	Financial frictions block			
μ	monitoring costs	beta	0.33	0.36
γ	entrepreneur's survival rate		0.01	calibrated steady state
	Foreign VAR			
$a11$	output persistence	normal	0.50	0.95
$a22$	inflation persistence	normal	0.00	-0.01
$a33$	interest rate persistence	normal	0.50	0.89
	Other relevant model parameters			
ρ_R	persistence in Taylor rule	beta	0.80	0.73
r_y	output gap weight in Taylor rule	truncated normal	0.12	0.07
r_π	inflation weight in Taylor rule	normal	2.00	1.95
ρ_γ	persistence in entrepreneurial survival rate	beta	0.85	0.74
ξ_d	Calvo price rigidity in domestic prices	beta	0.50	0.78
ξ_x	Calvo price rigidity in export	beta	0.60	0.73
ξ_{mc}	Calvo price rigidity in import inputs for consumption	beta	0.65	0.64
ξ_{mi}	Calvo price rigidity in import inputs for investment	beta	0.60	0.60
	Recalibrated parameters			
ε_g	government share on output		0.29	calibrated

[5] Available at www.dynare.org, Koop [6] mentions underlying computational details.

Ω_c	foreign share in consumption	0.15	calibrated
Ω_i	foreign share in investment	0.93	calibrated
Ω_x	foreign share in export	0.55	calibrated

Table 1 Model parameters – priors and posteriors

Unlike in the original Swedish estimation, parameter measuring banks' monitoring costs, μ, has its posterior mean only slightly higher than the prior value of 1/3, which suggests that Czech banks do not face severe costly state verification as in Swedish case. Concerning the foreign VAR block, our results often referred to parameter $a11$ being greater than 1, which would mean that foreign GDP is unstable. To overcome this difficulty, we set an extra constraint in Dynare indicating that a11 parameter be in absolute value smaller than 1. In this case posterior estimate ended up at 0.95. Our prior belief that response of interest rates on the output gap is not significant comparing to response on inflation seems to be true because the ratio of these prior weights went down from 0.12/2 to 0.07/1.95. This of course amplifies response of interest rates on inflation, which is not surprising for a country in the inflation targeting regime.

To match Czech specifics, certain parameters needed to be recalibrated even though these parameters were not further estimated. This is especially the case of foreign share in investment that is calibrated much higher than in the Swedish case (0.93 versus 0.43).

4 Monetary policy implications

To demonstrate the impact of financial frictions, we have carried out a pair of *ex post* forecast exercises. We take known data up to 2008Q3 and replicate forecasts in turn for the model with financial frictions and also for a model in which financial frictions are turned off. In the end we compare both forecasts to reality. Figure 3 shows an *ex post* prediction of the interest rate, beginning in 2008Q3 and running 8 quarters ahead. In this pair of scenarios, outlook of foreign exogenous variables is taken according to known reality and forecast simulation is calculated forward as unexpected with respect to the foreign outlook[6]. We do not exert any other expert judgments leaving both predictions genuinely model implied and comparable.

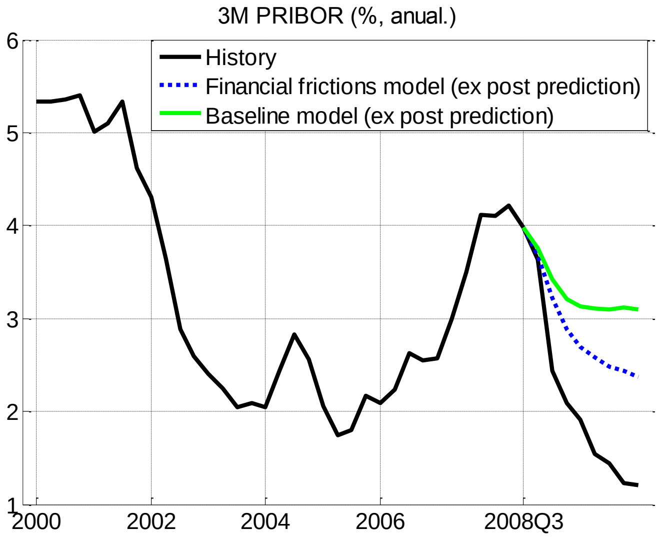


Figure 3 *Ex post* prediction of the interest rate

Both models overshoot realized development of the interest rate during the (pre-) crisis period. However, the model with financial frictions suggests more significant interest rate cut comparing to the model in which financial frictions block is neglected from. Effect of financial frictions, which mostly results from elevated interest rate spread, deepens throughout the course of time – the difference between both models is 60 basis points at the end of the forecast horizon.

[6] Rational expectations modeling approach allows us to generate a forecast based on future expected path of relevant variables, which we do not attempt here.

From technical point of view, financial frictions are switched off in the model as follows. Parameter of monitoring costs of banks is fixed at zero, which results in the fact that banks know with certainty what the current status of the obligor's property is without need of further costly inspection. Furthermore, the persistence of net worth of entrepreneurs is turned off allowing greater flexibility of entrepreneurial borrowing because shocks to the net worth no longer propagate to future and net worth of the entrepreneurs is thus more likely to wander around its steady state value more closely.

During the pre-crisis period both models suggest more or less identical trajectory of the interest rate. Figure 4 illustrates gradual dispersion between both forecast scenarios as the interest rate spread began to rise. This exercise was calculated as a recursive forecast. Since early 2009 the model with financial frictions captures the reality with substantially greater success.

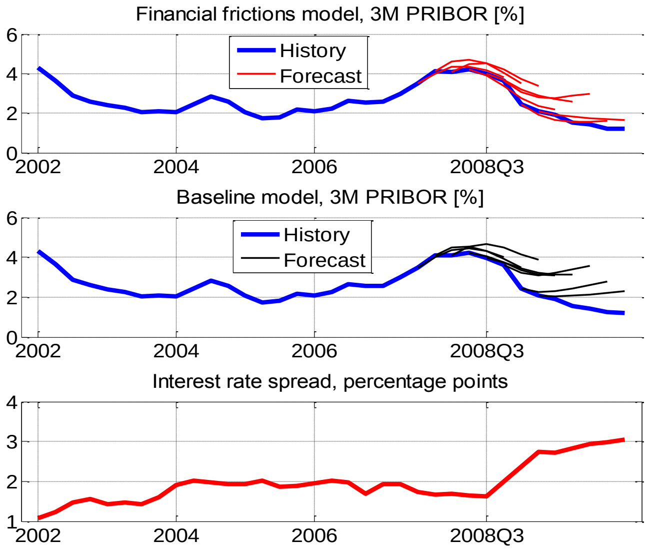


Figure 4 Recursive forecast exercise

Acknowledgements

We thank Lawrence Christiano, Mathias Trabandt and Karl Walentin for letting us experiment with their model code. This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

References

[1] Adolfson, M., Laséen, S., Linde, J. and Villani, M.: The Role of Sticky Prices in an Estimated Open Economy DSGE Model: A Bayesian Investigation. *Journal of the European Economic Association Papers and Proceedings* **3** (2005), 444-457.

[2] Bernanke, B., Gertler, M. and Gilchrist, S.: The Financial Accelerator in a Quantitative Business Cycle Framework. In: *Handbook of Macroeconomics* (Taylor, J. and Woodford, M.). Elsevier, Amsterdam, 1999, 1341-1393.

[3] Christiano, L., Eichenbaum, M., Evans, C.: Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy* **113** (2005), 1-45.

[4] Christiano, L., Trabandt, M., Walentin, K.: Introducing Financial Frictions and Unemployment into a Small Open Economy Model, Federal Reserve Bank of Atlanta CQER Working Paper No. 10-04 (2010).

[5] Dib, A. and Christensen, I.: Financial Accelerator in an Estimated New Keynesian Model. *Review of Economic Dynamics*, **11** (2008),155-178.

[6] Koop, G.: *Bayesian Econometrics*, Wiley (2003).

[7] Smets, F., and Wouters, R.: An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association* **1** (2003), 1123-1175.

Predictive modeling and measures of performance

Hana Skalská[1]

Abstract. The goal of predictive modeling is to design the model and estimate parameters of the best fitting, smallest and interpretable model which can be used for prediction. Discrimination, classification and prediction, scoring algorithms and scoring systems have become increasingly popular in real applications particularly in the past ten years. This article deals with discrimination (supervised classification) between two classes (binary classification) when outcome variable Y can take only two values. Brief overview of classification methods is done and then attention is focused on some problems associated with learning empirical classification model. The article surveys some problems of predictive classification modeling in real applications (overfitting, detection of changes in population, assessment of performance, feature selection with multiple comparison, problem of covariate and confounding variables). The influence of covariate variable and confounder variable on the quality of discrimination model is presented with an example based on real data.

Keywords: binary classification, misclassification, covariate variable, confounder.

JEL Classification: C35, C38
AMS Classification: 65C20

1 Predictive modeling and classification

Predictive model describes the relationship between dependent variable Y (outcome, response) and a set of independent variables, multivariate random vector $\boldsymbol{X}$. The goal of predictive modeling is to design the model and estimate parameters of the best fitting, smallest and interpretable model which can be used for prediction. Model is built on information from data (learning sample). An empirical estimate of predictive quality (error) of the model is usually based on learning (training) sample with the use of methods reducing bias such as ten-fold validation or one-leave-out (jackknife) method, or on independent test sample.

The expected benefit of predictive modeling follows from the possibility of generalization the results of modeling process (better understanding the process) and from the possibility to use the model in decision making. Prediction of future behavior for a new object is based on the model and known information given on explanatory variables measured on this object. The uncertainty about future behavior can be reduced if model is valid and applied under the conditions of stationary distribution.

Outcome variable can be quantitative (then solution is based on regression modeling). When outcome variable is binary or dichotomous than the task becomes to the problem of separation (discrimination) between distinct classes (groups) of objects sampled from several known populations. The value of outcome variable can be understood as an indicator or label of the class. To mention some examples, a bank can classify borrowers as being in risk of default or not in risk within pre-specified time horizon (borrower will be in default: "yes", "no"), applicants for credit are classified into "good", "bad" risk classes [4]. Marketing company would classify households to those responding or non-responding to sales offer, sent to them by mail (household will respond: "yes", "no"). Also corporate bankruptcy prediction model is an important issue [11], as the bankruptcy can bring economic losses to employees, customers, management and others (corporate is in risk bankruptcy: "yes", "no").

Discrimination, classification and prediction, scoring algorithms and scoring systems have become increasingly popular in real applications particularly in the past ten years. Their importance results partly from increasing possibilities to utilize data stored in different information system, partly from the necessity to take full advantage from information stored in distributed information systems, and from an effort to use data for decision support and risk minimization. Predictive methods allow to the users to analyze information from the past experience (stored in data) and use this information in sophisticated way for decision about future. This fundamental exploratory data mining task requires selection of appropriate data and providing other necessary initial steps (cleaning, transformation, checking data for consistency and treating inconsistencies). From descriptive analysis of data one can derive the facts important for selection of candidate explanatory variables with discriminatory power to separate distinct classes of objects and predict outcome.

[1] University of Hradec Kralove, Department of Informatics and Quantitative Methods. Rokitanskeho 62, 500 03 Hradec Kralove, hana.skalska@uhk.cz.

This article deals with discrimination (supervised classification) between two classes (binary classification) when outcome variable Y can take only two values i = 0, 1. Brief overview of classification methods is done and then attention is focused on some problems associated with learning empirical classification model: evaluation the quality of the model, problem of multiple selection procedure on explanatory variables, and possible influence of confounder variable.

1.1 Discrimination and classification

Binary discrimination assumes two groups of different objects A_0 (negative objects) or A_1 (positive objects). Objects are described with multivariate vector of quantitative explanatory variables $\boldsymbol{X}^T = [X_1, X_2, ..., X_r]$ and outcome variable Y must be known for each object. Objects are described with the pair $\{\boldsymbol{X}, Y\} \in \mathcal{X} \times \{0,1\}$. Multivariate random vector $\boldsymbol{X}$ represents observation in a space $\mathcal{X}$ of all possible observations $\boldsymbol{x}$ in multivariate space.

Discrimination task is specified under probability space of all possible measurements on $\{\boldsymbol{X},Y\}$ described with marginal probabilities $P = (\pi_1, F_0, F_1)$, where $\pi_1 = P(Y = 1)$ is the probability of positive objects group. Marginal distributions of multivariate random vectors $F_0(x) = F(x|y = 0)$, $F_1(x) = F(x|y = 1)$ are conditioned by outcome variable and describe dependence of Y on $\boldsymbol{X}$. Discrimination assumes finding an optimal classification function $z(x) : X \to \{0,1\}$ which will split sample space R of all possible sample outcomes into two distinct subsets R_0, and R_1, $R_0 \cup R_1 = X$. Classification model (classifier) is used for allocation (classification) of new objects into one of pre-defined groups.

Training sample consists of n_0 objects randomly selected from A_0 and n_1 objects randomly selected from A_1. Supervised classification (discrimination and classification) is based on training set of data with information about explanatory variables and true class labels known for each object. In the set of explanatory variables there must be some variables which are able to discriminate numerically the groups.

Different multivariate classification techniques are described [1, 7, 10] as linear discrimination (LD), quadratic discrimination (QD), logistic regression (LR), classification and regression trees (C&RT, C6.0), neural networks (NN), Naïve Bayes (NB), Super vector machine (SVM), K-nearest neighbors (K-NN), genetic algorithms (GA). Other methods are from the family of mixed linear programming solutions [8], or models are based on copulas and distributions assumptions [9].

Models are trained with regard to the best separation between classes and may be used as a tools for allocation new objects (new observations) into one of predefined groups. There are no domain specific methods of classification and usually predictive quality of different classification models is compared. Each technique can demonstrate some strengths and disadvantages, limitations and assumptions. For example LD assumes equality in covariance matrices, LR assumes approximately normal distribution in explanatory variables and careful diagnostic of leverage and influential points, C&RT is designed to maximize the entropy of the split subsets and is popular because of easy interpretation of the model, NN are sensitive to overfitting the model and they can't explain the results easily, NB simplify reality by assumption of mutually independent explanatory variables, GA can be integrated with NN, SVM can describe the broad group of tasks but solution given with many super vectors is difficult to interpret. Inside these algorithms many tools like boosting and bagging were developed for increasing the precision of the models. Other sophisticated methods were developed with the aim to get unbiased estimate of predictive quality (cross-validation, jackknife, bootstrap).

1.2 Important considerations in classification models

The quality of classification procedure is measured with misclassification errors or expected costs of misclassification. The predictive quality (the ability of classifier to classify correctly new observation) depends not only on the precision of classification model. If model is correct than predictive quality depends mostly on sample variability. In the time when model is applied than the quality of prediction depends also on degree of agreement conditions under which model is applied with conditions under which model was estimated. Stability in conditions should be evaluated [3]. Stationary distribution and prior probabilities assumptions should be checked in a statistical framework. Kolmogorov – Smirnov (K – S) test and Hellinger distance can indicate change in sample space distribution, Friedman test or Kruskal – Wallis analysis can indicate change in location. Sample selection bias caused by a covariate or confounding variable can be detected from failure correlation matrix [1] which measure association of failures in K – S test with each variable.

Understanding the domain in which classification model is built and validity of assumptions when model is applied are equally important as understanding the principles of classification techniques, its assumptions and principles of statistical inference. Steps leading to designing discrimination model can be summarized:

- Determination of potentially the best discriminating (explanatory) variables (candidate variables),
- decision about the technique of explanatory variable selection (selection from the set of candidate variables),
- selection of the best classification procedure criteria (minimal misclassification errors, minimal expected misclassification costs),
- assumption considerations, checking conditions under which model is built (usually prior probabilities and misclassification costs are specified),
- selection of explanatory variables (from the set of candidate variables) and building appropriate model(s),
- estimation of unbiased apparent error rate (APER), given as a fraction of observations in the training sample that are misclassified by the sample classification function,
- diagnostic for outliers, leverage points, influential observation, deleting them and recalculation the estimates
- evaluation and explanation (interpretation) of the model, comparison of the performance with other potential classification models,
- selection the best classification model,
- at the time of application the model, evaluation the stability of conditions is necessary.

2 Difficulties with empirical modeling

Discrimination consists in building an empirical classification model based on data (learning from data). As the "true" model is not known then either different models are trained on the same data and then compared each other or the same algorithm is trained repeatedly on the same data with different settings. Resulting models have to be evaluated and compared according their complexity, interpretability, confidence in parameter estimates, risk of misclassification errors and predictive quality.

2.1 Multiple comparisons procedure on candidate variables

Predictive quality of the model can be influenced with the process of selection explanatory variables into the model. Decision about the best discriminating variable is done on the base of maximum score value of test statistic for individual variables. Statistically significant variable with the highest value of test statistic is included into the model. The problem is that the test statistic criterion depends on sample and its importance is evaluated on the base of comparison with the critical value of one sample test statistic. If there are more than one candidate variables then candidate variable with the highest test criterion is selected and its test criterion is compared with critical value. The problem is that distribution of maximum value of test criterion is different from distribution of test criterion of individual variables.

Assume k candidate variables with potential discrimination ability. Then on sample data an evaluation function is calculated with score values $s_1, s_2, \ldots, s_k$ on each candidate variables and on this sample. Decision whether to add candidate c_{smax} with score value $s_{max} = \max\{ s_1, s_2, \ldots, s_k \}$ is based on comparison with critical value or based on p – value. Sample score s_{max} is an estimate of the population score S_* of the variable with the best overall discrimination. Jensen and Cohen proved [6] that S_{max} is biased estimator of S_* and that bias depends on number k, candidate variables. Bias $\varepsilon = P(S_{max} \geq s) - P(S_i \geq s)$ for all nonnegative values s (possible values of test statistic, score values), S_i represents unknown score statistic of individual variable.

Bias ε increases with increasing number of candidate items, with score variables on statistics that are mutually independent, with small size of the sample and with equality in expected values of test statistic of individual candidates:

1. bias increases with number of items which are compared. Items that are compared can be candidate variables, models (candidates of being the best model), or values on individual variable being candidates to became the "best binning" limit, or cut off – point (in the process of splitting a variable to categories),
2. bias increases if $S_1, S_2, \ldots, S_k$ approach to independence;
3. bias increases when size of the sample decreases. In sample with small sample size, bias is more serious than in large samples, because of sample statistics S_{max} has high standard error,
4. bias increases when expected values $E(S_1), E(S_2), \ldots, E(S_k)$ approach to equality (if they have similar discriminating ability).

Sampling distribution of S_{max}, population maximum score value is not known and expected value of the statistic depends on number of candidate variables. Therefore maximum score values resulting from comparison based different numbers of components are not comparable each other without adjusting to k. This implication can be generalized, algorithms is more likely to find higher scoring candidate variable on the set with higher number of candidate variables. Similar inference can be made to any procedure selecting the best candidate from a set of candidate models. Also any numerical variable with higher number of distinct values is more likely to be selected as discriminating variable than any variable with a few distinct values.

2.2 Empirical evidence of the problem

The problem of comparison maximal score values in multiple comparison procedure is known from inferential statistic and is presented empirically as a result of randomized experiment in Figure 1 and in Table 1. Public available data Housing Data Set were used from data repository UCI [2]. Data set consists in 20 potentially exploratory variables and one outcome variable MEDV. Later on MEDV was not used in experiment. Random experiment consists in generating 100 random vectors (new explanatory variables) with binary outcomes 0, 1 in ratio 1:2 (corresponding to the ratio of zeros and ones in original MEDV). All outcome variables in randomized experiment were independent on potentially exploratory variables. Then 100 replications were done with procedure logistic regression with stepwise selection of variables into the model. Empirical distribution of test criterion (maximum value) on candidate variables at the first step was observed.

Different number of candidate variables was examined, Figure 1 presents $k = 20$ and $k = 5$ candidate variables. Although average values of test criterions do not differ each other, distribution of s_{max} for $k = 20$ variables has different location (Figure 1a) than that from $k = 5$ variables (Figure 1b). There is a chance of selecting non significant variable into the model. This chance increases with k, number of candidate variables. Distribution of test criterions for individual variables and for maximum (most right on figures) in each experiment is visualized.

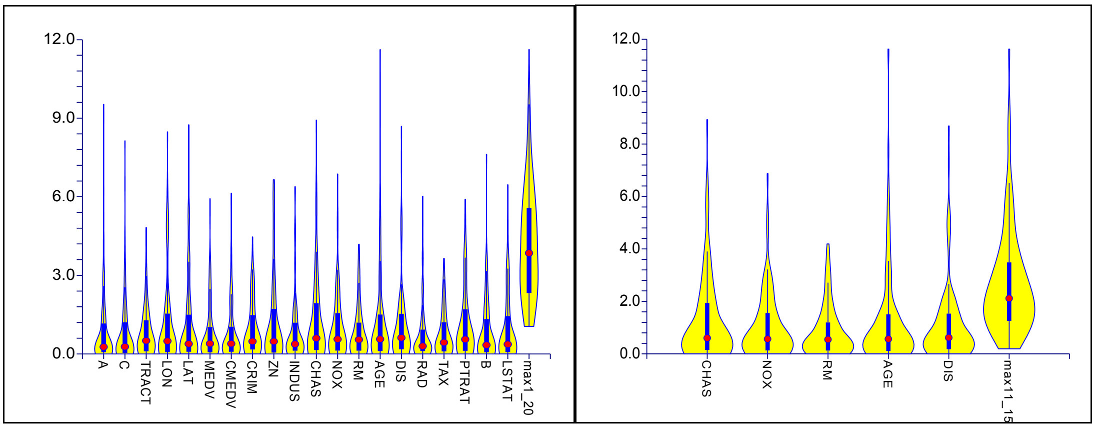


Figure 1 Violin plot (centres indicate median with confidence interval). Source: Author

a) $k = 20$ candidate variables and s_{max} distribution b) $k = 5$ candidate variables and s_{max} distribution

More detailed description of results of experiment can be seen in Table 1 with different settings for k.

Number of candidate variables k	Average value of test criterion	Max value (within 100 replications)	Median value of test criterion
20	4.2	11.6	3.8
10	3.4	9.5	3.1
5	3.7	11.6	2.7
3	1.6	6.0	1.6
2	1.5	7.6	1.1

Table 1 Descriptive characteristics of test criterion. Source: Author

n = 100 replications of experiment with real exploratory data and random outcome variable.

These facts are of practical importance. The necessity follows from this evidence about careful selection of candidate variables (competitors). Variable with highest number of distinct values has highest chance to be selected as important variable. Some hierarchical models (for example decision trees) on lower levels of the tree have larger bias ε than on higher levels where decision is based on larger number of observations. When there are two candidate variables for splitting criteria than highest chance of being selected has variable with the highest number of variants. Highest variability and increasing bias can increase a risk of misclassification. Some algorithms allows to target some variable (important for explanation of the model) into the model on the base of knowledge from the user (when importance of variable is known from domain knowledge). Targeted variable can reduce bias of the model caused by including other variable, as a result of incorrect rejecting null hypothesis for this variable.

2.3 Reduction the bias in multiple comparison

Two solutions are recommended as the best one. The first solution is based on constructing a reference distribution for S_{max} given by the randomization process (bootstrap). Generating random samples will produce sample distribution of S_{max} under the null hypothesis and will help to set up critical value of the test from these results.

The second solution follows from Bonferroni adjustment. This adjustment assumes that scores S_i are mutually independent and identically distributed. Then with k candidate items $P\left(S_{max} \geq critical\ value\right) = 1 - \left(1 - p\right)^k$. Value p is based on the sample distribution of a single score S_i. If α error is given and one would like to keep the same error in the multiple comparison procedure, then probability of the false rejection of H_0 in one experiment should equal $p = 1 - exp\left[\left(ln\left(1 - \alpha\right)/k\right)\right]$, to keep the same overall error.

As an example, to keep experiment wise error no more than α = 0,05 than it is necessary to set up individual error in each experiment p = 0,01. If there are k = 5 candidate items and p = 0,0051 if there are k = 10 candidate items etc.

2.4 Performance measures, covariate and confounding variables

Errors in binary classification can be described by confusion matrix as a ratio of false positive (fp) objects and ratio of false negative (fn) objects. Allocation of new object is based on discrimination score or on probability that object belong to the group A_0 (A_1). Common metrics of performance of classification rule are ROC curve, and AUC, area under ROC curve. ROCCH, left upper convex hull under the ROC space of different classification models can be used as a robust classification model under changing conditions (changing cost of misclassification and prior probabilities). Dual representation to the ROC curve is given with NEC (normalized expected costs). NEC can be used as a dual representation to ROC, but NEC evaluates and compares models with the use of expected costs of misclassification. ROC curve became very popular as it describes an association between false positive ratio and true positive ratio when cut – off point changes over all possible values of discrimination variable score.

Concept of covariate adjustment and confounder variable is known in etiologic studies [5] but this problem is relatively new in the field of predictive models. Covariate variable Z is variable which influences other variables. On Figure 2 there are results from field of medical data (sample size n = 115). Covariate variable Z (age) on horizontal axes is categorized into three categories. Thought it was known in that domain that age influences other variables, the detailed analysis proved statistically significant difference in slopes on regression lines.

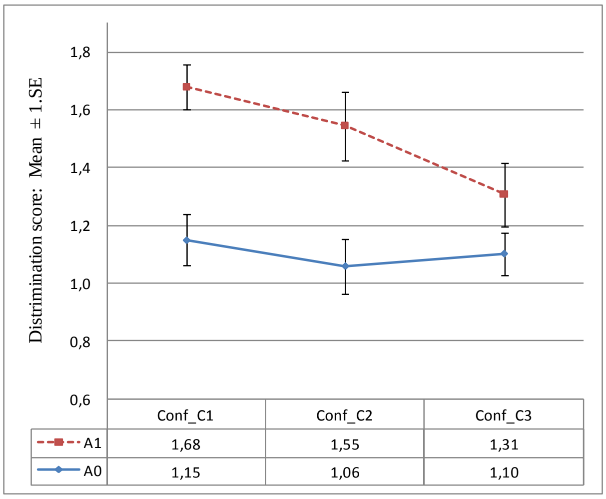


Figure 2 Discrimination ability is changing in dependence on confounder

Interaction between Z (here age), discrimination variable and outcome (groups of diseased and not diseased) was found. When Z is associated with discriminating variable and with binary outcome, it is confounder varia-

ble. Statistically significant negative association was found between decrease in probabilities of group A_1 and Z values (with estimates 0.55, 0.39, and 0.30 in categories ConfC1, Conf_c2, and Conf_C3 respectively), therefore Z is confounder variable.
Confounder variable as numerical value (not categorized) was not "successful" in the process of selection of variables important for classification. The discrimination model (without *Z*) reaches overall AUC in the sample 0.73 with 95 % confidence interval from 0.64 to 0.81, what indicates fairly good discrimination. But more detailed analysis in misclassification and analysis of influence of variable *Z* on discrimination suggested stratification of the sample as on Figure 2. After stratification the discrimination model failed in the category C3 of confounder variable (Conf_C3). This finding has very important consequences in reducing risk of error in prediction the outcome (A_1 group).

3 Conclusion

Understanding the domain in which classification model is built and validity of assumptions when model is applied are equally important as understanding the principles of classification techniques, its assumptions and principles of statistical inference. There are two problems related to the training empirical discrimination model. One problem is related to multiple comparison procedure and inductive decisions. Results of randomisation experiment suggest that warnings about careful setting type I. error in multiple comparisons are important. Behaviour of sample distribution of maximal value of test criterion from a set of candidate items produces bias in distribution. Knowledge of factors that increase bias is important in the process of selection of candidate variables.

Metrics used for comparison the predictive quality of different models and measures of predictive quality under changing conditions are frequently based on ROC, AUC as the summary statistics ROC curve and on other quantitative measures (ROCCH, NEC, etc.). Non detected covariate or confounder variables can shift the quality of the model.

Acknowledgements

The assistance received by using UCI Machine Learning Repository is highly appreciated.

References

[1] Cieslak, D. A., and Chavla, N., V.: A framework for monitorin classifiers' performance: When and why failure occurs? *Knowledge and Information Systems* **18** (2009), 83-109.
[2] Frank, A., and Asuncion, A.: *UCI Machine Learning Repository* [http://archive.ics.uci.edu/ml]. University of California, School of Information and Computer Science, Irvine, CA, 2010.
[3] Hand, D. J.: Classifier technology and the illusion of progress. *Statistical Science* **21** (2006), 1-14.
[4] Hand, D.J., Henley, W. E.: Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* **160** (1997), 523-541.
[5] Janes, H., and Pepe, M. S.: Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: An old concept in a new setting. *Am. J. Epidemiol.* **168** (2008), 89-97.
[6] Jensen, D.D., and Cohen, P.R.: Multiple Comparisons in Induction Algorithms. *Machine Learning* **38** (2000), 309-338.
[7] Johnson, R. A., and Wichern, D. W.: *Applied multivariate statistical analysis*. Pearson Education, Inc., Upper Saddle River, 2007.
[8] Jurík, P.: Mixed Integer Linear Programming in Credit Scoring System. In: *Proceedings of 26th International Conference on Mathematical Methods in Economics 2008* (Řehořová, P., Maršíková K., and Hubínka, Z., eds.), ČSOV and University of South Bohemia in České Budějovice, České Budějovice, 2010, 499-504.
[9] Pardubická, E., Lasáková, K., Gurný, P.: Prediction of the probability of default for the portfolio of three significant Czech banks. In: *Proceedings of 28th International Conference on Mathematical Methods in Economics 2010* (Houda, M., Friebelová, J., eds.), ČSOV and TU Liberec, Liberec, 2008, 230-236.
[10] Salzberg, L. S.: On comparing classifiers: Pitfalls to avoid and a recommended approach. *Data Mining and Knowledge Discovery* **1** (1997), 317-328.
[11] Sung T. K., Chang, N., and Lee G.: Dynamics of modeling in data mining. *Journal of Management of Information Systems – Special section: Data mining* **16** (1999), 65-85.

Real-World Applications of Goal Programming

Veronika Skočdopolová[1]

Abstract. Goal programming is quite popular approach for solving not only multi-objective decision problems. This paper deals with practical applications of goal programming models. In the introductory part a brief summary of multi-objective methods which use the goal programming methodology is given. The main part of the paper is devoted to goal programming applications in real decision-making problems. At first a model for planning officer accessions to the U.S. Navy is presented. Then two models connected with education (scheduling of algebra classes, and busing) are introduced. These models belong to the beginnings of goal programming applications in 70's and early 80's. In the last part of the paper model for white masses production optimisation and model for multi-step beef ration optimisation are presented as current applications of goal programming.

Keywords: goal programming, linear programming, officer accessions planning, educational problems, white masses production, beef-fattening diet

JEL Classification: C44
AMS Classification: 90C29, 90C90

1 Introduction

The aim of this paper is to make a review of basic methods based on the goal programming methodology and applications that use the goal programming methodology.

Goal programming is a methodology usually used for multi-criteria optimisation. This methodology uses minimisation of deviations from goals in opposition to the methods that maximise values of the criteria. Goal programming is both a modification and extension of linear programming. Goal programming and linear programming have very much in common – both require transformation of the real-world problem to a model in a specific format; the objective function is optimised subject to a set of constraints; variables have to be non-negative; objective function and constraints are linear; both are an instrument of decision-making. Nevertheless, there are also some differences between goal and linear programming. The goals are formulated as constraints, but they are not restrictive, they represent a desirable condition. Goal programming enables to have more conflicting goals in the model. In the goal programming model the deviational variables are defined beside the decision variables. The deviational variables (slack or surplus) enable all goals to be defined as equations[2]. The objective function of goal programming model usually does not contain the decision variables. It is a kind of penalisation function, which minimise the deviations from goals according to priorities of these goals.

Next part of the paper is devoted to the goal programming methods based on the distance of the achievements from the given goals. In the chapter 3 there are described five real applications of goal programming methodology – planning of officer accessions, two educational problems (problem of scheduling instruction, an busing), which are applications from 70's and 80's; my own application of goal programming methodology in a complex model of white masses production, and a current application– optimisation of beef fattening diet. The last part is a conclusion.

2 Goal Programming[3]

Goal programming (GP) is based on assumption that the main decision-maker objective is to satisfy his or her goals. Another basic principle that is used in GP is optimisation; decision-maker wants to choose the best solution from all the possible solutions. In this case we speak about Pareto-optimal solution, which means we cannot improve value of one of the criteria without making other criterion worse. We can find elements of optimisation in GP in case we try to get as near as possible to optimistically set goals. Last but not least, GP uses the balance principle, which is based on minimisation of maximal deviation from set goals.

More detailed description of the methods mentioned below and other methods based on goal programming methodology can be found in [3] or [4].

---

[1] University of Economics, department of econometrics, W. Churchill sq. 4, 130 67, Prague 3, skov01@vse.cz

[2] Based on [4] and [7].

[3] This chapter is based on [3].

Let us assume that the problem has generally K goals that we denoted with indices $k = 1, 2, \ldots, K$. Then we define n variables $x = (x_1, x_2, \ldots, x_n)$. These variables are factors that determine the solution and the decision-maker can influence them. All the variables have to fulfil the condition

$$\mathbf{x} \in X \ ,$$

where X is the set of feasible solution satisfying all of the constraints including non-negativity constraints. Every criterion is represented by objective function $f_k(\mathbf{x})$. The decision-maker fixes the set of goals g_k he wants to meet. For the k-th goal we can formulate following constraint

$$f_k(\mathbf{x}) + d_k^- - d_k^+ = g_k \,,$$

where d_k^- is negative deviation from the k-th goal, which represents the underachievement of the k-th goal, d_k^+ is positive deviation from the k-th goal, which shows the overachievement of the k-th goal. Both deviations are non-negative and only one of them can be positive. The decision-maker has to declare which of these deviations is for him unwanted; which of them he/she wants to penalise. It is possible to penalise the negative or positive deviation separately but also their sum. An example of unwanted negative deviation can be to reach a certain profit, where every lower value is bad. An example of unwanted positive deviation can be exceeding given limit of costs. Penalisation of deviations sum is used, when the decision maker wants to reach exact value of the criterion, e.g. he/she has to spend an exact amount of money dedicated to some project.

General goal programme can be formulated as follows:

Minimise

$$z = f(\mathbf{d}^-, \mathbf{d}^+) \,,$$

subject to

$$\mathbf{x} \in X \,, \tag{2.1}$$

$$f_k(\mathbf{x}) + d_k^- - d_k^+ = g_k \,, \qquad k = 1, 2, \ldots, K, \tag{2.2}$$

$$d_k^-, d_k^+ \geq 0 \,, \qquad k = 1, 2, \ldots, K, \tag{2.3}$$

where $z = f(\mathbf{d}^-, \mathbf{d}^+)$ is a general achievement function. It is worth mentioning that the model should include also the constraint $d_k^- \cdot d_k^+ = 0$, $k = 1, 2, \ldots, K$, which ensure that one of the deviations would be zero. Nevertheless this constraint is fulfilled automatically due to the nature of achievement function minimisation.

2.1 Lexicographic Goal Programming

Characteristic feature of lexicographic goal programming problem is existence of L of priority levels denoted by indices $l = 1, 2, \ldots, L$. For each level there is defined the unwanted deviations that is necessary to minimise. Lexicographic goal programme can be generally formulated lexicographical minimisation of

$$z = \left[h_1(\mathbf{d}^-, \mathbf{d}^+), h_2(\mathbf{d}^-, \mathbf{d}^+), \ldots, h_L(\mathbf{d}^-, \mathbf{d}^+)\right]$$

subject to (2.1) – (2.3), where $h_l(\mathbf{d}^-, \mathbf{d}^+)$ is the achievement function for l-th priority level. If we assume that the goal programme is linear and separable, then we can assume the form for l-th priority level as follows:

$$h_l(\mathbf{d}^-, \mathbf{d}^+) = \sum_{k=1}^{K} \left(\frac{u_k^l d_k^-}{n_k} + \frac{v_k^l d_k^+}{n_k} \right),$$

where u_k^l is the preferential weight connected with minimisation of the negative deviation from the k-th goal and v_k^l is the preferential weight of positive deviation from the k-th goal. Preferential weights enable decision-maker to set, which deviation is unwanted in each of priority levels. Denotation n_k stands for normalisation constant of

k-th criterion. This constant provides scaling of all goals onto the same units of measurement. The normalisation constant can be e.g. optimal value of the criterion.

Lexicographic minimisation of the achievement function is a series of L sequential optimisation, where the minimisation of deviations placed in higher priority level is more important than the minimisation of deviations placed in lower priority level. The minimisation in the lower priority level has a reduced feasible set, because the minimal values of the higher priority level optimisations must be maintained.

Lexicographic goal programming is practical for situations where the decision-maker can easily define order of the goals and does not want to make direct trade-off comparisons between the goals.

2.2 Weighted Goal Programming

The weighted goal programming allows direct trade-off between the unwanted deviations. This can be done by placing the deviations in a single criterion weighted normalised objective function. If we assume, that this objective function is linear, then the weighted goal programme can be generally formulated as follows:

Minimise

$$z = \sum_{k=1}^{K} \left(\frac{u_k d_k^-}{n_k} + \frac{v_k d_k^+}{n_k} \right),$$

subject to (2.1) – (2.3),where the symbols definitions are identical to those introduced for the lexicographical GP with the difference, that there are no more used the indices of priority levels.

2.3 Chebyshev Goal Programming

The third variant of GP I would like to mention in this paper is based on the Chebyshev means of distance measuring. As opposed to the sum of all deviations in weighted or lexicographic GP, the maximal deviation from any goal is minimised. Therefore, the Chebyshev GP is often named Minmax goal programming. Via this method the decision-maker gets solution with a better balance between the achievements of the set of goals than via e.g. the lexicographic GP, where some of the goals are preferred or than weighted GP, where the decision variables together make the objective function lowest with the expense of a very poor value in some of the goals. Last but not least, Chebyshev GP can find optimal solutions for linear models that are not located at extreme points in decision space, as opposed to the two previous mentioned variants of GP.

The Chebyshev goal programme can be generally formulated as follows:

Minimise

$$z = \lambda,$$

subject to

$$\mathbf{x} \in X,$$

$$f_k(\mathbf{x}) + d_k^- - d_k^+ = g_k, \qquad k = 1, 2, \ldots, K,$$

$$\frac{u_k d_k^-}{n_k} + \frac{v_k d_k^+}{n_k} \le \lambda, \qquad k = 1, 2, \ldots, K,$$

$$d_k^-, d_k^+ \ge 0, \qquad k = 1, 2, \ldots, K,$$

where λ is the maximal deviation from amongst the set of goals.

3 Practical applications

Goal programming is quite wide used optimisation technique. In this part of the paper I would like to give short information about some practical applications of goal programming. First of them is a model for planning officer accessions for U.S. Navy that was published in 1980 [1].

3.1 Planning officer accessions

The aim of this model is to assure required number of officers in different career speciality areas in the Navy. There is a variety of commission sources, where the officers are secured to meet requirement in different career speciality areas such as surface warfare, aviation warfare or submarine warfare. These sources have different capacities and costs. The officer population in the model is described by their number in each of a set of possible state at each time period. The officer's state is determined by factors such as warfare community (aviation, surface, etc.), commissioning programme and years of commissioned service (YCS). Within each community officer inventory requirements are specified by the number needed at several experience levels. Transition rates among the experience levels are used to project expected flows between states in following time periods in Markovian fashion to which are also added new accessions into the system. Variables of the model $y_{ijk}(t)$ are inventory of officers in community *i*, from source *j*, with YCS *k*, at beginning of time period *t* and $x_{ij}(t)$ are accessions to community *i* from source *j* in time period *t*. In the model there also deviation variables are defined. The goals $G_{im}(t)$ set in the model are to achieve number of officers in community *i*, for experience model *m*, in time period *t*. The set of goal constraints are formulated as follows:

$$\sum_{j=1}^{J}\sum_{k=l_i(m)}^{u_i(m)} y_{ijk}(t) + g_{im}^{-}(t) - g_{im}^{+}(t) = G_{im}(t)\,,$$

where $g_{im}^{-}(t)$ and $g_{im}^{+}(t)$ are the negative (shortfall) and positive (surplus) deviations for community *i*, experience level *m*, in time period *t*, *J* is the number of commissioning sources in model, $u_i(m)$ and $l_i(m)$ are the upper and lower limits for YCS in experience level *m*. Set of constraints that follow the limits of budget, training capacities and so on is not necessary to be described here. It can be found in [1]. The objective function of the model minimises weighted deviations from the goals subject to the constraints. The paper [1] also presents an illustrative example, where it is shown what happened when one of the sources is omitted. The described model has been developed and implemented in coordination with the U.S. Navy Deputy Chief of Naval Operations and has been used on a regular basis during the annual accession planning. Nevertheless, this model can be used also in manpower planning in other fields, not only in the military field.

3.2 Educational problems

Another applications where presented by R. A. Van Dusseldorp et al. in 1976 [7]. It was a set of three applications to educational problems; I will focus on two of them. The first one is the problem of scheduling instruction. There are 60 students in a high school enrolled in algebra. They can be thought through three different types of group classes – 60 (large), 30 (medium) or 15 (small) students in a class or in individual consultations. The problem is to decide how much of time the student should spend in each type of class subject to some constraints. These constraints are all specified as goal constraints – students should spend around 250 minutes per week in algebra class or individual consultation (positive and negative deviation variable is defined); students should have not more than 60 minutes per week of large group classes (positive deviation variable is defined), at least 40 minutes per week of small group classes (negative deviation), and at least 10 minutes per week they have to spend on individual consultations (negative deviation). Then there is a limit of teacher time – 1070 minutes per week. Beside the deviation variables, decision variables that stand for minutes spent at each type of class are defined. The objective function is minimisation of weighted sum of the deviation variables, where the weights are priorities of each goal. The model has been formulated as follows [7]:

Minimise

$$z = p_1 d_1^- + p_2 d_5^+ + p_3 d_2^+ + p_4 d_1^+ + p_5 d_4^- + p_6 d_3^-\,,$$

subject to

$$\begin{array}{lllllllllll}
T_L + T_M + T_S + T_I - d_1^- + d_1^+ & & & & = & 250, & \\
T_L - d_2^+ & & & & = & 60, & \\
T_S + d_3^- & & & & = & 40, & \\
T_I + d_4^- & & & & = & 10, & \\
T_L + 2T_M + 4T_S + 60T_I - d_5^- & & & & = & 1070, & \\
 & & & T_k & \geq & 0, & k \in \{L, M, S, I\}, \\
 & & & d_i^+,\ d_i^- & \geq & 0, & i = 1, 2, \ldots, 5,
\end{array}$$

where p_j, $j = 1, 2, \ldots, 6$, are the priorities of goals ordered from the highest priority p_1 to the lowest p_6, d_i^+ (or d_i^-), $i = 1, 2, \ldots, 5$, are the positive (or negative) deviations from i-th goal and T_k, $k \in \{L, M, S, I\}$, is the amount of minutes each student spend in large (L), medium-sized (M), small (S) group class or at individual consultation (I). The results of this model see in Table 1:

Type of class	Minutes per week
Large group	60
Medium group	155
Small group	25
Individual	10

Table 1 Results of the goal programming model [7]

According to this solution each student spends exactly 250 minutes per week with algebra and the teacher teach exactly 1070 minutes per week. All the conditions are met, except the one of minutes spend in small group classes, but this condition had the lowest priority. This was a multi-objective model. We can also try to solve this problem with basic linear model where the goal constraints expressed by equation are replaced by inequality constraints. The linear model can be formulated as follows:

$$
\begin{array}{rcrcrcrcrl}
T_L & + & T_M & + & T_S & + & T_I & = & 250, & \quad (3.1)\\
T_L & & & & & & & \leq & 60, & \quad (3.2)\\
 & & & & T_S & & & \geq & 40, & \quad (3.3)\\
 & & & & & & T_I & \geq & 10, & \quad (3.4)\\
T_L & + & 2T_M & + & 4T_S & + & 60T_I & \leq & 1070, & \quad (3.5)\\
 & & & & & & T_k & \geq & 0, & k \in \{L, M, S, I\} \quad (3.6)
\end{array}
$$

where the notation is the same as in the previous model. This model is just a system of inequalities that does not have any feasible solution. If we insist on the total amount of 250 minutes per week that each student spend in algebra classes, we can set as the objective the total amount of teacher time used to teach algebra. If we omit the constraint (3.5) and add the minimisation objective function $z = T_L + 2T_M + 4T_S + 60T_I$, we get results in Table 2:

Type of class	Minutes per week
Large group	60
Medium group	140
Small group	40
Individual	10

Table 2 Results of the basic linear model

In this case, teacher spends 1100 minutes with algebra teaching that is 30 minutes more than the requirement. All the constraints are fulfilled as equations. This example shows the difference in usage of basic linear model and goal programming model.

The second application described by Van Dusseldorp is busing. The problem of busing has two requirements. The first of them is "to achieve a specified percentage range composition by group. The other is to minimise transportation costs via minimisation of total busing distance." [7, p. 11] We have three schools and three zones with students. We know the student population by groups, school capacities, and the busing distances. In the model there are used four priorities – to have all students assigned to a school, not to exceed school capacities, to achieve a student composition, and to minimise the total busing distance. These priorities form the constraint of the model. The objective function is minimisation of weighted sum of deviation variables. With problem of busing using linear approach have dealt Stimson and Thompson [6], but they could consider only one of the objectives – students' composition or busing distance.

3.3 White masses production

An example of application of goal programming approach in a complex model is optimisation of white mass production [5]. The term *white mass* is used for yoghurts, acidified milk drinks and other milk derivatives. The aim of this model is to minimise total production cost of the whole process with respect to all constraints. The company produces white masses from several ingredients such as cream or skimmed milk. The ingredients for white masses production can be purchased or produced from raw milk. During the production of ingredients the content of fat and protein has to be in balance. In this case the deviation variables which represent the surplus or lack of fat and protein were used. In comparison with pure goal programming model, in this model the sum of deviation variables is not minimised, but the deviation should be less than a given value.

3.4 Beef-fattening diet

The authors showed in [8] some reasons, why to deal with beef ration optimisation. The most important reason is of course costs of producing each ration. Standards of composition of each ration given by the European Union also have to be followed. The authors also mentioned the influence of stock-raising on the environment, because the stock-raising is one of the big sources of greenhouse gas. A two-phase optimisation approach based on mathematical programming was described in [8]. The first phase utilises standard linear blending problem to minimise the costs of ration production with respect to its composition. The resulting value of minimal costs is then targeted as a cost goal in the second phase. In the second phase, there was used the weighted goal programming approach. The authors also defined intervals for the deviations from goals. They implemented this two-phase model to an application created in MS Excel and they used the MS Excel Solver for solving the problem.

4 Conclusions

This paper is just a very brief view of goal programming theory and its applications. Goal programming was formulated in 50's. Since that time it became very popular. There are many methods and models using goal programming approach. The aim of future research is to compile summary of methods based on goal programming approach with review of its practical applications. Linear programming is a computationally feasible problem (see e.g. [2]). Another aim of future work might be to prove whether this holds also for goal programming.

Acknowledgements

The research is supported by the Internal Grant Agency of the University of Economics (grant no. F4/18/2011).

References

[1] Bres, E. S., et al.: A Goal Programming Model for Planning Officer Accessions. *Management Science* **26** (1980), 773–783.
[2] Cerny, M., Sindelarova, I.: Some Pictures and Scripts for Teaching IPMs. In: *Proceedings of the 26th International Conference on Mathematical Methods in Economics 2008* (Rehorova, P., Marsikova, K., Hubinka, Z.), Technical University Liberec, 2008, 59–67.
[3] Jones, D., Tamiz, M.: *Practical Goal Programming*. International Series in Operations Research & Management Science. Springer, 2010.
[4] Lee, S. M.: *Goal Programming for Decision Analysis*. Auerbauch Publishers, Philadelphia, 1972.
[5] Skocdopolova, V., Jablonsky, J.: Optimisation of Production Process in Food-Processing Company. In: *Mathematical Methods in Economics 2009* (Brozova, H., Kvasnicka R.). Czech University Life Sciences Prague, 2009, 290–295.
[6] Stimson, D. H., Thompson, R. P.: Linear Programming, Busing and Educational Administration. *Socio-Economic Planning Sciences* **8** (1974), 195-206.
[7] Van Dusseldorp, R. A. et al.: Applications of Goal Programming to Education. Presented at 14 Annual International Convention of the Association for Educational Data Systems, Phoenix, Arizona, May 3–7, 1976. Available at http://www.eric.ed.gov/PDFS/ED127661.pdf [14.4.2011]
[8] Zgajnar, J., et al.: Multi-step beef ration optimization: application of linear and weighted goal programming with penalty function. *Agricultural and Food Science*, **19** (2010), 193–206.

Separable Utility Functions in Dynamic Economic Models

Karel Sladký[1]

Abstract. In this note we study properties of utility functions suitable for performance evaluation of dynamic economic models under uncertainty. At first, we summarize basic properties of utility functions, at second we show how exponential utility functions can be employed in dynamic models where not only expectation but also the risk are considered. Special attention is focused on properties of the expected utility and the corresponding certainty equivalents if the stream of obtained rewards is governed by Markov dependence and evaluated by exponential utility functions.

Keywords: Utility functions, decision under uncertainty, dynamic economic models, Markov reward chains, exponential utility functions, certainty equivalent

JEL classification: C44,D81
AMS classification: 91B16,90C40

1 Economic decisions and utility functions

Economic decisions are usually based on the outcome, say ξ, as viewed by the decision maker. Usually the decision maker has no complete information on the problem, his decisions are made under some kind of uncertainty. In general, the decision under uncertainty in its most simple form consists of three nonempty sets $\mathcal{I}$, $\mathcal{A}$ and $\mathcal{X}$ and a function $f : \mathcal{I} \times \mathcal{A} \mapsto \mathcal{X}$. In particular, $\mathcal{I} = \{1, 2, \dots, N\}$ characterizes the uncertainty of the problem, elements of the set $\mathcal{I}$ are called states of the considered system (sometimes called states of nature or of the world); $\mathcal{A} = \{1, 2, \dots, K\}$ is the set of possible decisions (or actions) and $\mathcal{X}$ is the set of outcomes of the decision problem equipped with a complete and transitive relation $\preceq$ on $\mathcal{X}$. As concerns the relation $\preceq$ that determines decision maker's preference among elements of the set of outcomes (and also preference among actions). In particular, if for $\xi_1, \xi_2 \in \mathcal{X}$ (resp. $a_1, a_2 \in \mathcal{A}$) it holds $\xi_1 \preceq \xi_2$ (resp. $a_1 \preceq a_2$) it means that outcome ξ_2 (resp. decision a_2) is at least as preferable as outcome ξ_1 (resp. decision a_1). By completeness of the relation we mean that every two elements of $\mathcal{X}$ (resp. of $\mathcal{A}$) are related, i.e. given any $\xi_1, \xi_2 \in \mathcal{X}$ there are three possibilities: either $\xi_1 \preceq \xi_2$ but not $\xi_2 \preceq \xi_1$, then we write $\xi_1 \prec \xi_2$; or $\xi_2 \preceq \xi_1$ but not $\xi_1 \preceq \xi_2$, then we write $\xi_2 \prec \xi_1$; or both $\xi_1 \preceq \xi_2$ and $\xi_2 \preceq \xi_1$, then we write $\xi_2 \sim \xi_1$. By transitivity we mean that $\xi_1 \preceq \xi_2$ and $\xi_2 \preceq \xi_3$ implies $\xi_1 \preceq \xi_3$ for any three elements $\xi_1, \xi_2, \xi_3 \in \mathcal{X}$; the same also holds for any three elements of the action set $\mathcal{A}$.

Furthermore, in many decision problems on choosing action (decision) $a \in \mathcal{A}$ the outcome $\xi_j \in \mathcal{X}$ occurs only with (individual's subjective) probability p_j^a for $j = 1, 2, \dots, N$ (where $\sum_{j=1}^{N} p_j^a = 1$), which is familiar to the decision maker (stochastic model). In this case we speak about lottery or prospect, and let $\mathcal{Y}$ with generic y be the set of all lotteries or all probability distribution on $\mathcal{Y}$. Obviously, if the decision maker had a complete ranking of all lotteries on the set of outcomes, then he could obtain a complete ranking of all decisions in $\mathcal{A}$. To this end the decision maker can replace condition on complete ordering of the set $\mathcal{X}$ by complete ordering of the set $\mathcal{Y}$ (see Axiom 1). Moreover, under some other technical assumptions (Axioms 2,3 specified in the sequel), ordering of decision may be expressed by numerical function, called the utility function.

Axiom 1. (Preference Ordering) The decision maker has a preference ordering defined on $\mathcal{Y}$ which is a transitive and complete ordering.

[1]Institute of Information Theory and Automation of the AS CR, Department of Econometrics, Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic, e-mail: sladky@utia.cas.cz

Axiom 2. **(Continuity)** If for $y_1, y_2, y_3 \in \mathcal{Y}$, $y_1 \preceq y_2 \preceq y_3$, there exists an $\alpha \in [0,1]$ such that $\alpha y_1 + (1-\alpha) y_3 \sim y_2$.

Axiom 3. **(Independence)** If for $y_1, y_2 \in \mathcal{Y}$, $y_1 \preceq y_2$, then for all $\alpha \in (0,1]$ and all $y \in \mathcal{Y}$

$$\begin{array}{lll} \text{if } \; y_1 \prec y_2 & \text{then also} & \alpha y_1 + (1-\alpha) y \preceq y_2 + (1-\alpha) y \\ \text{if } \; y_1 \sim y_2 & \text{then also} & \alpha y_1 + (1-\alpha) y \sim y_2 + (1-\alpha) y \end{array}$$

Theorem. Under Axioms 1–3 there exists a real-valued function $u : \mathcal{X} \mapsto \mathbb{R}$, called utility function, such that for all $y_1, y_2 \in \mathcal{Y}$

$$y_1 \preceq y_2 \;\text{ if and only if }\; \mathsf{E}^{y_1}[u(\xi)] \le \mathsf{E}^{y_2}[u(\xi)]$$

($\mathsf{E}^{y}[\cdot]$ is the expectation with respect to probability distribution induced by $y \in \mathcal{Y}$).

Furthermore, u is unique up to a positive linear transformation, i.e. if $\tilde{u}$ is another function with the above property, there exists a positive scalar α_1 and a scalar α_2, such that $\tilde{u}(\xi) = \alpha_1 u(\xi) + \alpha_2$.

For the *proof* of this fundamental theorem see e.g. [1], [2], [10].

In words: Economic decisions based on the outcome, say ξ, may be represented by an appropriate utility function, say $u(\xi)$ assigning a real number to each possible outcome ξ. Utility function $u(\xi)$ must be monotonically increasing, i.e. we assume that larger values of outcome are preferred, and unique up to a positive linear transformation. Furthermore, in economic models we assume that utility functions are concave.

In case of stochastic models outcome ξ is a random variable and we consider expectation of utilities assigned to (random) outcomes, i.e. the value $U(\xi) := \mathsf{E}u(\xi)$. Certain (or certainty) equivalent, say $Z(\xi)$, is then defined by $u(Z(\xi)) := \mathsf{E}u(\xi)$ (in words, certainty equivalent is the value, whose utility is the same as the expected utility of possible outcomes). Additional important concepts in the utility theory are that of the

coefficient of absolute risk aversion defined by $R^a(\xi) := -\frac{u''(\xi)}{u'(\xi)}$, along with

coefficient of relative risk aversion defined by $R^r(\xi) := \xi R^a(\xi) = -\xi \cdot \frac{u''(\xi)}{u'(\xi)}$.[1]

For handling real life models decision makers must be able to express $u(x)$ in a concrete form. Typical utility functions are:

- Linear function: $$u(x) = a + bx \;\text{ where }\; b > 0$$
- Quadratic function: $$u(x) = a + bx - cx^2 \;\text{ where }\; b > 0,\; c > 0.$$
- Logarithmic function: $$u(x) = a + b\ln(x+c) \;\text{ where }\; b > 0,\; c \ge 0.$$
- Fractional function: $$u(x) = a - \frac{1}{x+b} \;\text{ where }\; b > 0,\; c > 0$$
- The function: $$u(x) = \begin{cases} x^{1-a} & \text{for } \; 0 < a < 1 \\ \ln x & \text{for } \; a = 1 \\ -x^{1-a} & \text{for } \; a > 1 \end{cases}$$

  Observe that then $R^r(x) = a$, i.e. coefficient of relative risk aversion is constant; this utility function belong to the CRRA (Constant Relative Risk Aversion) utility functions.
- Exponential function: $$u(x) = -\mathrm{e}^{-ax} \;\text{ with }\; a > 0$$

Introducing the so-called risk aversion coefficient $\gamma \in \mathbb{R}$ exponential utility functions, as well as linear utility functions, assigned to a random outcome ξ can be also written in the following more compact form

$$u^{\gamma}(\xi) = \begin{cases} (\operatorname{sign} \gamma)\, \exp(\gamma \xi), & \text{if } \gamma \ne 0 \\ \xi & \text{for } \gamma = 0. \end{cases} \tag{1}$$

[1] (here $u'(\xi) = \frac{\mathrm{d}u(\xi)}{\mathrm{d}\xi}$, $u''(\xi) = \frac{\mathrm{d}^2 u(\xi)}{\mathrm{d}\xi^2}$).

Observe that exponential utility function considered in (1) is separable what is very important for sequential decision problems, i.e. $u^\gamma(\xi^{(1)}+\xi^{(2)}) = \text{sign}(\gamma)\, u^\gamma(\xi^{(1)}) \cdot u^\gamma(\xi^{(2)})$ for $\gamma \neq 0$ and if $\gamma = 0$ for the resulting linear utility function $u(x) = bx$ we have $u^\gamma(\xi^{(1)}+\xi^{(2)}) = u^\gamma(\xi^{(1)}) + u^\gamma(\xi^{(2)})$. Unfortunately, considering stochastic models, in contrast to exponential utility functions, linear utility functions cannot reflect variability-risk features of the problem. Obviously $u^\gamma(\cdot)$ is continuous and strictly increasing, and *convex* for $\gamma > 0$, so-called risk seeking case, and *concave* for $\gamma < 0$, so-called risk aversion case.

Furthermore, exponential utility functions

- are the most widely used non-linear utility functions, cf. [3],
- in most cases an appropriately chosen exponential utility function is a very good approximation for general utility function, cf. [6].

If exponential utility (1) is considered, then for the corresponding certainty equivalent $Z^\gamma(\xi)$ given by

$$u^\gamma(Z^\gamma(\xi)) = \mathsf{E}[(\text{sign}\ \gamma)\ \exp(\gamma\xi)]$$

we have

$$Z^\gamma(\xi) = \begin{cases} \frac{1}{\gamma}\ \ln\{\mathsf{E}\,[\exp(\gamma\xi)]\}, & \text{if } \gamma \neq 0 \\ \mathsf{E}[\xi] & \text{for } \gamma = 0. \end{cases} \tag{2}$$

Observe that if ξ is constant then $Z^\gamma(\xi) = \xi$, if ξ is nonconstant then by Jensen's inequality

$$\begin{aligned} Z^\gamma(\xi) &> \mathsf{E}\xi \quad &&(\text{if } \gamma > 0,\ \text{the risk seeking case}) \\ Z^\gamma(\xi) &< \mathsf{E}\xi \quad &&(\text{if } \gamma < 0,\ \text{the risk averse case}) \\ Z^\gamma(\xi) &= \mathsf{E}\xi \quad &&(\text{if } \gamma = 0,\ \text{the risk neutral case}) \end{aligned}$$

The following facts will be useful in the sequel:

1. For $U^{(\gamma)}(\xi) := \mathsf{E}u^\gamma(\xi)$, i.e. $U^{(\gamma)}(\xi) := \mathsf{E}\exp(\gamma\xi)$, the Taylor expansion around $\gamma = 0$ reads

$$U^{(\gamma)}(\xi) = 1 + \mathsf{E}\sum_{k=1}^{\infty} \frac{(\gamma\xi)^k}{k!} = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot \mathsf{E}\xi^k. \tag{3}$$

Observe that in (3) the first (resp. second) term of the Taylor expansion is equal to $\gamma\mathsf{E}\xi$ (resp. $\frac{1}{2}(\gamma^2)\mathsf{E}\xi^2$). In particular, if for random variables ξ, ζ with $\mathsf{E}\xi = \mathsf{E}\zeta$ it holds $\mathsf{E}\xi^2 < \mathsf{E}\zeta^2$ (or equivalently $\text{var}\,\xi < \text{var}\,\zeta$) and $\mathsf{E}\xi^k$ are uniformly bounded in k then there exists $\gamma_0 > 0$ such that $U^{(\gamma)}(\xi) < U^{(\gamma)}(\zeta)$ for any $\gamma \in (-\gamma_0, \gamma_0)$.

2. In economic models (see e.g. [1], [10]) we usually assume that the utility function $u(\cdot)$ is increasing (i.e. $u'(\cdot) > 0$), concave (i.e. $u''(\cdot) < 0$) with $u(0) = 0$ and $u'(0) < \infty$ (so called the Inada condition).

Since a positive linear transformation of the utility function $u^\gamma(\xi)$ preserves the original preferences (see the Theorem, cf. also [1],[10]) we shall also consider the utility functions

$$\bar{u}^\gamma(x) \;=\; 1 - \exp(\gamma x), \quad \text{where}\ \ \gamma < 0 \quad (\text{the risk averse case}) \tag{4}$$

$$\tilde{u}^\gamma(x) \;=\; \exp(\gamma x) - 1, \quad \text{where}\ \ \gamma > 0 \quad (\text{the risk seeking case}) \tag{5}$$

and the function $\bar{u}^\gamma(x)$ satisfies all above conditions imposed on a utility function in economy theory. Observe that the Taylor expansions of $\bar{u}^\gamma(x)$ and of $\tilde{u}^\gamma(x)$ read

$$\bar{u}^\gamma(x) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{|\gamma|^k}{k!} \cdot x^k, \ \ \text{where } \gamma < 0, \quad \tilde{u}^\gamma(x) = \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot x^k, \ \ \text{where } \gamma > 0 \tag{6}$$

and if $x = \xi$ is a random variable for the expected utilities we have

$$\bar{U}^\gamma(\xi) \;:=\; \mathsf{E}\bar{u}^\gamma(\xi) = 1 - U^{(\gamma)}(\xi) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{|\gamma|^k}{k!} \cdot \mathsf{E}\xi^k \tag{7}$$

$$\tilde{U}^\gamma(\xi) \;:=\; \mathsf{E}\tilde{u}^\gamma(\xi) = U^{(\gamma)}(\xi) - 1 = \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \cdot \mathsf{E}\xi^k. \tag{8}$$

Illustrative example. Consider an individual that may repeatedly bet \$1 on the toss of a fair coin or not bet at all. f he bets and guesses correctly he wins \$2, if he does not guess correctly, he losses \$1 and if he decides not to bet he receives compensation \$1. Here the state space $\mathcal{I}$ consists of two elements H (head) and T (tail). The action set $\mathcal{A}$ consists of three elements $\mathcal{A} = \{a_0, a_1, a_2\}$ where decision a_0 is not to bet, decision a_1 is bet on heads and decision a_2 is bet on tails. Then the set of outcomes $\mathcal{X} = \{0, 1, 2\}$ and the values of the function $f : \mathcal{I} \times \mathcal{A} \mapsto \mathcal{X}$ are given by $f(H, a_0) = 1$, $f(T, a_0) = 1$, $f(H, a_1) = 2$, $f(T, a_1) = 0$, $f(T, a_2) = 2$, $f(H, a_2) = 0$. Moreover, we may consider also some kind of lottery in the decision process, in particular we may assume that the selected decision occurs in accordance with a given probability mechanism.
Now we need a ranking among decisions that is consistent in a well-defined sense with our ranking of outcomes. Moreover, according to Theorem the ranking should be determined by a numerical function $u(\cdot)$ that maps the set of decisions $\mathcal{A}$ to the set of real numbers such that $a_i \preceq a_j$ if and only if $u(a_i) \preceq u(a_j)$ for all $a_i, a_j \in \mathcal{A}$ and $i \neq j$. If $u(\cdot)$ is linear, i.e. $u(x) = x$ then $\mathsf{E}^{a_0} u(\xi) = 1$, $\mathsf{E}^{a_1} u(\xi) = \mathsf{E}^{a_2} u(\xi) = \frac{1}{2}(2 + 0) = 1$, hence then the resulting expectation of the linear utility function is independent of the selected decision, i.e. $a_0 \sim a_1 \sim a_2$. However, using the exponential utility function $u^\gamma(x)$ optimal decision depends on the value of the risk aversion coefficient $\gamma \neq 0$. In particular, again $\mathsf{E}^{a_0} u(\xi) = 1$, but $\mathsf{E}^{a_1} u(\xi) = \mathsf{E}^{a_2} u(\xi) = \frac{1}{2}(\mathrm{e}^{\gamma 2} + 1)$. Hence if $\gamma > 0$ then $a_0 \prec a_1 \sim a_2$ and if $\gamma < 0$ then $a_1 \sim a_2 \prec a_0$.

Up to now we have studied properties of utility functions for models of "static" (stochastic) systems where the uncertainty is represented by the decision maker's ignorance of the "current state" of the system along with possibly probabilistic behavior of the outcome. In the sequel we focus attention on systems that develop over time and the decision maker need not have complete information of the state of the system and where additional decision can be taken if the individual guesses correctly.

2 Separable utility functions in stochastic dynamic models

2.1 Complete information on state

Consider a family of models for decision under uncertainty formulated in Section 1 specified by nonempty sets $\mathcal{I}$, $\mathcal{A}$ along with a family of $\mathcal{X}^{(i)}$ (for $= 1, 2, ...N$) with individual probabilities p_{ij}^a for $i, j = 1, 2, \ldots, N; a = 1, 2, ..., K$ (of course, $\sum_{j=1}^N p_{ij}^k = 1$), familiar to the decision maker along with his (or her) knowledge of the current state of the system. This represents a Markov decision chain $X = \{X_n, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, \ldots, N\}$, finite set $\mathcal{A}_i = \{1, 2, \ldots, K\}$ of possible decisions (actions) in state $i \in \mathcal{I}$ and the following transition and reward structure:

p_{ij}^a : transition probability from $i \to j$; $\quad r_{ij}^a$: one-stage reward for a transition from $i \to j$

Policy controlling the chain is a rule how to select actions in each state. In this note, we restrict on stationary policies, i.e. the rules selecting actions only with respect to the current state of the Markov chain X. Then a policy, say π, is determined by some decision vector f whose ith element $f_i \in \mathcal{A}$ identifies the action taken if the chain X is in state $X_n = i$; hence also the transition probability matrix $\boldsymbol{P}(f)$ of the Markov decision chain. Observe that the ith row of $\boldsymbol{P}(f)$ has elements $p_{i1}^{f_i}, \ldots, p_{iN}^{f_i}$ and that $\boldsymbol{P}^*(f) = \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} [\boldsymbol{P}(f)]^k$ exists. In what follows, $\boldsymbol{R}(f) = [r_{ij}^{f_i}]$ is the transition reward matrix, i.e. $\boldsymbol{R}(f)$ is an $N \times N$ matrix of one-stage rewards (for details see e.g. [7]).

If the chain starts in state i and policy $\pi \sim (f)$ is followed, let $\xi_i^{(n)}(\pi)$ (abbreviated as $\xi^{(n)}$) be the total reward received in the n next transition of the Markov chain X and $\xi_{X_m}^{(m,n)}(\pi)$ be the total reward received in the $n-m$ next transition if the Markov chain was after m first transition in state X_m. Then for the expected (exponential) utility $U_i^\pi(\gamma, n)$, the certainty equivalent $Z_i^\pi(\gamma, n)$ and its mean value $J_i^\pi(\gamma)$ we have

$$U_i^\pi(\gamma, n) := \mathsf{E}_i^\pi[\exp(\gamma \xi^{(n)})] = \mathsf{E}_i^\pi \exp[\gamma\,(r_{i,X_1} + \xi_{X_1}^{(1,n)})], \tag{9}$$

$$Z_i^\pi(\gamma, n) := \frac{1}{\gamma} \ln\{\mathsf{E}_i^\pi\,[\exp(\gamma \xi^{(n)})]\} \qquad \text{for } \gamma \neq 0, \tag{10}$$

$$J_i^\pi(\gamma) := \lim_{n\to\infty} \frac{1}{n} Z_i^\pi(\gamma, n) \tag{11}$$

and hence for the expectation of the utility functions $\bar{u}^\gamma(\xi^{(n)})$ and $\tilde{u}^\gamma(\xi^{(n)})$ we have (cf. (8),(8))

$$\bar{U}_i^\pi(\gamma, n) := 1 - U_i^\pi(\gamma, n), \qquad \tilde{U}_i^\pi(\gamma, n) := U_i^\pi(\gamma, n) - 1. \tag{12}$$

Conditioning in (9) on X_1, since policy $\pi \sim (f)$ is stationary, from (9) we immediately get the recurrence formula

$$U_i^\pi(\gamma, n+1) \;=\; \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot \mathrm{e}^{\gamma r_{ij}}\cdot U_j^\pi(\gamma,n) = \sum_{j\in\mathcal{I}} q_{ij}^{f_i}\cdot U_j^\pi(\gamma,n) \quad \text{with } \; U_i^\pi(\gamma,0)=1 \tag{13}$$

or in vector notation and by iterating

$$\boldsymbol{U}^\pi(\gamma,n+1) \;=\; \boldsymbol{Q}(f)\cdot \boldsymbol{U}^\pi(\gamma,n) = (\boldsymbol{Q}(f))^n\cdot \boldsymbol{e} \qquad \text{with } \; \boldsymbol{U}^\pi(\gamma,0)=\boldsymbol{e}, \tag{14}$$

where $\boldsymbol{Q}(f) = [q_{ij}^{f_i}]$ with $q_{ij}^{f_i} := p_{ij}^{f_i}\cdot \mathrm{e}^{\gamma r_{ij}}$, $\boldsymbol{U}^\pi(\gamma,n)$ is the vector of expected utilities with elements $U_i^\pi(\gamma,n)$ and $\boldsymbol{e}$ is a unit (column) vector.

Observe that $\boldsymbol{Q}(f)$ is a nonnegative matrix, and by the Perron–Frobenius theorem (cf. [4]) the spectral radius $\rho(f)$ of $\boldsymbol{Q}(f)$ is equal to the maximum positive eigenvalue of $\boldsymbol{Q}(f)$. Moreover, if $\boldsymbol{Q}(f)$ is irreducible (i.e. if and only if $\boldsymbol{P}(f)$ is irreducible) the corresponding (right) eigenvector $\boldsymbol{v}(f)$ can be selected strictly positive, i.e.

$$\rho(f)\,\boldsymbol{v}(f) = \boldsymbol{Q}(f)\cdot\boldsymbol{v}(f) \qquad \text{with} \quad \boldsymbol{v}(f) > 0. \tag{15}$$

Moreover, under the above irreducibility condition it can be shown (cf. e.g. [5], [9]) that there exists decision vector $f^*\in\mathcal{A}$ such that

$$\boldsymbol{Q}(f)\cdot\boldsymbol{v}(f^*) \;\le\; \rho(f^*)\,\boldsymbol{v}(f^*) = \boldsymbol{Q}(f^*)\cdot\boldsymbol{v}(f^*), \tag{16}$$

$$\rho(f) \;\le\; \rho(f^*) \equiv \rho^* \qquad \text{for all } \; f\in\mathcal{A} \tag{17}$$

and decision vector $\hat{f}\in\mathcal{A}$ such that

$$\boldsymbol{Q}(f)\cdot\boldsymbol{v}(\hat{f}) \;\ge\; \rho(\hat{f})\,\boldsymbol{v}(\hat{f}) = \boldsymbol{Q}(\hat{f})\cdot\boldsymbol{v}(\hat{f}) \tag{18}$$

$$\rho(f) \;\ge\; \rho(\hat{f}) \equiv \hat{\rho} \qquad \text{for all } \; f\in\mathcal{A}. \tag{19}$$

In words, $\rho(f^*)\equiv\rho^*$ (resp. $\rho(\hat{f})\equiv\hat{\rho}$) is the maximum (resp. minimum) possible positive eigenvalue of $\boldsymbol{Q}(f)$ over all $f\in\mathcal{A}$.

If the Perron eigenvectors $\boldsymbol{v}(f^*) = \boldsymbol{v}^*$, $\boldsymbol{v}(\hat{f}) = \hat{\boldsymbol{v}}$ are strictly positive, there exist positive numbers $\alpha_1 < \alpha_2$ such that $\alpha_1\hat{\boldsymbol{v}} \le \boldsymbol{e} \le \alpha_2\boldsymbol{v}^*$ and hence by (16), (18) and by (10),(11)

$$\alpha_1\hat{\rho}^n\hat{\boldsymbol{v}} \le \boldsymbol{U}^\pi(\gamma,n) \le \alpha_2(\rho^*)^n\boldsymbol{v}^* \tag{20}$$

$$n\ln(\hat{\rho}) + \ln(\alpha_1\hat{v}_i) \le \gamma Z_i^\pi(\gamma,n) \le n\ln(\rho^*) + \ln(\alpha_2 v_i^*) \tag{21}$$

$$\gamma^{-1}\ln(\hat{\rho}) \le J_i^\pi(\gamma) \le \gamma^{-1}\ln(\rho^*) \tag{22}$$

From (20),(21),(22) we can see that the asymptotic behavior of $\boldsymbol{U}^\pi(\gamma,n)$ heavily depends on ρ^*, $\hat{\rho}$, and that the maximum, resp. minimum, growth rate of each $U_i^\pi(\gamma,n)$ is independent of the starting state. Similarly, $J_i^\pi(\gamma)$ (mean value of the corresponding certainty equivalent $Z_i^\pi(n,\gamma)$ growing linearly in time) is independent of the starting state and bounded by $\ln(\hat{\rho})$ and by $\ln(\rho^*)$.

2.2 Incomplete information on state

In what follows we assume that the decision maker has no information of the current state of the system, but he knows current values of the obtained rewards. Moreover, he can also employ results concerning optimal policy obtained in subsection 2.1, i.e. the decision maker knows optimal actions in each state of the system. His knowledge of the system structure and optimal control policy along with information of current values of obtained rewards ("signalling information") may help him to selected optimal or suboptimal policy. In what follows we sketch how to handle such problems on examples slightly extending our illustrative example and reformulate it in terms of Markov decision chains (MDC).

Illustrative example: Formulation as MDC. Consider a Markov decision chain with state space $\mathcal{I} = \{H, T\}$, actions a_1, a_2 in each state and the following transition and reward structure:

$p_{HH}^{a_1} = \frac{1}{2}, r_{HH}^{a_1} = 2;\;\; p_{HT}^{a_1} = \frac{1}{2}, r_{TT}^{a_1} = 0;\quad p_{HH}^{a_2} = \frac{1}{2}, r_{HH}^{a_2} = 0;\;\; p_{HT}^{a_2} = \frac{1}{2}, r_{HT}^{a_2} = 2;$
$p_{TT}^{a_1} = \frac{1}{2},\;\; r_{TT}^{a_1} = 0;\;\; p_{TH}^{a_1} = \frac{1}{2},\;\; r_{TH}^{a_1} = 2;\quad p_{TH}^{a_2} = \frac{1}{2}, r_{TH}^{a_2} = 0;\;\; p_{TT}^{a_2} = \frac{1}{2}, r_{TT}^{a_2} = 2.$

Recall that a_1, a_2 is bet on heads, tails respectively; we ignore decision a_0 not to bet. In what follows $q_{ij}^a = p_{ij}^a \exp(\gamma r_{ij}^a)$ for $a = a_1, a_2$ and $i, j = H, T$, in particular, $q_{ij}^a = \frac{1}{2}\exp(\gamma 2)$ or $q_{ij}^a = \frac{1}{2} \cdot 1$.
Hence each row of the resulting 2×2 matrix $\boldsymbol{Q}(\cdot) = [q_{ij}^k]$ contains a single element $\frac{1}{2}\exp(\gamma 2)$ and $\frac{1}{2}$. Hence the spectral radius of each $\boldsymbol{Q}(\cdot)$ equals $\frac{1}{2}[\exp(\gamma 2) + 1]$ and constant vector is the corresponding right eigenvector. Since decision a_0 (not to bet) brings unit reward, if the risk aversion coefficient γ is positive, the decision maker should prefer betting on heads or on tails, if γ is negative optimal decision is not to bet.

Extended illustrative example: Formulation as MDC. Suppose that in the considered Illustrative Example an individual can extend options after his betting on heads and guessing correctly. The additional action is to bet \$1 on heads and toss of an unfair coin (probability of head $\frac{1}{3}$, probability of tail $\frac{2}{3}$). If he guesses correctly receives \$3 and can repeat such bet.
To this end consider a Markov decision chain with state space $\mathcal{I} = \{H, T, \bar{H}\}$ and slightly modified transition and reward structure of the previous example by replacing transition from state H if action a_1 (instead of $p_{HH}^{a_1} = \frac{1}{2}, r_{HH}^{a_1} = 2$) by $p_{H\bar{H}}^{a_1} = \frac{1}{2}, r_{H\bar{H}}^{a_1} = 2$ and define transitions and rewards in state $\bar{H}$ as:

$p_{\bar{H}\bar{H}}^{a_1} = \frac{1}{2}, r_{\bar{H}\bar{H}}^{a_1} = 2;$ $p_{\bar{H}T}^{a_1} = \frac{1}{2}, r_{\bar{H}T}^{a_1} = 0;$ $p_{\bar{H}H}^{a_2} = \frac{1}{2}, r_{\bar{H}H}^{a_2} = 0;$ $p_{\bar{H}T}^{a_2} = \frac{1}{2}, r_{\bar{H}T}^{a_2} = 2;$
$p_{\bar{H}\bar{H}}^{a_3} = \frac{1}{3},$ $r_{\bar{H}\bar{H}}^{a_3} = 3;$ $p_{\bar{H}T}^{a_3} = \frac{2}{3}, r_{\bar{H}T}^{a_3} = 0.$

Obviously, if action a_3 is not selected in state $\bar{H}$ the problem can be treated as the previous one. To this end we focus attention on policies selecting decision a_3 and starting with decision a_1. Then the spectral radius of the corresponding 3×3 matrix $\boldsymbol{Q}(\cdot) = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2}e^{\gamma 2} \\ \frac{1}{2} & \frac{1}{2}e^{\gamma 2} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3}e^{\gamma 3} \end{bmatrix}$ is greater than $\frac{1}{2}[\exp(\gamma 2) + 1]$ (observe that $\frac{2}{3} + \frac{1}{3}e^{\gamma 3} > \frac{1}{2} + \frac{1}{2}e^{\gamma 2}$ for each $\gamma > 0$).

Conclusions

In this note basic facts concerning decision under uncertainty along with typical utility functions are summarized. Special attention is focused on properties of the expected utility and the corresponding certainty equivalent if the stream of obtained rewards is governed by Markov dependence and evaluated by exponential utility functions.

Acknowledgement. This research was supported by the Czech Science Foundation under Grants P402/11/0150, P402/10/1610 and P402/10/0956.

References

[1] J. von Neumann and O. Morgenstern: Theory of Games and Economic Behaviour. Third Edition, Princeton Univ. Press, Princeton, NJ 1953.

[2] D. P. Bertsekas: Dynamic Programming and Optimal Control, Volume 1. Third Edition. Athena Scientific, Belmont, Mass. 2007.

[3] J. L. Corner and P. D. Corner: Characterization of decision in decision analysis practice. J. Oper. Res. Soc. *46* (1995), 304–314.

[4] F. R. Gantmakher: The Theory of Matrices. Chelsea, London 1959.

[5] R. A. Howard and J. Matheson: Risk-sensitive Markov decision processes. Manag. Sci. *23* (1972), 356–369.

[6] C. W. Kirkwood: Approximating risk aversion in decision analysis applications. Decision Analysis *1* (2004), 51–67.

[7] M. L. Puterman: Markov Decision Processes – Discrete Stochastic Dynamic Programming. Wiley, New York 1994.

[8] K. Sladký and M. Sitař: Risk sensitive and mean variance optimality in Markov decision processes. In: Proceedings of the 26th International Conference Mathematical Methods in Economics 2008 (P. Řehořová, K. Maršíková, eds.). Faculty of Economics, Technical University of Liberec, Liberec 2008, pp. 452–461.

[9] K. Sladký: Growth rates and average optimality in risk-sensitive Markov decision chains. Kybernetika *44* (2008), 206–217.

[10] A. Takayama: Analytical Methods in Economics. Harvester Wheatsheaf, Hertfordshire 1994.

Impact of Asymmetric Shocks and Structural Differences between the Czech economy and Euro Area 12

Martin Slanicay[1]

Abstract. Asymmetric shocks and structural differences are regarded as the main causes of a potential suboptimality of common monetary policy. In this paper I present a new approach for analysis of asymmetric shocks and structural differences between two countries, using historical shock decomposition. An advantage of this method is that it jointly examines the impacts of asymmetric shocks and structural differences. This method covers adjustment processes to asymmetric shocks as well as the frequency with which the shocks occur.

I demonstrate this method using a New Keynesian DSGE model of two economies, presented in Kolasa [14], which I estimate on the data of the Czech economy and Euro Area 12. The model is estimated with Bayesian techniques using seven key macroeconomic variables in both economies: GDP, consumption, investment, prices, real wages, nominal interest rate and internal exchange rate defined as prices of non-tradable goods relative to prices of tradable goods. The behavior of the model is driven by seven structural shocks in both economies: tradable and non-tradable productivity shocks, labor supply shocks, investment efficiency shocks, consumption preference shocks, government spending shocks and monetary policy shocks. The model allows for structural parameters to differ between economies and for correlations between corresponding shocks in both economies.

Results suggest that impacts of structural shocks are more correlated than it seems from estimated correlations between shocks alone. The largest correlations are between impacts of shocks on inflations and interest rates. I also find that an evolution of variables in the Czech economy is relatively more influenced by shocks in productivity while the evolution of variables in Euro Area 12 is relatively more influenced by investment efficiency shocks. It also seems that the monetary policy of the ECB is relatively more discretionary than the monetary policy of the CNB, nevertheless it seems that the CNB follows the ECB in their discretionary policy.

Keywords: New Keynesian, DSGE model, Bayesian estimation, asymmetric shocks, structural differences, shock decomposition.

JEL classification: E32

1 Introduction

Asymmetric shocks and structural differences are regarded as the main causes of a potential suboptimality of common monetary policy. Asymmetry of shocks can be defined as differences in timing, magnitude or persistence of structural macroeconomic shocks. Structural differences are then defined as differences in propagation mechanisms of structural shocks. In case of asymmetric shocks and (or) structural differences in a monetary union, applied monetary policy facing structural shocks must be suboptimal for some countries.

A lot of economic research is devoted to these issues because of their important role in evaluating

[1]Masaryk University, Faculty of Economics and Administration, Department of Economics, Lipová 41a, Brno 602 00, e-mail: slanicay@mail.muni.cz

benefits and costs of common currency. Pauer [15] provides a non-technical overview of a role of asymmetric shocks in a debate about benefits and costs of common currency. Several papers discuss the presence and relative importance of asymmetric shocks and structural differences. Jondeau and Sahuc [12] test the structural heterogeneity within Euro Area and find that asymmetric shocks are the main sources of a different behavior of countries in Euro Area, while structural differences play almost no role. Kolasa [13], [14] investigates the structural heterogeneity between Poland and Euro Area and find that the main sources of the structural heterogeneity are the volatility and synchronization of shocks hitting both economies.

Several authors try to determine to what extent are the shocks within EU asymmetric. Bayoumi and Eichengreen [4] find that shocks are significantly more idiosyncratic across EU countries than across US regions, which may indicate that the EU will find it more difficult to operate a monetary union. On the other hand, Verhoef [16] shows that within EMU the symmetries of demand and supply shocks increases over time.

Some studies are dealing with an adjustment process to asymmetric shocks. Alexius and Post [1] examine how floating exchange rates respond to asymmetric shocks and find that exchange rates display some stabilizing properties but can mainly be characterized as disconnected from the rest of the economy. Amisano, Giammarioli and Stracca [2] examine the adjustment process to asymmetric shocks in Italy and argue that joining EMU does not alter the adjustment process to idiosyncratic demand and cost push shocks and that the EMU system is not hit by idiosyncratic monetary shocks anymore. Driver and Wren-Lewis [10] try to quantify the costs imposed by asymmetric shocks under EMU compared to free floating. Their results suggest that these costs are significantly higher under EMU than under free floating.

Many authors examine structural differences alone. Benigno and López-Salido [5] find differences in inflation dynamics between Germany on the one side and France, Italy, Netherlands and Spain on the other side. They discover that inflation dynamics in Germany is characterized by forward-looking nature of price setting and average duration of prices about 5 quarters while the other group of countries is characterized more by backward-looking nature of price setting and average duration of prices about 8 quarters. Fabiani and Morgan [11] examine differences in the relationship between wage growth, inflation and tightness of the labor market across Germany, France, Italy, Netherlands and Spain. They provide empirical evidence that there exist large differences even across these "core" countries. Angeloni and Ehrmann [3] try to explain why differences in national inflation rates and growth rates arise within the Euro Area. They find out that the main explanation can be due to differences in inflation persistence. Campa and Gonzlez Minguez [7] investigate exchange rate pass-through in Euro Area countries. They find substantial differences across Euro Area countries in the way how a common exchange rate movement gets transmitted into prices. Most of these differences are caused by a distinct degree of openness of each country to non-euro area imports rather than by the heterogeneity in the structure of imports. Demertzis and Hugues Hallett [9] investigate differences in unemployment rates in Europe. They show that disparities in unemployment rates are mainly due to differences in labor market fundamentals causing natural rate of unemployment to differ. Asymmetric shocks and policy differences, both causing differences in unemployment gap, play limited role in explaining unemployment disparities. Cecchetti [8] provides the evidence that differences in a financial structure influence a transmission mechanism of the monetary policy. He shows that countries with poorer direct capital access, less concentrated and less healthy banking systems display a greater sensitivity of inflation and output to policy changes.

Many papers mentioned above examine solely asymmetric shocks or structural differences. This paper tries to examine impacts of asymmetric shocks and structural differences jointly. Many papers examine to what extent there are asymmetric shocks or structural differences between two countries. One can argue that asymmetric shocks and structural differences are not important alone, but due to their impact on the economy. It is more important to know how much different behavior of economy they cause. This paper tries to answer this question, using historical shock decomposition. Some papers analyze adjustment process to asymmetric shocks. This analysis is meaningful but it tells us nothing about frequency of these shocks. Method used in this study provides results about impact of asymmetric shocks and structural differences. These results incorporate adjustment processes to asymmetric shocks, as well as the frequency of shocks.

2 Model

The goal of this paper is to analyze impact of asymmetric shocks and structural differences between two countries. I decided to use a two-country model where both countries are modeled in the same way. It is also important to use a complex model, which realistically replicates the behavior of economies regarding their responses to various shocks. Namely, the model should contain all parts of national accounts. Asymmetric shocks are also analyzed because of their importance for monetary policy. New Keynesian DSGE models are regarded as a benchmark for monetary policy analysis and for this purpose are widely used by central banks. For these reasons, I decided to use a New Keynesian DSGE model of two economies, presented in Kolasa [13], [14]. Derivation of this model from microfoundations, as well as its log-linear form can be found in Kolasa [14]. In this section I restrict my description of this model to a brief non-technical overview of its structure.

The model assumes 5 types of representative agents in both economies. Households consume tradable and non-tradable goods produced by firms. There is an assumption of habit formation in consumption and an assumption that consumption of a final tradable good requires consumption of ω units of nontradable distribution services. Households also trade bonds and their intertemporal choice about consumption is influenced by preference shocks. Households supply labor and set wages on the monopolistically-competitive labor market. Their labor supply is influenced by labor supply shocks and their wage-setting is subject to a set of labor demand constraints and to Calvo constraint on the frequency of wage adjustment, see Calvo [6]. According to the Calvo constraint, every period each household resets its wage with the probability $1 - \theta_w$ and with the probability θ_w partially adjust its wage according past inflation. Households also accumulate capital which they rent to firms. Capital accumulation is subject to investment-specific technological shocks and adjustment costs.

There are two types of firms in the model, producers of tradable goods and producers of non-tradable goods. Both of them employ a Cobb-Douglas production function with constant returns to scale. Productivity in both sectors is influenced by productivity shocks. Firms hire labor on the labor market and sell their goods on the monopolistically-competitive good markets. They set prices on the good market subject to a set of demand constraints and to the Calvo constraint on the frequency of price adjustment, see Calvo [6]. According to the Calvo constraint, every period each firm resets its price with the probability $1 - \theta_p$ and with the probability θ_p partially adjust its price according past inflation.

Fiscal authority collects lump-sum taxes which they use for government expenditures and transfers to households, so that the state budget is balanced each period. The government expenditures consist only of domestic non-tradable goods and are modeled as a stochastic process - government expenditures shock. Given our assumptions about households, Ricardian equivalence holds in this model. Monetary authority behaves according a backward-looking Taylor rule and deviations from this rule are explained as monetary shocks. The model is completed with an assumption of a complete bond market and an assumption of goods and labor markets clearing.

3 Estimation

The model is estimated using quarterly data of the Czech economy and Euro Area 12 economy from the 1.Q 2000 to 4.Q 2010. Data series are downloaded from the web database of Eurostat. I use following 14 time series (seven for each economy): real GDP, consumption, investment, HICP, nominal wage index, short-term interest rate (3-month rates) and internal exchange rate defined as prices of non-tradable goods (services and energy) relative to prices of tradable goods (others). Except for the nominal interest rates, all observables are seasonally adjusted and expressed as demeaned log-differences. Nominal interest rates are demeaned and expressed as quarterly rates.

The behavior of the model is driven by seven structural shocks in both economies, such as productivity shocks in tradables and nontradables, labor supply shocks, investment efficiency shocks, consumption preference shocks, government spending shocks and monetary policy shocks. Except for the monetary shocks, which are modeled as IID process, all shocks are represented by AR1 process. I allow for correlations between corresponding shocks in both economies, e.g. between domestic preference shocks and foreign preference shocks.[1] I also allow for different values of structural parameters in both countries. Estimated correlations between corresponding shocks in both economies are presented in Table 1.

[1]For shocks represented by AR1 process it means that I allow for correlations between the innovations in these shocks.

shocks	correlation
productivity shocks in tradables	0.32
productivity shocks in nontradables	0.23
investment efficiency shocks	0.38
preference shocks	0.14
labor supply shocks	0.07
government expenditures shocks	0.11
monetary shocks	0.58

Table 1: Correlations between Corresponding Shocks

The largest symmetry is between shocks connected to a technology and between monetary shocks. Other shocks can be regarded as asymmetric. Nevertheless all estimated correlations are rather low, which suggests that there exists a large asymmetry between shocks in the Czech economy and Euro Area 12. There are two possible problems with this interpretation. Firstly, except for the monetary shocks the correlations are not between whole shocks (represented by AR1 process), but only between the innovations in these shocks. Secondly, structural shocks are not interesting because of themselves, but because of their impact on the economy. It is more interesting to know how large is the asymmetry between the impacts of these shocks on domestic and foreign variables.

4 Impact of Shocks

If we want to know how the structural shocks influenced the behavior of both economies, we have to make a historical shock decomposition of the main macroeconomic variables in both economies. I restrict my analysis to three most important variables: output, inflation and interest rate. Shock decomposition displays individual contributions of shocks to the deviations of the examined variable from its respective steady state. Evolution of the examined variable is in every period decomposed into the contributions of structural shocks. If we put together contributions of a particular shock to the development of the examined variable in every period, we get trajectory of impact of this kind of shock on the evolution of the examined variable in the examined period. Because of using a linear model, it is also possible to sum up contributions of several shocks in every period. For example, we can calculate the joint effect of the domestic and foreign preference shocks on a particular variable simply by summing up their contributions to the evolution of this variable, displayed by shock decomposition.

Using shock decompositions, we can calculate correlations between trajectories of shock's impact on a particular variable in a domestic and foreign economy.[2] I view it as more meaningful to compare combined effect of domestic and foreign shock of the same type on domestic and foreign variables. One can argue that both shocks can have common cause or that the shock in the foreign economy can induce the same type of shock in the domestic economy, and vice versa. I decided to examine joint effect of all productivity shocks instead of separate effects of productivity shocks in tradables and nontradables, because both shocks can have the same cause. Table 2 displays correlations between impact of shocks on domestic variables and their foreign counterparts. For example, cor=0.53, which corresponds to the row "preference shocks" and to the column "output" and subcolumn "cor", is the correlation between the joint effect of the domestic and foreign preference shocks on the domestic output and the joint effect of the domestic and foreign preference shocks on the foreign output. Correlations in the Table 2 can be viewed as measures of the asymmetric impact of shocks on the particular variable in both economies. I also calculate how much each type of shock contributes to explaining the evolution of a particular variable. These values are presented in column "IoDV" ("IoFV") of the Table 2, expressed as percentage shares of particular shocks in the development of the domestic (foreign) variable.

We can see that the correlations between impacts of structural shocks significantly differ across examined variables. For example, the impact of shocks in productivity on interest rates is highly correlated (cor=0.89) while impact of these shocks on output is almost uncorrelated (cor=0.1). We can also see that, in most cases, the correlations between impacts of structural shocks are higher than correlations between

[2] I consider the Czech economy as the domestic economy and Euro Area 12 as the foreign economy.

	output			inflation			interest rate		
shocks	cor	IoDV	IoFV	cor	IoDV	IoFV	cor	IoDV	IoFV
productivity shocks	0.1	20%	6%	0.69	30%	27%	0.89	30%	19%
inv. efficiency shocks	0.91	16%	33%	0.55	10%	15%	0.88	9%	21%
preference shocks	0.53	11%	12%	0.63	1%	5%	0.95	2%	8%
labor supply shocks	0.5	13%	8%	0.72	21%	20%	0.77	17%	13%
gov. expenditures shocks	0.3	24%	27%	0.56	3%	5%	0.17	5%	7%
monetary shocks (IID)	0.19	12%	9%	0.66	11%	18%	0.78	8%	24%
initial conditions	-0.02	4%	7%	0.94	23%	10%	0.7	28%	7%
overall index	0.39			0.73			0.77		

Table 2: Impact of Shocks (cor - correlation of shock's impact, IoDV - Impact on Domestic Variable, IoFV - Impact on Foreign Variable)

shocks alone. It suggests that the shocks are less asymmetric than it seems from estimated correlations of shocks, presented in Table 1.

In Table 2 we can also see some interesting structural differences between the Czech economy and Euro Area 12, concerning impact of particular shocks. It seems that evolution of domestic variables (output, inflation, interest rates) is relatively more influenced by shocks in productivity than by investment efficiency shocks while evolution of variables in Euro Area 12 is relatively more influenced by investment efficiency shocks. It also seems that the monetary policy of the ECB is more discretionary than the monetary policy of the CNB, because monetary shocks play much larger role in explaining interest rates in Euro Area (24% in EA vs. 8% in the CR). One can also argue that the big correlation between impacts of monetary shocks on interest rates (cor=0.78) suggest that the CNB follows the ECB in their discretionary policy, i. e. when the ECB deviates from the Taylor rule than the CNB deviates as well and in the same direction.

Problematic feature of results about impact of particular shocks, presented in the Table 2, is their fragmentation. For this reason I calculate an overall index of impacts of all shocks for each variable. This overall index is a weighted average of all correlations of shock's impact where weights are set as the percentage share of particular shocks in the development of the examined variable in the domestic economy. These overall indexes are presented in the last row of the Table 2. Values of these indexes have similar interpretation as correlation coefficients. The biggest values of this overall index are for impact of shocks on inflation (0.73) and interest rates (0.77). One can argue that these variables are the most important in our analysis. The asymmetric shocks and structural differences are analyzed with respect to monetary policies. Monetary policies of the CNB and the ECB are primary focused on the inflation targeting and are mainly conducted by setting interest rates. Results suggest that impact of shocks on these variables is rather symmetric between the Czech economy and Euro Area 12.

5 Conclusion

In this paper I present a new approach for analysis of asymmetric shocks and structural differences between two countries, using historical shock decomposition. An advantage of this method is that it examines impact of asymmetric shocks and structural differences jointly. This method covers adjustment process to asymmetric shocks as well as their frequency. I demonstrate this method on a New Keynesian DSGE model of two economies, presented in Kolasa [13], [14], which I estimate on the data of the Czech economy and Euro Area 12, using Bayesian techniques. Results suggest that impacts of structural shocks are much more correlated than it seems from estimated correlations between shocks alone. The biggest correlations are between the impact of shocks on inflation and interest rates. I also find interesting structural differences between the Czech economy and Euro Area 12. It seems that evolution of variables in the Czech economy is relatively more influenced by shocks in productivity than by investment efficiency shocks while evolution of variables in Euro Area 12 is relatively more influenced by investment efficiency shocks. It also seems that monetary policy of the ECB is relatively more discretionary than monetary policy of the CNB, nevertheless it seems that the CNB follows the ECB in their discretionary policy.

Further research will be focused on the dynamics and time changes of these results, using sequential estimations. An important question is whether there is a convergence between business cycles of the Czech economy and the Euro Area 12. Sequential estimations will show us whether correlations between impacts of shocks are increasing over time. Consequently, I will examine whether percentage shares of structural shocks in the development of variables in domestic and foreign economy converge to each other. I also plan to estimate this model on the data of Slovenia and Slovakia in order to determine whether joining EMU increased synchronization of business cycles.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

References

[1] Alexius, A., Post, E.: Exchange Rates and Asymmetric Shocks in Small Open Economies, *Empirical Economics* **35(3)** (2008), 527–541.

[2] Amisano, G., Giammarioli, N., Stracca, L.: EMU and the Adjustment to Asymmetric Shocks: the Case of Italy, European Central Bank, Working Paper No. 1128, 2009.

[3] Angeloni, I., Ehrmann, M.: Euro Area Inflation Differentials, European Central Bank, Working Paper No. 388, 2004.

[4] Bayoumi, T., Eichengreen, B.: Shocking Aspects of European Monetary Unification, NBER Working Paper No. 3949, 1992.

[5] Benigno, P.,Lopez-Salido, J. D.: Inflation Persistence and Optimal Monetary Policy in the Euro Area, *Journal of Money, Credit and Banking* **38(3)** (2006), 587–614.

[6] Calvo, G.: Staggered Prices in a Utility Maximizing Framework. *Journal of Monetary Economics*, **12(3)** (1983), 383-398.

[7] Campa, J. M., Gonzalez Minguez, J. M.: Differences in Exchange Rate Pass-through in the Euro Area, *European Economic Review* **50(1)** (2006), 121–145.

[8] Cecchetti, S. G: Legal Structure, Financial Structure, and the Monetary Policy Transmission Mechanism, *Economic Policy Review* **5(2)** (1999), 9–28.

[9] Demertzis, M., Hugues Hallett, A.: Asymmetric Transmission Mechanisms and the Rise in European Unemployment: A Case of Structural Differences or of Policy Failure? *Journal of Economic Dynamics and Control* **22(6)** (1998), 869–886.

[10] Driver, R. L., Wren-Lewis, S.: European Monetary Union and Asymmetric Shocks in a New Keynesian Model, *Oxford Economic Papers* **51(4)** (1999), 665–688.

[11] Fabiani, F., Morgan, J.: Aggregation and Euro Area Phillips Curves, European Central Bank, Working Paper No. 213, 2003.

[12] Jondeau, E., Sahuc, J. G.: Testing Heterogeneity within the Euro Area, Banque de France, Documents de Travail 181, 2007.

[13] Kolasa, M.: Structural Heterogeneity or Asymmetric Shocks? Poland and the Euro Area through the Lens of a Two-country DSGE Model, National Bank of Poland, Working Paper No. 49, 2008.

[14] Kolasa, M.: Structural Heterogeneity or Asymmetric Shocks? Poland and the Euro Area through the Lens of a Two-country DSGE Model, *Economic Modelling* **26(6)** (2009), 1245–1269.

[15] Pauer, F.: Will Asymmetric Shocks Pose a Serious Problem in EMU? Oesterreichische Nationalbank, Working Paper 23, 1996.

[16] Verhoef, B. A.: The (A)Symmetry of Shocks in the EMU, Netherlands Central Bank, DNB Staff Reports 106, 2003.

Price formation in random matching model

Rostislav Staněk [1]

Abstract. This paper is dealing with a question how are monetary prices determined in the random matching model. When answering this question current literature assumes that money or good is indivisible and solves this problem in Nash axiomatic bargaining framework. This paper provides a random matching model where prices are determined in sequential bargaining process and the indivisibility restrictions are not imposed. Determination of prices in this model does not depend on the marginal values but it is determined by the probability of finding a new trading partner. There is a price dispersion in the model when agents are heterogenous.

Keywords: decentralized market, random matching, bargaining

JEL classification: D40, C72, C78
AMS classification: 91A10, 91A40, 91A80,

1 Introduction

A major concern in modern monetary theory has been the development of models where money is essential because of the frictions in the decentralized exchange. The most known and the most used are the random matching models [5]. These models are capable to explain the essentiality of money because they assume that trade takes place on the decentralized market with absence of double coincidence of wants, limited monitoring and limited commitment (see e.g. [3], [4]). But because of the absence of the centralized market, there arise a problem of price formation. Moreover, there is a problem to determine the endogenous distribution of money holdings. In this paper, I set the problem of the distribution of money holdings aside and I focus on the formation of prices. Several solutions to this problem can be found in the literature:

- Trejos-Wright [7] solution assumes that goods are divisible but money is indivisible and the traders always exchange some amount of good for one unit of money. The amount of good exchanged is determined by the Nash axiomatic bargaining solution.
- Green-Zhou [1] solution assumes, on the other hand, that money is divisible but good is indivisible. The price for one unit of good is again given by the Nash axiomatic bargaining solution.
- Lagos-Wright [2] solution proposes a framework where the trade alternates between the centralized and the decentralized market. Price formation on the decentralized market then depends on the questionable assumption that agents are able to trade on the centralized market.

I consider these solutions unsatisfactory in two respects. First, they impose doubtful restrictions on the exchange process such as indivisibility of good, indivisibility of money or the existence of a centralized market. Second, all of these solutions use Nash axiomatic bargaining concept. Instead of that, it is possible to use the sequential model of bargaining which is advantageous in several ways. The choice of bargaining power of trading partners is not given exogenously. Rather it is determined endogenously by the matching process. The sequential bargaining model also enables for the possibility that the bargaining breaks down.

The aim of this paper is to analyze price formation in the model with decentralized trade. In particular, I address two questions: "What factors determine prices?" and "Is there any price dispersion?" In order to answer these questions, I provide a random matching model where prices are determined by the sequential bargaining. Simultaneously, the model allows for divisibility of money and goods and enables

[1] Masaryk University/Economic Faculty, Lipová 41a, Brno, 602 00, the Czech Republic, e-mail: rostanek@mail.muni.cz

only decentralized exchange. The bargaining process is modeled by a modification of the Rubinstein bargaining game with alternating offers [6]. My model differs from the Rubinstein model in two important respects. First, my model combines the Rubinstein bargaining game with the baseline random matching framework of Kiyotaki and Wright model [5]. Second, in the players bargain over the division of a pie of a fixed size in the Rubinstein game. Whereas in my model, bargaining determines not only the division of the pie but also the size of the pie.

2 Model environment

In this section, I describe the environment of the model. The environment is similar to many random matching models. In particular, consider an economy with M infinitely-lived agents and N perishable and indivisible types of goods, where $N \geq 3$. There are N types of agents. An agent of type i produces only good i and consumes good $(i+1) \mod N$. Hence, trade is necessary for consumption to take place. Economic agents of each type are either sellers or buyers. Each seller of type i enters the market with the intention to sell the good i. Each buyer of type i wants to buy a good of type $i+1$. Denote S the total number of sellers and B the total number of buyers. Suppose that the proportion of buyers of the type i to the total number of buyers is the same for all types of buyers, i.e. $\frac{B_i}{B} = \frac{1}{N}$. The same holds for all types of sellers.

The environment is characterized by the decentralized trade. At each date, each agent in the market is matched with positive probability with an agent of the opposite type. If the agent is matched with a new partner, he must abandon his current trading partner. The probabilities can be derived from the matching technology. Assume that there are M new matches in each period. Then, $\alpha = \frac{M}{S}$ denotes the probability that the seller is matched with a new buyer and similarly $\beta = \frac{M}{B}$ denotes the probability that the buyer is matched with a new seller in each period. Note that α and β can be interpreted as a measurement of competition on the supply side and demand side respectively. Further, denote $P_t(i)$ as the agent who is in agent's i match in period t. As we can see, the agent produces or consumes only if she meets the proper trading partner. Hence, if the agent i is matched with the agent $i-1$, then the agent i is able to produce and the agent $i-1$ is able to consume. Otherwise an agent is idle. Each buyer is able to consume in pairwise meeting with probability $\sigma = \frac{1}{N}$ and each seller is able to produce with the same probability. We can interpret the parameter σ as a measurement of a degree of specialization. If the parameter σ is low, then a lot of different types of goods is produced in the economy and each agent demands a smaller fraction of them. So, the lower the value of σ, the more specialized the economy is.

Agents are supposed to be a expected utility maximizers. Each seller's preferences are represented by the utility function $v_{s,i}(q_i, m_i)$, i.e. she is concerned about the amount of good i she produces and the amount of money she holds. Each buyer's preferences are represented by a utility function $v_{b,i}(q_{i+1}, m_i)$, i.e. buyer is concerned about the amount of good $i+1$ she consumes and the amount of money she holds. I assume that preferences of all agents are quasilinear in money. This means that the one-period payoff function of the seller i that produces the quantity q_i of his good and sells this quantity at price p can be written as

$$v_{s,i}(q_i, m_i + p) - v_{s,i}(0, m_i) = m_i - c_i(q_i) + p - c_i(0) - m_i = p - c_i(q_i), \tag{1}$$

where $c(q_i)$ is the cost function with common properties $c'(q_i) > 0$, $c''(q_i) > 0$ and $c(0) = 0$. Alternatively, one-period payoff function of buyer i that buys the amount q_{i+1} at price p can be written as

$$v_{b,i}(q_{i+1}, m_i - p) - v_{b,i}(0, m_i) = u_i(q_{i+1}) + m_i - p - u_i(0) - m_i = u_i(q_{i+1}) - p, \tag{2}$$

where $u(q_{i+1})$ represents the buyer's willingness to pay with properties $u'(q_{i+1}) > 0$, $u''(q_{i+1}) < 0$ and $u(0) = 0$. If the agent does not trade, then he obtains the utility of zero. Time is discrete and infinite, indexed by integers. Each agent discounts future payoff by the discount parameter $\delta \in (0, 1)$.

During the match the members of any matched pair bargain over the quantity of goods exchanged q and the price for this quantity p. At the start of the match a random device selects who makes the first offer. Denote (p^B, q^B) as the offer made by the buyer and (p^S, q^S) as the offer made by the seller. In the event of acceptance, the transaction is completed at the agreed upon quantity and price. In the event of rejection, the pair continues bargaining with probability $(1-\alpha)(1-\beta)$. In this case, the agent who rejected the offer makes an counteroffer. With probability $\alpha(1-\beta)$ the seller is matched with a new partner and the buyer remains unmatched. With probability $\beta(1-\alpha)$ the buyer starts bargaining with a new partner while the seller remains unmatched and with probability $\alpha\beta$ both agents start bargaining with new trading partners.

3 Equilibrium with homogenous agents

In this section, I present an equilibrium of the model where all agents are homogenous. Homogenous agents do not differ in their tastes or production possibilities. So, assume that $c_i(q_i) = c_j(q_i)$ for all j and $u_i(q_{i+1}) = u_j(q_{i+1})$ for all j. Define a strategy of an agent as a function that assign to every possible history of events within a match either an offer or a response to an offer. This means that an agent cannot condition her behavior on the events that occurs in any period before she started bargaining with his current partner. Rubinstein [6] shows that bargaining game with alternating offers has a unique subgame perfect equilibrium where player's strategies are stationary, i.e. each agent uses the same rule of behavior in every period, and the agreement is reached immediately. Therefore, I restrict my attention on the stationary strategies of the following type

- the seller offers the pair (q^{S*}, p^{S*}) and accepts an offer (q^B, p^B) such that $p^B - c(q^B) \geqq \delta((1-\alpha)(1-\beta)(p^{S*} - c(q^{S*})) + (1-(1-\alpha)(1-\beta))U^S)$ where U^S is the expected payoff of a seller in the event of breakdown,
- the buyer offers the pair (q^{B*}, p^{B*}) and accepts an offer (q^S, p^S) such that $u(q^S) - p^S \geqq \delta((1-\alpha)(1-\beta)(u(q^{B*}) - p^{B*}) + (1-(1-\alpha)(1-\beta))U^S)$ where U^B is the expected payoff of a seller in the event of breakdown.

Now, I show that both players offer the quantity that maximizes the gains from trade. Then, I calculate the equilibrium prices and discuss their determinants.

Consider a match with a single coincidence of wants. Suppose that the seller makes an offer first. Subgame perfect equilibrium requires that the buyer accepts every offer (p^S, q^S) such that her expected payoff $u(q^S) - p^S$ is equal or greater than the expected payoff from rejecting the offer. The seller thus solves the following problem

$$\max p^S - c(q^S) \quad \text{s.t.} \quad u(q^S) - p^S = \delta((1-\alpha)(1-\beta)(u(q^B) - p^B) + (1-(1-\alpha)(1-\beta))U^B). \tag{3}$$

It is obvious that the condition is binding, otherwise the seller can increase its profit by offering a slightly higher price. The expected payoff U^B is determined by the strategies of other players and the matching probabilities. Hence, it is clear that it does not depend on the offer (q^S, p^S). The solution of this problem is therefore given by two following equations.

$$u'(q^S) = c'(q^S), \tag{4}$$

$$u(q^S) - p^S = \delta((1-\alpha)(1-\beta)(u(q^B) - p^B) + (1-(1-\alpha)(1-\beta))U^B). \tag{5}$$

By analogy, buyer solves the following problem

$$\max u(q^B) - p^B \quad \text{s.t.} \quad p^B - c(q^B) = \delta((1-\alpha)(1-\beta)(p^S - c(q^S)) + (1-(1-\alpha)(1-\beta))U^S). \tag{6}$$

The solution of this problem is again given by two equations

$$u'(q^B) = c'(q^B), \tag{7}$$

$$p^B - c(q^B) = \delta((1-\alpha)(1-\beta)(p^S - c(q^S)) + (1-(1-\alpha)(1-\beta))U^S). \tag{8}$$

As we can see, the seller as well as the buyer offer the quantity q^* that solves the equation (4) or (7) and maximizes the gains from trade. From the properties of functions $u(q)$ and $c(q)$ we know that this solution is unique.

Now, I shall find the equilibrium prices. To find equilibrium prices we need to calculate expected utility in the event of breakdown. Denote V^{US} as the expected utility of the seller that is not matched or is matched with an inappropriate trading partner, i.e. $P_t(i) \neq i-1$. Denote V^{UB} as the expected utility of the buyer that is not matched or is matched with the trading partner $P_t(i) \neq i+1$. Similarly, V^{MS} denotes the expected utility of the seller that is matched with the trading partner $P_t(i) = i-1$ and V^{MB} is expected utility of a buyer matched with the trading partner $P_t(i) = i+1$. Expected utility U^S and U^B can be stated as follows

$$U^S = \frac{\delta(\alpha\sigma V^{MS} + (\beta(1-\alpha) + \alpha(1-\sigma))V^{US})}{1-(1-\alpha)(1-\beta)}, \tag{9}$$

$$U^B = \frac{\delta(\beta\sigma V^{MB} + (\alpha(1-\beta) + \beta(1-\sigma))V^{UB})}{1-(1-\alpha)(1-\beta)}. \tag{10}$$

The expected utility V^{US} and V^{UB} can be expressed recursively as

$$V^{US} = \delta(\alpha\sigma V^{MS} + (1-\alpha\sigma)V^{US}), \tag{11}$$
$$V^{UB} = \delta(\beta\sigma V^{MB} + (1-\beta\sigma)V^{UB}). \tag{12}$$

Finally, because the agents reach agreement at the start of every match with single coincidence of wants, the following two equations hold

$$V^{MS} = \frac{1}{2}(p^S + p^B - c(q^S) - c(q^B)), \tag{13}$$
$$V^{MB} = \frac{1}{2}(u(q^S) + u(q^B) - p^S - p^B). \tag{14}$$

The equilibrium prices p^{S*} and p^{B*} can be find by straightforward algebra by solving equations (5), (8) (9), (10), (11), (12), (13) and (14). The solution of these equations does not provide an easy survey. But we can get the main insight into the problem of determination of equilibrium prices when we suppose that the agents are almost completely patient. The fact that the discount factor is less than 1 creates a friction in addition to the frictions given by matching process. If we want to know how the decentralization of the trade affects equilibrium prices we need to consider the case when the agents are almost completely patient. In this case, the equilibrium prices are given by the following equations

$$\lim_{\delta\to 1} p^{S*} = \frac{\alpha u(q^*) + \beta c(q^*)}{\alpha+\beta}, \quad \lim_{\delta\to 1} p^{B*} = \frac{\alpha u(q^*) + \beta c(q^*)}{\alpha+\beta}. \tag{15}$$

In contrast to the centralized market, the price is unrelated to marginal costs or marginal willingness to pay. In this decentralized market model, the price is given by the matching probabilities. The agent that has a greater probability to find a suitable partner bargains more favorable price. If we relate the equilibrium price to the number of buyers and sellers, we get the equilibirum price $\frac{Bu(q^*)+Sc(q^*)}{B+S}$. If there is the same number of sellers and buyers in the market, then the price is $\frac{u(q^*)+c(q^*)}{2}$ and it halves the total surplus from trade between the buyer and the seller. If the number of buyers in the market exceeds the number of sellers, then the price raises and the seller gets greater share of the surplus. Note also, that the price does not depend on the degree of specialization in the limit (this is not generally true if $\delta \ll 1$).

The equilibrium prices leads to the following expected payoffs from a single-coincidence of wants match

$$V^{MS*} = \frac{(u(q^*) - c(q^*))(1-\delta(1-\sigma\alpha))}{2-\delta(2-\alpha\sigma-\beta\sigma)}, \quad V^{MB*} = \frac{(u(q^*) - c(q^*))(1-\delta(1-\sigma\beta))}{2-\delta(2-\alpha\sigma-\beta\sigma)}.$$

The expected gains from trade depends on the parameters α, β and σ. The effect of matching probabilities is the same as discussed above. The greater is the degree of specialization σ, the smaller is the expected utility of an agent with smaller matching probability. Consider, for example, a market where the number of sellers exceed the number of buyers. In this case, sellers have smaller bargaining power and the price is relatively low. Above that, this competition effect is supported by the specialization effect and the price will be even lower when the degree of specialization increases. But the specialization effect disappears as agents become more patient. Finally, consider the case when discount factor goes to 1 and relate the probabilities to the population frequencies of buyers and sellers. Then, the expected gains from trade are as follows

$$\lim_{\delta\to 1} V^{MS*} = \frac{B(u(q^*) - c(q^*))}{B+S}, \quad \lim_{\delta\to 1} V^{MB*} = \frac{S(u(q^*) - c(q^*))}{B+S}.$$

Again, we can see that the expected gains from trade fundamentally depends on the number of sellers and buyers.

4 Equilibrium with heterogenous agents

We have seen that there is no price dispersion when the agents are homogenous. But is this still true when the agents are heterogenous? In this section, I abandon the assumption of homogenity of agents. For the sake of simplicity, I assume that agents differ only in their production skills. Assume that the cost function of agent i is $K_i c(q_i)$ where K is a random variable with $E(K) = 1$. I call K_i as the production efficiency parameter. It is clear that the lower K_i, the more productive the agent is. In all other respects the model is the same as in the previous section.

We can find subgame perfect equilibirum strategies in the same way as in the model with homogenous agents. Consider a match with a single coincidence of wants, where the seller's cost function is $K_i c(q_i)$. The seller's and buyer's problems when making an offer are given by the following expressions respectively

$$\max p^S - K_i c(q^S) \quad \text{s.t.} \quad u(q^S) - p^S = \delta((1-\alpha)(1-\beta)(u(q^B) - p^B) + (1-(1-\alpha)(1-\beta))U^B), \tag{16}$$
$$\max u(q^B) - p^B \quad \text{s.t.} \quad p^B - K_i c(q^B) = \delta((1-\alpha)(1-\beta)(p^S - K_i c(q^S)) + (1-(1-\alpha)(1-\beta))U^S). \tag{17}$$

The solutions of these two problems are given by the two constraints and the following equation

$$u'(q) = K_i c'(q) \tag{18}$$

It is obvious that the trading partners always bargain the effective amount of production which maximizes their total surplus. Notice that the effective amount of production q^* depends on the parameter of production efficiency K_i. If the seller is more productive, then the quantity exchanged is greater.

The seller's and buyer's expected payoff in the event of breakdown is defined by equations (9) and (10) respectively. If the seller or buyer remain unmatched, then their expected payoffs are given by equations (11) and (12) respectively. The seller's expected payoff from single coincidence of wants meeting is

$$V^{MS} = \frac{1}{2}(p^S + p^B) - K_i c(q^*). \tag{19}$$

Note that the seller always meets the buyer with the same utility function. Because the seller's strategy is stationary, she offers the same quantity and price in every match. The buyer's expected payoff from single coincidence of wants meeting is

$$V^{MB} = u(q^{E*}) - \frac{1}{2}(p^{ES} - p^{EB}), \tag{20}$$

where q^{E*} is the amount of good bargained in the match with average seller, i.e. the seller with $K_i = 1$, p^{ES} and p^{EB} denotes the prices offered in the match with average seller by the seller and buyer respectively. The equilibrium price and the expected gains from trade obtained in the match with the seller with production efficiency K_i are given by constraints in problems (16) and (17), equations (9), (10), (11), (12), (19), (20) and by the prices p^{ES} and p^{EB}. Obviously average prices p^{ES} and p^{EB} are the same as equilibrium prices from the previous section. So, prices and expected payoffs from trade can be calculated. I present the results for agents that are almost completely patient

$$\lim_{\delta\to 1} p^{S*} = u(q^*) - \frac{\beta u(q^{E*}) + \beta c(q^{E*})}{\alpha+\beta}, \qquad \lim_{\delta\to 1} p^{B*} = u(q^*) - \frac{\beta u(q^{E*}) + \beta c(q^{E*})}{\alpha+\beta}, \tag{21}$$
$$V^{MS} = u(q^*) - K_i c(q^*) - \frac{\beta u(q^{E*}) + \beta c(q^{E*})}{\alpha+\beta}, \qquad V^{MB} = \frac{\beta u(q^{E*}) + \beta c(q^{E*})}{\alpha+\beta}. \tag{22}$$

We can see that the conclusions made in previous section regarding the influence of the number of buyers and sellers on the equilibrium price are still valid. The higher the number of buyers, the higher the price and vice versa. Note that the buyer's expected gain from trade V^{MB} does not depend on the efficiency of the seller in the match. Consider that a buyer is matched with a seller with production efficiency parameter $K_i < 1$. Then, the negotiated amount of production is q^* such that $q^* > q^{E*}$. The total surplus is also greater compared to the match with average seller. But the buyer obtains the same surplus as in the match with average seller and all surpluses originating from more efficient production goes to the seller. This also means that less efficient sellers obtain lower surpluses permanently. Moreover, there is an upper bound on the production efficiency parameter $\overline{K} = \frac{u(q^*)}{c(q^*)} - \frac{\beta u(q^{E*}) + \beta c(q^{E*})}{(\alpha+\beta)c(q^*)}$ such that sellers with $K_i > \overline{K}$ are not able to trade at all.

We are primarily interested in the prices bargained by different sellers. It is clear that p^{S*} or p^{B*} are higher when the buyer meets a more efficient seller. But we have to keep in mind that this is the price for the whole amount q^*. This does not tell anything about the price for a unit of production. In order to analyze dispersion in unit prices, I have to compare p^{E*}/q^{E*} and p^*/q^*. Suppose that $q^* > q^{E*}$. From the equations (15) and (21) we know that the price change induced by the change of equilibrium quantity is $u(q^*) - u(q^{E*})$. When we consider an infinitely small change in the equilibrium quantity, we obtain after some algebra that the unit price of the more efficient seller $q^* > q^{E*}$ is lower than the unit price of average seller p^{E*}/q^{E*} if and only if

$$\frac{\partial u(q)}{\partial q}\frac{q^{E*}}{p^{E*}} < 1. \tag{23}$$

Note that the condition can be restated as $p^{E*}/q^{E*} > u'(q^{E*})$. This means that the price of efficient seller is lower if and only if the average unit price is greater than the marginal willingness to pay. If this is the case and the buyer is matched with a more efficient seller, then the bargained quantity is above average. If the unit price was the same as the average unit price, then the surplus of the buyer would be lower because she would buy additional quantity for a higher price than is her marginal willingnes to pay. But we know that buyer's surplus remains the same in all matches. Hence, the unit price of the efficient seller must be lower than the average unit price. Naturally, if the condition (23) is fulfilled, then the unit price of the less efficient seller is above average. On the other hand, if the average unit price is lower than the marginal willingness to pay, then the unit price of a more efficient seller is above average and the unit price of a less efficient seller is below average. Finally, if the average unit price is equal to the marginal willingness to pay and the degree of heterogenity is small, then the unit price of all sellers is the same and no price dispersion occurs. This analysis also shows that in the decentralized market the decrease in the cost function does not straightforwardly indicate that the seller will charge higher or lower price. But this depends on his productivity and average market price.

5 Conclusion

In the conclusion, I emphasize the main findings of this paper. I have shown that the price formation in the random matching model with decentralized trade, divisible money and divisible good is significantly different from the price formation in the model with the centralized market. The traders exchange efficient amount of good in every match. But equilibrium prices are completely unrelated to marginal costs or to the marginal willingness to pay. The equilibrium price is rather determined by the opportunity to find a new and possibly more favorable trading partner. This opportunity is mainly determined by the number of buyers and sellers in the market. The degree of specialization affects the equilibrium price only if the agents are not patient. All of these factors should be taken into consideration when one uses the Nash axiomatic bargaining concept and defines the bargaining power of players. There exists a price dispersion in the random matching model when the agents are heterogenous. Interestingly, the price charged by more efficient sellers can be above or below average price. This depends on the fact whether the average price on the decentralized market is above or below the hypothetical centralized market price.

References

[1] Green, E., and Zhou, R.: A rudimentary model of search with divisible money and prices, *The Journal of Economic theory* **81** (1998), 252–271.

[2] Lagos, R., and Wright, R.: A unified framework for monetary theory and policy, *The Journal of Political Economy* **113** (2005), 463–484.

[3] Kocherlakota, N.: Money is memory, *Journal of Economic Theory* **81** (1998), 232–251.

[4] Kiyotaki, N., and Moore, J.: Evil is the root of all money, *American Economic Review* **92** (2002), 62–66.

[5] Kiyotaki, N., and Wright, R.: A search theoretic approach to monetary economics, *American Economic Review* **83** (1993), 63–77.

[6] Rubinstein, A.: Perfect Equilibrium in a Bargaining Model, *Econometrica* **50** (1982), 97–109.

[7] Trejos, A., and Wright, R.: Search, bargaining, money nad prices, *The Journal of Political Economy* **103** (1995), 118–141.

Calendar Anomalies in the Central European Stock Exchanges during the Financial Crisis

Daniel Stavárek[1], Tomáš Heryán[2]

Abstract. The aim of the paper is to estimate the day of the week effect in the stock markets in the Czech Republic, Hungary and Poland over the period 2006 – 2011. We separately estimate a modified GARCH-M (1,1) model for five sub-periods capturing individual phases of the financial crisis. We found mixed evidence on the day of the week presence. The anomalies in returns were revealed particularly in the stabilization phase and post-crisis period; however, they are opposite to usual findings presented in previous research. We did not discover any solid evidence of the day of the week effect in volatility of returns.

Keywords: day of the week effect, stock market, financial crisis, Central Europe.

JEL Classification: C32, G01, G14
AMS Classification: 62P20, 91B44

1 Introduction

Behaviour of stock markets belongs to the most widely analyzed and discussed topics in financial literature. Fama [4] postulated that securities markets were extremely efficient in reflecting information about individual stocks and about the stock market as a whole. He states that the Efficient Market Hypothesis (EMH) can be divided into three categories depending on the nature of the information subset of interest: (i) strong form, (ii) semi-strong form, (iii) weak form. However, many studies found empirical evidence against validity of the semi-strong and weak forms of EMH. These contradictions are considered as anomalies. If the anomalies appear regularly in trading with stocks and can influence stock market returns they are usually referred to calendar anomalies. Calendar anomalies rest on the basic assumption that the past behaviour of a stock's price is rich in information pertaining to its future behaviour. In other words, the study of calendar anomalies suggests that investors could use these results on anomalies to predict stock market movements on given days [7]. The most important examples of calendar anomalies in stock markets are day of the week effect, twist of the Monday, turn of the month, turn of the year and holiday periods.

The present paper examines the day of the week effect in three Central European emerging stock markets (Czech Republic, Hungary and Poland).This calendar anomaly refers to the variation of stock market returns by day of the week. The existing empirical evidence suggests that Monday is generally considered to exhibit negative returns whereas Friday exhibits positive and the highest returns in a week. The phenomenon of day of the week effect has been extensively researched in many stock markets around the world using various techniques and approaches. However, the issue has been particularly addressed in the world's leading developed and emerging stock markets. The stock markets covered in this study have been so far researched rarely. The existing studies generally found none or little evidence of the day of the week effect in the respective stock markets.

Tonchev and Kim [8] searched for various anomalies in the Czech, Slovak and Slovenian capital markets in 1999 – June 2003 applying OLS regression along with GARCH estimation. As regards to the day of the week effect, they only found weak evidence in mean for Slovenia, but in the opposite direction. Apolinario *et al.* [1] used GARCH and TGARCH models on July 1997 – March 2004 data from 13 European stock markets including the Czech Republic. The Czech market as one of the two did not exhibit any evidence of the day of the week effect. Chukwuogor-Ndu [3] used a battery of parametric and non-parametric tests on daily returns of 15 European stock markets in the period 1997 – 2004. Application of the Kruskal-Wallis test provided no evidence of the day of the week effect in the Czech Republic's stock market. Bubák and Žikeš [2] focused on the same period 1997 – 2004 and investigate the seasonality and nontrading effect in stock market indices of the Czech Republic, Germany, Hungary and Poland. They used the PAR-PGARCH model and revealed significant day of the week effects in the mean of returns on the Czech and Polish index, and significant seasonality in the volatility of the Hungarian index. Yalcin and Yucel [9] investigated anomalies in the stock markets of 20 emerging economies including all markets analyzed in the present paper. They used the EGARCH-M model and found that the

[1] Silesian University, School of Business Administration, Department of Finance. Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, e-mail: stavarek@opf.slu.cz.
[2] Silesian University, School of Business Administration, Department of Finance. Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic, e-mail: stavarek@opf.slu.cz.

variance of Friday returns in the Czech Republic is significantly lower than that of Wednesdays. For Hungary, Monday coefficients are negative and significantly different from those of Wednesdays. For Poland, Tuesdays and Thursdays have higher variance than other days of the week. There was found a little evidence of the day of the week effect in the Czech and Hungarian markets in [5]. The authors investigated 18 European markets and concluded that the effect is rather local than global phenomenon.

The aim of the paper is to estimate the day of the week effect in the stock markets in the Czech Republic, Hungary and Poland over the period of five years (April 2006 – March 2011). The present study substantially contributes to the existing literature as it covers the most recent period and the markets that have not been in the centre of researchers' interest. Since the analyzed period includes the global financial crisis the paper also reveals the influence of the financial crisis and its phases on presence and character of the day of the week effect.

The paper is structured as follows. In Section 2, we introduce and describe the dataset and specify the model used in estimation. In Section 3, we present and discuss the results obtained from estimations. Section 4 concludes the paper with summary of crucial findings.

2 Data and model specification

This paper employs the daily closing values of the stock market main indices in the Czech Republic, Hungary and Poland. Namely, we use the Prague Stock Exchange Index (PX), the Budapest Stock Exchange Index (BUX) and the Warsaw Stock Exchange Index (WIG). The time series of the indices' closing values were collected from the Patria financial database. We consistently use five observations per week and the returns for nontrading days are calculated using the closing price indices from the last trading day. This approach is followed in order to avoid possible bias from the loss of information due to public holidays.

The period under estimation starts on April 2006 and end on March 2011. Hence, we have 1303 daily returns for each stock market in total. The whole estimation period was divided into five sub-periods to capture individual phases of the financial crisis. The pre-crisis period (Period 1) is from April 2006 to March 2007. The phase of crisis initialization (Period 2) starts in April 2007 and ends on 14 March 2008. The crisis culmination (Period 3) lasts from 17 March 2008 to end of March 2009. The phase of crisis stabilization (Period 4) covers the period from April 2009 to March 2010. Finally, the post-crisis phase (Period 5) starts in April 2011 and ends on 31 March 2011.

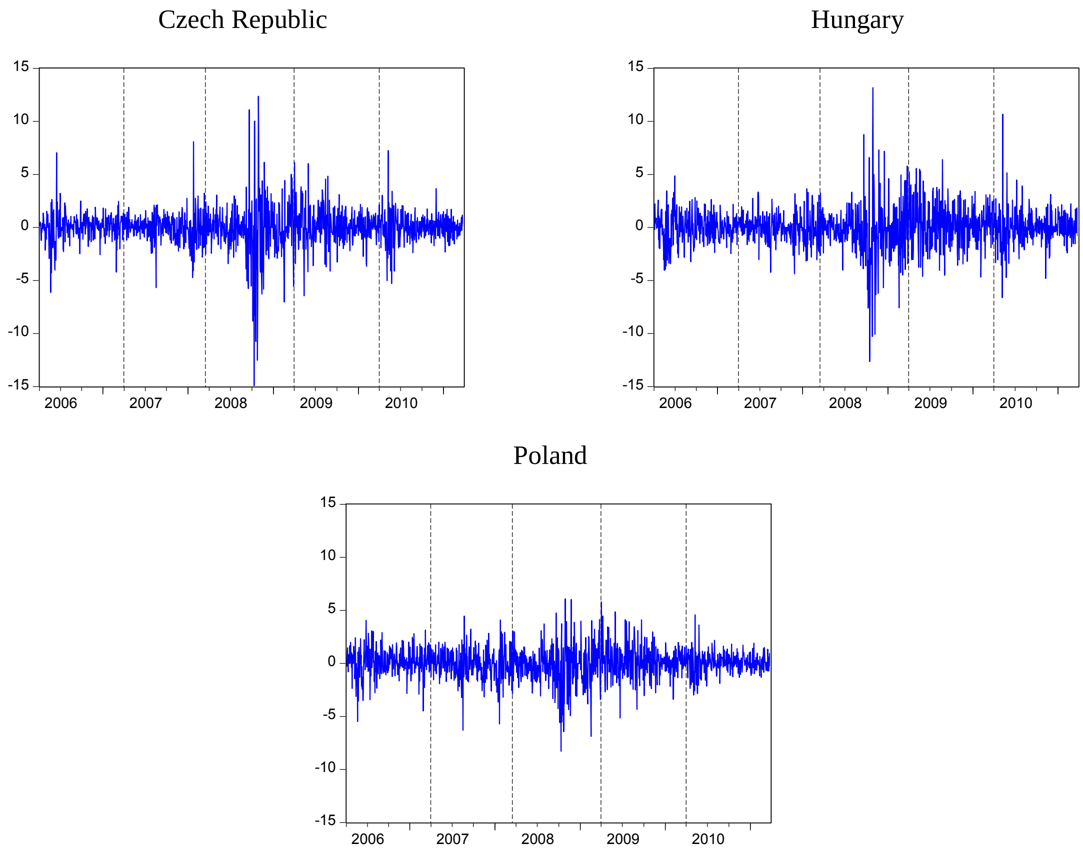


Figure 1 Stock market indices' daily returns (in %) in the financial crisis phases

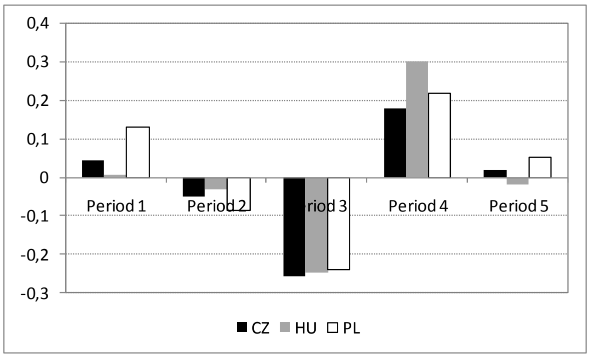
Following the standards used in literature, daily returns are calculated as first difference in natural logarithms and then multiplied by 100 to approximate percentage changes:

$$R_t = \ln\left(I_t / I_{t-1}\right) \times 100 \tag{1}$$

where I_t and R_t refer to the stock market index closing value and the daily return on day t, respectively. The calculated daily returns are depicted in Figure 1. The vertical lines in graphs delimitate the five phases of the financial crisis.

Although a higher volatility and serious fluctuations of daily returns are observable in all stock markets during the phase of crisis culmination the Polish market seems to be least affected by the crisis. By contrast, one can observe many extreme daily returns considerably exceeding the usual levels in the Czech and Hungarian markets. All markets demonstrate a clear tendency to restoration of standard behavior patterns in the course of Period 4 and Period 5. The differences among the phases of financial crisis can be also documented by comparison of the average daily returns and standards deviations. The graphs in Figure 2 show the average daily return and standard deviation for all markets in individual periods. Whereas the standard deviation in Period 1 and Period 2 are very similar in all markets the average daily returns differ and Period 2 exhibits negative returns. Culmination of the financial crisis brought to the stock markets substantial volatility and remarkably negative average daily returns. The stabilization and post-crisis phases are characteristic of gradually decreasing of volatility to the pre-crisis levels. The rebound of returns in Period 4 was followed by stagnation in Period 5. Hence, none of the individual periods is similar to the others and, subsequently, the estimations of the day of the week effect are conducted separately for each period.

Average logarithmic daily return



Standard deviation of daily returns

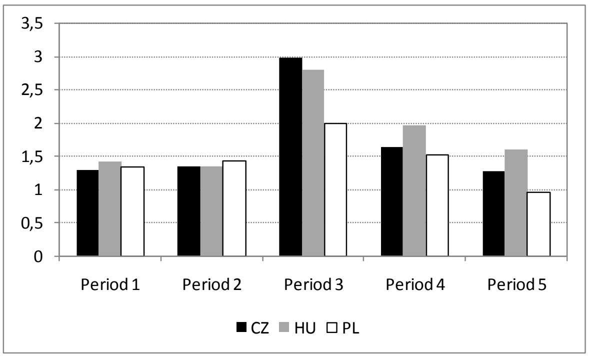


Figure 2 Average logarithmic daily return and standard deviation in individual periods

The standard approach how to estimate the day of the week effect is in mean returns through the conventional OLS regression methodology, and appropriately defined five dummy variables, one for each day of the week. However, this approach suffers from a serious problem with the variance of residuals that is not constant and possibly time-dependent. In this respect, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model is able to capture the time-varying variability in the variance of the residuals. This approach has the advantage that the conditional variance can be expressed as a function of past errors. These models assume that the variance of the residual term is not constant through time.

We used the GARCH-M (1,1) model in the following specification:

$$R_t = \alpha_0 + \alpha_{TUE} TUE + \alpha_{WED} WED + \alpha_{THU} THU + \alpha_{FRI} FRI + \lambda h_t + \varepsilon_t \tag{2}$$

$$h_t^2 = \alpha + \delta_{TUE} TUE + \delta_{WED} WED + \delta_{THU} THU + \delta_{FRI} FRI + \beta \varepsilon_{t-1}^2 + \gamma h_{t-1}^2 \tag{3}$$

where R_t represents daily returns of an examined index. *TUE, WED, THU* and *FRI* are the dummy variables for Tuesday, Wednesday, Thursday and Friday, while we exclude the Monday's dummy variable from the equation to avoid the dummy variable trap. Further, h_t is the conditional variance, ε_t denotes the residual term and λ is a measure of the risk premium, as it is possible that the conditional variance (proxy for risk) can affect stock market returns. If λ is positive the risk-averse agents must be compensated to accept the higher risk.

We followed suggestions from [6] and consider the day of the week effect also in the volatility equation. Hence, we included the dummy variables also in the variance equation. Such a modification of the standard GARCH-M specification accounts for the possible stationary effects within the variance equation. Our approach finally leads to joint estimates of the day of the week effect, not only in the mean but also in the conditional variance.

3 Results of the day of the week estimation

We estimated the GARCH-M (1,1) model according to the specification in (2) and (3) for all the countries and periods analyzed. The obtained results are presented in Tables 1-3. The upper part of the table summarizes results for the mean equation and the lower part reports the results for the variance equation.

As regards to the day of the week effect in the mean equation, the individual meaning for each one of the dummy variables could reveal the presence of an atypical yield during a day of the week with respect to that of Monday. Although we found some evidence of the effect presence the results are rather mixed and one can reveal only few common patterns for all three stock markets. In addition, the significance and sign of individual dummy variables differ across the periods.

	Period 1	**Period 2**	**Period 3**	**Period 4**	**Period 5**
α_0	-0.0552	-0.1497	-0.0065	0.3189	0.2576***
α_{TUE}	0.0892	0.1717	-0.4871***	-0.4016	-0.4430**
α_{WED}	0.3056	0.3166***	0.4818***	-0.2624	-0.1046
α_{THU}	0.3332	0.1304	-0.0661	-0.4769***	-0.1585
α_{FRI}	0.0582	0.3092***	0.0172	-0.6648**	-0.3394**
λ	0.0346	0.0273	-0.0076	0.0803	0.0037
Variance equation					
α	1.1012**	-0.4328*	-1.1800	-0.5152	0.3260
δ_{TUE}	-1.0505	0.7422**	1.2545	0.6080	-0.3814
δ_{WED}	-1.2304**	0.4169	1.4414	0.4897	0.1499
δ_{THU}	-0.5207	0.5713***	1.0767	0.5449	-0.6997
δ_{FRI}	-1.3041**	0.6136	2.4269	1.0938	-0.5566
β	0.1879***	0.2059*	0.1885*	-0.0392*	0.1064**
γ	0.6420*	0.8024*	0.8292*	1.0103*	0.8713*

Table 1 Estimation of GARCH-M model for the Czech Republic's stock market
Note: *,**,*** denote significance at 1%, 5% and 10% respectively

	Period 1	**Period 2**	**Period 3**	**Period 4**	**Period 5**
α_0	0.0581	-0.2057	-0.1799	0.6324	0.2244
α_{TUE}	0.0207	0.1050	0.2259	-0.9819*	-0.2146
α_{WED}	0.0871	0.1756	0.4817	-0.8400**	0.1264
α_{THU}	0.2295	0.2035	0.1257	-0.9220*	-0.2447
α_{FRI}	0.0086	0.3086	0.0146	-1.0326*	-0.2744
λ	-0.0390	0.0619	-0.0159	0.0930	-0.0561
Variance equation					
α	0.3156	0.0901	-0.4191	-0.1756	-0.3990
δ_{TUE}	-0.1279	0.1424	0.4114	1.0556	0.9587
δ_{WED}	-0.3687	0.2789	1.1826	-0.7112	0.8910
δ_{THU}	0.4372	0.3761	-0.3175	0.7416	0.1365
δ_{FRI}	-1.1367***	0.4671	1.1203	0.1603	0.4993
β	0.1021***	0.3296**	0.1499*	0.0704**	0.1135*
γ	0.8598*	0.5040*	0.8549*	0.9164*	0.8455*

Table 2 Estimation of GARCH-M model for the Hungary's stock market
Note: *,**,*** denote significance at 1%, 5% and 10% respectively

	Period 1	Period 2	Period 3	Period 4	Period 5
α_0	0.1633	-0.1011	-0.1886	0.4715**	0.2978**
α_{TUE}	-0.0800	0.1819	0.0100	-0.3194	-0.4305*
α_{WED}	-0.0906	0.0497	0.1231	-0.2389	-0.0821
α_{THU}	0.0471	-0.0324	-0.0203	-0.5551*	-0.2856***
α_{FRI}	0.0163	0.1321	-0.1435	-0.7288*	-0.1990
λ	-0.0397	0.0407	0.0210	0.1628	-0.0321
Variance equation					
α	0.2000	-0.1428	0.9970	0.2189	0.0836
δ_{TUE}	0.1066	0.4628	-0.9153	-0.4426**	0.0971
δ_{WED}	-0.4769	0.1341	-1.3423	-0.7811***	-0.1141
δ_{THU}	1.0982***	0.5810	-1.6628	0.4164	-0.1182
δ_{FRI}	-1.2511***	0.1471	-0.5987	-0.0579	-0.1831
β	0.0560***	0.0971***	0.1228**	0.0778***	0.0610**
γ	0.8916*	0.8453*	0.8579*	0.8951*	0.9067*

Table 3 Estimation of GARCH-M model for the Poland's stock market
Note: *,**,*** denote significance at 1%, 5% and 10% respectively

However, there is almost no evidence of the day of the week effect during the first three periods. In the pre-crisis period and the phases of crisis initialization and culmination the return in the Central European markets was independent of the day of the week. The only exception is the Wednesday's return in the Czech Republic in Period 2 and Period 3 that seems to be (at 10% significance level) higher than the yields on Monday. By contrast, the day of the week effect is significantly present in all markets during the Period 4 capturing the crisis stabilization. We discovered that Thursday's and Friday's dummy variables are significant and negative in all markets suggesting that returns on these two days are lower than those of Monday. Negative and significant coefficients were also estimated for Tuesday's and Wednesday's dummies in Hungary; and positive and significant coefficient was revealed for constant in the Poland's model. In the last estimation period, we revealed a common feature of the Czech and Polish markets. In particular, we found significant and positive coefficient of constant and negative coefficients for Tuesday's and Friday's dummy variables. Such results are, however, inconsistent with the usual concept of the day of the week effect that has been empirically revealed in most studies where average Monday returns are usually significantly lower and average Friday returns significantly greater than the average returns for the other days of the week.

In the variance equation of the modified GARCH (1,1) model we allow the conditional variance to change for each day of the week. This is the way how we can examine the day of the week effect in volatility. The highest volatility as suggested by estimated coefficients occurred during the crisis culmination. However, the day of the week effect was not detected in volatility in this period. We did not find any evidence of the effect in the Hungarian stock market volatility at all. Some evidence was revealed in the remaining two markets; however the coefficients are often significant only at 10% level. The statistically significant coefficients are very scattered across the periods and days and, hence, we are not able to come to any robust conclusion on the day of the week effect in the stock returns volatility. The coefficients β and γ, i.e. coefficients of the lagged value of the squared residual term and the lagged value of the conditional variance respectively, are positive and significant. The sum of β and γ is generally lower than one. Thus, our results suggest that conditional variances are positive and not explosive in our sample.

We used Ljung-Box Q statistics and ARCH-LM test for possible autocorrelation and remaining ARCH effects in the estimation residuals. We applied 5, 10, 15 and 20-day lags and the results (not reported here but available upon request) suggest that that autocorrelation and ARCH effects are not present in the corresponding residuals. Thus, our estimations do not suffer from any problem with model specification.

4 Conclusion

In this paper, we estimated the day of the week effect in three Central European stock markets over the course 2006 – 2011. In order to examine the influence of the financial crisis the whole time span was divided into five periods reflecting individual phases of the crisis. Estimations of the modified GARCH-M (1,1) model provided a mixed evidence on presence of the day of the week effect in the stock market returns. Significantly abnormal returns were revealed particularly in the crisis stabilization phase and post-crisis period. However, the results are inconsistent with usual conclusions reported in literature as we revealed significantly greater returns on Monday and significantly lower returns on Thursday and Friday. During the pre-crisis period and phases of initialization and culmination the stock markets did not reflect the day of the week effect and the results for each day do not differ significantly from the other days of the week. We examined the day of the week effect also in volatility of returns but no solid evidence of the effect presence was found.

Acknowledgements

Support of the Czech Science Foundation within the project GAČR 403/11/2073 and Silesian University Grant System within the project SGS/25/2010 is greatly acknowledged.

References

[1] Apolinario, R. M. C., Santana, O. M., Sales, L. J., and Caro, A. R.: Day of the week effect on European stock markets. *International Research Journal of Finance and Economics* **2** (2006), 53-70.

[2] Bubák, V., and Žikeš, F.: Seasonality and the Nontrading Effect on Central-European Stock Markets. *The Czech Journal of Economics and Finance* **56** (2006), 1-7.

[3] Chukwuogor-Ndu, C.: Stock market returns analysis, day-of-the-week effect, volatility of returns: Evidence from European financial markets 1997-2004. *International Research Journal of Finance and Economics* **1** (2006), 112-124.

[4] Fama, E. F.: Efficient capital markets: A review of theory and empirical work. *The journal of Finance* **25** (1970), 383-417.

[5] Högholm, K., Knif, J., and Pynnönen, S.: Common and local asymmetry and day-of-the-week effects among EU equity markets. *Quantitative Finance* **11** (2011), 219-227.

[6] Kiymaz, H., and Berument, H.: The day of the week effect on stock market volatility and volume: International evidence. *Review of Financial Economics* **12** (2003), 363-380.

[7] Lim, S. Y., Mun Ho, C., and Dollery, B.: An empirical analysis of calendar anomalies in the Malaysian stock market. *Applied Financial Economics* **20** (2010), 255-264.

[8] Tonchev, D., and Kim, T. H.: Calendar effects in Eastern European financial markets: evidence from the Czech Republic, Slovakia and Slovenia. *Applied Financial Economics* **14** (2004), 1035-1043.

[9] Yalcin, Y., and Yucel, E. M. The Day-of-the-Week Effect on Stock-Market Volatility and Return: Evidence from Emerging Markets. *Czech Journal of Economics and Finance (Finance a úvěr)* **56** (2006), 258-277.

Using linguistic fuzzy modeling for MMPI-2 data interpretation

Jan Stoklasa[1], Jana Talašová[2]

Abstract. Psychological diagnostics is a crucial psychological activity. It involves systematic acquisition of a large amount of information, data classification, interpretation and final derivation of conclusions. It is desirable to develop systems able to speed up the process and reduce the risk of errors.

This paper considers possibilities of linguistic fuzzy modeling for psychological data analysis and evaluation; perspectives of knowledge transfer are discussed. We describe the process of conversion-symptoms identification based on data provided by MMPI-2 (Minnesota Multiphasic Personality Inventory). Linguistic fuzzy rules are introduced to represent the expert knowledge of the process in three stages – protocol validity, data appropriateness, and "conversion V" obviousness.

Finally, a fuzzy-rule-base aggregation of the three evaluations of a MMPI-2 profile is introduced. Sugeno's fuzzy inference algorithm is used. A fuzzy classification of conversion-symptom presence into three categories (present, possibly present and not present) is performed in this step. The model is implemented in Excel.

Keywords: Linguistic fuzzy modeling, MMPI-2, psychological diagnostics, fuzzy classification.

JEL Classification: C44
AMS Classification: 91E10

1 Introduction

Psychological diagnostics is usually the first step of any psychological intervention. Thorough analysis of all the data obtained by various diagnostic methods (test methods and clinical methods) is needed to gain a valid understanding of client's current state and situation. Unfortunately the amount of data can easily exceed the analytical capacities of a diagnostician. If we take into account that the client's are available in many different forms – linguistic descriptions, pictures, numbers or intervals (results of some diagnostic methods), scales, and even subjective impressions – the aggregation and interpretation of such data becomes a nontrivial task. A diagnostician also needs to be aware of the context and usually employs his expert knowledge and experience in this process.

Once we see the process from this perspective, various problems arise. As the amount of data to be processed and interpreted grows, so does the room for mistakes and misinterpretations. The time consumption of this process is also a point to be considered. Any tool of error elimination that would reduce the time of data processesing and interpretation would be most welcome. In this paper we introduce a linguistic fuzzy model for a particular psychodiagnostic method and present one particular diagnosis that meets these requirements.

In this paper we are presenting a tool for psychologists for conversion-symptoms identification based on the MMPI-2 results. The international classification of diseases – 10th revision (see [11]) denotes dissociative (conversion) disorders as the F44 category. In our application, we narrow the scope and consider only the subcategories F44.4 – F44.7. This group of disorders is called dissociative motor and sensory disorders and can be roughly characterized by neurological symptoms such as numbness or paralysis with no underlying neurological causes.

Minnesota Multiphasic Personality Inventory second revision (MMPI-2) and its previous version are the most widely used psychological inventories for psychopathology assessment worldwide (see [2]) and in the Czech Republic as well (see [4,8]). The MMPI was developed by Hathaway and McKinley [3] in 1940, later Netík adapted the second revision into Czech in 2002 (see [4]). This method was chosen for our research because it provides various means of validity assessment, is widely used, and much research has been done since 1940

[1] Palacky University in Olomouc/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, , 77146 Olomouc, Czech Republic, jan.stoklasa@upol.cz;
Palacky University in Olomouc/Faculty of Arts, Dept. of Psychology, Křížkovského 10, 771 80, Olomouc, Czech Republic, dzoni@seznam.cz.
[2] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz.

concerning its validity and reliability. When a client answers the 567 items of MMPI-2, the method then provides 567 answers to these items, 10 clinical scales scores, more than 7 validity scale scores and over 70 content and supplementary scales scores. In this paper we propose a linguistic fuzzy model to help the diagnostician recognize the presence of conversion symptoms. The process uses data provided by MMPI-2 and diagnostic criteria suggested by Greene in [2]. It also draws on diagnostician's expert knowledge and experience (these are implemented in the fuzzy rules). Requirements imposed on the model by psychologists were: a) to suggest interpretation of the data in the framework of conversion-symptoms identification, b) comprehensible environment and results; and c) results justification (to provide more than just the highest level of data aggregation). We also wanted to consider possible research and knowledge transfer possibilities.

2 Used mathematical apparatus

In this paper we use the concept of fuzzy sets and linguistic fuzzy modeling introduced by Zadeh in [12, 13]. Let U be a nonempty set (the universe). 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 $A(\cdot)$ is called a membership function of the fuzzy set A. R will denote the set of all real numbers.

The height of a fuzzy set A is a real number $\mathrm{hgt}(A) = \sup_{x \in U} \{A(x)\}$. A union of fuzzy sets A and B on U is a fuzzy set $A \cup B$ on U with a membership function for all $x \in U$ given by $(A \cup B)(x) = \max\{A(x), B(x)\}$. An intersection of fuzzy sets A and B on U is a fuzzy set $A \cap B$ on U with a membership function for all $x \in U$ given by $(A \cap B)(x) = \min\{A(x), B(x)\}$. Let *A* be a fuzzy set on *U* and *B* be a fuzzy set on *V*. Then the Cartesian product of *A* and *B* is the fuzzy set $A \times B$ on $U \times V$ with the membership function defined for all $(x, y) \in U \times V$ by $(A \times B)(x, y) = \min\{A(x), B(y)\}$.

A fuzzy number C is a fuzzy set on R satisfying three conditions: 1) the kernel of C, $\mathrm{Ker}(C) = \{x \in \mathsf{R} \mid C(x) = 1\}$, is a nonempty set; 2) all the α-cuts of the fuzzy set C, $C_\alpha = \{x \in \mathsf{R} \mid C(x) \geq \alpha\}$, are closed intervals for all $\alpha \in (0,1]$; and 3) the support of the fuzzy set C, $\mathrm{Supp}(C) = \{x \in \mathsf{R} \mid C(x) > 0\}$, is bounded.
If the $\mathrm{Supp}(C) \subseteq [a,b]$, we call C a fuzzy number on the interval $[a,b]$. The family of all fuzzy numbers on the interval $[a,b]$ is denoted by $F_N([a,b])$.

Let $A_1, A_2, ..., A_n \in F_N([a,b])$, then we say that $A_1, A_2, ..., A_n$ form a fuzzy scale on $[a,b]$ if these fuzzy numbers form a Ruspini fuzzy partition (see [5]) on $[a,b]$ (i.e. $\sum_{i=1}^{n} A_i(x) = 1$, for all $x \in [a,b]$) and are numbered in accordance with their ordering.

A linguistic variable is a quintuple $(X, T(X), U, M, G)$ where X is the name of the linguistic variable, $T(X)$ is the set of its linguistic values, U is the universe, on which the mathematical meanings of the linguistic terms are defined, G is a syntactical rule (grammar) for generating linguistic terms from $T(X)$, and M is a semantic rule (meaning), that assigns to every linguistic term $\mathcal{A} \in T(X)$ its meaning $M(\mathcal{A})$ as a fuzzy set on U. A linguistic variable $(X, T(X), [c,d], M, G)$, $T(X) = \{\mathcal{A}_1, \mathcal{A}_2, ..., \mathcal{A}_n\}$ is called a linguistic scale if $A_1 = M(\mathcal{A}_1), A_2 = M(\mathcal{A}_2), ..., A_n = M(\mathcal{A}_n)$ are fuzzy numbers forming a fuzzy scale on $[c,d]$.

Let $(X_j, T(X_j), [c_j, d_j], M_j, G_j)$, $j = 1, ..., m$, and $(Y, T(Y), [c,d], M, G)$, be linguistic variables (usually linguistic scales). Let $\mathcal{A}_{ij} \in T(X_j)$ and $M_j(\mathcal{A}_{ij}) = A_{ij} \in F_N([c_j, d_j])$, for all $i = 1, ..., n$, $j = 1, ..., m$. Let $\mathcal{B}_i \in T(Y)$ and $M(\mathcal{B}_i) = B_i \in F_N([c,d])$, $i = 1, ..., n$. Then the following scheme is called a linguistically defined function (base of fuzzy rules).

$$
\begin{array}{l}
\text{If } X_1 \text{ is } \mathcal{A}_{11} \text{ and } \ldots \text{ and } X_m \text{ is } \mathcal{A}_{1m}\text{, then } Y \text{ is } \mathcal{B}_1. \\
\text{If } X_1 \text{ is } \mathcal{A}_{21} \text{ and } \ldots \text{ and } X_m \text{ is } \mathcal{A}_{2m}\text{, then } Y \text{ is } \mathcal{B}_2. \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\text{If } X_1 \text{ is } \mathcal{A}_{n1} \text{ and } \ldots \text{ and } X_m \text{ is } \mathcal{A}_{nm}\text{, then } Y \text{ is } \mathcal{B}_n.
\end{array}
\tag{1}
$$

Using the approach of Sugeno & Yasukawa [7], we consider the rule base (1) and an m-tuple of crisp input values (a_1, a_2, ... , a_m). By entering these observed values into the linguistically defined fuzzy function, we get the output $b^{SY} = \left(\sum_{i=1}^{n} h_i \cdot b_i\right) / \left(\sum_{i=1}^{n} h_i\right)$, where $h_i = \min\{A_{i1}(a_1), A_{i2}(a_2), ..., A_{im}(a_m)\}$ and b_i is the center of gravity of B_i, defined by the formula $b_i = \int_{y \in V} B_i(y) \cdot y \, dy / \int_{y \in V} B_i(y) \, dy$, i=1,...,n. The advantage of this approach is that the linguistic terms B_i, i=1,...,n, can be used for the linguistic description of the output. If b^{SY} lies in the intersection of supports of two neighboring fuzzy numbers $B_{i_k}, B_{i_{k+1}}$ then the output b^{SY} can be characterized as being $B_{i_k}(b^{SY})$ percent of B_{i_k} and $B_{i_{k+1}}(b^{SY})$ percent of $B_{i_{k+1}}$.

The same can be also done by using Takagi-Sugeno fuzzy controller (presented in [9]), where the consequent parts of the rules are modeled by constant functions. The fuzzy controller of Takagi & Sugeno [9] considers a rule base in the following form:

$$\begin{array}{l} \text{If } x_1 \text{ is } A_{11} \text{ and ... and } x_m \text{ is } A_{1m}\text{, then } y = g_1(x_1, ..., x_m). \\ \text{If } x_1 \text{ is } A_{21} \text{ and ... and } x_m \text{ is } A_{2m}\text{, then } y = g_2(x_1, ..., x_m). \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \text{If } x_1 \text{ is } A_{n1} \text{ and ... and } x_m \text{ is } A_{nm}\text{, then } y = g_n(x_1, ..., x_m). \end{array} \tag{2}$$

Here $x_1, x_2, ..., x_m$ are the input variables, $A_{i1}, A_{i2}, ... A_{in}$ are fuzzy numbers on intervals $[c_j, d_j]$ for all j=1, ..., m, and $y = g_i(x_1, ..., x_m)$ describes the control function for the i-th rule, i=1, ..., n. Let us consider again an m-tuple of crisp input values $a_1, a_2, ..., a_m$, $a_j \in [c_j, d_j]$ for all j=1,2,...,m. The output of Takagi-Sugeno fuzzy controller is computed as $b^{TS} = \sum_{i=1}^{n} h_i \cdot g_i(a_1, a_2, ..., a_m) / \sum_{i=1}^{n} h_i$, where h_i, i=1,2,...,n, are defined in the same way as in the Sugeno-Yasukawa algorithm. If $y = g_i(x_1, ..., x_m) = b_i$, $b_i \in \mathbb{R}$ for all i=1,2,...,n, we speak about the Sugeno fuzzy controller; its input-output function is in the form $b^S = \sum_{i=1}^{n} (h_i \cdot b_i) / \sum_{i=1}^{n} h_i$. If we take b_i as representatives of $B_i = M(B_i)$, i=1,2,...,n, that were used in the fuzzy rule base (1), we get the same using the Sugeno algorithm as before by the Sugeno-Yasukawa algorithm (Sugeno & Yasukawa represent B_i's by their centers of gravity, in the following text we will use elements of kernels of triangular fuzzy numbers B_i).

3 Methods

Greene in [2] and Netík in [4] suggest diagnostic criteria for detecting the presence of conversion symptoms. We combine these with the expert knowledge of one skilled diagnostician to construct a four phase linguistic fuzzy model that meets all the requirements given in the introduction section. The expert, drawing on his experience with MMPI-2 and conversion patients, softened the criteria formulated in Greene [2], thus creating a linguistic description of appropriate scale values which are represented by fuzzy numbers in the model. We have identified four phases of MMPI-2 data assessment.

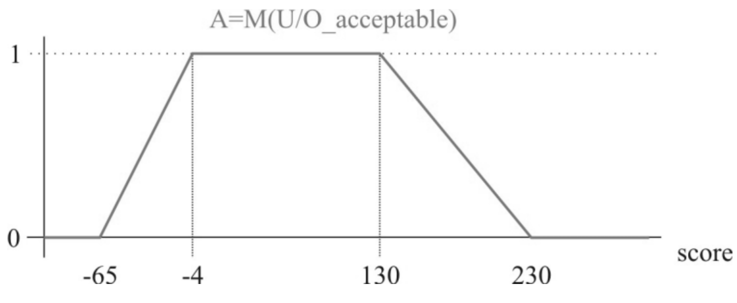


Figure 1 A fuzzy number representing the meaning of "acceptable scores" of the U/O reporting validity scale .

Validity assessment is based on 7 validity scales (?, TRIN, VRIN, U/O reporting, L, F, Fb – see [2] for details). For each of these scales we define the meaning of the linguistic term "acceptable scores" by a fuzzy number (the universe of this fuzzy number is given by all the possible values of the respective scale) – Figure 1 provides an example. Validity rate of a particular MMPI-2 protocol is then determined through (3), where $?', TRIN', ..., Fb'$ are fuzzy singletons representing the respective scale scores. Validity rate is a real number from $[0,1]$ and the following holds: $\text{validity rate} = 1 - \text{invalidity rate}$.

$$\text{hgt}\left\{\left(?' \times TRIN' \times ... \times Fb'\right) \cap \left(M(?\_acceptable) \times M(TRIN\_acceptable \times ... \times M(Fb\_acceptable)\right)\right\} \tag{3}$$

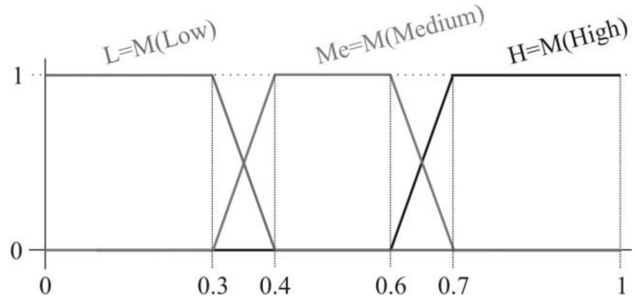


Figure 2 Fuzzy scale for validity rate interpretation.

In order to describe the resulting validity rate linguistically, we define a linguistic scale (see Figure 2). This way we are able to interpret for example the validity rate 0.34 as being 60% low and 40% medium.

The next step of the diagnostic process is to assess the MMPI-2 protocol “at first sight”. We set up a one-rule “filter” that distinguishes between MMPI-2 protocols that can indicate converse symptoms and those that are not supporting such diagnosis at all. This discrimination is based on relationships among the 10 clinical scales scores. We call this step **data appropriateness determination**. Again, acceptable relationships are described linguistically and a fuzzy number meaning is assigned to each of them. The resulting appropriateness rate is from $[0,1]$ and can be interpreted linguistically using the fuzzy scale from Figure 3.

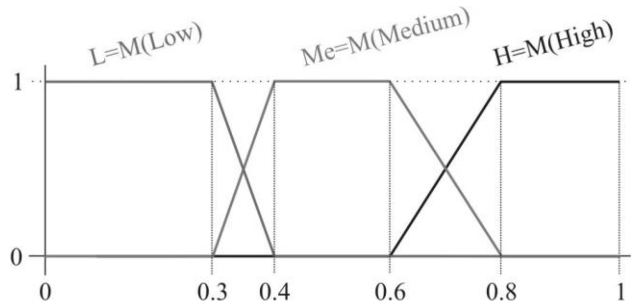


Figure 3 Fuzzy scale for appropriateness and converse V obviousness rate interpretation.

The most specific identification of converse symptoms (see [2]) is the so called “Converse V”, which is such a configuration of three clinical scales scores (Hypochondrias (*Hs*), Depression (*D*) and Hysteria (*Hy*)), where both *Hs* and *Hy* scores are above the *D* score, all are “clinically significant” and none of the scores Hs and Hy is “too large”. In other words the plot of these scales scores should resemble the shape of the letter V. We have transformed this description into the following:

1. (*Hs – D*) is *significant* and (*Hy – D*) is *significant*.
2. (*Hs – D*) is *very_significant* or (*Hy – D*) is *very_significant*
3. *Hs_Hy_ratio* is *acceptable,* where

$$Hs\_Hy\_ratio = \begin{cases} \dfrac{\max\left(Hs-D, Hy-D\right)}{\min\left(Hs-D, Hy-D\right)}, & \text{if } \min\left(Hs-D, Hy-D\right) \neq 0, \\ 100 & \text{else.} \end{cases}$$

Figure 4 shows obvious and indistinct “converse V” shape described by these three conditions. Again an **obviousness rate** from $[0,1]$ is obtained and can be linguistically interpreted using the fuzzy scale in Figure 3.

Validity rate	Appropriateness rate	Converse V obviousness rate	Conversion symptoms presence
High	High	High	Present
High	High	Medium	Present
High	Medium	High	Present
...	...	...	...
Medium	Medium	High	Possibly present
...	...	...	...
Anything	Anything	Low	Not present

Table 1 A part of the rule base for conversion symptoms presence determination.

From the validity, appropriateness and converse V obviousness rates we can now determine, whether conversion symptoms are present or not. We have 11 fuzzy rules available (see Table 1, where the meaning of the linguistic term “Anything” is described by the fuzzy set $AN = M(Anything)$; $AN(x) = 1$ for all $x \in U$; U is the universe on which the meanings of linguistic terms of the respective linguistic variable are defined on). With the three real values of the validity, appropriateness and converse V obviousness rates as inputs, we use a modified Sugeno & Yasukawa approach to fuzzy control to derive outputs. We see this as a fuzzy classification problem. We have 3 classes (Not_present – nr. 0, Possibly_present – nr. 1, Present – nr. 2). Our modification is that the consequent parts of the rules are represented not by centers of gravity, but by the number of the class. Numbers of the classes form a cardinal scale. This allows us to perform fuzzy classification (we accept partial membership to two neighboring classes) and obtain a number from $[0, 2]$ that can again be interpreted linguistically.

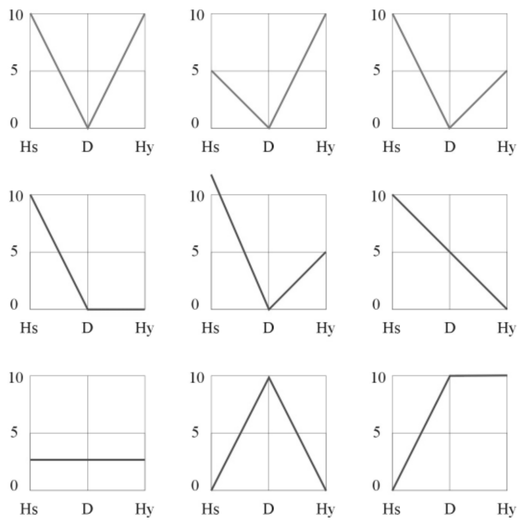


Figure 4 Plots of possible Hs, D and Hy scores configurations. The top row depicts examples of obvious converse V shapes (obviousness rate = 1), the middle and bottom rows depict examples of indistinct converse V shape (obviousness rate = 0).

4 Results

We have presented a linguistic fuzzy model for the purposes of psychological diagnostics. The above-described linguistic fuzzy model has been implemented in MS Excel (see Figure 5). It uses 17 MMPI-2 scale scores (10 clinical scales and 7 validity scales) as inputs and provides 1 overall output – it determines, whether the conversion symptoms are i) present, ii) possibly present, or iii) not present. Results on lower levels of information aggregation are also available: protocol validity rate, data appropriateness rate and converse V obviousness rate. Finally, to fully support the justification of the diagnosis, important segments of all the antecedent parts of linguistic rules and their fulfillment rates are also provided. The model reflects the experience and knowledge of one particular expert diagnostician as well as the diagnostic and interpretational guidelines contained in [2,4]. At present we are testing the model on 250 MMPI-2 protocols.

5 Discussion

We have managed to successfully capture expert knowledge and present it in a form that can be understood by psychologists not familiar with linguistic fuzzy modeling. Although our approach introduces to the process some level of uncertainty, which is inherent in the linguistic description of expertly defined rules, the results of the testing seem promising. There are still some small discrepancies between the results of our model and those obtained by strictly following the criteria e.g. in [2] or [4]. These may have many causes, including the need for general revision of some diagnostic recommendations, new norms and verification of validity of these recommendations for Czech population, and even specificity of our sample of 250 MMPI-2 protocols. At present we are fine-tuning the model to minimize these discrepancies.

Once the fine-tuning phase is completed, the presented model may prove useful in various areas. The practical diagnostic application is obvious. However, inclusion of expert knowledge into a formalized process and presenting the results in an intelligible way is the first step in knowledge transfer based on linguistic fuzzy modeling. Teaching and professional training programs may benefit from such a tool as well. We may also consider its use in research – formalized and software implemented expert knowledge (or experience) can be tested and verified on large samples. Our work suggests that interdisciplinary applications of linguistic fuzzy modeling are not only possible, but even desirable.

Input data

"?"	L	F	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma	Si	O/U	Fb	VRIN	TRIN	**Converse symptoms:**	**0,187**
0	62	46	60	57	71	60	53	53	54	52	50	49	-42	43	47	53		

Validity verification		Rule fulfillment
1.	"?" score is **acceptable**.	1
2.	TRIN score is **acceptable**.	1
3.	VRIN score is **acceptable**.	1
4.	U/O reporting score is **acceptable**.	0,37705
5.	L score is **acceptable**.	1
6.	F score is **acceptable**.	1
7.	Fb score is **acceptable**.	1
	Validity rate	**0,37705**

Data appropriateness determination		Rule fulfillment
1.	(D-Sc) is **big_enough**.	1
2.	(Pt-Mf) is **acceptable** and (Pt-Pa) is **acceptable** and (Pt-Ma) is **acceptable**.	1
3.	D is **elevated**.	0,4
4.	Hy-max{Pd,Mf,Pa,Pt,Sc,Ma,Si} is **clearly_positive**.	1
	Appropriateness rate	**0,4**

Converse V obviousness assessment		Rule fulfillment
1.	(Hs-D) is **significant** and (Hy-D) is **significant**.	**0,6**
2.	(Hs-D) is **very_significant** or (Hy-D) is **very_significant**.	**1**
3.	Hs_Hy_ratio is **acceptable**.	**0**
	Obviousness rate	**0**

Figure 5 MS Excel implementation results.

Acknowledgements

The presented research was supported by the grant *PrF_2011_022 Mathematical models and structures* obtained from the Internal Grant Agency of the Palacky University in Olomouc.

References

[1] Dubois, D., Prade, H. (Eds.).: *Fundamentals of fuzzy sets*. Kluwer Academic Publishers, Dordrecht, 2000.

[2] Greene, R. L.: *The MMPI-2: An interpretive manual*. Allyn and Bacon, Boston, 2000.

[3] Hathaway, S. R., McKinley, J. C.: A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of psychology*, 10, 249-254, 1940.

[4] Netík, K.: *The Minnesota Multiphasic Personality Inventory - 2: první české vydání*. Testcentrum, Praha, 2002.

[5] Ruspini, E.: A new approach to clustering. *Inform. Control*, 15, 1969, pp. 22-32.

[6] Sugeno, M.: An introductory survey on fuzzy control. *Information Sciences*, 36, 1985, pp. 59-83.

[7] Sugeno, M. & Yasukawa, T.: A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on fuzzy systems*, 1 (1), 1993, pp. 7-31.

[8] Svoboda, M., Řehan, V. et al. *Aplikovaná psychodiagnostika v České republice*. Psychologický ústav FF MU v Brně, Brno, 2004. [Applied psychodiagnostics in the Czech Republic]

[9] Takagi, T., & Sugeno, M.: Fuzzy identification of systems and its application to modeling and control. *IEEE Transactions on systems, man and cybernetics.*, 1 (15), 1985, pp. 116-132.

[10] Talašová, J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování. Univerzita Palackého v Olomouci, Olomouc, 2003. [Fuzzy methods of multicriteria evaluation and decision making]

[11] World health organisation.: *Mezinárodní klasifikace nemocí - 10. revize: Duševní poruchy a poruchy chování* (3rd edition). Psychiatrické centrum Praha, Praha, 2006. [International Classifications of Diseases – 10th revision: Mental and behavioural disorders (3rd edition)]

[12] Zadeh, L. A.: Fuzzy Sets. *Inform. Control*, 8, 1965, pp. 338-353.

[13] Zadeh, L. A.: The concept of linguistic variable and its application to approximate reasoning. *Information sciences*, Part 1: 8, 1975, pp. 199-249, Part 2: 8 1975, pp. 301-357, Part 3: 9 1975, pp. 43-80.

Wheat production function analysis using panel data

Tereza Suchánková[1]

Abstract. The paper focuses on the analysis of wheat production function. The significance of classical production factors as well as climate and fertilizer factors was tested. The analysis consists on application of classical Cobb-Douglas production function and its comparison to the constant elasticity of substitution (CES) production function. First, Kmenta's linear approximation to the 2-input CES function was applied and afterwards also the Taylor n-input linear approximation was used and tested. The parameters of production function were estimated using available panel data from 14 Czech regions in the period from 2000 to 2009 by OLS and GLS methods – depending on the form of individual effects used in the model. The convenience of the application of fixed or random individual effects was tested and discussed. One of the conclusions from the analysis is the ascertainment that the Cobb-Douglas production function is not applicable to agricultural plant products – especially wheat.

Keywords: CES production function, agriculture, panel data.

JEL Classification: C2, C3, C5
AMS Classification: 90C15

1 Introduction

Production or yield of agricultural crops is strongly influenced by weather factors. For this reason it is advantageous to capture its considerable influence on the harvest into the production function. The influence of climatic factors in the farm area or in the region (mainly represented by rainfall and average temperature for a given period) on the real harvest will be tested.

Wheat production function (generally production function of any agricultural crop) should, however, primarily depend on the amount of available land as another exogenous variable. The influence of capital, i.e. consumption of fixed capital - machinery and equipment will be tested as the second production factor. To analyze the production function we expect to include the non-objectified technological progress to capture of output growth independently of the production factors. We will dynamize the production function by this way.

The thesis will analyze the relevant explanatory variables and then the most convenient expression of the resulting wheat production function. The analysis will target two basic forms of production function - with a unit elasticity of substitution and constant elasticity of substitution, i.e. Cobb - Douglas production function (PFCD), respectively the production function with constant elasticity of substitution (PFCES). The analysis will focus on testing the suitability of the use of CES function for two explanatory variables and the possibility of using multifactorial CES function in the analysis of wheat production. Both types of production function will be analyzed primarily in their linearized form because of the application of panel data. We use and compare Kmenta's linear approximation for two inputs and Taylor approximation to the n-input CES function. The various expressions of production function will be tested in terms of their degree of conformity with reality and of course the statistical significance of the estimated parameters. Using existing panel data we will include the influence of individual differences of each region in the model, either in form of fixed or random effects.

2 Theoretical background

Production function allows to some extent to explain the output value generated by either firm, industry or the whole economy based on diverse combinations of factors determining the existing technology. Detailed explanations of the Cobb – Douglas production function used in the analysis can be found in [5]. The same literature explains the principles of the CES production function, including well-known Kmenta's approximations for two inputs. Detailed description can be found in [6]. Description of the Taylor approximation of multifactorial CES function can be found in [3] and [4]. Theoretical background and estimation methods for working with panel data can be found for example in [2] or [7] – fixed and random effects are sufficiently detailed.

Properties of wheat not only for its cultivation are well described in [1].

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech Republic, tereza.suchankova@vse.cz.

3 Estimation and production function analysis

Wheat production function model assumes that dependent variable, ie the harvest of wheat in metric tons, is affected by three types of explanatory variables. The first type of variable are the traditional factors of production - land (sown area in hectares), than the capital, expressed as consumption of fixed capital (machinery and equipment) and labor factor. In the case of wheat production function the influence of labor is neglected because of poor predictability of statistical data due to the type and character of observed data. While the influence of capital on the wheat harvest, which could be called into question in a similar manner, will be tested because we assume that if the wheat is in the Czech Republic the most cultivated crop in a long term and covers almost 50% of the total area of arable land (from ČSÚ statistics), a significant portion of capital will be consumed in its production. The second type of variable is represented by the climate variables, the average month temperature and average month rainfall. The third type expresses values that enhance the yield of the crop, i.e. industrial and farm fertilizers and pesticides.

To estimate the production function we utilize the available annual panel data reflecting the differences in the various regions of the Czech Republic for the period between 2000 and 2009.

Table 1 shows depending of the wheat harvest explained by the explanatory variables in the form of correlation coefficients. The explanatory variables in the table below correspond to the factors mentioned above in the text (L – land, C – capital, T3-T7 – average month temperature from March to July and S3-S7 – average month rainfall also from March to July, the period between wheat sprouting and harvesting). The correlation coefficients in the table below represent the values for the three selected regions to illustrate the various impacts of explanatory variables on the wheat harvest.

Region	L	C	T3	T4	T5	T6	T7	S3	S4	S5	S6	S7
Středočeský	0,53	0,70	-0,24	-0,05	-0,58	-0,72	-0,40	0,14	0,29	-0,01	0,44	0,11
Jihočeský	0,70	0,68	-0,44	-0,17	-0,46	-0,72	-0,53	0,42	0,03	-0,20	0,17	0,13
Jihomoravský	0,80	0,66	-0,39	-0,13	-0,58	-0,73	-0,35	0,44	0,45	0,38	0,16	0,15

Table 1 Correlation coefficients between wheat harvest and the explanatory variables

The values of correlation coefficients presented in Table 1 indicate the possibility of existence of some dependency between the wheat harvest and explanatory factors. The influence of labor, capital and some climatic factors may be important.

The following subsection will focus on estimating the wheat production function based on assumptions of Cobb – Douglas production function. The other types of production function for modeling wheat harvest will follow. In all of the following models was assumed the influence of non-objectified technological progress which is expressed through a proxy variable time. The econometric software EViews 5.0. and Gretl were used to estimate the parameters and statistical characteristics.

3.1 Application of Cobb – Douglas production function

The influence of the aforementioned explanatory variables was first expressed using the classic Cobb - Douglas production function (PFCD). To capture the individual effects caused by the different nature of each region the form of fixed and then random effects was used in the model.

As stated above, we assume that the wheat harvest is influenced by a land production factor, capital consumption and climate factors such as average monthly temperature and rainfall in the region. Given that these variables together are in sum 12 and given the length of time series and the number of regions, if we would include them all at once into the model, the estimated parameter of variables that really have an important impact on the dependent variable would have not too significant t-test or may be even insignificant. For this reason, we will examine the impact of these variables independently. First, the statistical significance and explanatory power of classical production factors (in this case land and capital) will be tested and then the effect of fluctuations of the average monthly temperature and precipitation. The results from this partial analysis would determinate which variables are really important to explain the variability of wheat harvest and which are less. The following equation shows the well-known PFCD with two production factors.

$$Y_{it} = a_i e^{gt} L_{it}^{\alpha} C_{it}^{\beta} e^{u_{it}}, \qquad (1)$$

where Y_{it} is a harvest of wheat in the region i and in the time t, a_i represents the level of achieved technology in the region i, g the non-objectified technological progress – the parameter for the proxy variable time t, L_{it} is the sawn land in hectares in the region i and in the time t and C_{it} is the consumption of fixed capital in the region i and the time t. The coefficients α and β are the elasticities of output (harvest) with respect to the land input or to the capital input respectively. The sum of these elasticities gives the information about the returns to scale. The u_{it} is the stochastic disturbance term. For more information about the individual effects of each region, it is useful to log-transform the model in order to obtain the form linearized in parameters. Then, we have:

$$\ln Y_{it} = \ln a_i + gt + \alpha \ln L_{it} + \beta \ln C_{it} + u_{it}. \tag{2}$$

The Table 2 shows the OLS estimation results and the values reveal that all estimated parameters are significantly different from zero at 1% confidence level. On the other hand, the coefficients for production factors land and capital do not satisfy the basic assumptions of PFCD. Their value doesn't lie in the interval (0, 1). Coefficient of determination and its adjusted version is very high and close to one.

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
t	0,026056	0,002119	12.2936	0.0000***
$\ln(L_{it})$	1,184706	0,047388	24.9999	0.0000***
$\ln(C_t)$	1,447745	0,124945	11.5871	0.0000***
$\ln(a_i)$ – constant	-65,55102	4,468126	-14.6708	0.0000***
R-squared	0,995570	Adjusted R-squared		0.994994

Table 2 The OLS estimation results – PFCD – land and capital

The Table 3 reflects the results of the same OLS estimation of PFCD but only with the average monthly temperature variables. The expression of PFCD is in this case similar to the equation (1), in linearized form to the expression (2). Instead of variables for land and capital, there are 5 temperature variables from March to July.

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
t	0,023705	0,005702	4,1576	0,0001***
$\ln(T3_{it})$	-0,023798	0,032034	-0,7429	0,4590
$\ln(T4_{it})$	-0,34205	0,081415	-4,2013	0,0001***
$\ln(T5_{it})$	-0,96309	0,130278	-7,3925	0,0000***
$\ln(T6_{it})$	-0,68832	0,202096	-3,4059	0,0009***
$\ln(T7_{it})$	-0,70539	0,064050	-11,0131	0,0000***
$\ln(a_i)$ – constant	-27,8657	11,65020	-2,3919	0,0183**
R-squared	0.984964	Adjusted R-squared		0,982583

Table 3 The OLS estimation results – PFCD – average monthly temperature

The values in the Table 3 show that almost all estimated parameters are significantly different from zero at 1% confidence level except the variables designing the average monthly temperature in March which is not significant at all. Unfortunately, every temperature variable breaks the basic assumptions of PFCD because their values lie below zero and so they affect the wheat harvest negatively. Coefficient of determination and its adjusted version is very high and close to one like in the precedent case.

The results from estimation for the PFCD with average monthly rainfall variables are similar; there is no need to state them.

For all of these estimations we used the fixed effects approach. The aim of this partial analysis was to determinate the importance of different variables and the usability of PFCD for modeling wheat harvest. Discussion about the application of fixed or random effects will take place in the next subsection.

As shown above, the selected production and climatic factors very well explain the fluctuations in the harvest. Unfortunately, we can also conclude that the use of the PFCD to estimate wheat harvest is inappropriate, almost all of the explanatory variables violate basic assumptions of this production function. All this implies that *wheat production function has not the character of unit elasticity of substitution*. The suitability and applicability of production function with constant elasticity of substitution (PFCES) will be tested in the following subsection.

3.2 Application of production function with constant elasticity of substitution

The aim of the analysis in this subsection is to examine the appropriateness and applicability of PFCES to model wheat harvest. First, we focus on PFCES for 2 inputs - the linear approximation was formulated by Kmenta in [6] and then the general PFCES for n inputs whose linear approximation formulated Hoff in [3]. Both linearizations are based on Taylor series expansion.

CES production function for 2 inputs variables

CES production function is more general than PFCD. The elasticity of substitution may not be the unit, only need to be constant. This function assumes returns to scale not necessarily equal to one. To capture the individual effects is necessary to use the linearized form of the function. Given the issue and the number of relevant explanatory variables CES function for two inputs is very restrictive, but we will try to demonstrate the function application suitability using a limited set of inputs. First, as the PFCD case, to estimate the parameters it will be used only the traditional factors of production. In this case, the CES function has the following form:

$$Y = c\left[\gamma L^{-\rho} + (1-\gamma)C^{-\rho}\right]^{-r/\rho} e^{u}, \tag{3}$$

where c is the parameter of efficiency of the production process, γ is the distribution parameter depending on the units of both factors, r is the degree of homogeneity and ρ is the substitution parameter. The parameters can be estimated for example by nonlinear least squares.

As mentioned above, to incorporate the fixed or random individual effects, the linear form of the function is needed. Linear approximation is given by Kmenta in [6]:

$$\ln Y_{it} = \beta_1 + \beta_2 \ln L + \beta_3 \ln C + \beta_4 \left[\ln\left(\frac{L}{C}\right)\right]^2 + u \tag{4}$$

Using the estimated parameters β the estimates of the initial parameters of PFCES can be obtained.

$$c = e^{\beta_1}, \quad r = \beta_2 + \beta_3, \quad \gamma = \frac{\beta_2}{\beta_2 + \beta_3}, \quad \rho = \frac{-2\beta_4}{\beta_2 \beta_3}(\beta_2 + \beta_3). \tag{5}$$

After integration of non-objectified technological progress and fixed individual effects, the linearized 2- input CES function take the following form:

$$\ln Y_{it} = \alpha_i + \beta_1 + gt + \beta_2 \ln L_{it} + \beta_3 \ln C_{it} + \beta_4 \left[\ln\left(\frac{L_{it}}{C_{it}}\right)\right]^2 + u_{it}, \tag{6}$$

where α_i is a deviation from the constant, representing the influence of each individual region i, β_1 is a constant, g the non-objectified technological progress – the parameter for the proxy variable time t, β_2 and β_3 are the elasticity coefficients of the explanatory variables of land L and capital C and finally the coefficient β_4 expresses the elasticity of correction to the first part of the model which corresponds to the linearized form of PFCD. The compliance of the random component with Gauss - Markov assumptions is expected. The following Table 4 shows the estimates of coefficients and values of basic statistical characteristics.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
t	0,02641	0,00199	13,2670	0,0000***
$\ln(L_{it})$	1,08485	0,02597	41,776	0,0000***
$\ln(C_{it})$	1,55337	0,04851	32,020	0,0000***
$\ln(L_{it}/ C_{it})$^2	0,04861	0,01001	4,8572	0,0000***
β_1 – constant	-66,3335	4,20265	-15,784	0,0000***
R-squared	0,995808	Adjusted R-squared		0,995224

Table 4 The OLS estimation results – PFCES – land and capital

On the basis of the estimated coefficients (all statistically significant at 1% confidence level) the coefficients of the non-linear production function can be calculated. The translog approximation may be used only if the parameter ρ is around zero. In our case ρ= -0,076 so the linearization may be used. The value of ρ induces that the elasticity of substitution is higher than 1. The coefficient β_4 (0,04861) is statistically significant - if not, it would be more appropriate to use the classic PFCD. This production function can be characterized by increasing returns to scale r = 2,64. The Table 4 shows that wheat production is explained by the model for almost 100%. Influence of the non-objectified technological progress constitutes about 2.6% growth of wheat production per year.

The existence of important differencies between the regions in a form of different intercepts was tested by F – test. The null hypothesis assumes that all groups are characterized by one intercept. The test statistics had a value F(13,122) = 9,1088 with p-value = 7,277e-013 so the null hypothesis can be refused. There exist the important differencies between the regions – the model with fixed individual effects is convenient. The incorporating the individual character of each region by fixed effects gave more convincing results than the random effects (Hausman test showed that in random effects model the parameters estimated by GLS were not consistent). The following Table 5 shows fixed effects for each region.

Region	Fixed effect α_i	Region	Fixed effect α_i
Jihočeský	-0,212	Pardubický	0,003
Jihomoravský	-0,245	Plzeňský	-0,133
Karlovarský	0,187	Hl. m. Praha	0,473
Královéhradecký	0,035	Středočeský	-0,369
Liberecký	0,198	Ústecký	-0,052
Moravskoslezský	0,046	Vysočina	-0,136
Olomoucký	0,078	Zlínský	0,125

Table 5 Fixed individual effects - regions

The Wald test of heteroscedasticity with p-value = 0,3428 cannot reject the null hypothesis about the homoscedasticity of errors. Based on these facts we can conclude that the use of CES function to explain the wheat harvest by production factors land and capital is appropriate. Suitability of CES function to estimate the wheat harvest on the evolution of climatic variables is tested below. As an example, the combination of average temperature in June and average rainfall in July as explanatory variables is given.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
t	0,02402	0,00278	8,6467	0,0000***
$\ln(T6_{it})$	-0,56791	0,10288	-5,5200	0,0000***
$\ln(S7_{it})$	0,11564	0,03573	3,2364	0,0016***
$\ln(T6_{it}/ S7_{it})$^2	-0,03331	0,01349	-2,4684	0,0150**
β_1 – constant	-45,3819	5,59988	-8,1041	0,0000***
R-squared	0,587745	Adjusted R-squared		0,530299

Table 6 The OLS estimation results – PFCES – average temperature in June and average rainfall in July

As seen in Table 6 all estimated parameters are statistically significant at 1% confidence level except β_4 (5% confidence level). The parameter ρ is 0,459, its positive value indicates the elasticity of substitution less than 1. The Table 6 shows that wheat harvest is explained in 59% by the model. Influence of non-objectified technological progress constitutes about 2.4% growth of wheat production per year. The parameter β_2 represent the negative influence of higher temperatures in June to wheat harvest. Contrary, the rainfall in July had a positive effect.

The value of F-test (7,2351) with p-value = 2,559e-10 demonstrates the existence of important differencies between the regions in a form of different intercepts - the model with fixed individual effects is convenient. Like in the previous model, the fixed individual effects gave better results than the random effects – Hausman test proved that in random effects model the parameters estimated by GLS aren't consistent. The Wald test of heteroscedasticity with p-value = 0,8340 cannot reject the null hypothesis about the homoscedasticity of errors. All the results mentioned above show on *the appropriateness of the CES production function to model the wheat harvest*. The CES function with climate variables demonstrates the need to incorporate more than 2 inputs. In the next subsection the multivariable CES function will be studied.

CES production function for n inputs variables

General form of CES production function for n inputs is given by Kmenta in [6]:

$$Y = \gamma \left(\sum_{k=1}^{n} \beta_k x_k^{-\rho} \right)^{-r/\rho}; \quad \sum_{k=1}^{n} \beta_k = 1, \tag{7}$$

where x_k is an input variable, $k = 1,..,n$. Hoff in [4] states that "when the CES function with n inputs is approximated by a first-order Taylor approximation around $\rho = 0$, the results is still translog function as in the two-input case, but the restrictions on the translog parameters are not equal to the two-input restrictions, but an extension of these". The outline of the proof is in [3]. The translog form for ρ in neighborhood of zero is given below:

$$\ln Y = \ln \gamma + \sum_{i=1}^{n} \alpha_i \ln x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} \ln x_i \ln x_j , \tag{8}$$

where α_i and α_{ij} are the parameters which obey the following restrictions:

$$\alpha_k = r\beta_k , \quad \sum_{k-1}^{n} \alpha_k = r, \quad \alpha_{ij|i \neq j} = \frac{-2\alpha_{ii}}{1 + \sum_{´k \neq (i,j)} \frac{\alpha_k}{\alpha_j}} = \frac{-2\alpha_{jj}}{1 + \sum_{´k \neq (i,j)} \frac{\alpha_k}{\alpha_i}}. \tag{9}$$

The CES production function with 3 explanatory variables (land, capital and the average temperature in June) was tested at first. As the results of previous models showed, the influence of these variables on the wheat harvest is the most important. The estimation results for this function revealed that most parameters were not statistically significant. Compared to the Kmenta's approximation for two inputs PFCES the number of variables considerably increased. The parameters are not significant also for this reason. This fact was confirmed by estimating function for 4 variables (the average monthly rainfall for July was added). The estimation of linearized form of PFCES took advantage from the existence of individual regional effects. But to determine whether the PFCES with n inputs is suitable for wheat harvest estimation we attempted to estimate the nonlinear form of PFCES using nonlinear OLS and then using the maximum likelihood method. Both of these methods failed to estimate the function that can't lead to convergence. Both used econometric softwares led to the same results.

4 Conclusion

The paper presents the analysis of wheat production function. The significance of classical production factors as well as climate factors was tested. It was shown that wheat production function has not the character of unit elasticity of substitution and that the Cobb-Douglas production function is not appropriate. Then, Kmenta's linear approximation to the 2-input CES function was applied and the results demonstrated its suitability for the wheat case. When the individual effects were tested, the fixed ones gave more convincing results. On the other hand, the significance of climate variables indicates the need to incorporate more than 2 inputs but the application of multifactorial CES function in its neither nonlinear nor approximate linear form didn't lead to any useful results. The conclusions mentioned above induce the question whether the variable elasticity of substitution production function weren't also a suitable approach to model wheat harvest.

Acknowledgements

This work was supported by Internal Grant Agency of the University of Economics in Prague - IGA F4/18/2011.

References

[1] Carvet, B. F.: *Wheat: science and trade*. Wiley-Blackwell, Ames, 2009.
[2] Greene, W.H.: *Econometric analysis, 5th ed.* Prentice Hall, Upper Saddle River, New York, 2003.
[3] Hoff, A.: The translog approximation of the constant elasticity of substitution production function with more than two input variables. *Danish Research Institute of Food Economics Working Paper* **14** (2002).
[4] Hoff, A.: The linear approximation of the CES function with n Input Variables. *Marine Resource Economics* **19** (2004), 295-306.
[5] Hušek, R.: *Aplikovaná ekonometrie.* Vysoká škola ekonomická v Praze, Praha, 2001.
[6] Kmenta, J.: On estimation of the CES production Function. *International Economic Review* **8** (1967),80-89.
[7] Koop, G.: *Introduction to Econometrics.* John Wiley & Sons Ltd., England, 2008.

Methods for Output Gap Detection

Kvetoslava Surmanová, Andrea Furková[1]

Abstract. Economic development of each country has several phases. Analyze the economic cycles means to deal with the identification of individual phases of the economy and to detect factors that cause them. One of them is the output gap which is the main indicator of inflationary pressures in the economy. Determination of the output gap is important from a macroeconomic perspective, particularly for the creation of fiscal and monetary policies. The output gap is defined as the difference between the actual and potential output. Since potential output value is immeasurable as the output gap, we can say that its quantification is based on several assumptions that can have an economic or statistical aspect. In literature we can meet with a number of methods that enable its detection. This paper discusses the possible ways to estimate the output gap, showing their shortcomings. In conclusion, we present some of them applied to the Slovak economy.

Keywords: Output Gap, Trend model, Hodrick–Prescott Filter, Bend-Pass Filter, Unobserved components model.

JEL Classification: C22, E32
AMS Classification: 90C15

1 Introduction

The right decisions on economic polices can be implemented only with a good knowledge of the current state of the economy. The correlation of inflation and potential output is very important indicator of the economy. How are this two interrelated variables determine is crucial for the correct formation of monetary policy. Thus, in particular monetary authorities seeking on price stability focus their attention on the potential product. Unlike the real product, the potential output is an unobservable variable. The methods of estimation are used on the detection of the potential output. The difference between the actual value of the gross domestic product (GDP) and the potential output is the output gap. It follows that also quantification of the output gap is a process that is associated with problems and many inaccuracies.

Examining the output gap is mainly concerned with the staff of national banks. Also in Slovakia the potential output issue is dealt by institutions such as Financial Policy Institute of the Ministry of Finance (IFP MOF) or the National Bank of Slovakia (NBS). Team of IFP MOF in 2005 published a study which focused on measurement of the potential output and its exploitation from financial policy point of view, namely the determination of the structural fiscal balance. In 2010 IFP MOF published another study, which was aimed to build the model using multi-dimensional Kalman filter and NAIRU quantification. In the last years NBS has published several studies on this issue, e.g. [1], [2].

There are some important areas of output gap exploitation:
- measurement of structural fiscal imbalances;
- economy cycles identification (output gap estimates thus identify the cyclical position of the economy);
- provides information on short and medium term outlooks of the economy;
- help to assess the scope for non-inflationary growth.

2 The brief overview the output gap quantification methods

In practice are used several methods to detect the output gap. All methods can be divided into two groups (statistics and economics). Their classification depends on the input data sets that are used in estimation of the output gap and the methods difficulty. Although this issue is wide, the most appropriate methods of examination are still seeking[2].

[1] University of Economics in Bratislava/Faculty of Economics Informatics, Department of Operation Research and Econometrics, Dolnozemska 1,Bratislava, surman@euba.sk, furkova@pobox.sk.

[2] Report Economic Policy Committee [6] shows that in 1999 was prepared working group of expert, chaired by Mr. Cotis, to review the common estimation methods used by the European Commission and other national and international institutions.

2.1 Statistics Methods

These methods are based on time series analysis. This group contains Model of deterministic trend, Hodrick-Prescott filter and Unobserved components model, Band-pass filter and Beveridge-Nelson decomposition.

Model of deterministic trend is considered the easiest way to detect the potential output and the output gap. The most commonly used are linear or quadratic trends. This approach divides the output into two components – the trend (the potential product) and the cycle (the output gap), appropriate analytical formulation can be written as follows:

$$y_t = \beta_0 + \sum_{i=1}^{k} \beta_i t^i + c_t , \tag{1}$$

where logarithm of seasonal adjusted GDP in a nominal prices is denoted y_t, β_0 is a constant, *t* is a time trend and c_t is the output gap. If $i = 1$ this model is know as linear trend model. If $i = 2$ this model is know as quadratic trend model. Assumption of the linear trend has some limitation, therefore is not widely used in practice. The quadratic trend model is an alternative or can be used the linear trend model with the break points.

- The advantage of models of deterministic trend: non intensive input data requirements.
- Disadvantages:
  - method is suitable only for advanced economies with constant development of the potential product;
  - these methods are not recommended for transition economies;
  - unclear economic interpretation.

Hodrick-Prescott filter smooths the observed time series; it means HP filter purify time series from the trend component. The method was first used in a working paper published by Hodrick and Prescott in 1997 [4] and was aimed to analyze Postwar U.S. business cycles. HP filter is considered to be the decomposition of the time series into trend and cyclical component (output gap):

$$y_t = y_t^p + c_t . \tag{2}$$

Process HP filter computes the smoothed series y_t^p of y_t by minimizing the variance of y_t around y_t^p, subject to a penalty that constrains the second differences of y_t^p.

$$\left\{ y_t^p \right\}_{t=0}^{n+1} = \arg\ \min \sum_{t=1}^{n} \left\{ \left(y_t - y_t^p \right)^2 + \lambda \left[\left(y_{t+1}^p - y_t^p \right) - \left(y_t^p - y_{t-1}^p \right) \right]^2 \right\} , \tag{3}$$

where λ[3] is a positive value, smoothing parameter, or penalty parameter.

- Advantages of HP filter:
  - simple and easy repeatable;
  - often used in practice.
- Disadvantages:
  - purely mechanical method;
  - there are deviations from reality at the beginning and the end of the studied sample;
  - impossibility of deeper economic analysis.

Unobserved components model (UC model) famous as State space model (Sspace model). General formulation of UC model consists from two equations – “observation” equation (4) and “state” equation (5):

$$y_t = c_t + H_t \cdot x_t + v_t , \tag{4}$$

$$x_t = d_t + F_t \cdot x_{t-1} + w_t , \text{ for } t = 1, \ldots, n \tag{5}$$

where
y_t is an $(n \times 1)$ vector of measure variable,
x_t is an $(m \times 1)$ vector of possibly unobserved state variable,
c_t, d_t, H_t and F_t are conformable vectors and matrices,
v_t a w_t – are independent random effects which meet the criteria of white noise.

[3] Hodrick and Prescott suggest value for $\lambda = 100$ (for annual data), $\lambda = 1600$ (for quarterly data) and $\lambda = 14400$ (for monthly data).

This type of model can be used if the observed time series is composed of components that are for some reason possibly unknown, for example trend, seasonal component, etc. To solve this kind of model can be used Kalman filter. For its implementation we need to know the values of parameters c_t d_t, H_t and F_t. We must also know the starting value of state vector x_t. There are known several types of UC models for example Watson, Clark, Harvey and Jaeger models. Because time series of GDP logarithm is integrated of order I (1), we'll present only Watson model:

$$
\begin{aligned}
y_t &= y_t^p + c_t , \\
y_t^p &= \varphi + y_{t-1}^p + e_{1t} , \\
c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + e_{2t} .
\end{aligned}
\qquad (6)
$$

Watson model decomposes the real product trend component y_t^p and output gap c_t. “State” equations describe the nature of the trend component and the output gap characters. The trend component follows a random walk with drift φ. The output gap is defined as AR(2) process.

- Advantages of UC model:
  - non intensive input data requirements;
  - possibility of using assumptions about the behavior of unknown components.
- Disadvantage:
  - problems with the determination of baseline parameters and starting values of state vector.

Band-pass filter is a method that also allows quantifying the output gap. In the national economies, there are situations where the actual product fluctuates around potential. The adaptation mechanisms and anticyclical economic policy caused the oscillations. Then the output gap is form of oscillations, which vary in size and periodicity depending on the strength of the factors that cause convergence of the actual product to the potential.

Band-pass method is based on the separation of frequency cycles from the actual product, which yields the potential product. Consider with time series y_t. By application of BP filter for this time series, which is defined as a weighted average of past, present and future values of time series, can be obtained filtered time series yf_t:

$$
yf_t = B[y]_t = \sum_{u=-\infty}^{\infty} b_u y_{t-u} = y_t \sum_{u=-\infty}^{\infty} b_u L^u \qquad (7)
$$

With Fourier Transform of the original time series, we can define profit function for this filter. The profit function defines the passing weight of the frequency from original time series to filtered time series. If we want to maintain full time series fervency, the filter will be zero. Conversely, the value of the filter is equal to one if we want to completely suppress frequency.[4]

2.2 Economics Methods

This group includes methods of production function, multi-dimensional generalized HP filter, a structural VAR model and Unobserved components model, multidimensional Beveridge-Nelson decomposition, multidimensional filter HP and more. These methods are complex and provide comprehensive information about the output gap, because its application requires further information. These methods[5] will be not examined in this paper and will be the object of our next research.

3 The output gap – application of statistical methods

In this part, we present the selected application possibility of calculating the output gap of Slovak Republic by statistical methods. We analyze quarterly data form 1996 to 2010. The data are from the Statistical Office of the Slovak Republic. All analysis are based on the natural logarithm of time series seasonal adjusted GDP. We denote this time series GDP as *lgdp*. All analyzes are made in econometrics program Eviews 5.

3.1 Trend model

In this section as the first was estimated a linear trend model[6]:

[4] More about method BP filtre show [3].

[5] For more see [1], [2], [5], etc.

[6] Values in parentheses are test statistics of Student distribution.

$$lg\hat{dp}_t = 8{,}513 + 0{,}022 \cdot T_t$$
$$(682{,}4) \quad (61{,}9)$$

Resulting values of the residuals from this estimated model are the values of output gap (GAP_TREND_1). Based on the plot (Figure 1) we can see that the output gap in the last two years have a considerable variations and the curve deviates from the zero axis a lot. Therefore the original model trend is extended to break points and the estimated model trend can be written in the following form:

$$lg\hat{dp}_t = 8{,}493 + 0{,}022 \cdot T_t + 0{,}007 \cdot T2003_t - 0{,}007 \cdot T2006_t - 0{,}024 \cdot T2008_t \,,$$
$$(1011{,}1) \quad (48{,}3) \qquad (5{,}6) \qquad\qquad (-3{,}3) \qquad\qquad (-7{,}0)$$

Where $T2003_t = 0$ for $t \le 2003q1$ other $T2003_t = T_t - 29$, $T2006_t = 0$ for $t \le 2006q2$ other $T2003_t = T_t - 42$ and $T2008_t = 0$ for $t \le 2008q4$ other $T2008_t = T_t - 52$. From this model we can obtain the values of output gap (GAP_TREND_2). Figure 1 shows the behavior of both output gaps. Both curves have almost homogeneous process till year 2000. In the next reporting period, the course process is different. Due to smaller differences between the real and potential output especially in recent years the output gap from trend model with three breaks (years 2003, 2006 and 2008) seems to be more appropriate. In the first three years there is oscillation of real output around potential output. The peak of the economic cycle is evident in 1998q4 (in October 1998 the Slovak government was changed and it was beginning a period of a package of measures, which dampened of domestic demand components). Due to these extensive changes we can notice the period of recession from 1998q4 to 1999q4. In year 2000 there was a revival of domestic demand and this fact is evident from our curve (Figure 1). There is a smaller peak in 2001q4 and both output gaps oscillate around the zero axis. Figure 1 shows next break point in 2004q4 (see curve GAP_TREND_1), there are visible large differences between actual and potential outputs values. The change occurred in 2008, it was the beginning of the economic crisis and the economy went into recession again. From the curve GAP_TREND_2 is evident that economy starts recovering from the crisis in early 2009 and from 2010 actual output is above potential output.

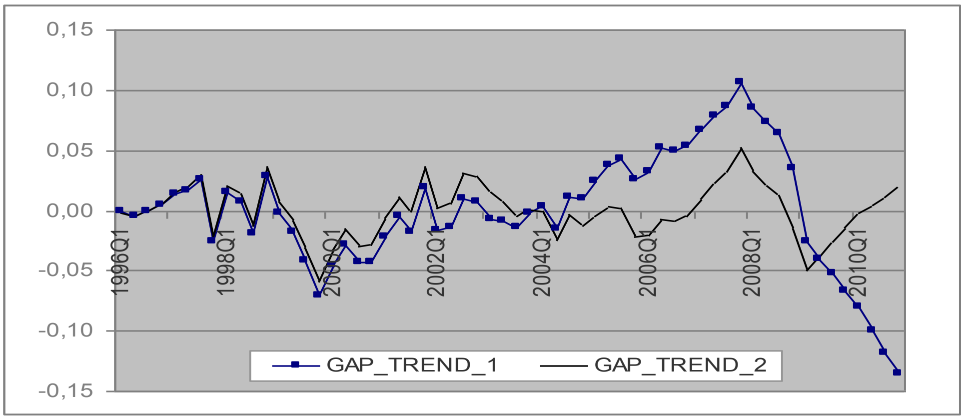


Figure 1 The Output gap – Linear Trend

3.2 HP filter

Figure 2 shows *the behavior of output gaps f*rom HP filter. In our analysis we set 3 alternatives for smoothing parameter λ. The initial value for this parameter was $\lambda = 1600$ according to the [4]. Following [2] we chose smaller values for λ. We can notice, that smaller value of parameter λ led to the reduction of the differences between the actual and potential outputs. From the economic cycles point of view, the output gap development is the similar to the results from the trend models. Figure 2 shows the 3 peaks of economics cycles in 1998q4, 2001q4 and 2007q4. The results correspond with the actual development of the Slovak economy.

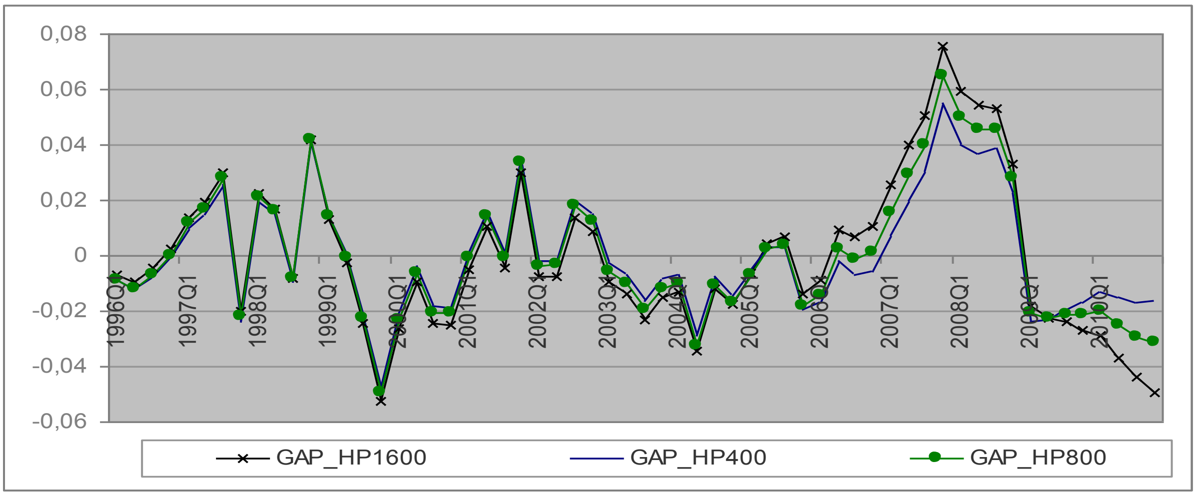


Figure 2 The Output gap – HP filter

3.3 UC model and Band-pass filter

First, we estimated production function by Kalman filter. We have defined a model according to the relation (6). At the beginning we entered the input parameters in the process of Kalman filter: $\varphi = 0{,}022$[7], $e_{1t} = 0$ a $e_{2t} = 0$. Having defined random variance components at a constant equals to 0.000001 we tried to minimize the uncertainty of this dynamic system.[8] In Figure 4 (curve Sspace) the resulting curve of the output gap can we see. The curve of the output gap Sspace model takes a similar trend as the output gap GAP_TREND_1 (Figure 1).

In this part were applied 2 BK filters[9]. Application of Baxter-King filter (published in 1995) requires setting lag *K*. According to [2] we chose *K = 12*, this fact cause losing the first and last K observations. Application of Christiano-Fitzgerald filter requires the filtered time series specification, in respect to the stationary or I (1) process and the selection of the removal of the trend or drift adjustment must be done. In comparison with before mentioned filters, BP filter provides values of output gap directly.

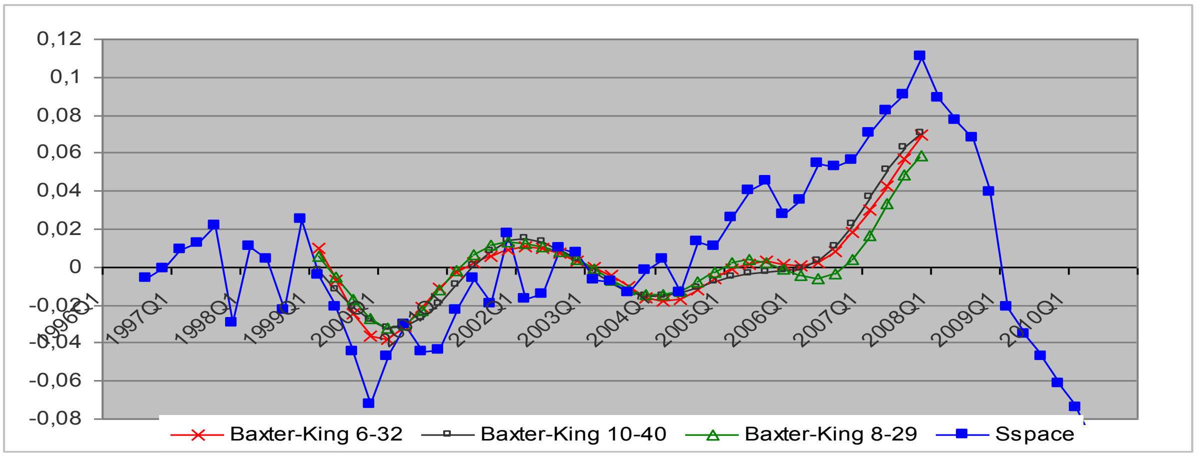


Figure 4 The Output gap - Sspace model and Baxter-King filter

[7] The trend component (output gap) follows random walk with drift φ. The drift is the quarterly growth rate of the potential output. Analogy of this parameter is the slope of the trend model (see part 3.1). Therefore, the input parameter value is set to 0,022.

[8] We chose a very small number in order to the filtered values were in the immediate vicinity of the observed variable y_t.

[9] EViews allows carrying 3 types BK filter: Baxter-King, Christiano-Fitzgerald fixed length with symmetric and Christiano-Fitzgerald with asymmetric full sample.

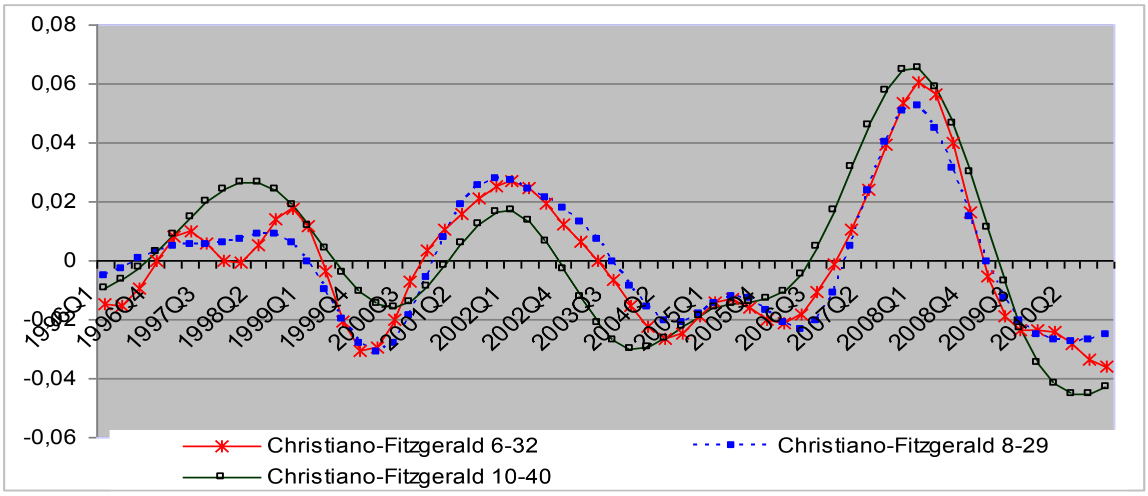


Figure 5 The Output gap: Christiano-Fitzgerald filter

Conclusion

The output gap estimation provides economic policy makers to identify business cycles and follow-up to prevent non-ferrous fluctuations. In this study we developed statistics methods for detection the output gap. Our results suggest that all methods identified the same economic cycles. All methods are based on time series of GDP. The comparison the obtained results had led to the conclusion that the shape of the output gap curves differ only at the end of observation period. It is therefore important for further analysis to take into account this fact and avoid mistakes that arise from the estimation of the output gap. The best methods appear deterministic trend models and the UC model because the output gap in comparison with other approaches since 2003 takes positive values, which corresponds to the real economic development.

In both cases this trend remained the same till to the economic crisis in 2008. This mentioned methods do not exploit the economic background information and it is their major disadvantage. The economic background information inclusions lead to the better-quality of estimation and accurate estimation of the output gap. Despite this disadvantage these methods are often primarily used for their simplicity.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0595/11 "Analysis of Business Cycles in the Economies of the Euro Area (with Regards to the Specifics of the Slovak Economy) Using Econometric and Optimization Methods"

References

[1] Antoničová, Z., and Huček, J.: Viacrozmerný filter s nemerateľnými stavmi, In: *BIATEC*, ročník 13, 6/2005.
[2] Benčík, M.: Metódy detekcie nerovnováhy v reálnej ekonomike SR. Výskumná štúdia 2/2008, http://www.nbs.sk/_img/Documents/PUBLIK/08_ben1.pdf.
[3] Christiano, L. J. and Fitzgerald, T. J.: *The Band Pass Filter*, Working paper 9906, Federal Reserve Bank of Cleveland, 1999, http://www.clevelandfed.org/Research/workpaper/1999/Wp9906.pdf.
[4] Hodrick, R. J., and Prescott, E. C.: Post-War U.S. Business Cycles: An empirical investigation. In: *Journal of money, credit and banking*, 29, 1-16, 1997.
[5] Output Gap and NAIRI Estimates within State-Space Framework: An Application to Slovakia. Financial Policy Institute The Ministry of Finance of the Slovak Republic, Economic Analysis 16, 2010, www.finance.gov.sk/ifp.
[6] Report on Potential Output and the Output Gap. Economic Policy Committee, Brusel, 2001, http://europa.eu/epc/pdf/finaloutput_en.pdf.
[7] Szomolányi, K., Lukáčik, M.: Model hospodárskych cyklov a reálne ekonomiky. In Nové trendy v ekonometrii a operačním výzkumu: [elektronický zdroj]. - Bratislava : Vydavateľstvo EKONÓM, 2010. ISBN 978-80-225-3126-9, s. [1-5].
[8] EViews 5 User´s Guide
[9] www.statistics.sk

Development and the cyclicality of government spending in the Czech Republic

Irena Szarowská[1]

Abstract. This paper aims to provide direct empirical evidence on business cycle relations between GDP and government spending in the Czech Republic. Government spending plays an important role in a fiscal policy as a possible automatic stabilizer. We analyzed annual data on government spending in compliance with the COFOG international standard. We use cross-correlation on cyclically filtered adjusted time series over the period 1995-2008. The cyclical properties of GDP and government spending function were, in average, found as weakly correlated. However, we report considerable differences in correlations across the spending functions. The lowest correlation coefficient (0.06) was found for recreation, culture and religion and the highest average was reported for economic affairs (-0.51). As regards to using government spending as the stabilizer, total government spending, general public services, defense, economic affairs and education spending were negative correlated and it confirms countercyclical relation between these spending functions and GDP. It is in line with theory suggestion. On the other hand, the highest spending function (social protection) correlated weak positive and it mean procyclical development.

The results of Johansen cointegration test proved the existence of long-run relationship between GDP and total government spending, public order and safety and economic affairs.

Keywords: government spending, cyclicality, economic growth, correlation, cointegration.

JEL Classification: C32, H50, E62
AMS Classification: 90C15

1 Introduction

The economy of the country is greatly influenced by the level and the structure of government spending. The government spending is an important tool for national governments to mitigate the uneven economic development and economic shocks across individual countries. Government spending plays important role in a fiscal policy of each country as a possible automatic stabilizer as from a Keynesian perspective, there is a view that government spending should act as a stabilizing force and move in a countercyclical direction. Procyclical fiscal policy is conversely policy expansionary in booms and contractionary in recessions. Serven [13] points that procyclical fiscal policy is generally regarded as potentially damaging for welfare: it can raise macroeconomic volatility, depress investment in real and human capital, hamper growth, and harm the poor. If expansionary fiscal policies in “good times” are not fully offset in “bad times”, they may also produce a large deficit bias and lead to debt unsustainability and eventual default. If a government respect a basic prescription that fiscal tools should function counter-cyclical, the optimal fiscal policy involves a decreasing of government spending in “good times” and a increasing of government spending in “bad times.” Contrary to the theory (it implies that government spending is countercyclical), a number of recent studies found evidence that government spending is procyclical. See Hercowitz and Strawczynski [7], Alesina et al., [2], Rajkumar and Swaroop [12] or Ganeli [4] for more details. Talvi and Vegh [14] show that fiscal procyclicality is evident in a much wider sample of countries. Lane [10] finds procyclicality in a single-country time series study of Irish fiscal policy. As Fiorito and Kollintzas [3] document for G7 countries, the correlation between government consumption and output indeed appears to show no pattern and be clustered around zero. Lane [11] also shows that the level of cyclicality varies across spending categories and across OECD countries. Abbot and Jones [1] test differences in the cyclicality of government spending across functional categories. Their evidence from 20 OECD countries suggests that procyclicality is more likely in smaller functional budgets, but capital spending is more likely to be procyclical for the larger spending categories. Many of researches like Gavin et al. [5], Gavin and Perotti [6] focuse on Latin America. Previously published studies are weakly supported by the data particularly in emerging and post-transition economies in which results can vary. We would like to eliminate the literature gap in this field and analyze government spending in the Czech Republic. The aim of the paper is to provide direct empirical evidence on busi-

[1] Silesian University in Opava, School of Business Administration in Karvina, Department of Finance, Univerzitni nam. 1934/3, 733 40 Karvina, Czech Republic, email: szarowska@opf.slu.cz.

ness cycle relation between Gross Domestic Product (GDP) government spending (G) and estimate long-run relationship between these variables in the Czech Republic.

We follow Abbot and Jones [1] and apply the cross-correlation technique and cointegration on annul data of GDP and government spending in compliance with the COFOG international standard during the period 1995-2009 from Eurostat. The paper is organized as follows. In the next section, we describe the dataset and empirical techniques used. In Section 3, we present the results of government spending development and cross-correlation. In Section 4, we estimate long- run relationship between output and government spending. In Section 5, we conclude with a summary of key findings.

2 Data and Methodology

The dataset consists of annual data on GDP and government spending in compliance with the COFOG international standard during the period 1995 – 2008. Although data from 2009 are available we prefer to work with a consistent dataset that excludes observations from a crisis period. All the data were collected from the Eurostat database. The series for GDP and total government spending and its subcomponent are adjusted at constant prices. We converted all series into logs and applied the Hodrick-Prescott filter with smoothing parameter 100 to each series with the aim to isolate the cycle component of time series. We apply cross-correlation to all combinations of GDP – category of government spending. Johansen cointegration test and the error correction model (ECM) are used to estimate the long-run relationship between output and government spending predicted by, for example, Wagner´s Law. Most of the results are calculated in econometric program Eviews 7.

Many studies point out that using non-stationary macroeconomic variable in time series analysis causes superiority problems in regression. Thus, a unit root test should precede any empirical study employing such variables. We decided to make the decision on the existence of a unit root through Augmented Dickey–Fuller test (ADF test). The equation (1) is formulated for the stationary testing.

$$\Delta x_t = \delta_0 + \delta_1 t + \delta_2 x_{t-1} + \sum_{i=1}^{k} \alpha_i \Delta x_{t-i} + u_t \tag{1}$$

ADF test is used to determine a unit root x_t at all variables in the time t. Variable Δx_{t-i} expresses the lagged first difference and u_t estimate autocorrelation error. Coefficients $\delta_0, \delta_1, \delta_2$ and α_i are estimated. Zero and the alternative hypothesis for the existence of a unit root in the x_t variable are specified in (2). The result of ADF test, which confirms the stationary of all time series on the first difference, is available on reguest.

$$H_0\text{: } \delta_2 = 0, H_\varepsilon\text{: } \delta_2 < 0 \tag{2}$$

The cross-correlation assesses how one reference time series correlates with another time series, or several other series, as a function of time shift (lag). Consider two series x_i and y_i where i = 0, 1, 2, …, N-1. The cross correlation r at delay d is defined as:

$$r = \frac{\sum_i \left[(x_i - m_x) * (y_{i-d} - m_{y)} \right]}{\sqrt{\sum_i (x_i - m_x)^2 \sqrt{(y_{i-d} - m_y)^2}}} \tag{3}$$

where m_x and m_y are the means of corresponding series.

The Hodrick-Prescott (HP) estimates an unobservable time trend for time series variables. Let y_t denote an observable macroeconomic time series. The HP filter decomposes y_t into a nonstationary trend g_t and a stationary residual component c_t, that is:

$$y_t = g_t + c_t \tag{4}$$

We note that g_t and c_t are unobservables. Given an adequately chosen, positive value of λ, there is a trend component that will minimize:

$$\min \sum_{t=1}^{T} (y_t - g_t)^2 + \lambda \sum_{t=2}^{T} \left[(g_{t+1} - g_t) - (g_t - g_{t-1}) \right]^2 \tag{5}$$

The first term of the equation is the sum of the squared deviations which penalizes the cyclical component. The second term is a multiple λ of the sum of the squares of the trend component's second differences. This second term penalizes variations in the growth rate of the trend component. The larger the value of λ, the higher is the penalty. Hodrick and Prescott advise that, for annual data, a value of λ = 100 is reasonable.

The Johansen method [8] applies the maximum likelihood procedure to determine the presence of cointegrating vectors in non-stationary time series as a vector autoregressive (VAR):

$$\Delta x_t = C + \sum_{i=1}^{K} \chi_i \Delta x_{t-i} + \pi Z_{t-1} + \eta_t \tag{6}$$

where x_t is a vector of non-stationary (in log levels) variables and C is the constant term. The information on the coefficient matrix between the levels of the Π is decomposed as $\Pi = \alpha \cdot \beta'$, where the relevant elements the α matrix are adjustment coefficients band the β matrix contains the cointegrating vectors. Johansen and Juselius [9] specify two likelihood ratio test statistics to test for the number of cointegrating vectors. The first likelihood ratio statistics for the null hypothesis of exactly r cointegrating vectors against the alternative $r+1$ vectors is the maximum eigenvalue statistic. The second statistic for the hypothesis of at most r cointegrating vectors against the alternative is the trace statistic. Critical values for both test statistics are tabulated in Johansen–Juselius [9]. If the variables are non-stationary and are cointegrated, the adequate method to examine the issue of causation is the Error Correction Model (ECM), which is a Vector Autoregressive Model VAR in first differences with the addition of a vector of cointegrating residuals. Thus, this VAR system does not lose long-run information.

3 Development and the cyclicality of government spending

Government spending can help in overcoming the inefficiencies of the market system in the allocation of economic resources. It also can help in smoothing out cyclical fluctuations in the economy and influences a level of employment and price stability. Thus, government spending plays a crucial role in the economic growth of a country. We used government spending in compliance with the COFOG international standard (Classification of the Functions of Government) in our analysis. Total government spending is divided into 10 basic divisions:

- G10: General public services
- G20: Defense
- G30: Public order and safety
- G40: Economic affairs
- G50: Environment protection
- G60: Housing and community amenities
- G70: Health
- G80: Recreation; culture and religion
- G90: Education
- G100: Social protection

3.1 The structure of government spending and its development

Firstly we analyzed the structure of government spending in a period 1995-2009. Results in Table 1 show the share of government spending by functions, their average on total spending during the whole period and the share of total government spending on GDP. Data confirm unstable and cyclical development of total government spending on GDP. In 1995, a high government spending was connected with privatization and transformation process. Five spending functions, on average, account for more than 84% of the total spending: social protection, economic affairs, health, general public services and education. Table 1 shows that social protection (G100) was the largest item of government spending from 1996, economics affairs (G40) were on the second and health spending (G70) on the third place till the year 2004. From 2005 the second and the third position has changed.

	1995	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**average**
G10	8.1	10.2	9.9	9.3	10	9.9	9.7	10.3	11	10.9	12	10.1	10.2	10.4	10.2
G20	3.4	3.8	3.9	3.5	3.9	4.1	3.6	3.4	4.1	3.1	3.7	2.9	2.8	2.6	3.4
G30	4.8	5.8	5.7	5.1	5.6	5.6	5	4.6	4.7	4.8	4.9	4.9	4.9	4.8	5.1
G40	37	18	19.9	22	19.5	17.5	20.9	19.3	17.6	16.7	15.4	16.2	16.1	16.8	19.3
G50	1.9	2.9	2.6	2.5	2.1	2.2	2.2	2.1	2.4	2.4	2.6	2.6	2.4	2.3	2.3
G60	1.8	2.9	2.5	2.8	2.4	2.6	2.7	1.4	2.6	3.5	3.6	3.6	2.7	2.6	2.7
G70	10.8	14.7	13.5	13.6	13.9	13.7	13.6	13.5	13.5	16.3	16	16.4	16.7	16.8	14.7
G80	2.1	3.1	2.6	2.6	2.4	2.4	2.5	2.7	2.7	2.7	2.7	3.1	2.9	2.9	2.7
G90	7.9	9.7	9.8	9.4	9.4	9.9	9.9	11.1	11	10.7	10.6	11.3	10.9	10.9	10.2
G100	21.9	28.9	29.7	29.2	30.7	32	30.1	31.5	30.3	28.9	28.5	29	30.2	30	29.4
G as % GDP	54.5	42.6	43.2	43.2	42.3	41.8	44.4	46.3	47.3	45.1	45	43.8	42.5	42.9	44.8

Table 1 Development of government spending function

The social protection spending G100 is the highest spending function and it takes nearly 1/3 of all government spending. It contains, for example, spending on sickness and disability, old age, survivors, family and children, unemployment, housing, social exclusion and R&D social protection.

3.2 The cyclicality of government spending

As was already noted, government spending is a possible automatic stabilizer. From this point of view, government spending should move in a countercyclical direction. We decided to assess the relationship between GDP and government spending and we analyzed the correlation between cycle components of GDP and all government spending functions. Figure 1 shows GDP and total government spending G before and after using HP filter.

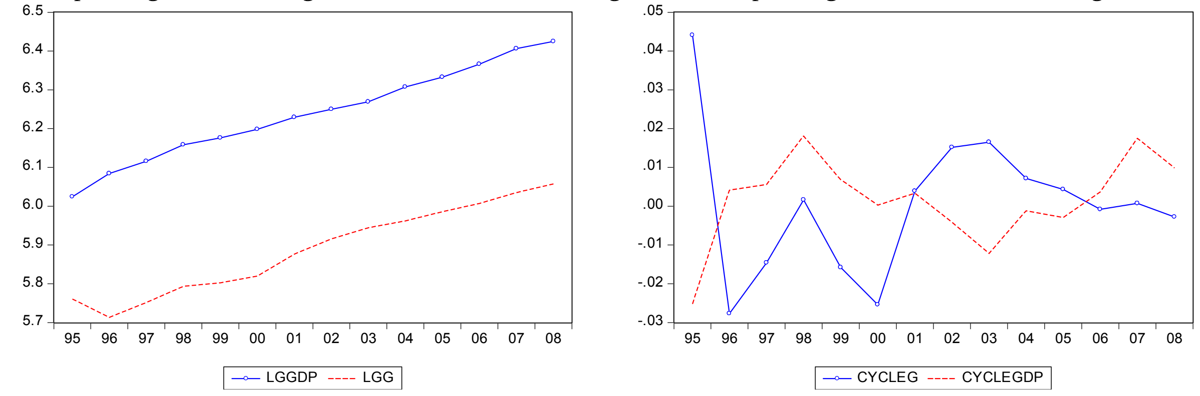


Figure 1 Development of GDP and G

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. The correlation coefficient can vary from -1 to +1. The correlation coefficient -1 indicates perfect negative correlation, and +1 indicates perfect positive correlation. Its value smaller 0.4 means weak correlation, from 0.4 to 0.7 moderate correlation and higher than 0.7 express strong correlation. A positive correlation coefficient indicates the procyclicality of government spending, negative value means that variables are countercyclical and value close to zero express acyclicality. We run cross-correlations for all possible combinations of GDP and government spending. The results are reported in Table 2. Here we present coefficients with no lag / lead; all results are available on request.

	Correlation coefficient	**Correlation**	**Cyclicality**
G10: General public services	-0.4320	moderate negative	countercyclical
G20: Defense	-0.5148	moderate negative	countercyclical
G30: Public order and safety	0.2479	weak positive	procyclical
G40: Economic affairs	-0.5184	moderate negative	countercyclical
G50: Environment protection	0.1410	weak positive	procyclical
G60: Housing and community amenities	0.1591	weak positive	procyclical
G70: Health	0.3272	weak positive	procyclical
G80: Recreation; culture and religion	0.0639	no correlation	acyclical
G90: Education	-0.3797	weak negative	countercyclical
G100: Social protection	0.3329	weak positive	procyclical
Total G	-0.6331	moderate negative	countercyclical

Table 2 Cyclicality of government spending

The results indicate significant difference across spending functions. We note that 70% of the correlation coefficients are lower than 0.4 in absolute value indicating a weak connection of spending to GDP. Total G, general public services, defense, economic affairs and education were negative correlated and it confirms countercyclical relation between these spending functions and GDP. It is in line with theory recommendation. Contrary to the theory, the correlation coefficients of the highest spending functions (social protection and health) were weak positive and it reports procyclical development of these sub-categories of government spending and GDP. The lowest correlation coefficient (0.06) was found for recreation, culture and religion and the highest average was reported for economic affairs (-0.51), except the coefficient for total government spending (-0.63).

4 Long- run relationship between government spending and GDP

We also analyzed the long-term relationship between GDP and all government spending functions. The Johansen cointegration test, which is also used in this paper, is nowadays frequently used for testing cointegration. Assumption for implementation of cointegration is done by the fact that time series are stationary at first difference.

Individual series are non-stationary, but their common cointegration movement in a long time lead (for example as a result of various market forces) to some equilibrium, though it is possible that in the case of short time periods there is a misalignment of such a long balance. The aim of cointegration test is to determine the number of cointegration relations *r* in the VAR models. It is also necessary to identify an optimal time lag. The optimal time lag is one period (year) ind it was found with using Akaike information criterion applied to estimation of the non-differenced VAR model. The results of Johansen cointegration test proved the existence of the long-run positive relationship between GDP and total government spending, public order and safety and economic affairs. Findings of test indicated no cointegration between GDP and other spending functions. Cointegration equations have the form expressed in (7), (8) and (9).

$$\Delta GDP = 1.083\, \Delta G - 0.134 \quad (7)$$
$$(0.131)^*$$

$$\Delta GDP = 1.243\, \Delta G30 + 0.530 \quad (8)$$
$$(0.0226)^*$$

$$\Delta GDP = 1.7433\, \Delta G40 - 2.7241 \quad (9)$$
$$(0.2198)\ ^*$$

A symbol Δ means difference of log variables: *GDP*, total government spending G, Public order and safety spending *G30* and economic affairs spending *G40*. A symbol * denotes significance at standard 5% level. The above equation shows that increase of total government spending by 1% is connected with increase GDP by 1.08%. We can find similar relationship between GDP and G30 (1.24%) and GDP and G40 (1.78%).

The cointegration regression considers only the long-run property of the model, and does not deal with the short-run dynamics explicitly. Therefore, ECM is used to detect these fluctuations as it is an adequate tool to examine the short-run deviations necessary to the achievement of long-run balance between the variables. Here, the optimal number of lag is one as was found. We define the ECM for GDP and total government spending in (10) and (11).

$$\Delta GDP_t = \alpha_0 + \omega_1\, (GDP_{t-1} - \gamma G_{t-1}) + \alpha_1\, \Delta GDP_{t-1} + \alpha_2\, \Delta G_{t-1} + u_{1t,} \quad (10)$$

$$\Delta G_t = \beta_0 + \omega_2\, (GDP_{t-1} - \gamma G_{t-1}) + \beta_1\, \Delta GDP_{t-1} + \beta_2\, \Delta G_{t-1} + u_{2t,} \quad (11)$$

In (10) and (11), GDP_t and G_t are cointegrated with cointegrating coefficient γ, α_0 and β_0 are constants of the model, ω_1 and ω_2 note the coefficients of cointegration equition, u_{1t} and u_{2t} mean residual components of long-term relationship. The ECM equations are similar for G30 and G40 spending functions. The model specification was tested by several residual components tests. We used the autocorrelation LM-test based on Lagranger multipliers, the normality test, and heteroskedasticity test. The performed tests reject the existence of all three phenomena. The results of the ECM for all thee founded cointegration are reported in Table 3. Standard errors are in parenthesis.

Cointegration between	Dependent variable	ω_1 resp. ω_2	GDP_{t-1}	G_{t-1}	α_0 resp. β_0
GDP and G	GDP_t	-0.0581 (0.1498)	0.1661 (0.2941)	-0,1389 (0.1414)	0.0368* (0.0115)
	G_t	0.260305 (0.2122)	0.2003 (0.4165)	-0,0389 (0.2003)	0.03599* (0.0163)
GDP and C30	GDP_t	-0.5465* (0.2353)	-0.0467 (0.1878)	0.7594** (0.3348)	0.0346* (0.0092)
	$G30_t$	1.1608* (0.3149)	0.3390*** (0.2473)	-0,0389 (0.2003)	-0.0067 (0.0124)
GDP and C40	GDP_t	0.0879*** (0.0524)	-0.1400 (0.2493)	0.0217 (0.0337)	0.0330* (0.00826)
	$G40_t$	0.7623* (0.2167)	-0.0153 (1.0311)	0.1946*** (0.1405)	0.0281 (0.0342)

Table 3 The error correction models

Symbols *, ** and *** denote significance at the 1%, 5% and 10% level. The findings report that ECM does not provide significant results for short- run relationship between GDP and G. In the case of G30 (G40), the ECM through lagged values explains convergence to long-run relationship in the context of short-run shocks and dynamics. Generally, we proved long-run relationship between GDP and G (resp. G30, G40) and value of coefficient suggest that government spending tends to follow GDP (adjusting coefficients for G, resp. G30, G40 are higher than for GDP) and it adapts to GDP changes.

5 Conclusion

The aim of this paper was to provide direct empirical evidence on business cycle relations between GDP and government spending in the Czech Republic from 1995 to 2009. Government spending plays important role in a fiscal policy as it can help to reduce cyclical fluctuations in the economy.

Although many studies suggest government spending is procyclical despite the recommendations of the theory, our research does not prove that. The results confirm cyclical development of total government spending on GDP in the Czech Republic during 1995-2008. Five spending functions, on average, account for more than 84% of the total spending: social protection, economic affairs, health, general public services and education. The cyclical properties of GDP and government spending function were, in average, found as weakly correlated. However, we report considerable differences in correlations across the spending functions and some correlation coefficients are sufficiently high. The lowest correlation coefficient (0.06) was calculated for recreation, culture and religion and the highest value was reported for economic affairs (-0.51). As regards to using government spending as a stabilizer, total government spending, general public services, defense, economic affairs and education spending were negative correlated and it confirms countercyclical relation between these spending functions and GDP. It is in line with theory suggestion. On the other hand, the highest spending function (social protection) correlated weak positive and it suggests procyclical movement of these spending functions. We also analyzed the long-term relationship between GDP and all government spending functions. The results of Johansen cointegration test proved the existence of long-run positive relationship between GDP and total government spending, public order and safety and economic affairs spending functions. As findings verify, they tend to follow GDP and adapt to GDP changes. Tests indicated no cointegration between GDP and other government spending functions.

Acknowledgements

This paper ensued thanks to the support of the grant GAČR 403/11/2073 "Procyclicality of financial markets, asset price bubbles and macroprudential regulation".

References

[1] Abbott, A., and Jones, P.: Procyclical government spending: Patterns of pressure and prudence in the OECD. *Economics Letters* **111** (2011), 230–232.
[2] Alesina, A. et al.: Why is fiscal policy often procyclical? *Journal of the European Economic Association* **6**, no. 5 (2008), 1006–1030.
[3] Fiorito, R., and Kollintzas, T.: Stylized facts of business cycles in the G7 from a real business cycles perspective. *European Economic Review* **38** (1994), 235–269.
[4] Ganelli, G.: The International Effects of Government Spending Composition. *Economic Modelling* **27**, no. 3 (2010), 631-40.
[5] Gavin, M. et al.: *Managing Fiscal Policy in Latin America and the Caribbean: Volatility, Procyclicality, and Limited Creditworthiness*. IDB Working Paper No. 326, 1996.
[6] Gavin, M., and Perotti, R.: Fiscal policy in Latin America. *Macroeconomics Annual* **12** (1997), 11–70.
[7] Hercowitz, Z., and Strawczynski, M.: Cyclical Ratcheting in Government Spending: Evidence from the OECD. *Review of Economics and Statistics* **86**, no. 1 (2004), 353-61.
[8] Johansen, S.: Cointegration and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* **59**, no.6 (1991), 1551–1580.
[9] Johansen, S. and Juselius, K.: (1990): Maximum Likelihood Estimation and Inference on Cointegration, with Applications. *Oxford Bulletin of Economics and Statistics* **52** (1990), 169–210.
[10] Lane, P. R.: International Perspectives on the Irish Economy. *Economic and Social Review* **29**, no. 2 (1998), 217-22.
[11] Lane, P. R.: The Cyclical Behaviour of Fiscal Policy: Evidence from the OECD. *Journal of Public Economics* **87**, no. 12 (2003), 2661-75.
[12] Rajkumar, A. S., and Swaroop, V.: Public Spending and Outcomes: Does Governance Matter? *Journal of Development Economics* **86**, no. 1 (2008), 96-111.
[13] Serven, L. *Macroeconomic Uncertainty and Private Investment in LDCs: an Empirical Investigation*. World Bank Working Paper No. 2035, 1998.
[14] Talvi, E., and Vegh, C. A.: Tax base variability and procyclical fiscal policy in developing countries. *Journal of Development Economics* **78** (2005), 156–190.

The β-convergence of the EU27 Countries

Karol Szomolányi, Adriana Lukáčiková and Martin Lukáčik [1]

Abstract. In the paper we deal with the convergence processes of the EU economies using the *β-convergence* concept. The neoclassical growth theory predicts that a lower level of real GDP per person would match up with a higher subsequent growth rate of real GDP per person. The data for EU27 countries are used to estimate a relation of the growth rates of real GDP per person against levels of real GDP per person. If there is a convergence process, the relation will be negative. Finally we use dummy regional variables to identify the clubs of countries with the different steady state level assuming that their technology level is the same.

We find that there has been a stronger unconditional (absolute) convergence of EU27 countries in the period 2003-2009 than it was in the period 1996-2002. The average yearly convergence rate has been about 1.3% for last seven years. Italy, Hungary, Portugal and Spain converge towards the lower steady state level than European average, while Luxembourg converges towards the higher level. These differences between these countries have enlarged in last 7 years.

Keywords: neoclassical growth model, *β*-convergence, EU27 countries

JEL Classification: C21, O47
AMS Classification: 62H12

1 Introduction and Brief Literature Overview

We verify the convergence of the EU27 countries using the *β*-convergence concept. Barro [2], chapter 5, Baumol [7], DeLong [8], Barro [1], Barro and Sala-i-Martin [3], [4], [5] and [6], chapter 11 used this concept to verify the convergence of the economies. These sources are well-known and used by nearly all reputable macroeconomic books. Therefore it is not necessary to discuss the problem in detail.

We try to answer if there has been an unconditional or absolute convergence of EU27 countries and if there have been clubs of countries with different steady state levels. We will find the unconditional convergence that has strengthened during last 7 years. However – assuming that all European countries have the same technology level – we will find that Italy, Hungary, Portugal and Spain form the club of countries with lower steady state level than European average. On the other hand, Luxembourg converges towards the higher steady state level. The differences in the steady state levels of the real GDP per person have enlarged in last 7 years.

According to the concept, convergence applies if a poor economy tends to grow faster than a rich one, so that the poor country tends to catch up to the rich one in terms of levels of per capita income or product. Therefore we estimate coefficients of the equation with the level of the real GDP per person in initial time of a studied period as exogenous variable and with the growth of the real GDP per person during a period as an endogenous variable. In the 2nd section we define several specifications of the convergence equations. We can use more explanatory variables that capture information about different steady state income levels and shocks in different European countries to yield consistent estimations of the convergence parameter.[2] In the 3rd section we deal with data and methodology which we use in the estimations. Then in the 4th section we estimate the specification with dummy variables which correspond to the countries with probably different steady state levels of the real GDP per person and, finally, the last section concludes.

2 Specification

We estimate the coefficients of the equations:

$$g_{i,t,t+T} = \alpha + \beta y_{i,t} + u_{i,t,t+T} \qquad \begin{array}{l} i \in \{\text{Belgium, Bulgaria, ..., United Kingdom}\} \\ t \in \{1996, 2003\}, t+T \in \{2002, 2008, 2009\} \end{array} \tag{1}$$

[1] University of Economics, Department of Operations Research and Econometrics, Dolnozemská cesta 1, 852 35, Bratislava, szomolan@euba.sk.

[2] See Barro and Sala-i-Martin [6] , chapter 11.

$$g_{i,t,t+T} = a + \beta y_{i,t} + \sum_{k=1}^{n} \gamma_k s_{i,k,t} + v_{i,t,t+T} \qquad i \in \{\text{Belgium, Bulgaria, ..., United Kingdom}\} \quad t \in \{1996, 2003\}, t+T \in \{2002, 2008, 2009\} \tag{2}$$

where $y_{i,t}$ is log of the real GDP per person of country i at time t, $g_{i,t,T}$ is the growth rate of the real GDP/person of country i for T years:

$$g_{i,t,t+T} = y_{i,t+T} - y_{i,t}$$

We study convergence in several periods. The full range period is period with t = 1996 and $t + T$ = 2009. Taking in the account the financial crisis, this period could be sort of problematic. Therefore – to insure the robustness of the estimates – we made also estimations in the period with t = 1996 and $t + T$ = 2008. Furthermore we divided the full range period into two periods with t = 1996 and $t + T$ = 2002 and with t = 2003 and $t + T$ = 2009.

If the β coefficient in the univariate specification (1) turns out to be negative, we will conclude that poor economies tend to grow faster than rich economies so that "absolute convergence" applies. However, the neoclassical growth model predicts a multivariate specification such as (2), where the sum term captures information about the steady state level of the country i, or the different aggregate shocks in the countries.

3 Data and Methodology

All data we use are gathered from the portal of the Eurostat. We use GDP in 2000 euros and the population size of each country in years 1996, 2002, 2003, 2008 and 2009.[3] Table 2 shows linear least-squares estimates in the forms (1) and (2) for 25-27 EU countries for various periods. The rows of table 2 correspond to the different time periods. The column named *Basic Equation* of the table refers to the equation with only one explanatory variable (1). Following Barro and Sala-i-Martin [6], chapter 11, we determine the steady state level in the (2) specification by different geographical position of the countries and the different aggregate shocks by different growth in the main economic sectors in the countries. The column named *Equations with Regional Dummies* adds four geographical dummies, corresponding to four European regions: East, North, South and West, i.e. there are estimates of the β parameters of the specification (2) with n = 4 dummy variables corresponding to the four mentioned regions. The distribution of the countries into the regions is in the table 1. The last column adds an additional variable to the specification (2) as an attempt to hold aggregate shocks constant. The variable denoted $s_{i,5,t}$ is calculated as

$$s_{i,5,t} = \sum_{j=1}^{7} \omega_{i,j,t} \left(y_{j,t+T} - y_{j,t} \right) \tag{3}$$

where $\omega_{i,j,t}$ is the weight of sector j in state i's personal income in time t and $y_{j,t}$ is the log of national average of personal income per worker in sector j in time t. The seven sectors used in analysis are NACE: agriculture, construction, finance, industry, manufacturing, services and trade.[4]

Region	Countries in Region
East	Czech Republic, Hungary, Poland, Slovakia, Slovenia
North	Denmark, Estonia, Finland, Latvia, Lithuania, Sweden
South	Bulgaria, Cyprus, Greece, Italy, Malta, Romania
West	France, Ireland, Portugal, Spain, United Kingdom

Table 1 The distribution of the countries into the regions

Each cell in table 2 contains the estimates of β. The asterisks denote the statistical significance of the parameter at the 10 percent level (*) and at the 5 percent level (**). Except the estimation in the period 1996-2002 all estimates are statistically significant at the 5 percent level. Similarly the three models for the period 1996-2002 are the overall significant at the only 10 percent level. As results of estimates for periods 1996-2008 and 1996-2009 in basic equations and equations with regional dummies were similar, we did not realized the estimate of the equations with regional dummies and structural variable for the shortened period.

[3] We miss the real GDP of Malta in 1996.
[4] See http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/; the database is not full therefore we miss the values of the variable for Malta 1996 and 2003 and for Greece in 1996.

Period	Basic Equations	Equations with Regional Dummies	Equations with Regional Dummies and Structural Variable
1996-2009	-0.147737 **	-0.162475 **	-0.179214 **
1996-2008	-0.164068 **	-0.184547 **	-
1996-2002	-0.037901 *	-0.054038 **	-0.082054 *
2003-2009	-0.090424 **	-0.089922 **	-0.086456 **

Table 2 The estimates of the β coefficients

We can estimate the average yearly convergence rate as[5]:

$$\textit{average convergence rate} = -\frac{\log(\beta+1)}{T} \tag{4}$$

The values of the estimated average yearly convergence in each period and for each specification are computed in the table 3:

Period	Basic Equation	Equations with Regional Dummies	Equations with Regional Dummies and Structural Variable
1996-2009	1.1419%	1.2665%	1.4107%
1996-2008	1.3795%	1.5693%	-
1996-2002	0.5520%	0.7936%	1.2231%
2003-2009	1.3540%	1.3461%	1.2918%

Table 3 The estimates of the average yearly convergence values

4 Clubs of Countries with Different Steady State Level

We graph the convergence lines from the univariate regressions using the (1) specifications in the figure 1 (period 1996-2009) and figure 2 (period 2003-2009). We can determine countries that could converge towards the higher (their position is above convergence line) or the lower (their position is below convergence line) steady state level. The both graphs imply that Italy, Hungary, Portugal and Spain could converge towards the lower steady state level and Luxembourg could converge towards the higher steady state level all time.

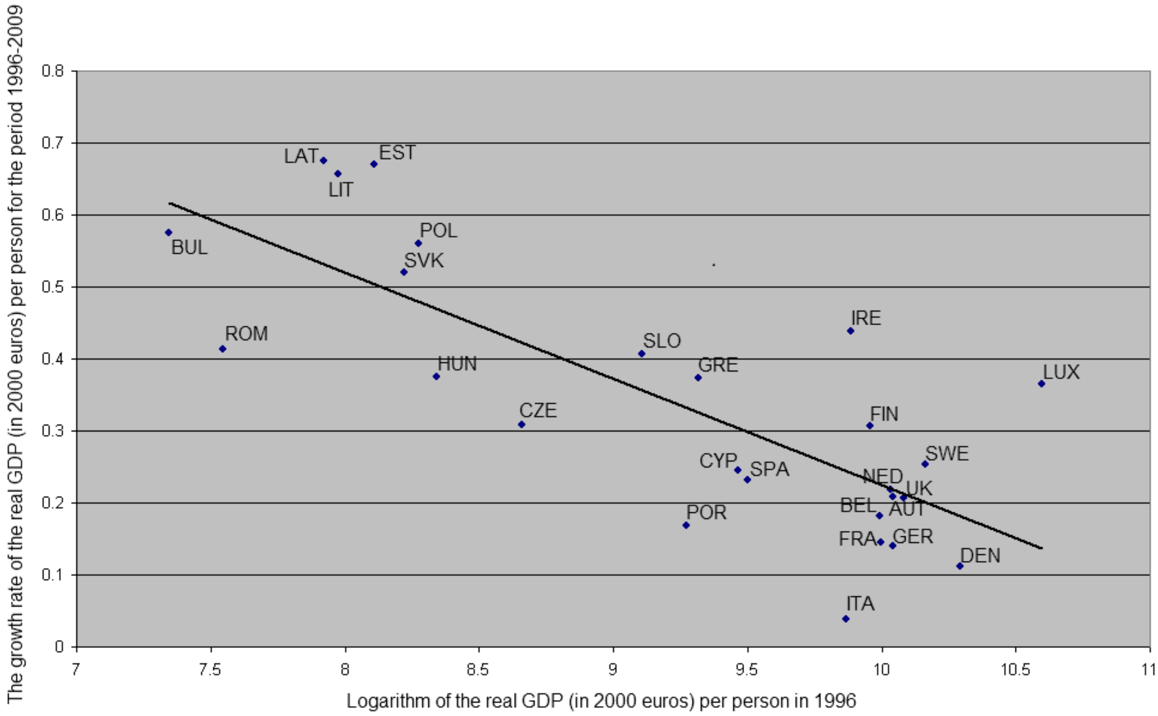


Figure 1 Convergence line, 1996-2009 period

5 See Barro and Sala-i-Martin [6], chapter 11

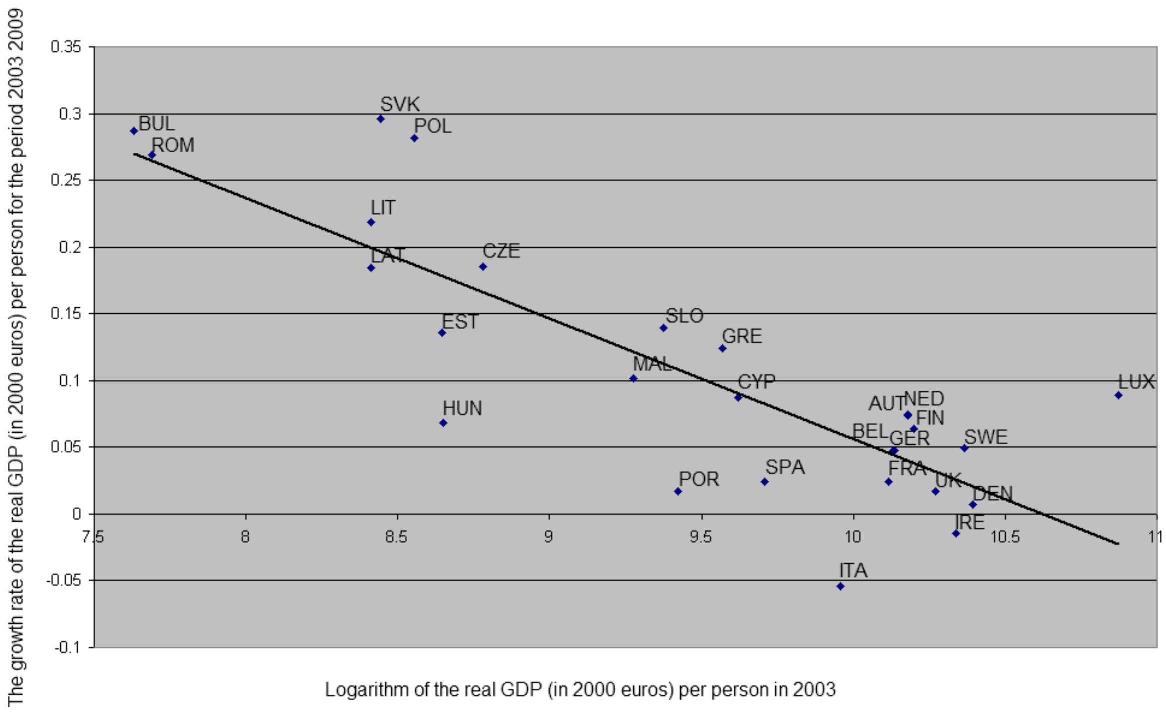


Figure 2 Convergence line, 2003-2009 period

Comparing the both figures it seems that the Czech Republic and Romania have improved their positions towards the convergence line and Slovakia and Poland have improved their position outwards the convergence line in the last seven years. On the other hand Ireland and Baltic countries (Estonia, Latvia and Lithuania) have made worse their position towards the convergence line in the last seven years.

We can statistically verify whether the countries have different steady state level as it has an European average – as it comes from both figures 1 and 2 . We estimated the equations in the forms:

$$g_{i,1996,2009} = 0.1733 - 0.1786 y_{i,1996} - 0.1487 pish + 0.2206 lux + 0.2420 s_{i,5,1996,2009} + v_{i,1996,2009} \quad (5)$$
$$(4.7035)(0.0332) \qquad (0.0546) \qquad (0.1226) \qquad (0.6450) \qquad \mathrm{R}^2 = 0.7586$$

$$g_{i,2003,2009} = 3.4349 - 0.0865 y_{i,2003} - 0.1043 pish + 0.1360 lux - 0.3402 s_{i,5,2003,2009} + v_{i,2003,2009} \quad (6)$$
$$(1.8735)(0.0124) \qquad (0.0187) \qquad (0.0420) \qquad (0.2672) \qquad \mathrm{R}^2 = 0.9015$$

where *pish* is dummy variable with ones for Italy, Hungary, Portugal and Spain and zeros for other countries, *lux* is dummy variable with one for Luxembourg and zeros for other countries, $s_{i,5,t,t+T}$ is variable computed by the equation (3), $g_{i,t,t+T}$ is the growth rate of the real GDP per person and $y_{i,t}$ is the initial level of the real GDP per person. The values in parenthesis are standard deviations of the estimates of the coefficients. The coefficient by *lux* in 1996-2009 equation is statistically significant only at the 10 percent level.

Using equations (5) and (6) the average European convergence rate is 1.41% (1996-2009) or 1.29% (2003-2009). It comes from the neoclassical growth theory – assuming that the technology level is the same in the Europe – that a coefficient multiplying a dummy variable, corresponding to a club of countries, is the product of the minus β coefficient (a coefficient by $y_{i,t}$ in the equation) and the difference between logarithms of the real GDP per person steady state level of this club and the European average.[6] Therefore we can estimate the relative difference between the real GDP per person steady state levels of *pish* countries (or Luxembourg) and the European average.

$$\left(y^*_{pish} - y^*\right)_{1996,2009} = \frac{-0.1487}{0.1786} = -0.8323 \quad (7)$$

[6] See Barro and Sala-i-Martin [6], chapters 2 and 11.

$$\left(y^*_{pish} - y^*\right)_{2003,2009} = \frac{-0.1043}{0.0865} = -1.2060 \quad (8)$$

$$\left(y^*_{lux} - y^*\right)_{1996,2009} = \frac{0.2206}{0.1786} = 1.2346 \quad (9)$$

$$\left(y^*_{lux} - y^*\right)_{2003,2009} = \frac{0.1360}{0.0865} = 1.5714 \quad (10)$$

The asterisk in the upper index denotes the steady state level; the lower index in the form t, $t + T$ denotes the equation that was used for estimates of (5) or (6). Using the Wald coefficient restriction test the difference of the Luxembourg steady state level from the European average is not statistically significant at the 10 percent level; the other estimates are significant at the 5 percent level.

The real GDP per person steady state level of Italy, Hungary, Portugal and Spain has been by 0.83 % (for the period 1996-2009) or 1.21 % (for the period 2003-2009) lower than the European average. On the other hand, using the estimate for the period 2003-2009 the Luxembourg real GDP per person steady state level has been by 1.57 % higher.

5 Conclusion

We conclude that there has been a stronger unconditional (absolute) convergence of EU27 countries in the period 2003-2009 than it was in the period 1996-2002. The average yearly convergence rate has been about 1.3% for last seven years. However, we found the different convergence club as well. Italy, Hungary, Portugal and Spain have been converging towards the lower steady state level and Luxembourg has been converging towards the higher steady state level. The differences have enlarged in last seven years.

Acknowledgements

This paper is supported by the Grant Agency of Slovak Republic - VEGA, grant no. 1/0595/11 "Analysis of Business Cycles in the Economies of the Euro Area (with Regards to the Specifics of the Slovak Economy) Using Econometric and Optimization Methods".

References

[1] Barro, R. J.: Economic Growth in a Cross Section of Countries. *Quarterly Journal of Economics* **106**, (1991), 407–443.
[2] Barro, R. J.: *Macroeconomics – A Modern Approach*, Thomson South-Western, Mason, 2008.
[3] Barro, R. J., and Sala-i-Martin, X.: Convergence across States and Regions. *Brookings Papers on Economic Activity* **1**, (1991), 107–182.
[4] Barro, R. J., and Sala-i-Martin, X.: Convergence. *Journal of Political Economy* **100**, (1992a), 223–251.
[5] Barro, R. J., and Sala-i-Martin, X.: Regional Growth and Migration: A Japan–United States Comparison. *Journal of the Japanese and International Economies* **6**, (1992b), 312–346.
[6] Barro, R. J., and Sala-i-Martin, X.: Economic growth, 2nd ed. The MIT Press, Cambridge, 2004.
[7] Baumol, W. J.: Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show. *American Economic Review* **76**, (1986), 1072–1085.
[8] DeLong, J. B.: Productivity Growth, Convergence, and Welfare: Comment. *American Economic Review* **78**, (1988), 1138–1154.
[9] Sorensen, P.B., and Whitta-Jacobsen, H.J.: *Introducing Advanced Macroeconomics: Growth and Business Cycles* 2nd ed. The McGraw Hill, New York, 2010.
[10] Surmanová, K., and Furková A.: Does the Entrance of the Country into European Union Influence an Inflation Convergence? In *Proceedings of Quantitative methods in economics: multiple criteria decision making XIII* (Pekár, J., and Lukáčik, M. eds.), Bratislava, 2006, 176–181.

Teaching OR and Econometrics: A Case for Emphasizing Links between Graphs and Matrices

Irena Šindelářová[1], Jan Zouhar[2]

Abstract. In a directed graph with weighted edges, there is a relationship between paths in the graph and the inversion of a certain matrix that contains information about the edge weights. To an OR/econometrics student, this relationship is sometimes mentioned in connection with structural analysis. However, as we show in the paper, this relationship can be used to easily derive some results in other fields connected with OR and econometrics – namely, in the study of Markov chains and structural equation modelling (SEM). Traditionally, these results are arrived at using the means of matrix algebra. As the use of graphical models of real-world problems seems to be very instructive and easy to grasp for students, we advocate laying an emphasis on the graphical approach in teaching the aforementioned problems.

Keywords: OR education, econometrics education, graph theory.

JEL classification: A22
AMS classification: 97M40, 97K30

1 Introduction

The mathematical representation of many problems in the field of OR and econometrics can start either with a matrix-form description of the problem, or with a depiction using a weighted directed graph – where both the matrix and the graph contain the same information. For example, the mathematical models in structural analysis and Markov chains are of this sort. As far as the task of finding the solutions to these problems is concerned, the initial model representation doesn't really matter; for the sake of computer realization of the resulting calculations, most of the results are typically expressed in matrix form. However, as we argue later on, in teaching these subjects, the choice of modelling framework *does* make a difference, and we are in favour of taking the graphical approach wherever possible.

In our experience, the use of graphs in OR and econometrics education at the Czech universities has witnessed a gradual decline over the recent years; in several areas, it gave way to the purely matrix-algebraic approach. As an instance, look at the theory of Markov chains: while the older textbooks (such as [11]) used to start the exposition of Markov chains with simple graphs describing the Markov states and their possible transitions, the more recent textbooks [6], [8] restrict themselves solely to matrix algebra in the treatment of the underlying theory.

The reasons for this seem quite clear: matrix notation is very compact (compared to graphs) and both the results and their derivations can easily be taken down in full generality – while graphs tend to be describing a specific instance of the problem in question. Therefore, the matrix approach allows for a very concise treatment of the subject, and in turn, cuts down on both paper consumption (for textbooks and students' lecture notes) and teachers' time needed to cover the topics in their lectures.

However, from the pedagogical point of view, graphs possess several advantages over matrices; we tried to identify the key ones in the list below:

- *Graphs help visualize the problem.* Visualization has long been accepted as one of the basic concepts in education. For some reason, mathematics teachers seem to be reluctant to step away from the rigorous symbolism of the traditional mathematical language towards the "soft" type of knowledge contained in images, as noted in [2]. Graph theory, however, can serve as a useful tool containing a

[1]University of Economics, Prague, Dept. of Econometrics, Nám. W. Churchilla 4, Praha 3, 130 67, e-mail: isin@vse.cz
[2]University of Economics, Prague, Dept. of Econometrics, Nám. W. Churchilla 4, Praha 3, 130 67, e-mail: zouharj@vse.cz

piece of "the best of both worlds" – in graph theory, visual concepts can be given a rigorous meaning. For this reason, the use of graph theory in education has been suggested in various fields, some of which have little to do with applied mathematics, see e.g. [9] and [10].

- *Graphs make mathematical proofs serve the right purpose in education.* As Hanna notes in [4], there has been an upsurge in papers on the role of proof in mathematical education since 1990s. Most of these papers stress the fact that the role of proofs differs slightly in research and in education. As Knuth points out in [7], in research questions, the main aim is to *verify* that a statement is true; in education, on the other hand, the proof should *explain why* the statement is true. In our opinion, starting the proofs from a graphical model is more efficient in creating mathematical understanding among students than is the matrix approach. One of the advantages of graphs lies in the visualization principle we discussed before. Another advantage comes from the fact that graphical models naturally use a bottom-up approach that is easy to understand: with the graphical approach, one inevitably has to start with a small-scale example for which a graph can easily be drawn, and only then generalize the results thereof; with matrices, on the other hand, one can proceed with a general statement of the problem right from the outset – which can be viewed as a drawback in education, as it allows for skipping the parts that motivate the subsequent mathematical procedures.
- *Graphs are closely related to the typical process of building a mathematical model.* In his study about effective ways of OR teaching, King [5] points out that the very first step in the mathematical modelling methodology that OR students should be taught is "identifying the key features of the problem in systems terms and, if possible, representing them diagrammatically." For many a problem, its rough diagrammatical depiction is just a small step away from a formal graphical model.

As we have already noted, even in cases where the graphical model is easily understood and a way of solving underlying problem is found, it is often convenient to express the result in terms of a formula that can be fed into various software packages for matrix computations. This is where the links between graphs and matrices come into play (the links we're calling for in the title). In the rest of the paper, we focus on a concrete example of such a link which, if mastered by the students, can both improve their understanding of some of the topics in OR and econometrics and enable them to translate the graphically-derived results into matrix notation and vice versa. We provide three quite different applications of this link: first application comes from structural analysis, second is the analysis of absorbing Markov chains, and third one is the covariance analysis in structural equation models (SEM).

2 Notation and terminology

Throughout the text, we only deal with *directed graphs* with *weighted edges*. An example of such a graph is given in Figure 1. As the example suggests, we consider only *simple* graphs, i.e. graphs without parallel edges pointing in the same direction, and we allow for *loops* (such as the one in node C). In text, we refer to the edge going from node i to node j as $i \to j$. A *path* is a sequence of nodes that can be traced along the edges in the graph (observing the edge directions). The *length of a path* is the number of edges traversed along the path (we also allow for paths of zero length consisting of a single node). The *weight of a path* is the product of the weights of the edges along the path. In Figure 1, the length of the path A $\to$ C $\to$ D is 2 and its weight is $1.2 \times 2.0 = 2.4$. A *trail* is similar to a path, only the edge directions needn't be observed; the *length* and *weight of a trail* are defined analogously. In Figure 1, C $\leftarrow$ A $\to$ B is a trail of length 2 and weight $1.2 \times 0.9 = 1.08$.

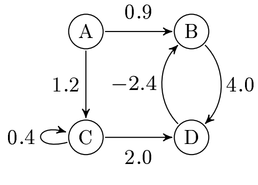


Figure 1: A directed graph with weighted edges.

The information contained in a graph can be represented using a weight matrix: *weight matrix of a graph* is a square matrix where each row and each column correspond to a node in the graph and

$$(i,j) \text{ element} = \begin{cases} \text{weight of } i \to j, & \text{if the graph contains } i \to j, \\ 0 & \text{otherwise.} \end{cases}$$

For instance, the weight matrix of the graph in Figure 1 looks as follows (we used alphabetical *ordering* of nodes here):

$$\begin{array}{c c} & \begin{array}{cccc} \text{A} & \text{B} & \text{C} & \text{D} \end{array} \\ \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} & \begin{pmatrix} 0 & 0.9 & 1.2 & 0 \\ 0 & 0 & 0 & 4.0 \\ 0 & 0 & 0.4 & 2.0 \\ 0 & -2.4 & 0 & 0 \end{pmatrix} \end{array}.$$

3 The missing link

Broadly speaking, the link we mentioned in the introduction part relates the powers of the weight matrix to weights of paths in the graph. In class, we typically present it in the form of two separate propositions.

Proposition 1. *Let $\mathcal{G}$ be a weighted directed graph with weight matrix $\boldsymbol{A}$. For any non-negative integer n, the (i, j) element of $\boldsymbol{A}^n$ equals the sum of the weights of all alternative paths of length n that lead from node i to node j in $\mathcal{G}$, i.e. all paths of the type* $i \underbrace{\rightarrow \ldots \rightarrow}_{n \text{ edges}} j$.

Although a rigorous proof of Proposition 1 by induction is not difficult, we suggest not doing it in class with students who are less mathematically inclined. Doing a small-scale example is typically sufficient in showing how and why the powers of the weight matrix correspond with path weights; a graph with 4 nodes and n going from 1 to 3 will do in most cases. Sometimes, though, students have trouble with seeing why the case $n = 0$ works as well, as they do not understand the *empty product* properly (by definition, the weight of a zero-length path is the empty product, thus equalling 1).

Proposition 2. *Let $\mathcal{G}$ be a weighted directed graph with weight matrix $\boldsymbol{A}$, such that the power series $\sum_{n=0}^{\infty} \boldsymbol{A}^n$ converges. Then the (i, j) element of $(\boldsymbol{I} - \boldsymbol{A})^{-1}$ equals the sum of the weights of all alternative paths leading from node i to node j in $\mathcal{G}$, i.e. all paths of the type* $i \rightarrow \ldots \rightarrow j$.

Proposition 2 is in fact a simple corollary of proposition 1, using the identity $(\boldsymbol{I} - \boldsymbol{A})^{-1} = \boldsymbol{I} + \boldsymbol{A} + \boldsymbol{A}^2 + \boldsymbol{A}^3 + \ldots$, which even a slower student can prove in no time (given that $\sum_{n=0}^{\infty} \boldsymbol{A}^n$ converges).

It's especially Proposition 2 that has many applications in OR and econometrics; the next sections will describe three quite different instances of its use.

4 Example one: Structural analysis

Consider the following classroom exercise in structural analysis:

The production processes carried out by a manufacturer are described by the assembly tree (or, *Gozinto graph*, see [3, p. 94]) in Figure 2. If the weight of edge $i \rightarrow j$ equals w, the manufacturer uses up w units of i to produce a unit of j. Parts A, B, C are purchased from suppliers and the remaining parts are assembled from A, B, C.

a) The purchase/assembly costs of parts A through H are given in Table 1; in case of parts D through H, these costs relate only to the assembly in the last stage of production and do *not* include the purchase/assembly costs of the part's components. Find the *total cost* connected with each part, i.e. the cost *including* the purchase/assembly of all components of a part.

b) The current stock (*inventory*) and the desired stock (*orders*) of parts A through H are given in Table 1. Find the amounts of all parts that need to be purchased/assembled in order to achieve the desired stock.

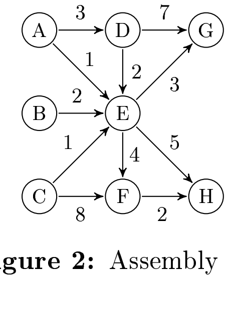


Figure 2: Assembly tree.

Part	A	B	C	D	E	F	G	H
Cost ($\boldsymbol{c}$)	2	6	3	9	4	10	15	7
Inventory ($\boldsymbol{i}$)	30	16	25	15	0	0	10	5
Orders ($\boldsymbol{o}$)	0	0	0	0	20	70	50	45

Table 1: Cost, inventory, and orders.

In the spring semester this year, we confronted our students with this problem in the form of a home assignment. For solving part *b*), nearly all students used the *Leontief formula* (see [3, p. 97]) they had been taught in class, i.e. they computed $(\boldsymbol{I}-\boldsymbol{A})^{-1}(\boldsymbol{o}-\boldsymbol{i})^{\top}$, where $\boldsymbol{A}$ is the weight matrix of the assembly tree. In part *a*), however, most of them did not take the matrix approach. Instead, they used the graph in order to trace, for instance, the number of parts A through E needed to produce a unit of G. The idea here is quite simple and easily seen from the graph, and can be expressed as follows:

> *The number of units of i needed to produce a unit of j equals the sum of the weights of all paths going from i to j in the assembly tree.*

For instance, the number of units of A needed to produce a unit of G equals the sum of the weights of paths A $\to$ D $\to$ G, A $\to$ D $\to$ E $\to$ G, and A $\to$ E $\to$ G, which gives $3\times 7+3\times 2\times 3+1\times 3=42$. If the students knew the link between paths and weight matrices from Proposition 2, they would have probably realized that this implies the following:

> *The number of units of i needed to produce a unit of j equals the (i,j) element of $(\boldsymbol{I}-\boldsymbol{A})^{-1}$.*

The solution of part a) is then easily computed as $\boldsymbol{c}(\boldsymbol{I}-\boldsymbol{A})^{-1}$. Note that the interpretation of the elements of $(\boldsymbol{I}-\boldsymbol{A})^{-1}$ can be derived from the Leontief formula as well [by using vectors of the type $(1,0,0,\ldots),(0,1,0,\ldots)$, etc., instead of $(\boldsymbol{o}-\boldsymbol{i})$]. However, tracing the paths in the graph seems to be a more natural thing to do than translating the problem into matrix manipulations.

5 Example two: Absorbing Markov chains

As we've already noted in the introduction, Markov chains (MCs) are one of the subjects where graphical models have become omitted from education recently. For the reasons we stated before, we don't think this is a good idea. In this example, we show how the link between graphs and matrices from Proposition 2 can help derive some of the theoretical results in a very comprehensible way.

Figure 3 shows a simple example of an absorbing MC. The nodes represent *states* of the MC, the edges are possible *transitions* between succesive time periods, and the edge weights are *conditional transition probabilities*. Here, A, B, C are the *transient states* of the MC, and D, E are the *absorbing states*. The weight matrix of the MC's graph is refered to as the *transition matrix*, typically denoted $\boldsymbol{P}$. Using a suitable ordering of the states, the transition matrix can be decomposed into blocks in the following way (zeros were replaced with blanks in the middle term):

$$\boldsymbol{P} = \begin{array}{cc} & \begin{array}{ccc|cc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} \end{array} \\ \begin{array}{c} \text{A}\\ \text{B}\\ \text{C}\\ \text{D}\\ \text{E} \end{array} & \left(\begin{array}{ccc|cc} & q_{\text{AB}} & & r_{\text{AD}} & \\ & & q_{\text{BC}} & & r_{\text{BE}} \\ q_{\text{CA}} & & & r_{\text{CD}} & r_{\text{CE}} \\ \hline & & & 1 & \\ & & & & 1 \end{array}\right) \end{array} = \begin{array}{rc} & \begin{array}{c|c} \text{trans.} & \text{absor.} \end{array} \\ \begin{array}{r} \text{transient states} \\ \text{absorbing states} \end{array} & \left(\begin{array}{c|c} \boldsymbol{Q} & \boldsymbol{R} \\ \hline \mathbf{0} & \boldsymbol{I} \end{array}\right) \end{array}.$$

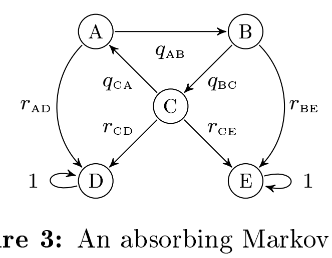


Figure 3: An absorbing Markov chain.

One of the characteristics of an absorbing MC is the conditional probability that the MC will finish in a particular absorbing state, given the current state of the MC. Let's assume that the current state in Figure 3 is A, and we're interested in calculating the probability of finishing in D. Looking at the graph, students typically come up with something like the following idea: "The probability of finishing in D is the sum of the probabilities of all alternative paths leading from A to D." A more general and precise statement of this is:

> *The probability that a MC finishes in absorbing state j given that the current state is i equals the sum of the weights of all paths leading from i to j and avoiding the loop in j.*

In order to use Proposition 2 and express this statement in matrix terms, we first need to drop the loops in absorbing states from the graph. The resulting weight matrix has the form

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{Q} & \boldsymbol{R} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

Now we can use Proposition 2 to conclude:

> *The probability that a MC finishes in absorbing state j given that the current state is i equals the (i,j) element of $(\boldsymbol{I} - \boldsymbol{A})^{-1}$.*

It's straightforward to check by direct multiplication that

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \begin{pmatrix} (\boldsymbol{I} - \boldsymbol{Q})^{-1} & (\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix},$$

where we used $\boldsymbol{I}$ as a generic symbol to denote identity matrices of appropriate ranks. The upper right block, $(\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R}$, is a matrix that contains the probabilities of finishing in a particular absorbing state when starting from a transient one (this is the result that is normally arrived at using purely algebraic means).

6 Example three: Paths and covariances in SEM

Structural equation modelling ranks among the disciplines of multivariate statistical analysis. Its roots date back to 1920s, when the American geneticist and statistician Sewall Wright started using what he called *path diagrams* to represent a qualitative type of information about a system of random variables; the aim was to explicitly incorporate prior knowledge about causality in correlation analysis (see [12]). The method that provides a computational link between path diagrams and correlations is called the *method of path coefficients*, described at length in [13]. In the 1960s and 1970s, the method has been adapted by numerous sociologists and economists, and gradually evolved into SEM; see [1] for an engaging review of the history of path analysis and SEM, together with a debate on the views on causal interpretation of the results. In this section, we show how the link between graphs and matrices from Section 3 enables us to easily derive a relationship that is analogous to the method of path coefficients, only that we consider covariances instead of correlations.

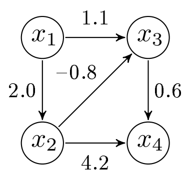


Figure 4: A path diagram.

A path diagram with four random variables x_1 through x_4 may look like the one in Figure 4. In practical applications, the edges typically represent causal links between the variables; for the sake of mathematical modelling, we treat the diagram as the definition of the following *linear functional relationships* between the variables: each node is a weighted sum of its parents (the weights being the edge weights) and a disturbance term. In Figure 4, this represents the following system of linear equations:

$$\begin{aligned} x_1 &= \varepsilon_1\,, \\ x_2 &= 2.0x_1 + \varepsilon_2\,, \\ x_3 &= 1.1x_1 - 0.8x_2 + \varepsilon_3\,, \\ x_4 &= \phantom{1.1x_1 - {}} 4.2x_2 + 0.6x_3 + \varepsilon_4\,, \end{aligned} \tag{1}$$

where ε_1 through ε_4 are the disturbance terms. For simplicity, we assume that the disturbances are pairwise uncorrelated and all have unit variances (however, the analysis can easily be extended for the case both assumptions are dropped). If we denote by $\boldsymbol{A}$ the weight matrix of the graph in Figure 4, and define vectors $\boldsymbol{x} = (x_1, \ldots, x_4)^\top$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_4)^\top$, we can rewrite (1) as

$$\boldsymbol{x} = \boldsymbol{A}^\top \boldsymbol{x} + \boldsymbol{\varepsilon}. \tag{2}$$

If, for brevity, we denote by $\boldsymbol{B}$ the matrix $(\boldsymbol{I}-\boldsymbol{A})^{-1}$, we can write the solution of (2) for $\boldsymbol{x}$ as $\boldsymbol{x}=\boldsymbol{B}^{\top}\boldsymbol{\varepsilon}$. Using this to calculate the variance-covariance matrix of $\boldsymbol{x}$ gives $\operatorname{var}\boldsymbol{x}=\operatorname{var}(\boldsymbol{B}^{\top}\boldsymbol{\varepsilon})=\boldsymbol{B}^{\top}\boldsymbol{B}$, where we used the assumption that $\operatorname{var}\boldsymbol{\varepsilon}=\boldsymbol{I}$. Therefore, we can express the covariance between x_i and x_j as

$$\operatorname{cov}(x_i,x_j)=\textstyle\sum_k b_{ki}b_{kj}, \tag{3}$$

where b_{ki} is the (k,i) element of $\boldsymbol{B}$. If we now apply Proposition 2 to the right-hand side in (3), the summands have an interesting interpretation: since b_{ki} sums the weights of all paths of the type $k\to\ldots\to i$ and b_{kj} does the same for paths $k\to\ldots\to j$, the term $b_{ki}b_{kj}$ sums the weights of all trails of the type $i\leftarrow\ldots\leftarrow k\to\ldots\to j$. Note that these trails contain no *head-to-head node*, i.e. a node of the type "$\ldots\to l\leftarrow\ldots$". This brings us to the following conclusion:

The covariance between x_i and x_j equals the sum of the weights of all trails between nodes x_i and x_j that contain no head-to-head nodes.

This relationship is closely related to the method of path coefficients (compare with [13, p. 193]). The latter allows for correlated disturbances; though we do not show it here, the analysis presented above can easily be extended to allow for a general variance-covariance matrix of the disturbance vector $\boldsymbol{\varepsilon}$.

7 Conclusion

In most of this paper, we focused on showing how a particular link between graphs and matrices can help students in understanding three quite different topics in OR and econometrics. Nevertheless, we'd like to stress that the primary concern of the paper is to promote graphical approaches in teaching *in general*, as we put forward in the introduction.

Acknowledgements

This work was supported by the Internal Grant Agency of the University of Economics, Prague, project no. F4/18/2011.

References

[1] Denis, D. J., Legerski, J.: Causal modeling and the origins of path analysis. *Theory & Science* [online], **7**(1) (2006). URL: <http://theoryandscience.icaap.org/content/vol7.1/denis.html>

[2] Eisenberg, T., Sheva, B.: On understanding the reluctance to visualize, *Zentralblatt für Didaktik der Mathematik*, **26**(3) (1994), 109–113.

[3] Fiala, P.: *Modely produkčních systémů.* Oeconomica, Praha, 2005.

[4] Hanna, G.: Proof, explanation and exploration: an overview, *Educational Studies in Mathematics*, **44**(1) (2000), 5–23.

[5] King, M.: A framework for teaching OR techniques, *The Journal of the Operational Research Society*, **39**(12) (1988), 1087–1093.

[6] Kořenář, V.: *Stochastické procesy.* Oeconomica, Praha, 2010.

[7] Knuth, E. J.: Secondary school mathematics teachers' conceptions of proof, *Journal for Research in Mathematics Education*, **33**(5) (2002), 379–405.

[8] Prášková, J., Lachout, P.: *Náhodné procesy I.* Karolinum, Praha, 2005.

[9] Šedivý, J.: Teaching of elements of logic by means of finite graphs, *Educational Studies in Mathematics*, **1**(4) (1969), 422–439.

[10] Tatsuoka, M. M.: Graph theory and its applications in educational research: a review and integration, *Review of Educational Research*, **56**(3) (1986), 291–329.

[11] Unčovský, L.: *Stochastické modely operačnej analýzy.* SNTL, Praha, 1980.

[12] Wright, S.: Correlation and causation, *Journal of Agricultural Research*, **20** (1921), 557–585.

[13] Wright, S.: The method of path coefficients, *Annals of Mathematical Statistics*, **5** (1934), 161–215.

Using Markov Chains in Project Management

Štěpánka Škulavíková[1], Petr Fiala[2]

Abstract. Project management is the discipline of planning, organizing, securing and managing resources to bring about the successful completion of specific project objectives. The project is defined as a unique, complicated process with the beginning, the end and limited budget and resources. Risk is a very important factor in project management. Projects are in accelerating world rhythm the right option of solving problems of lot of companies. Lot of professionals tried to find sophisticated way to improve techniques for project management in different ways. This paper is focused on idea, that project can be also seen as a system that undergoes transitions from one state to another one, so we can implement Markov chains. The question paper should answer is possibility of facing risks and uncertainty in project duration and cost predictions.

Keywords: project management, Markov chain, earned value analysis

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Project management is the discipline of planning, organizing, securing and managing resources to bring about the successful completion of specific project objectives. Projects are in accelerating world rhythm the right option of solving problems of lot of companies. Lot of professionals tried to find sophisticated way to improve techniques for project management in different ways. Nothing is permanent, everything is temporary, and that makes pressure on companies to finish new products or services faster, cheaper and definitely not to fail. Risk is a very important factor in project management. This study paper should show connection between stochastic methods and project management to analyze risk in projects. The paper is organized as follows. Section 2 will explain main terms of project management discipline. Basics of Markov chains are introduced in section 3. Section 4 is devoted to explanation of earned value analysis. , and association with project management.

2 Project Management

For effective managing increasing amount of projects in various dimensions, qualified managers use experienced methods to deal with organizing projects. A project is a unique endeavor to produce a set of deliverables within clearly specified time, cost and quality constraints (see [7]). Projects are different from standard business operational activities as they:

- Are unique in nature.
- Have a defined timescale.
- Have an approved budget.
- Have limited resources.
- Involve an element of risk.
- Achieve beneficial change.

Because of clear definition of projects, we can without problems say when managers can use their skills, to organize process of successful running of project by using special methods and tools. In that moment we speak about the project management. These days there is project management implemented in every big company. The necessity of dealing with dynamic environment is slowly getting into surroundings of middle and small companies. For today's business profit is a magic word. Searching for a possibility of less risk and most successful running company is way how to reach profit is natural aspect. That is exactly mission of project management theory in practice.

[1] University of Economics in Prague, Department of Econometrics, Czech republic, Prague 3,W. Churchill Sq. 4, 13067, stepanka.skulavikova@gmail.com
[2] University of Economics in Prague, Department of Econometrics, Czech republic, Prague 3,W. Churchill Sq. 4, 13067, pfiala@vse.cz

Project management is based on definition of projects and is closely correlated with project life cycle. Each project is facing to time limits, costs restricts, allocated resources and internal as well as external risks. If project misses one of the restrictions, impact is usually height. All those indicators are watched during all project live cycle. The project life cycle typically passes sequentially through four stages (see [2]): definition, planning, execution, and delivery. The starting point begins the moment the project is given to go-ahead. Project effort starts slowly, builds to a peak, and then decline to delivery of the project to the customer.

1. **Definition stage:** Specifications of the project are defined; project objectives are established; teams are found major responsibilities are assigned.
2. **Planning stage:** The level of effort increases, and plans are developed to determine what the project will entail, when it will be scheduled, whom it will benefit, what quality level should be maintained, and what the budget will be.
3. **Execution stage:** A major position of the project work takes place-both physical and mental. The physical product is produced (a bridge, a report, a software program). Time, cost, and specification measures are used for control. Is the project on schedule, on budget, and meeting specifications? What are the forecasts of each of these measures? What revisions/changes are necessary?
4. **Delivery stage:** Includes the two activities: delivering the project product to the customer and redeploying project resources. Delivery of the project might include customer training and transferring documents. Redeployment usually involves releasing project equipment/materials to other projects and finding new assignments for team members."

Tools for project management have a lot of specification. Despite high variety of projects, basic structure is for all of them the same. Managers have to solve always new problems, but still can find helpful information and advises from projects realized in past. One of the most important is information which helps for quality forecasting. We can identify lot of factors, which influent the quality of estimates. Most important factors are planning horizon, project duration, project structure and organization, human factor, padding estimates and organization culture. It is part of project managers' work to make a detail view of project in many aspects. Good quality project plan depends also on well done cost and risk analysis.

2.1 Cost Analysis

It is necessary to have detail cost analysis for our project. Anytime during project can clear cost analysis save whole project. It is obvious when and where is space to save some money and extremely important time schedule of all expected expenses. There is more kind of costs which have to be monitored. "In [7] says that cost management is the process by which costs/expenses incurred on the project are formally identified, approved and paid. Expense forms are completed for each set of related project expenses such as labor, equipment and materials costs. Expense forms are approved by the project manager and recorded within an expense register for auditing purposes."

Cost analyses are kept updating whole project life. They are part of weekly minutes and indispensable part of project reports. Important benefit of numbers from those analyses is on time notification about potential complications.

2.2 Risk Analysis

To deliver project successfully to the customer is risk analysis one of the key action. With perfect plan of all detail activities, resources and costs, we have still no guarantee of achievement. Reality shows that even really deep risk analysis does not have to protect us from everything, but can minimize impact of potential problems. There are a few rules which provide useful complex analysis. "In [7] a comprehensive risk plan includes a:

- list of the foreseeable project risks;
- rating of the likelihood of each risk occurring;
- description of the impact on the project should a risk actually occur;
- rating of the overall importance of each risk;
- set of preventative actions to be taken to reduce the likelihood of the risk occurring;
- set of contingent actions to be taken to reduce the impact should the risk eventuate;
- process for managing risk throughout the project."

If project manager fallows list of tasks above, gets strong tool to avoid critical moments in project life. Risk analysis should be kept updated till the end of project. Any of potential risks included in risk analysis should not ruin project under the hands of responsible manager. All the risks which we expect won't have as serious consequences as if we would skip them in our project plan preparation. Thanks contingent plan, manager won't be so surprise when problem come. Risk analysis is very extensive topic which could cover a separate paper.

3 Markov Chain

Lot of literature define project as a stochastic process. Project is seen as a process which goes from one possible situation to another. This idea is actually similar to the Markov chain definition. That open for us way of application theory of Markov chain in theory of project management. Markov chain can be defined like a set of states with set up probability of transition from any state to any other, or to itself.

A Markov chain can be describe as follows (see [3]): We have a set of states, $S = \{s_1, s_2, \ldots, s_n\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state in moment 0. Usually this is done by specifying a particular state as the starting state.

All probabilities of transition define Markov chain and are specified in transition matrix. The transition matrix:

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{nn} \end{bmatrix}$$

Prediction of the future state in moment m can be described by multiplying vector of state probabilities in moment m -1 by the transition matrix P:

$$p(m) = p(m-1)P = p(0)P^m$$

Markov Chain Classification

The subject of Markov chains is best studied by considering special types of Markov chains. Stochastic model known as a Markov chain is divided to two main branches absorbing and ergodic (see [3]).

The first type that we shall use is called an absorbing Markov chain. A state si of a Markov chain is called absorbing if it is impossible to leave it (i.e., $p_{ii} = 1$). A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step). In an absorbing Markov chain, a state which is not absorbing is called transient."

A second important kind of Markov chain we shall use is an ergodic Markov chain, defined as follows. A Markov chain is called an ergodic chain if it is possible to go from every state to every state (not necessarily in one move). In many books, ergodic Markov chains are called irreducible.

A Markov chain is called a regular chain if some power of the transition matrix has only positive elements.

4 Earned Value Analysis in Project Execution

Earned Value Analysis (EVA) is important tool used by many project managers to coordinate plan and real project situations. Some basics of EVA are summarized. For calculation of EVA is unavoidable to define basic terms.

4.1 Earned Value Analysis

Earned value analysis (see [5]) is an industry standard way to:

- measure project's progress (Time spent),
- forecast the completion date and final cost (Work Accomplished),
- project schedule and cost variance if any (Money spent).

So EVA is used anytime of project cycle and displays real worth of all factors of project health. EVA is a snapshot in time, which can be used as a management tool as an early warning system to detect deficient or endangered progress (see [8]). It ensures a clear definition of work prior to beginning that work. It provides an objective measure of accomplishments, and an early and accurate picture of the contract status. It can be as simple as tracking an elemental cost estimate breakdown as a design progresses from concept through to 100% construction documents, or it can be calculated and tracked using a series of mathematical formulae (see below). In either case, it provides a basis for course correction. It answers two key questions:

1. At the end of the project, is it likely that the cost will be less than, equal to or greater than the original estimate?
2. Will the project likely be completed on time?"

EVA gives us the most important answers to project managers, customers, project owner or top management from involved companies. The results are useful background for project team weekly-meetings. As project is finishing EVA results give higher quality knowledge about project situations with progressed stage.

4.2 EVA Calculation

EVA measures progress against a baseline (see [8]). It involves calculating three key values for each activity in the WBS:

1. The Planned Value (PV), (formerly known as the budgeted cost of work scheduled or BCWS)—that portion of the approved cost estimate planned to be spent on the given activity during a given period.
2. The Actual Cost (AC), (formerly known as the actual cost of work performed or ACWP)—the total of the costs incurred in accomplishing work on the activity in a given period. This Actual Cost must correspond to whatever was budgeted for the Planned Value and the Earned Value (e.g. all labor, material, equipment, and indirect costs).
3. The Earned Value (EV), (formerly known as the budget cost of work performed or BCWP)—the value of the work actually completed."
4. These three values described above are used in equations to describe project situation.

Cost Variance (*CV*): $CV = EV - CV$

Schedule Variance (*SV*): $SV = EV - PV$

A negative schedule variance (SV) calculated at a given point in time means the project is behind schedule, while a negative cost variance (CV) means the project is over budget (see [8]).

In converted form we get efficiency indicators. Those values are used to forecast total project realization estimates.

Cost performance index (CPI): $CPI = \frac{EV}{AC}$

Schedule performance index (SPI): $SPI = \frac{EV}{PV}$

One of the main advantages is simple calculation. Not all of the EVA is easy way to understand real project situations for project managers. Next advantage is lucidity of graphic visualization, which is the most powerful tool to explain data to not deep involved interested persons.

4.3 Markov Chains in Project Management

As we said above, we are able to make a snapshot of project in anytime. We are also able to identify phase, or an activity project is in. Thanks EVA and other analyses we can compare real situation to plan and count the variance. Depending on results of all analyses, project manager will make his decisions.

As a very helpful tool, Markov chains theory can be applied to process of project prediction. There are two different view of this combination. One sees project as a Markov chain with tasks as states, another one sees health situation as states of chain. Both methods will be described below.

Project Activities as a Markov Chain States

Based on definition of project activities, we can describe exact position of project anytime. There are tree possibilities for next step of project. Desired situation is project transits to the next planned activity, another possibility is repetition of activity just finished, and less progressive way to come back to activity previous. This theory was described by Hardie N. [4] He made some assumptions to make this model more realistic and easier to count:

- activity durations are all the same;
- repeating an activity takes the same time as the original activity
- the reversion probability is the same the same for each stage to each previous activity
- the project is no linear with no branches

This model doesn't copy real situation too much. That is not realistic to expect same activity duration for all tasks in project. Activities duration us usually changing during project life cycle. Linear project can be expected only for really small project. But we can find some project which will be at least close to described model. Smith and Eppinger [6] in their paper used same model. They use example of small project as a chain which is divided into the three stages that correspond to the cascading process of task completion. Task A is executed first, with a probability of 1.0 of moving on to Task B, which is then attempted and succeeds with probability of 0.8. Task A must then be repeated with a probability of 0.2. A and B then iterate until their results are compatible, at which time task C is executed for the first time. After task C, there is a probability of 0.5 that the design is complete, but with a probability of 0.5 there must be iteration among A, B and C to complete the project. This model allows skip activity B when we come back to activity A after finishing activity B. It means that activity B may not to be repeated again.

Situations as Markov Chain States

Model of applying Markov chains in EVA is presented in [1]. Te modified model is more realistic and universal one. As a stage of stochastic process are identified thirteen possible situations that project can get. Parameters of description those possibilities are budget spending and timing. Project can be in good condition and copy plan, or we can identify any deviation. In general deviations can be negative, positive, or zero. As a measurement are used CPI and SPI values from EVA analysis. Counting is based on geometrical distance of perfect match of plan and real situation and other deviation.

Midpoint is optimal situation without any deviation.

Geometrical midpoint:

$$[SPI; CPI] = [1; 1]$$

The closest four points are one of positive or negative deviation of money or time, or mix of budget and schedule problems. Points farther from midpoint represent deeper problems compare to planned prediction. Transition is expected only between two close points. For geometrical model visualization see Figure 1. Particular coordinates of points were calculated as simultaneous equations.

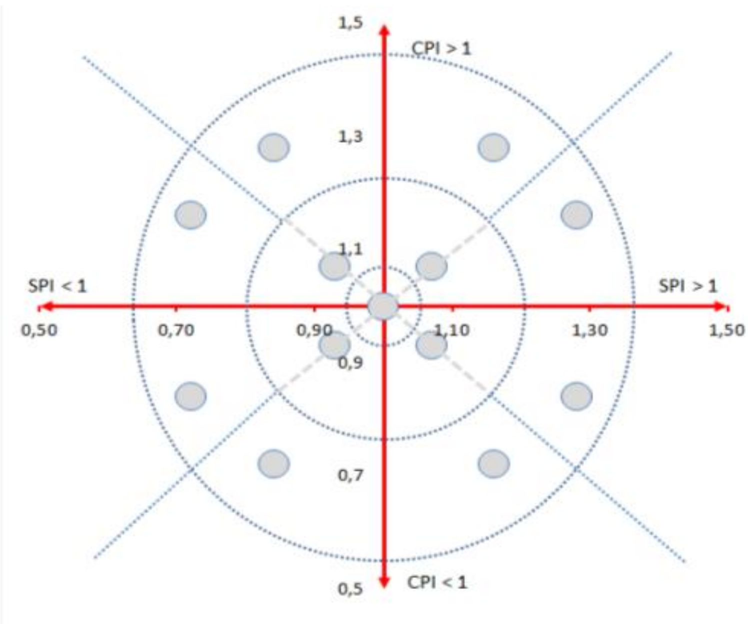


Figure 1 Geometrical interpretation of the model

Equation of circle:

$$(x - X)^2 + (y - Y)^2 = r^2$$

For this model is preferable directive equation of line:

$$y = kx + q$$

Based on CPI and SPI explanation was set up that circles will have midpoint as written above and radius will be 0.1 and 0.325. After that is possible to calculate exact points of model.

Also, important idea of model is that we cannot expect extreme change from serious budget overdraft to under budget situation in one step. The change is smooth. This fact is also described as an inverse relationship between distance between points and probability of transition. Model is centrally symmetrical. Probability of staying in same stage is same as probability of transition to another stage.

This probability can be described by formula:

$$p_{ij} = \frac{\sum_{i,j}^{N} p_{ij}}{N}; \sum_{j}^{N} p_{ij} = 1,$$

where N is the number of neighbor states in the diagram (Figure 1).

Model can be used anytime during life cycle. EVA indicate real situation of project and model helps with prediction of future states of project. Information support is programmed as an application in Microsoft Excel. Application environment combines easy information sharing, user friendly manipulation and facile using. Application allows users really comfortable and fast information for their decision making.

5 Conclusion

Risk as important factor influences prediction in projects. Application of Markov chains in modeling and analysis of risk can be useful. In the paper a simple using of Markov chains in earned value analysis is presented. The approach gives a simple instrument for analyzing the progress in project execution. The basic approach gives possibilities for generalizations. The area is promising for next scientific research.

Acknowledgement

The research project was supported by Grant No. IGA F4/16 /2011, Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Doležal, J. : Predikce v projektech s využitím Markovských řetězců. Dissertation thesis, Brno, 2010.

[2] Gray C. F., and Larson E. W.: *Project Management*, 2nd Editon, Post and Telecom Press, China 2010

[3] Grinstead, C., M, and Snell, J., L.: *Introduction to Probability*, an on-line textbook in pdf format.

[4] Hardie, N.: The prediction and control of project duration: a recursive model. In: *International Journal of Project Management* **19** (2001), 401–409.

[5] *sarawathibhat.blogspot.com* [online]. 2006 [cit. 2011-05-08]. Earned Value Analysis. WWW: <http://saraswathibhat.blogspot.com/2006/04/earned-value-analysis-eva.html >.

[6] Smith, R. P, and Eppiger, S. D.: A predictive Model of Sequential Iteration in Engineering Design, In: *Management Science*, Vol, 43, No. **8** (1997), 1104–1120.

[7] Westland, J.: *The Project Management Life Cycle*, Kogan Page Limited, Great Britain and the United States, 2006.

[8] *www.wbdg.org* [online]. 2005 [cit. 2011-05-06]. Earned Value Analysis. WWW: <http://www.wbdg.org/resources/value_analysis.php >.

A Simple Decision Problem of a Market Maker

Martin Šmíd[1]

Abstract. We formulate a simple decision model of a market maker maximizing an utility from his consumption. We reduce the dimensionality of the problem to one. We find that, given our setting, the quotes set by the market maker depend on the inventory of the traded asset but not on the amount of cash held by the market maker.

Keywords: market microstructure, market makers, decision problem, probability constraints, stochastic optimization

JEL classification: C51, G10
AMS classification: 90C15

1 Introduction

One of the key roles in price formation at today's financial markets is played by market makers (MMs) who are obliged to set buying and selling qoutes and trade for the prices they set. Clearly, as other economic agents, MMs are profit maximizers. The economic analysis of their behaviour is, however, quite complicated since the decision problems they face are usually intractable (see [1], [2] or [3]).

In the present paper, we suggest a rather simple version of such a decision problem. In particular, we assume the MM to maximize (an utility) from his consumption while keeping the probability of the bankruptcy (i.e. running out of the money or the traded asset) at a prescribed, perhaps very small level. We do not give analytic solution of the problem but we reduce its dimensionality. As a result of our analysis, we find that the quotes depend on the inventory of the traded assets but they do not depend on the amount of cash held by the MM.

Even if our model is only single-period one, it does not suffer from the logic of "scorched earth", i.e. the today's actions do not steel from the future to a great extent. The reason for this is that, by keeping the probabilities of crossing zero by the inventory processes very small, the model tends to "keep a distance" from the boundaries. Nevertheless, dynamization of our model could be a promising direction of a further research.

The paper is organized as follows: after a definition of the setting, the model is formulated and partially solved. A short conclusion is finishing the paper.

2 The setting

Let there be two types of agents: the market makers posting quotes (the best bid and ask) and the (liquidity) traders.

We assume the market makers to be homogeneous, forming an oligopoly, so that they may be treated as a single representative agent who sets the quotes A and B (the best ask, bid, respectively) in order to maximize an utility from their consumption. Denote U the corresponding utility function and assume that it is strictly increasing.

In reaction to the quotes, the traders post market orders, which we assume to be unit for simplicity.

Assume the traders to post orders with an intensity depending solely on a distance of the corresponding quote to a fair price $\Pi \in \mathbb{R}$: In particular, the intensity of the arrival of buy orders is assumed to be

$$\kappa(A - \Pi) \geq 0.$$

[1]Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic, Dpt. of Econometrics, Pod Vodárenskou věží4, Praha 8, 182 08, e-mail: smid@utia.cas.cz

while the intensity of the sell orders' arrival is given by

$$\lambda(\Pi - B) \geq 0.$$

Quite naturally, assume that both κ and λ are continuous strictly decreasing, defined on $[0, H_\kappa]$, $[0, H_\lambda]$ respectively, for some $H_\kappa, H_\lambda > 0$ and that

$$\kappa(H_\kappa) = \lambda(H_\lambda) = 0$$

(i.e. nobody wants expensive asset and nobody wants to give anything for free).

3 The decision problem

Denote M and N the amount of money, traded asset, respectively, held by the MM and assume that the MM maximizes the (utility from his) consumption while keeping the probability of running out of the money or the commodity at a prescribed (small) level α i.e. he solves

$$V(M, N, \Pi) = \max_{A,B,C} U(C) \tag{1}$$

$$A, B, C \geq 0, \qquad B \leq A, \qquad \mathbb{P}[M + AX - BY - C \leq 0] \leq \alpha, \qquad \mathbb{P}[N + Y - X \leq 0] \leq \alpha$$
$$X \sim \mathrm{Po}(\kappa(A - \Pi)), \qquad Y \sim \mathrm{Po}(\lambda(\Pi - B)), \qquad X \perp\!\!\!\perp Y.$$

Here, X and Y are random variables counting numbers of buy market orders, sell market orders, respectively, which were posted given the quotes are A and B.

To simplify further analysis, we shall assume that the intensities κ and λ are high enough so that X and Y may be approximated by normal variables, namely

$$X \dot\sim \mathcal{N}(\kappa(A - \Pi), \kappa(A - \Pi)), \qquad Y \dot\sim \mathcal{N}(\lambda(\Pi - B), \lambda(\Pi - B)) \tag{2}$$

Note that, given this assumption,

$$N + Y - X \dot\sim \mathcal{N}\left(N + \lambda(\Pi - B) - \kappa(A - \Pi), \lambda(\Pi - B) + \kappa(A - \Pi)\right),$$
$$M + AX - BY - C \dot\sim \mathcal{N}\left(M + A\kappa(A - \Pi) - B\lambda(\Pi - B) - C, A^2\kappa(A - \Pi) + B^2\lambda(\Pi - B)\right)$$

so the constraints may be approximated by

$$\kappa(A - \Pi) - \lambda(\Pi - B) - N + q\sqrt{\kappa(A - \Pi) + \lambda(\Pi - B)} \leq 0 \tag{3}$$

$$B\lambda(\Pi - B) - A\kappa(A - \Pi) - M + C + q\sqrt{A^2\kappa(A - \Pi) + B^2\lambda(\Pi - B)} \leq 0 \tag{4}$$

where q is the α-quantile of the standard normal distribution.

Proposition 1. *For problem (1) with the approximation (2), the following is true:*
(i) An optimal solution exists for any $M \geq 0, N \geq 0, \Pi \in \mathbb{R}^+$.
(ii) In optimum, (4) is fulfilled with "=".
(iii) In optimum, either $\lambda(\Pi - B) = 0$ or (3) is fulfilled with "=".
(iv) If $\hat{A}, \hat{B}, \hat{C}$ is the optimal solution of (1) then $\kappa(\hat{A} - \Pi), \lambda(\Pi - \hat{B})$ is an optimal solution of

$$v(N, \Pi) = \max_{K,L \geq 0} \left[A(K)K - B(L)L - q\sqrt{A(K)^2 K + B(L)^2 L}\right] \tag{5}$$

$$B(L) \leq A(K),$$
$$K - L - N + q\sqrt{K + L} \leq 0 \tag{6}$$

and $\hat{C} = v(N, \Pi) + M$. Here,

$$A(K) = A(K, \Pi) = \Pi + \kappa^{-1}(K), \qquad B(L) = B(L, \Pi) = \Pi - \lambda^{-1}(L).$$

(v) If $(\hat{K}, \hat{L})$ is the optimal solution of (5) then either $\hat{L} = 0$ or

$$\hat{L} = \Lambda(\hat{K}), \qquad \Lambda(K) = K - N + \frac{q^2}{2} + \frac{q}{2}\sqrt{q_\alpha^2 + 8K - 4N}$$

(vi) Denote

$$\eta(K) = \left[A(K)\left(K - q\sqrt{K}\right)\right], \qquad K_0 = N + \frac{q^2}{2}\left(1 - \sqrt{\frac{4N}{q^2} + 1}\right), \qquad K_L = \kappa(B(0)).$$

If the optimal solution $\hat{K}$ *of the problem*

$$\tilde{v}(N, \Pi) = \max_K \eta(K), \qquad 0 \le K \le K_0 \wedge K_L, \tag{7}$$

coincides with the solution of

$$\dot{v}(\Pi) = \max_K \eta(K), \qquad 0 \le K \le \infty \tag{8}$$

then $(\hat{K}, 0)$ *is the solution of (5).*
(vii) If $(\hat{K}, \hat{L})$, $\hat{L} > 0$ *is an optimal solution of (5) then* $\hat{K}$ *is an optimal solution of*

$$\dot{v}(N, \Pi) = \max_{K \ge 0}\left[A(K)K - B(\Lambda(K))\Lambda(K) - q\sqrt{A(K)^2 K + B(\Lambda(K))^2 \Lambda(K)}\right], \tag{9}$$

$$B(\Lambda(K)) \le A(K), \qquad K \ge \frac{N}{2} - \frac{q}{8}$$

and $\hat{L} = \Lambda(\hat{K})$.

Proof. (i) A strategy $A = \Pi + H_\kappa, B = \Pi - H_\lambda, C = 0$, producing $X \equiv 0$, $Y \equiv 0$, is clearly feasible. The existence of optimal solution then follows from the continuity of κ and λ.
(ii) If $\hat{A}, \hat{B}, \hat{C}$ were optimal and (4) was fulfilled with "$<$" for $A = \hat{A}$, $B = \hat{B}$, $C = \hat{C}$ then there would exist $\bar{C} > \hat{C}$ still fulfilling (4) which is a contradiction with the optimality of $\hat{A}, \hat{B}, \hat{C}$.
(iii) If $\hat{A}, \hat{B}, \hat{C}$ were optimal with $\hat{L} > 0$ (with the consequence that $B > \Pi - H_\lambda$) and if (3) held with "$<$" for $A = \hat{A}$, $B = \hat{B}$ then, from the continuity of the intensities, there would exist $\Delta > 0$ such that (3) is fulfilled for

$$A = \hat{A}, \qquad B = \hat{B} - \Delta, \tag{10}$$

for which, however, $AX - BY$ is less both in expectation and in variance than the same variable given $A = \hat{A}$, $B = \hat{B}$ which implies that (4) is fulfilled with $<$ given (10) yielding the existence of a feasible $\bar{C} > \hat{C}$, i.e. a contradiction to the optimality of $\hat{A}, \hat{B}, \hat{C}$.
(iv) Consider a problem

$$\max_{A, B \ge 0} g(A, B) \tag{11}$$

$$g(A, B) = A\kappa(A - \Pi) - B\lambda(\Pi - B) - q\sqrt{A^2\kappa(A - \Pi) + B^2\lambda(\Pi - B)}$$

$$B \le A, \qquad \kappa(A - \Pi) - \lambda(\Pi - B) - N + q\sqrt{\kappa(A - \Pi) + \lambda(\Pi - B)} \le 0$$

and note that g is the negative of the LHS of (4) without M and C hence, by (ii), for $\hat{A}, \hat{B}, \hat{C}$ optimal to (1) it has to hold that

$$g(\hat{A}, \hat{B}) + M - \hat{C} = 0 \tag{12}$$

We show that $\hat{A}, \hat{B}$ is then the optimal solution of (11): Indeed, if $\hat{A}, \hat{B}$ were not optimal to (11) there would exist $\bar{A}, \bar{B}$, $\bar{B} \le \bar{\ }$, fulfilling (3) such that $g(\bar{A}, \bar{B}) > g(\hat{A}, \hat{B})$. However, then $\bar{A}, \bar{B}, \bar{C}$, $\bar{C} = \hat{C} + (g(\bar{A}, \bar{B}) - g(\hat{A}, \hat{B})) > \hat{C}$ would fulfil both (3) and (4) which is a contradiction to the optimality of $\hat{A}, \hat{B}, \hat{C}$. Finally, note that (11) is equivalent to (5).
(v) Let $\hat{K} \ge 0$,$\hat{L} > 0$ be optimal. By (iii), (3) holds with "=" hence $\hat{L}$ has to solve the equation

$$(\hat{K} - L - N)^2 = q^2(L + \hat{K})$$

whose solutions are

$$L_{1,2} = \hat{K} - N + \frac{q^2}{2} \pm q\sqrt{\frac{q^2}{4} + 2\hat{K} - N}$$

A trivial calculation shows that $\hat{K} - L_1 - N < 0$ which proves that λ_1 is indeed a solution of (3) with "=". Once $\hat{K} - L_2 - N > 0$, L_2 cannot be a solution of (3) with "=" hence L_1 is unique. Assume now that $\hat{K} - L_2 - N \le 0$ i.e. both L_1 and L_2 are candidates for the value of $\hat{L}$. Denote $f(L) = \frac{\hat{K} - L - N}{\sqrt{\hat{K} + L}}$

and note that (3) holds with the equality iff $f(L) = -q$. The fact that $f'(L) = \frac{1}{2(\hat{K}+L)^{3/2}}\left(N - 3\hat{K} - L\right)$ proves that f is possibly increasing starting form zero up to some threshold and decreasing starting from the threshold hence, necessarily, f is increasing in L_2 and decreasing in L_1which implies, however, that $(\hat{K}, 0)$ fulfils (6) and, by arguments similar to those from (iii), the strategy $\hat{K}, \hat{L}$ is dominated by $(\hat{K}, 0)$ if $\hat{L} = L_2$. Therefore, it has to be $\hat{L} = L_1$.
(vi) Note first that (7) coincides with (5) with an additional constraint $L = 0$. Further, since the objective function of (5) is decreasing in L, any solution if (5) is dominated by $(\hat{K}, 0)$ given that $\hat{K}$ is a solution of (8). If $\hat{K}$, in addition, satisfies the conditions of (7), necessarily $(\hat{K}$,0) is the solution of (5).
(vii) The assertion follows from (v). □

The Proposition, we have just proved, gives us a directions for solving the problem (5) (and, consequently, (1)). The procedure is as follows

1. Solve (7) and (8). If the optimal solution $\hat{K}_0$ of (7) coincides with that of (8) then $\left(\hat{K}_0, 0\right)$ is the optimal solution of (5) and we may stop the procedure.
2. If the solutions of (7) and (8) differ, solve (9) and check the objective value reached its solution $(\hat{K}, \hat{L})$ with that of $(\hat{K}_0, 0)$. The solution with the higher objective value is the solution of (5).

Note also that, quite surprisingly, the optimal quotes depend only on the inventory of the traded asset but not on the inventory of money.

4 Conclusion

We have formulated a simple but realistic decision problem of a market maker and we reduced its solution to one-dimensional problems. Our result may be useful in further analysis of the microstructure effects at financial markets.

Acknowledgements

This work was supported by grants No. 402/09/0965, P402/10/1610 and P402/10/0956 of the Czech Science Foundation. I thank an anonymous referee for a correction of several my errors.

References

[1] B. Biais, L. Glosten, and Ch. Spatt. Market microstructure: A survey of microfoundations, empirical results, and policy implications. *Journal of Financial Markets*, 8:217–264, 2005.

[2] A. Madhavan. Market microstructure: A survey. *Journal of Financial Markets*, 3:205–258, 2000.

[3] A. Madhavan, and S. Schmidt. An analysis of changes in specialist inventories and qoutations. *J. of Finance*, vol. XLVIII (5). 1993

A model for evaluating creative work outcomes at Czech Art Colleges

Jana Talašová[1], Jan Stoklasa[2]

Abstract. The Register of Artistic Performances is currently being developed in CZ that will contain information on works of art originating from creative activities of art colleges and faculties. Outcomes in various fields of artistic production will be divided into 27 categories, based on their significance, size, and international reception (each criterion classifies into three classes), and each category will be assigned a score. The total score will provide a basis for allocating a part of the state-budget subsidy among art colleges.

The paper discusses the model used to determine scores for each category. The approach is based on Saaty's method, which expertly compares significances of all 27 categories. Creating Saaty's matrix of preference intensities for abstract categories, while maintaining acceptable consistency for such a large matrix, is a difficult task. In the paper we describe a procedure for obtaining required information from a team of persons responsible for different fields of artistic production. A search for solution to this problem has led to new interpretations of Saaty's matrix elements and its consistency condition.

Keywords: Multiple criteria evaluation, Saaty's method, work of art.

JEL Classification: C44
AMS Classification: 91B74

1 Register of Artistic Performances, Classification of Works of Art

The Register of Artistic Performances (RAP) is currently being developed in the Czech Republic that should contain information on works of art originating from creative activities of art colleges and faculties (see [6]). The RAP is conceived as an analogy to the register of R&D outcomes where information on outcomes of research institutions (including universities) has been collected for some years already. In both the registers the outcomes are stored under several categories. These categories are assigned scores. The sum of scores of all the outcomes of a given university is considered an indicator of its performance in the area of creative activity. These numeric values can then be used in decisions regarding one part of the total money to be allocated among universities from the state budget.

The structure of the evaluated categories used in the Czech model was inspired, to some extent, by the artistic categories in the Slovak Republic (see [7]). However, the mathematical model used to determine scores for each category in Slovakia is quite different.

For the purposes of registration of works of art originating from creative activities of the Czech art colleges and faculties, the whole area of artistic production is divided into seven fields: fine arts, design, architecture, theatre, film, literature, and music.

Each piece of art, regardless of the field, is categorized according to the following three criteria:
- Relevance or significance of the piece;
- Extent of the piece;
- Institutional and media reception/impact of the piece.

In each criterion, three different levels are distinguished (denoted by capital letters for easier handling):
- The criterion *Relevance or significance of the piece*:
  - A – a new piece of art or a performance of crucial significance;
  - B – a new piece of art or a performance containing numerous important innovations;
  - C – a new piece of art or a performance pushing forward modern trends.

[1] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz.

[2] Palacký University, Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jan.stoklasa@upol.cz.

- The criterion *Extent of the piece*:
  - K - a piece of art or a performance of large extent;
  - L - a piece of art or a performance of medium extent;
  - M - a piece of art or a performance of limited extent.
- The criterion *Institutional and media reception/impact of the piece*:
  - X – international reception/impact,
  - Y – national reception/impact,
  - Z – regional reception/impact.

The resulting category for a piece of art is given by a combination of three capital letters – e.g. AKX, BKY, or CLZ. There are 27 categories altogether. The decision concerning the relevance or significance of the piece (choice of A, B or C) rests upon expert assessment; the experts have at their disposal general definitions of each category and examples of works of art in each category – for all three levels of each criterion and for all 7 fields of artistic production – to assist them in the decision process. As for the extent of the piece (levels K, L, M), all the classes are specified for all the fields of art. As for the institutional and media reception/impact, lists of institutions corresponding to categories X, Y, Z are available for all fields.

Le us notice, there are interactions among the three mentioned criteria. The first one (expertly defined *Relevance or significance of the piece of art*) and the third one (*Institutional and media reception/impact of the piece*) partly overlap. That means, we are not allowed to set separately the weights of criteria and the scores of levels for each of them, and then calculate the scores of categories by means of the weighted average operation. It is necessary to set directly the scores of the categories that are described by the triples of criteria levels.

2 Determining scores for particular categories of artistic production

Saaty's method (see [2, 3, 4]) served as a basis for determination of scores for all 27 categories of artistic production. However obvious it was that this mathematical tool is the most appropriate for such a task, certain challenges concerning its use were also clearly apparent: (1) a difficulty for a team of experts to express preferences with respect to abstract categories; (2) a difficulty to reach acceptable consistency of Saaty's matrix under such a large number of categories; (3) a consensus within the group of experts (professional guarantors of particular fields of art). The proposed solution to these problems will be described in the following paragraphs.

Admittedly, expressing one's opinion on intensities of preferences with respect to abstract categories is difficult. Experts, professional guarantors of artistic fields, were first asked to provide specific (historical) examples of works of art in all categories in their field. (This step was also important to ensure, or to suggest modifications to ensure, that corresponding categories be really comparable in terms of evaluation across fields.) Next, professional guarantors of each field of art set their preferences concerning pairs of categories, while considering the representatives (examples) of each category to aid them in their decisions.

Although it was possible for each of these experts to express their preferences separately, and only then to derive the collective preferences (from the individual ones), we used a different approach. The collective preferences were set directly at a team meeting of experts. The reason was that art-college experts are not used to work with mathematical models and individual inputting of required data could prove difficult for them. Achieving consensus was also intentionally preferred over averaging different opinions.

Great effort was made to find the best way of converting expert preferences concerning the 27 categories of artistic production (represented in each field of art by specific examples) into a mathematical model in order to determine their scores. To facilitate the process of inputting required data by the experts and to achieve the necessary consistency of this input, the following two-step procedure was performed:

In the first step, we have determined the order of importance of the categories by the pairwise comparison method (see [2, 5]). This method employs a matrix of preferences and indifferences $P = \left\{ p_{i,j} \right\}_{i,j=1,\dots,27}$. For its elements it holds that:

$p_{i,j} = 1$, if the ith category is more important than the jth category;

$p_{i,j} = 0,5$, if the ith category is equally important as the jth category;

$p_{i,j} = 0$, if the jth category is more important than the ith category.

It is sufficient for the experts to fill in the upper right triangle of the matrix, that is, the elements $p_{i,j}, i < j,$ as $p_{i,i} = 0,5$ and $p_{j,i} = 1 - p_{i,j}$. The row sums $D_i = \sum_{j=1}^{27} p_{i,j}, i = 1,\dots,27$, determine the order of importance of the categories (their quasi-ordering, transitive and complete relation, that can be described as a linear ordering of classes of indifferent elements). We need to verify consistency of the preferences in the sense of transitivity, that is, whether it holds that $p_{i,k} \geq \max\left\{p_{i,.j}, p_{j,k}\right\}$ for all $i, j, k \in \{1,...,27\}$. If the matrix is not consistent, we make a minimum amount of changes necessary for it to become so. These changes are then consulted with the team of experts and if they are approved of, we can proceed. All the changes actually made while solving our problem are summarized in Tab 1.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	Prefe- rence	Preference order	
		AKX	AKY	AKZ	ALX	ALY	ALZ	AMX	AMY	AMZ	BKX	BKY	BKZ	BLX	BLY	BLZ	BMX	BMY	BMZ	CKX	CKY	CKZ	CLX	CLY	CLZ	CMX	CMY	CMZ			
1	AKX	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	26,5	1	AKX
2	AKY		0,5	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	25	2	AKY
3	AKZ			0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	25	2	AKZ
4	ALX				0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	23,5	4	ALX
5	ALY					0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	21,5	6	ALY
6	ALZ						0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	20,5	7	ALZ
7	AMX							0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	22,5	5	AMX
8	AMY								0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	18,5	9	AMY
9	AMZ									0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	17,5	10	AMZ
10	BKX										0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	19,5	8	BKX
11	BKY											0,5	1,0	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	16	11	BKY
12	BKZ												0,5	0,0	1,0	1,0	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	14	13	BKZ
13	BLX													0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	16	11	BLX
14	BLY														0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	12,5	15	BLY
15	BLZ															0,5	0,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	11,5	16	BLZ
16	BMX																0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	14	13	BMX
17	BMY																	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	10,5	17	BMY
18	BMZ																		0,5	0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	9	18	BMZ
19	CKX																			0,5	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	9	18	CKX
20	CKY																				0,5	1,0	0,0	1,0	1,0	1,0	1,0	1,0	6,5	21	CKY
21	CKZ																					0,5	0,0	1,0	1,0	1,0	1,0	1,0	5,5	22	CKZ
22	CLX																						0,5	1,0	1,0	1,0	1,0	1,0	7,5	20	CLX
23	CLY																							0,5	1,0	0,0	1,0	1,0	3,5	24	CLY
24	CLZ																								0,5	0,0	1,0	1,0	2,5	25	CLZ
25	CMX																									0,5	1,0	1,0	4,5	23	CMX
26	CMY																										0,5	1,0	1,5	26	CMY
27	CMZ																											0,5	0,5	27	CMZ

changes made to achieve consistency of the matrix resulting form the final order

change made from 0,5 to 1 or from 1 to 0,5 (**"small change"**)

change made from 0 to 1 (**"big change"**)

Table 1 The pairwise comparison matrix with highlighted changes.

In the second step, Saaty's matrix of preference intensities $S = \left\{s_{i,j}\right\}_{i,j=1,...,27}$ was constructed for categories numbered in ascending order according to their significance determined in the previous step. Again, it was sufficient to fill in the upper right triangle of the matrix. The elements $s_{i,j}, i < j,$ were set as follows:

$s_{i,j} = 1\ldots$ the ith and jth categories are **equally** important;

$s_{i,j} = 3\ldots$ the ith category is **slightly more important** than the jth category;

$s_{i,j} = 5\ldots$ the ith category is **strongly more important** than the jth category;

$s_{i,j} = 7\ldots$ the ith category is **very strongly more important** than the jth category;

$s_{i,j} = 9\ldots$ the ith category is **extremely more important** than the jth category.

It holds that $s_{i,i} = 1$ and $s_{j,i} = {}^{1}\!/\!_{s_{i,j}}$, for the intensity of preference $s_{i,j}$ expresses the ratio of preferences between both categories.

The traditional requirement for consistency in Saaty's method, that is $s_{i,k} = s_{i,j} \cdot s_{j,k}$ for all $i, j, k \in \{1,...,27\}$, is basically unachievable. For example, consider only four arbitrary objects that are linearly ordered according to their importance. If each of them is just slightly more important than the following one, then in the case of full consistency the first one would have to be 27 times more important than the fourth. But the maximum value available for expressing intensity of preference is nine (as is shown by psychological research [3], this is the highest number of levels of importance that human is able to distinguish). We have weakened the original requirement on consistency, which was too strong, and for the purposes of our work we have requested $s_{i,k} \geq \max\left\{s_{i,.j}, s_{j,k}\right\}$ for all $i, j, k \in \{1,...,27\}$. When the categories are numbered as to their importance, this requirement is easy to verify. In addition to the fact that the matrix S has to be reciprocal (i.e. $s_{i,i} = 1$ and $s_{i,j} = {}^{1}\!/\!_{s_{j,i}}$ for $i, j \in \{1,...,27\}$) in view of the above-mentioned condition, consistency means that the elements of S are nondecreasing from left to right and from bottom up. If the matrix, as set by the experts, is not consistent, we propose the minimum amount of changes necessary for it to become so – the team of professional guarantors either approve of these changes or make their own to achieve consistency. Tab. 2 illustrates the changes actually made in our application in order to remove inconsistencies from the original matrix S. (Tab. 2 contains also changes induced by re-dividing the pairs of indifferent categories having originated from the pair-wise comparison method.)

		1 AKX	2 AKY	3 AKZ	4 ALX	5 AMX	6 ALY	7 ALZ	8 BKX	9 AMY	10 AMZ	11 BKY	12 BKZ	13 BLX	14 BMX	15 BLY	16 BLZ	17 BMY	18 BMZ	19 CKX	20 CLX	21 CKY	22 CKZ	23 CMX	24 CLY	25 CLZ	26 CMY	27 CMZ
1	AKX	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	9	9	9	9	9	9	9	9	9	9
2	AKY		1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
3	AKZ			1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	7	9	9	9	9	9	9
4	ALX				1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9	9	9	9
5	AMX					1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9	9
6	ALY						1	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9	9
7	ALZ							1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7	9
8	BKX								1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
9	AMY									1	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
10	AMZ										1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
11	BKY											1	5	5	5	5	5	5	5	5	5	5	5	5	5	7	7	7
12	BKZ												1	3	5	5	5	5	5	5	5	5	5	5	5	7	7	7
13	BLX													1	5	5	5	5	5	5	5	5	5	5	5	7	7	7
14	BMX														1	3	3	3	5	5	5	5	5	5	5	7	7	7
15	BLY															1	3	3	5	5	5	5	5	5	5	7	7	7
16	BLZ																1	3	5	5	5	5	5	5	5	5	7	7
17	BMY																	1	5	5	5	5	5	5	5	5	7	7
18	BMZ																		1	5	5	5	5	5	5	5	5	5
19	CKX																			1	3	5	5	5	5	5	5	5
20	CLX																				1	5	5	5	5	5	5	5
21	CKY																					1	3	3	5	5	5	5
22	CKZ																						1	3	5	5	5	5
23	CMX																							1	5	5	5	5
24	CLY																								1	3	3	3
25	CLZ																									1	3	3
26	CMY																										1	3
27	CMZ																											1

3,5,7,9	změny vyplývající z rozdělení kategorií
3,5	změny respektující párové porovnávání sousedních kategorií
5	změny pro udržení konzistence vyvolané zadáním "červených hodnot"

Table 2 Saaty's matrix of preference intensities with highlighted changes.

Under the assumption that S is close enough to an ideally consistent matrix (i.e. matrix that fulfills $s_{i,k} = s_{i,j} \cdot s_{j,k}$ for all $i, j, k \in \{1, ..., 27\}$), the scores of 27 categories, representing their relative importance, are calculated by Saaty's method as components of the eigenvector corresponding to the largest eigenvalue.

The resulting scores of artistic categories can also be obtained from S in a different way. The columns of S can be interpreted as repeated measurements of the relative importances of the 27 categories. These measurements are performed by the team of experts who compare all the categories with the first one, than the second one, and so on until the 27th one. From the point of view of mathematical statistics, these are compositional data, i.e. data bearing only relative information (see [1]). Information contained in this data can be expressed by estimating its mean value. A proper estimator of the mean value of this kind of data is a vector whose components are geometric means of the corresponding components of vectors representing single measurements. The relative scores of all 27 categories can be also obtained by computing geometric means of the rows of Saaty's matrix (this calculation method is known as the logarithmic least squares method, see [2]). This weaker consistency of S ($s_{i,k} \geq \max\{s_{i,j}, s_{j,k}\}$ for all $i, j, k \in \{1, ..., 27\}$) is then a natural requirement that allows for an easy check on consistency of the expertly entered data. The facts that S has to be reciprocal and, with the categories ordered according to their importance, that the values of a well entered matrix S must be nondecreasing from left to right and from bottom up can serve as a good guiding principle for teams of experts in defining the preference intensities of pairs of categories.

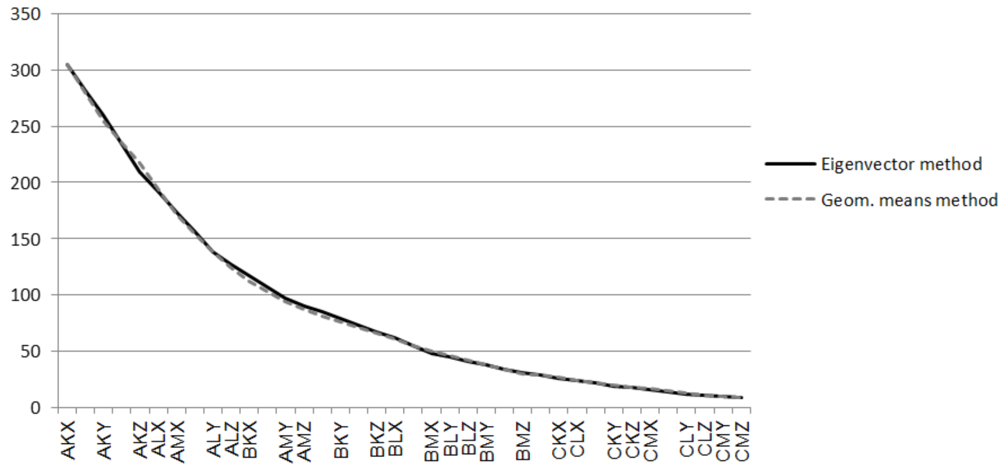


Figure 1 Graphical comparison of the eigenvector method with the geometric means method.

Category	Relevance or significance	Extent	Institutional reception	Eigenvector method	Geom. means method
AKX	Crucial significance and originality	Large	International	305	305
AKY	Crucial significance and originality	Large	National	259	254
AKZ	Crucial significance and originality	Large	Regional	210	217
ALX	Crucial significance and originality	Medium	International	191	194
AMX	Crucial significance and originality	Small	International	174	171
ALY	Crucial significance and originality	Medium	National	138	138
ALZ	Crucial significance and originality	Medium	Regional	127	124
BKX	Bearing many important inovations	Large	International	117	112
AMY	Crucial significance and originality	Small	National	97	94
AMZ	Crucial significance and originality	Small	Regional	90	87
BKY	Bearing many important inovations	Large	National	79	75
BKZ	Bearing many important inovations	Large	Regional	66	66
BLX	Bearing many important inovations	Medium	International	62	61
BMX	Bearing many important inovations	Small	International	48	50
BLY	Bearing many important inovations	Medium	National	44	46
BLZ	Bearing many important inovations	Medium	Regional	40	41
BMY	Bearing many important inovations	Small	National	37	38
BMZ	Bearing many important inovations	Small	Regional	31	30
CKX	Developing current trends	Large	International	26	26
CLX	Developing current trends	Medium	International	24	24
CKY	Developing current trends	Large	National	19	20
CKZ	Developing current trends	Large	Regional	17	18
CMX	Developing current trends	Small	International	16	16
CLY	Developing current trends	Medium	National	12	13
CLZ	Developing current trends	Medium	Regional	10	11
CMY	Developing current trends	Small	National	9	9
CMYŹ	Developing current trends	Small	Regional	8	9

Table 3 Scores obtained by the Saaty matrix eigenvector method and those determined as geometric means of rows of S.

Tab. 3 compares the scores determined by the Saaty matrix eigenvector method with those determined as geometric means of the rows. The scores are normalized so that the maximum is 305 (analogy to R&D outcomes evaluation). It is easy to see that the differences between these two methods are not significant, see Fig. 1. The Saaty matrix eigenvector method will be used in testing the model on the first real dataset, gathered by Czech art colleges and faculties for the years 2008 to 2010.

3 Conclusion

The Register of Artistic Performances and the methodology of evaluating artistic production originating from creative activities of art colleges and faculties are currently being pilot-tested in the Czech Republic. At present, our effort is focused on refining the triplets of class specification for all three criteria and for all the fields of art, and particularly on developing a most objective mechanism of expert classification of artistic production into 27 categories.

The mathematical model for score determination was developed in an effort to achieve the best possible conversion of preferences of the expert team into scores for different categories of artistic production. With Saaty's method serving as an appropriate basis, the solution to this problem required its implementation in a special procedure.

Acknowledgements

This research is conducted with the support of the Centralized Developmental Project C41 entitled *Evaluating Creative Work Outcomes Pilot Project* and financed by the Czech Ministry of Education.

References

[1] Aitchison, J. *The Statistical Analysis of Compositional Data*. *Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London , 1986.

[2] Ramík, J.: *Analytický hierarchický process (AHP) a jeho využití v malém a středním podnikání.* Slezská univerzita v Opavě, Karviná, 2000.

[3] Saaty, T.L.: *The Brain: Unraveling the Mystery of How it Works, The Neural Network Process* . 481 pp., RWS Publ., 1999.

[4] Saaty, T.L.: *The Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process.* Vol. VI of the AHP Series, 478 pp., RWS Publ., 2000.

[5] Talašová, J.. *Fuzzy metody vícekriteriálního hodnocení a rozhodování.* Vydavatelství Univerzity Palackého, Olomouc. 2003.

[6] Zelinský, M.(ed.): *Registr uměleckých výkonů*. Akademie múzických umění v Praze, 2010.

[7] Smernica č. 13/2008-R zo 16. októbra 2008 o bibliografickej registrácii a kategorizácii publikačnej činnosti, uměleckej činnosti a ohlasov. Ministerstvo školstva Slovenskej Republiky.

An Extended Model for the Design of City Mass Transport Network

Dušan Teichmann[1], Michal Dorda[2], Jindřich Frič[3]

Abstract. The presented article deals with the problem of the design of city mass transport lines network. The problem will be solved by means of linear programming methods. The fundamentals of this model consist in the known pieces of knowledge − the lines are identified through assigned vehicles. However, the results of original mathematical models often require further corrections for traffic realization. Therefore, mathematical modeling is not usually used to solve this problem. Nevertheless, there are many possible modifications of the models to solve such types of problems; the proposed models are universal. In addition, in many cases mathematical models enable us to find the optimal solution. The presented model has the following characteristics. The total costs for vehicle ride will be the optimization criterion. The proposed model will include the conditions which ensure better construction of timetable after the resolution of the model. Vehicles are allocated according to the acceptable intervals and these intervals are predetermined. The model will be formed for a company with a homogeneous vehicle fleet.

Keywords: linear programming, design public transport network

JEL Classification: C61
AMS Classification: 90C05

1 Introduction

The design of transport lines network of an urban public transport is among the basic transport theory problems. Without an effective conception, urban public transport is a service which does not fulfill the passengers' requirements. Urban public transport systems are seldom proposed by using mathematical methods. However, these mathematical methods enable us to overcome many constraints. By means of mathematical models an optimal solution to the problem can be found, unlike in the case of an intuitive approach. In addition, mathematical models facilitate the observation of many restrictions, even in the case of extremely complex problems.

2 Formulation of problem, theory for solution and the mathematical model

The transport network on which we shall run the urban public transport is represented by the graph $G[V,H]$. The vertices indicate the points in which the network branches or in which the transport lines end. The edges represent the communications between the defined points. The input set of lines L_0 for this transport network is known. The round journey time o_l for every transport line $l \in L_0$ is known, too, as well as the set of transport lines for every communication. This set of lines is represented by the symbol L_h. For every communication (edge) of the network, the decisive intensity of passengers for a chosen time unit and the capacity of one vehicle are known. We shall decide how many vehicles will be assigned for every transport line $l \in L_0$.

The initial models which served as inspiration for the following contemplations were published in [2] or [3]. Other models can be found in [4], [6], [7] or [8]. These proceedings are analogical to the proceedings in foreign literature. We can find a more detailed list of sources in [1]. Most models are restricted to the basic scheme, in which the problem of the line network design is dealt with in four basic phases: assigning the vehicles to the lines, determining the scheduling connections (time table), the vehicle scheduling and network time coordination [5]. The bases of all presented models are usually represented by the variables modelling a part of the vehicle capacity assigned to the lines. If a certain part of the vehicle fleet capacity (the vehicles or places) is assigned, the line runs. There is a limitation in all models concerning the fulfilling of the passengers' require-

[1] VŠB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, 70833 Ostrava-Poruba, the Czech Republic, dusan.teichmann@vsb.cz.
[2] VŠB-Technical University of Ostrava, Faculty of Mechanical Engineering, Institute of Transport, 17. listopadu 15, 70833 Ostrava-Poruba, the Czech Republic, michal.dorda@vsb.cz.
[3] CDV, v.v.i, Division of Traffic Safety and Traffic Engineering, Líšeňská 33a, 636 00 Brno, the Czech Republic, jindrich.fric@cdv.cz.

ments. Another limitation of some models is that the vehicle fleet capacity will not be exceeded. In case of a vehicle fleet composed of various types of transport (tram, buses), there is a requirement that each line is serviced by one type of transport only. Most often, the number of vehicles we need to satisfy the passengers' demands is minimized or the minimum from the proportions between the offered and the required number of places in a link (edge of graph) is maximized. However, the assignment itself of the vehicles to the lines cannot be sufficient. Another problem lies in proposing the time table − determining the headway serving the line, above all. Certain complications can occur there, as the previous models do not respect this restriction. It can happen quite easily that such a number of vehicles that does not ensure a proper (regular and integer) headway is assigned to the line. The solver of the problem often has to carry out additional adjustments of the result, which can considerably deteriorate the final solution. Therefore, our goal is to search for such a model that would solve also this secondary problem when assigning the vehicles to the lines.

For the sake of simplicity, let us think about a simple urban public transport network with one type of vehicles and all the lines running in a shuttle way. When modifying the model, we add the starting set of headways to the individual lines set. In urban public transport it is common to use certain standardized interval lengths. Let us think about the line $l \in L_0$ with the round journey time $o_l = 60$ minutes, supposing that in a middle-sized town, the headways of 10, 12, 15, 20, 30 and 60 minutes are held to be standard. Let us denote the set of suitable intervals for the line $l \in L_0$ with the symbol P_l . To keep this requirement, we have to use such a type of solution in which the round journey time o_l is divided by the line interval $l \in \mathrm{P}_l$. To reach the suitable headway in the running of $l \in L_0$ it must hold the equation:

$$10 v_{l1} + 12 v_{l2} + 15 v_{l3} + 20 v_{l4} + 30 v_{l5} + 60 v_{l6} = 60 \tag{1}$$

where v_{lp} is the number of vehicles assigned for the line $l \in L_0$ running in the suitable interval $l \in \mathrm{P}_l$ or generally speaking for the line $l \in L_0$:

$$\sum_{p \in P_l} t_{lp}\, v_{lp} = o_l \tag{2}$$

where the value t_{lp} represents the suitable headway length $p \in P_l$ of the line $l \in L_0$. Furthermore, it is necessary to fulfil the condition that maximally one interval will be applied for every transport line. We can ensure it by means of the binary variable denoted for example w_{lp} . If it holds $w_{lp} = 1$ after the solution the problem, then the interval $p \in P_l$ is applied for the line $l \in L_0$. If it holds $w_{lp} = 0$, it will mean the contrary. In this model the group of conditions will appear in the form (T is prohibitive constant):

$$\sum_{p \in P_l} w_{lp} \le 1 \text{ for } l \in L_0 , \tag{3}$$

$$v_{lp} \le w_{lp}\, T \text{ for } l \in L_0 \text{ a } p \in P_l . \tag{4}$$

If the condition (1) remained in this form, just one interval would be applied in each line $l \in L_0$. Nevertheless, this does not have to be true, as not every line has to run in the final traffic solution. However, it is evident that there must be the value of zero or o_l on the right side of the equation (1). The value of zero must appear on the right side of the equation in the case in which the line does not run, the value of o_l in the opposite case. To remove the above mentioned shortcoming we complete the condition into the form:

$$\left(10 v_{l1} + 12 v_{l2} + 15 v_{l3} + 20 v_{l4} + 30 v_{l5} + 60 v_{l6}\right) = 60 \sum_{p \in P_l} w_{lp} \text{ for } l \in L_0 . \tag{5}$$

The mathematical model (the version minimizing the total number of vehicles assigned to the network) has the form (k is vehicle capacity, q_h is decision intensity of travellers on link):

$$\min f(v) = \sum_{l \in L_0} \sum_{p \in P_l} v_{lp} , \tag{6}$$

subject to:

$$\sum_{l\in L_h}\sum_{p\in P_l}\frac{60\,k\,v_{lp}}{o_l}\geq q_h \quad \text{for } h\in H, \tag{7}$$

$$\sum_{p\in P_l} t_{lp} v_{lp} = o_l \sum_{p\in P_l} w_{lp} \quad \text{for } l\in L_0, \tag{8}$$

$$\sum_{p\in P_l} w_{lp} \leq 1 \quad \text{for } l\in L_0, \tag{9}$$

$$v_{lp} \leq w_{lp}T \quad \text{for } l\in L_0 \text{ and } p\in P_l, \tag{10}$$

$$w_{lp}\in\{0;1\} \quad \text{for } l\in L_0 \text{ and } p\in P_l, \tag{11}$$

$$v_{lp}\in Z_0^+ \quad \text{for } l\in L_0 \text{ and } p\in P_l. \tag{12}$$

The objective function (6) represents the number of vehicles assigned to the line routes in the network. The group of limiting conditions (7) ensures that a sufficient number of places will be offered in every link of the transport network. The group of limiting conditions (8) ensures the choice of a desired interval for each line route. The group of limiting condition (9) will ensure that maximally one interval will be chosen for each line. The group of limiting conditions (10) links the variables v_{lp} and w_{lp} (the value of T is the constant representing the total number of vehicles in the vehicle fleet). The groups of limiting conditions (11) and (12) determine the definition scope of variables.

3 The calculation experiments

In this chapter we will present the results of one of the experiments which were carried out with the proposed model. The defined network had 15 links; the input set of lines L_0 contained 15 lines. The lines can be served by the vehicles with the capacity of 100 places. A network set in this way corresponds to the line network of the urban public transport of a small town. The decisive passengers' intensities in individual network links and the round journey times of individual lines are presented in tables 1 and 2.

The link	1	2	3	4	5	6	7	8
The link vertices	1-2	2-3	2-7	3-4	3-5	3-7	4-5	5-6
The decisive passengers' intensity per hour	200	550	585	290	205	760	390	125
The link	9	10	11	12	13	14	15	
The link vertices	5-7	5-8	7-8	8-9	8-10	10-11	11-12	
The decisive passengers' intensity per hour	690	185	430	200	335	260	175	

Table 1 The input data about the network and decisive intensity in individual links

The transport line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
The round journey time of the line [min]	30	30	60	60	20	20	20	20	20	30	20	20	12	60	15

Table 2 The round journey times in individual lines in minutes

The solved network is presented in figure 1.

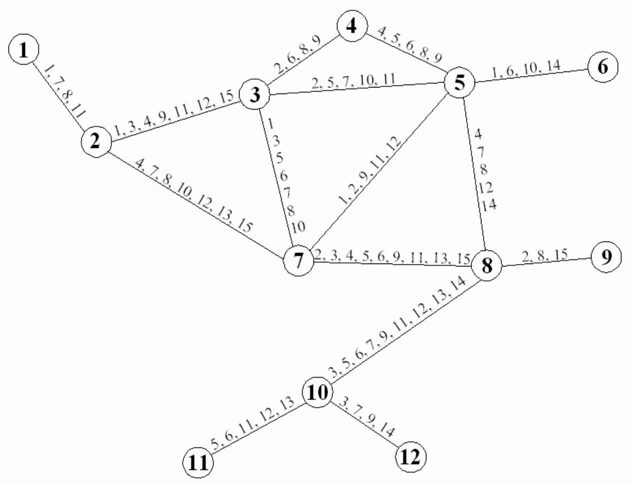


Figure 1 The input set of lines and their running in the transport network

The solving was executed in the optimization software Xpress-IVE; the results are summarized in table 3.

The line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
The number of assigned vehicles	2	2	2	6	1	0	0	0	0	0	0	0	0	0	0
The used headway	15	15	30	10	20	-	-	-	-	-	-	-	-	-	-

Table 3 The calculation results of the experiment

The calculation time for this network was under 0,1s.

The figure 2 shows the optimization calculation course.

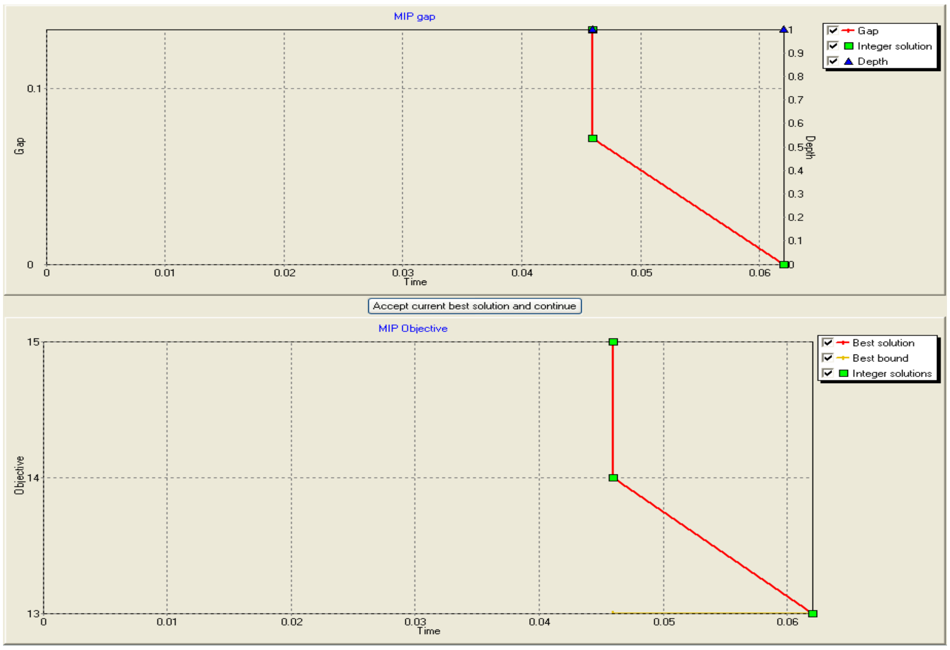


Figure 2 The optimization calculation course

4 Conclusion

The above presented article deals with the proposal of the line network of urban public transport by means of the linear model. The existing models that have been studied (as specified in the introductory parts of this article) present the solutions which usually fulfill the passengers' requirement. The vehicles are assigned to the lines without using suitable headways. This disadvantage is dismantled by means of the model proposed in this article. The calculation experiments that have been executed have confirmed its functionality.

References

[1] Blatoň, M., and Teichmann, D.: *Selected Methods Used Abroad to Address the Urban Public Transport Line Network Design*. Perners's Contacts. Pardubice: University of Pardubice, 2010, No. 17, pp. 7-14.

[2] Černá, A., and Černý, J.: *Theory of Management and Decision in Transportation Systems* (in Czech). Pardubice: IJP, 2004.

[3] Černý, J., and Kluvánek, P.: *Bases of Mathematical Transport Theory* (in Slovak). Bratislava: VEDA, 1991.

[4] Daněk, J., Plevný, M., and Teichmann, D.: *The Rational Operation of the Urban Transport Line Network by Minimization of the Needed Vehicles.* E+M Economics and Management. Liberec: Technical University of Liberec, 2010, no. 2, pp. 53-62.

[5] Janáček, J., and Koháni, M.: *Waiting Time Optimization with IP-Solver.* Communications. Žilina Univerzity of Žilina, 2010, no. 12, pp. 60-64.

[6] Janáček, J., and Širc, J.: *The Column Generation Technique for Public Transport Line Planning by IP-Solver.* Communications. Žilina: University of Žilina, 2010, No. 12, pp. 60-64.

[7] Jánošíková, Ľ., Blatoň, M., and Teichmann, D.: *Design of Urban Public Transport Lines as a Multiple Criteria Optimization Problem.* In: Proceedings of the conference "Urban Transport 2010", Cyprus, 2010, May 4-7, pp. 137-149.

[8] Peško, Š.: *Support of Operations Research Methods for Line System Design.* In: Proceedings of 6th International Conference "Urban Transportation Infrastructure", Žilina: University of Žilina, Žilina, 2008, October 22-23, p. 29.

Simulation Methodology for Financial Assets with Imprecise Data

Tomáš Tichý[1], Michal Holčapek[2]

Abstract. During last decades the stochastic simulation approach, both via Monte Carlo (MC) and Quasi Monte Carlo (QMC) has been vastly applied and subsequently analyzed in almost all branches of science. Very nice applications can be found in areas that rely on modeling via stochastic processes, such as finance. However, since financial quantities, opposed to natural processes, depend on human activity, their modeling is often very challenging. Many scholars therefor suggest to specify some parts of financial models by means of fuzzy set theory. In this contribution the recent knowledge of fuzzy numbers and their approximation is utilized in order to suggest fuzzy-MC simulation to modeling of returns of financial quantities, such as prices of stocks, commodities or exchange rates. Finally, three distinct types of potential fuzzy-stochastic models are suggested, including quantile estimation illustrations.

Keywords: Fuzzy variable, Stochastic variable, Fuzzy-stochastic variable, Financial models, Risk estimation

1 Introduction

During last decades the stochastic simulation approach, both via MC and QMC has been vastly applied and subsequently analyzed in almost all branches of science. Very nice applications can be found in areas that rely on modeling via stochastic processes. For example, in finance it is usually assumed that returns of almost all assets – the riskless bond is an exception – can be described by a stochastic differential equation based on a random term with a suitable probability distribution. However, opposed to processes in nature financial processes crucially depend on human behavior and are rather imprecise (vague) than random. Modeling is therefore more challenging. More importantly, it is often difficult to provide reliable parameter estimates for particular models.

As an alternative, many researchers suggest to specify a financial model or at least its parameters as a fuzzy number, ie. instead of real variables with a suitable distribution function we assume potentially infinite number of intervals, each with its membership function as given by a particular α-cut.

The application of fuzzy set theory in finance usually restricts to basic 'analytic' operations with fuzzy and crisp inputs and includes Black and Scholes or binomial model type option pricing problem (see eg. [15] or [20]), Value at Risk estimation [21] or portfolio selection process [17]. However, it is well known that many problems in finance cannot be solved analytically even if simple (geometric) Brownian motion is assumed or the implied volatility [3] is utilized (see eg. the problem of exotic option valuation or far tail risk estimation). It is therefore surprising that there are just a few papers on MC simulation of fuzzy random models and usually with a rather brief treatment of the simulation approach (see eg. [13]).

In the following sections we utilize the recent knowledge of fuzzy numbers and their approximation and suggest simulation approach to their modeling. We also provide three distinct types of financial models after fuzzyfication – (i) standard market model (Brownian motion) with fuzzy parameter, (ii) Brownian motion with fuzzy subordinator, and (iii) Brownian motion with fuzzyfied gamma subordinator. Finally, illustrative examples are provided, including quantile estimation.

[1]Technical University Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, 701 21 Ostrava, Czech Republic, e-mail: tomas.tichy@vsb.cz.

[2]University of Ostrava, Institute for Research and Applications of Fuzzy Modeling, 30. dubna 22, 701 03 Ostrava, Czech Republic, e-mail: michal.holcapek@osu.cz.

2 LU-fuzzy Number

Let $\mathbb{R}$ denotes the set of real numbers and $A : \mathbb{R} \to [0,1]$ be a mapping. We say that A is a *fuzzy number* if A is normal (ie. there exits an element x_0 such that $A(x_0) = 1$), convex (ie. $A(\lambda x + (1-\lambda)y) \geq \min(A(x), A(y)$ for any $x, y \in \mathbb{R}$ and $\lambda \in [0,1]$), upper semicontinuous and supp(A) is bounded, where supp$(A) = cl\{x \in \mathbb{R} \mid A(x) > 0\}$ and cl is the closure operator (see [9, 5]). Note that the most popular models of fuzzy numbers are the triangular and trapezoidal shaped models investigated by Dubois and Prade in [6]. Their popularity follows from the simple calculus as addition or multiplication of fuzzy numbers which can be established for them. This is also a reason why we can find many recent papers on the approximation of fuzzy numbers by the aforementioned models (see eg. [1, 2] and the references therein). In order to model fuzzy numbers we will use a more advanced model of fuzzy numbers based on the interpolation of given knots using rational splines that was proposed by Guerra and Stefanini in [8] and developed in [16]. This model generalizes the triangular fuzzy numbers and gives a broad variety of shapes enabling more precise representation of fuzzy real data. Nevertheless, the calculus is still very simple.

Recall that a piecewise rational cubic Hermite parametric function $P \in C^1[\alpha_0, \alpha_n]$, with parameters v_i, w_i, $i = 0, \ldots, n-1$, is defined for $\alpha \in [\alpha_i, \alpha_{i+1}]$, $i = 0, \ldots, n-1$ by

$$P(\alpha) = P_i(\alpha, v_i, w_i) = \frac{(1-\theta)^3 f_i + \theta(1-\theta)^2(v_i f_i + h_i d_i) + \theta^2(1-\theta)(w_i f_{i+1} - h_i d_{i+1}) + \theta^3 f_{i+1}}{(1-\theta)^3 + v_i\theta(1-\theta)^2 + w_i\theta^2(1-\theta) + \theta^3},$$

where the notations f_i and d_i are, respectively, the real data values and the first derivative values (slopes) at the knots $\alpha_0 < \alpha_1 < \cdots < \alpha_n$, $h_i = \alpha_{i+1} - \alpha_i$, $\theta = (\alpha - \alpha_i)/h_i$ and $v_i, w_i \geq 0$. The parameters v_i and w_i are called the tension parameters.[1] In this work, we will use a global monotonicity setting (cf. [16]):

$$v_i = w_i = \begin{cases} \frac{d_{i+1}+d_i}{f_{i+1}-f_i}, & \text{if } f_{i+1} \neq f_i; \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

A main reason for this assumption is a natural calculus which can be introduced for fuzzy numbers based on this type of piecewise rational cubic Hermite parametric functions. One can see that each such parametric function $P \in C^1[\alpha_0, \alpha_n]$ may be expressed in the matrix form consisting of parameters as follows:

$$\mathbf{P} = \begin{pmatrix} \mathbf{f} \\ \mathbf{d} \end{pmatrix} = \begin{pmatrix} f_{\alpha_0} & \cdots & f_{\alpha_n} \\ d_{\alpha_0} & \cdots & d_{\alpha_n} \end{pmatrix} \tag{2}$$

for a partition $\alpha_0 < \cdots < \alpha_n$ of the interval $[a_0, a_n]$. In what follows, we will identify each parametric function $P \in C^1[\alpha_0, \alpha_n]$, for which the presumption in (1) is satisfied, with a matrix $\mathbf{P}$ established above and simply write $P(\alpha) = \mathbf{P}(\alpha)$.

Now we may define a special case of LU-fuzzy numbers introduced in [8]. Note that our definition is slightly different than the original one, but the idea remains the same. Recall that an α-cut of a fuzzy number A is a set $A_\alpha = \{x \in \mathbb{R} \mid A(x) \geq \alpha\}$.

Definition 1. A fuzzy number A is an *LU-fuzzy number*, if there exist a partition $0 = a_0 < \cdots < \alpha_n = 1$ and $2 \times (n+1)$ matrices $\mathbf{A}^-$ and $\mathbf{A}^+$ (defined in (2)) such that

1. $A_\alpha = [\mathbf{A}^-(\alpha), \mathbf{A}^+(\alpha)]$ for any $\alpha \in [0,1]$,
2. $f^-_{\alpha_{i+1}} \geq f^-_{\alpha_i}$ and $f^+_{\alpha_{i+1}} \leq f^+_{\alpha_i}$ for any $i = 0, \ldots, n-1$, and $f^-_{\alpha_n} = f^+_{\alpha_n}$,
3. $d^-_{\alpha_i} \geq 0$ and $d^+_{\alpha_i} \leq 0$ for any $i = 0, \ldots, n$,
4. if $f^-_{\alpha_i} = f^-_{\alpha_{i+1}}$ (or $f^+_{\alpha_i} = f^+_{\alpha_{i+1}}$), then $d^-_{\alpha_i} = d^-_{\alpha_{i+1}}$ (or $d^+_{\alpha_i} = d^+_{\alpha_{i+1}}$).

An LU-fuzzy number will be denoted by $\mathbf{A} = (\mathbf{A}^-, \mathbf{A}^+)$ and the set of all LU-fuzzy numbers defined over a partition $0 = a_0 < \cdots < \alpha_n = 1$ will be denoted by $\mathcal{F}_{LU}(\alpha_0, \ldots, \alpha_n)$.

[1] For details, we refer to [7, 14].

Let $D \subseteq \mathbb{R}^m$ and $g : D \to \mathbb{R}$ be a real function which has all partial derivatives on the domain D, ie. $g'_{x_k}(a_1, \ldots, a_m) \in \mathbb{R}$ for any $(a_1, \ldots, a_m) \in D$ and $k = 1, \ldots, m$. A general procedura showing how to extend the function g to a function $\tilde{g} : \mathcal{D} \to \mathcal{F}_{LU}(\alpha_0, \ldots, \alpha_n)$, where $\mathcal{D} \subseteq \mathcal{F}_{LU}(\alpha_0, \ldots, \alpha_n)^m$ is a suitable domain, can be formulated within the two following steps:[2]

1. Put $\mathbf{m} = \{1, \ldots, m\}$, consider $\pi : \mathbf{m} \to \{-, +\}$ and define
$$\mathbf{B}^{\pi(1),\ldots,\pi(m)} = \begin{pmatrix} f_{\alpha_0}^{\pi(1),\ldots,\pi(m)} & \ldots & f_{\alpha_n}^{\pi(1),\ldots,\pi(m)} \\ d_{\alpha_0}^{\pi(1),\ldots,\pi(m)} & \ldots & d_{\alpha_n}^{\pi(1),\ldots,\pi(m)} \end{pmatrix}, \tag{3}$$
where (for any $k = 0, \ldots, n$)
$$\begin{aligned} f_{\alpha_k}^{\pi(1),\ldots,\pi(m)} &= g(f_{1\alpha_k}^{\pi(1)}, \ldots, f_{m\alpha_k}^{\pi(m)}), \\ d_{\alpha_k}^{\pi(1),\ldots,\pi(n)} &= g'_{x_1}(f_{1\alpha_k}^{\pi(1)}, \ldots, f_{n\alpha_k}^{\pi(n)}) d_{1\alpha_k}^{\pi(1)} + \cdots + g'_{x_n}(f_{1\alpha_k}^{\pi(1)}, \ldots, f_{n\alpha_k}^{\pi(n)}) d_{n\alpha_k}^{\pi(n)}. \end{aligned}$$

2. Denote $\mathbf{B}_{\alpha_k} = (\mathbf{B}_{\alpha_k}^-, \mathbf{B}_{\alpha_k}^+)$ the pair of k-th columns of $\mathbf{B}^-$ and $\mathbf{B}^+$ and define
$$\tilde{g}(\mathbf{A}_1, \ldots, \mathbf{A}_m) = \mathbf{B} = (\mathbf{B}^-, \mathbf{B}^+) \tag{4}$$
such that, for any $k = 0, \ldots, n$, we have
$$(\mathbf{B}_{\alpha_k}^-, \mathbf{B}_{\alpha_k}^+) = (\min_{\pi:\mathbf{m}\to\{-,+\}} \mathbf{B}_k^{\pi(1),\ldots,\pi(m)}, \max_{\pi:\mathbf{m}\to\{-,+\}} \mathbf{B}_k^{\pi(1),\ldots,\pi(m)}), \tag{5}$$
where min (and analogously max) is defined by
$$\min\left\{\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix}\right\} = \begin{pmatrix} a \\ b \end{pmatrix} \text{ if and only if } a \le c \text{ or } (a = c \text{ and } b \le d).$$

Example 1. One can simply check that

1. $(\mathbf{A}^-, \mathbf{A}^+) + (\mathbf{B}^-, \mathbf{B}^+) = (\mathbf{A}^- + \mathbf{B}^-, \mathbf{A}^+ + \mathbf{B}^+)$,
2. $k(\mathbf{A}^-, \mathbf{A}^+) = (k\mathbf{A}^-, k\mathbf{A}^+)$ and $-k(\mathbf{A}^-, \mathbf{A}^+) = (-k\mathbf{A}^+, -k\mathbf{A}^-)$ for $k \ge 0$,
3. $\exp[(\mathbf{A}^-, \mathbf{A}^+)] = (\exp[\mathbf{A}^-], \exp[\mathbf{A}^+])$ with
$$\exp[\mathbf{A}^-]_{\alpha_k} = \begin{pmatrix} \exp[f_{\alpha_k}^-] \\ \exp[f_{\alpha_k}^-] d_{\alpha_k}^- \end{pmatrix} \text{ and } \exp[\mathbf{A}^+]_{\alpha_k} = \begin{pmatrix} \exp[f_{\alpha_k}^+] \\ \exp[f_{\alpha_k}^+] d_{\alpha_k}^+ \end{pmatrix},$$

where the usual addition of matrices and the usual scalar multiplication are applied.

3 Fuzzy random variable

We follow the approach to fuzzy random variable proposed by Kwakernaak [11, 12] and later formalized in a clear way by Kruse and Meyer [10]. In this approach, a fuzzy random variable is viewed as a fuzzy perception/observation/report of a classical real-valued random variable. The model is stated as follows (we use the LU-fuzzy numbers to represent the values of fuzzy random variables).

Definition 2. Given a probability space $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathcal{F}_{LU}(v, w)$ is said to be a *fuzzy random variable* if for all $\alpha \in [0, 1]$ the two real-valued mappings $\inf X_\alpha : \Omega \to \mathbb{R}$, $\sup X_\alpha : \omega \to \mathbb{R}$ (defined so that for all $\omega \in \Omega$ we have that $X_\alpha(\omega) = [\inf(X(\omega))_\alpha, \sup(X(\omega))_\alpha]$) are real-valued random variables.

Since each LU-fuzzy number is representable by matrices $\mathbf{A}^-$ and $\mathbf{A}^+$, we can define fuzzy random variable using random matrices as follows.

[2] Let us stress that a careful choice of the domain $\mathcal{D}$ has to be given to ensure the correctness of the extended mapping $\tilde{g}$.

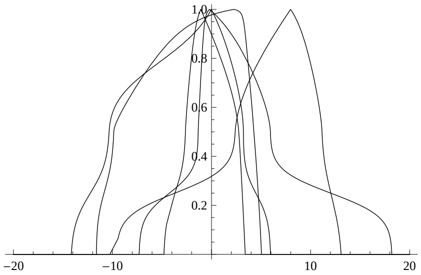


Figure 1: Pseudo-random LU-fuzzy numbers

Definition 3. Given a probability spaces $(\Omega, \mathcal{A}, P)$, a mapping $X : \Omega \to \mathcal{F}_{LU}(v, w)$ is said to be a *fuzzy random variable* (or FRV for short) if there exist a partition $0 = \alpha_0 < \cdots < \alpha_n = 1$ of the interval $[0, 1]$ and mappings

$$\mathbf{F}^-, \mathbf{F}^+, \mathbf{D}^-, \mathbf{D}^+ : \Omega \to \mathbb{R}^{n+1}$$

such that $p_i \circ \mathbf{F}^-$, $p_i \circ \mathbf{F}^+$, $p_i \circ \mathbf{D}^-$, $p_i \circ \mathbf{D}^+$, where p_i denotes i-th projection, are real-valued random variables for any $i = 1, \ldots, n+1$ and $X(\omega)$ is determined by

$$\begin{aligned} \mathbf{A}^-(\omega) &= \begin{pmatrix} \mathbf{F}^-(\omega) \\ \mathbf{D}^-(\omega) \end{pmatrix} = \begin{pmatrix} p_1 \circ \mathbf{F}^-(\omega) & \ldots & p_{n+1} \circ \mathbf{F}^-(\omega) \\ p_1 \circ \mathbf{D}^-(\omega) & \ldots & p_{n+1} \circ \mathbf{D}^-(\omega) \end{pmatrix} \\ \mathbf{A}^+(\omega) &= \begin{pmatrix} \mathbf{F}^+(\omega) \\ \mathbf{D}^+(\omega) \end{pmatrix} = \begin{pmatrix} p_1 \circ \mathbf{F}^+(\omega) & \ldots & p_{n+1} \circ \mathbf{F}^+(\omega) \\ p_1 \circ \mathbf{D}^+(\omega) & \ldots & p_{n+1} \circ \mathbf{D}^+(\omega) \end{pmatrix}. \end{aligned} \tag{6}$$

Definition 4. We say that two FRVs X and Y are *independent (identically distributed)*, if $p_i \circ \mathbf{F}^-_X$, $p_i \circ \mathbf{F}^+_X$, $p_i \circ \mathbf{D}^-_X$, $p_i \circ \mathbf{D}^+_X$ and $p_i \circ \mathbf{F}^-_Y$, $p_i \circ \mathbf{F}^+_Y$, $p_i \circ \mathbf{D}^-_Y$, $p_i \circ \mathbf{D}^+_Y$ are independent (identically distributed), respectively, for any $i = 1, \ldots, n+1$.

Note that using the interpolation we obtain random variables $u^-(\alpha)$ and $u^+(\alpha)$ under the probability space $(\Omega, \mathcal{A}, P)$. Hence, the FRVs introduced above are FRVs in the Kruse-Meyer sense [10].

On Fig. 1, we can see five pseudo-randomly generated LU-fuzzy numbers defined under the normal distribution (the kernels, i.e., the points with the membership degree equal to 1, are determined from $N(0, 4)$, further values are determined in such way that the difference between f_{α_i} and $f_{\alpha_{i+1}}$ is a random value from $N(0, 2)$).

4 Financial returns with imprecise data

Finally, in this section we apply the fuzzy set theory to built up models for financial returns with imprecisely specified input data (fuzzy data). For this purpose we assume three distinct models: (i) standard market model (Brownian motion) with fuzzy parameter, (ii) Brownian motion with fuzzy subordinator, and (iii) Brownian motion with fuzzyfied gamma subordinator.[3]

4.1 Standard market model

Assume a standard market model, ie. Brownian motion with drift μ and standard deviation σ:

$$X(t) = \mu t + \sigma \sqrt{t} \varepsilon, \tag{7}$$

where $\varepsilon \sim N(0, 1)$. This is a standard market model for (log-) returns of financial asset prices, such as stocks, commodities or exchange rates. Since it can be difficult to obtain a reliable estimate of standard deviation σ, we can define it as an LU-fuzzy number.

[3]Due to the lack of space fuzzy histograms for ten α-cuts of financial returns (X) assuming all three suggested models will be available at the time of the conference at one of the author's web pages, http://homel.vsb.cz/ tic02.

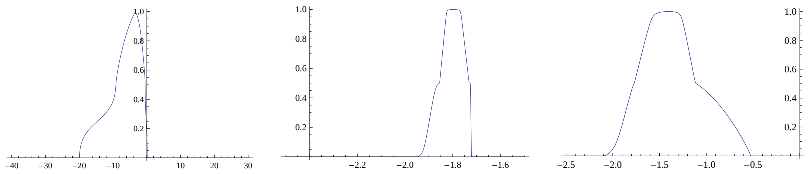


Figure 2: Quantiles for particular models ($q = 0.1$)

Since model (7) is based on the assumption of normal (Gaussian) returns, which is rarely fulfilled by empirical observations, it rather serves only as a building block for more complex (and reliable) models. One of the approaches is to evaluate the Brownian motion in a random time – a proxy to market activity (see eg. [4] for comprehensive review and or for potential applications and implications for risk management purposes). Bellow, we suggest two ways how to reformulate in finance quite popular variance gamma model with parameters θ and ϑ and gamma subordinator ν:

$$X(t) = \theta g(t) + \vartheta\sqrt{g(t)}\varepsilon, \tag{8}$$

where $g \sim Gamma(t/\nu, \nu)$ (ν describes the variance of the subordinator) and is independent of $\varepsilon \sim N(0,1)$.

4.2 Brownian motion with fuzzy subordinator

A common assumption is to evaluate the Brownian motion (7) in random time specified by non-negative subordinator. As an alternative, we can assume that in:

$$X(t) = \mu t + \theta(x_{LU}(t) - t) + \sigma\sqrt{x_{LU(t)}}\varepsilon \tag{9}$$

the subordinator, $x_{LU}(t)$, is a non-negative LU-fuzzy number centred around t.

4.3 Brownian motion with fuzzyfied gamma subordinator

As another alternative, we can follow variance gamma model (8) with fuzzyfication of gamma variables (the subordinator), ie. we first simulate variable $g \sim Gamma(1/\nu, \nu)$, which is used as an input to obtain LU-fuzzy number, g_{LU} (in line with Ex. 1 the LU-fuzzy number is centred around random gamma variable). In turn, fuzzy number g_{LU} is used as the subordinator of Brownian motion:

$$X(t) = \mu t + \theta(g_{LU}(g(t)) - t) + \sigma\sqrt{g_{LU}(g(t))}\varepsilon \tag{10}$$

4.4 Quantile estimation

In finance, it is often important to know the quantiles, ie. what is the limit result (return, loss, price, etc.) for a given probability, or alternatively, what is the probability of a given limit value of return or price. The estimated quantiles are commonly used within riskmangement (internal, capital regulation, etc.) and are referred to as Value at Risk (VaR). Next, the quantiles can be used to calculation of expected shortfalls or even option price. Illustrations of quantiles for $q = 0.1$ are provided in Fig. 2 for all three models.

5 Conclusion

Many issues of financial modeling and decision making require some knowledge about the future states. However, sometimes it is very difficult to get reliable parametrization of stochastic models. In this contribution we have formulated alternative approach to modeling of financial quantities based on fuzzy random values. This approach includes interval and crisp approach as a special case.

Suggested models of financial returns can be incorporated into classic problems of financial modeling, such as portfolio VaR estimation and option valuation. However, depending on the type of the problem, additional operations with LU-fuzzy numbers might be useful.

Acknowledgement

The support provided by VSB-TU Ostrava under the SGS project SP2011/7 and Institutional Research Plan MSM 6198898701 is kindly announced.

References

[1] Ban, A. I.: On the nearest parametric approximation of a fuzzy number – revisited. *Fuzzy Sets Syst.* **160** (21): 3027–3047, 2009

[2] Ban, A. I.: Triangular and parametric approximations of fuzzy numbers – inadvertences and corrections. *Fuzzy Sets Syst.*, **160** (21): 3048–3058, 2009

[3] Benko, M., Fengler, M, Hardle, W, Kopa, M.: On Extracting Information Implied in Options. *Computational Statistics* **22** (4): 543–553, 2007.

[4] Cont, R., Tankov, P.: *Financial Modelling with Jump Processes* Chapman & Hall/CRC press, 2004.

[5] Dubois, D., Prade, H.: Operations on fuzzy numbers. *Int. J. Syst. Sci.*, **9**: 613–626, 1978.

[6] Dubois, D., Prade, H.: *Fuzzy sets and systems. Theory and applications.* Academic Press, 1980.

[7] Gregory, J. A.: Shape preserving spline interpolation. *Comput. Aided Design*, **18**: 53-57, 1986

[8] Guerra, M. L., Stefanini, L.: Approximate fuzzy arithmetic operations using monotonic interpolations. *Fuzzy Sets Syst.* **150** (1): 5–33, 2005

[9] Klir, G. J., Yuan, Bo: *Fuzzy Sets and Fuzzy Logic: Theory and Applications.* Prentice Hall, 1995.

[10] Kruse, R., Meyer, K. D.: *Statistics with vague data.* D. Reidel Publishing Company, 1987.

[11] Kwakernaak, H.: Fuzzy random variables – I. Definitions and theorems. *Inf. Sci.* **15**: 1–29, 1978.

[12] Kwakernaak, H.: Fuzzy random variables – II. Algorithms and examples for the discrete case. *Inf. Sci.* **17**: 253–278, 1979.

[13] Nowak, P., Romaniuk, M.: Computing option price for Levy process with fuzzy parameters. *European Journal of Operational Research* **201**, 206-210, 2010.

[14] Sarfraz, M.: A rational cubic spline for the visualization of monotonic data. *Comput. Graphics* **24**: 509–516, 2000.

[15] Simonelli, M. R.: Fuzziness in valuing financial instruments by certainty equivalents. *European Journal of Operational Research* **135** (2), 296-302, 2001.

[16] Stefanini, L., Sorini, L., Guerra, M. L.: Parametric representation of fuzzy numbers and application to fuzzy calculus. *Fuzzy Sets Syst.* **157** (18): 2423–2455, 2006.

[17] Tanaka, H., Guo, P., Turksen, I. B.: Portfolio selection based on fuzzy probabilities and possibility distributions. *Fuzzy Sets and Systems* **111**: 387-397, 2000.

[18] Tichý, T.: Posouzení odhadu měnového rizika portfolia pomocí Lévyho model. *Politická ekonomie* **58** (4): 504–521, 2010.

[19] Tichý, T.: Posouzení vybraných možností zefektivnění simulace Monte Carlo při opčním oceňování. *Politická ekonomie* **56** (6): 772–794, 2008.

[20] Zmeškal, Z.: Application of the fuzzystochastic methodology to appraising the firm value as a European calls option. *European Journal of Operational Research* **135** (2): 303-310, 2001.

[21] Zmeškal, Z.: Value at risk methodology under soft conditions approach (fuzzystochastic approach). *European Journal of Operational Research* **161** (2): 337-347, 2005.

Acceleration of evacuation problem solving by branch and bound method with assistance of rapid excluding of branches

Ľubomír Toman[1]

Abstract. This paper deals with solving of the evacuation problem, particularly of the evacuation vehicle assignment problem which comprises its key part from the operation research point of view. The problem is solved by iterative approach commonly used in fuzzy optimization for different fixed values of total evacuation time. The branch and bound method is used for solving the problem. In this paper is shown the approach which tries to accelerate of branch and bound method convergence to a final solution with assistance of rapid excluding as many branches as possible. The branch excluding is used at the beginning of the solving process.

Keywords: branch and bound method, branch excluding, acceleration of solving process, evacuation, combinatorial problem.

JEL Classification: C61
AMS Classification: 90C57

1 Introduction

The evacuation belongs to basic means which are used for protection of people lives and health. When an emergency occurs and people who are located in any area are endangered by some thread the evacuation must be performed. The target of the evacuation is to evacuate people who are located in the endangered municipalities to the predetermined refuges in a minimal time. The vehicles located in the fleets can be used for execution of an evacuation.

The evacuation problem consists of few parts. The first part is based in determining the possible sets of refuges, fleets and endangered municipalities for a particular emergency. For example, in the case of the particular river overflows its banks it can be determined the refuges, fleets and municipalities. These predetermined sets comprise input data for other evacuation problem parts. Second part divides the group of inhabitants to the municipalities with smaller number of inhabitants [2]. The last part of the evacuation problem is the vehicle assignment problem (VAP) [9] which is the complex combinatorial problem. The target of this problem is to assign appropriate number of vehicles from the fleets to the municipalities so that the total evacuation time is minimal. The iterative approach [1] for solving VAP seems to be good way for solving this problem according to the results from [1] [3]. The branch and bound method is used for solving the problem. This paper deals with how to improve efficiency of this method.

Iterative approach

As have been said above the target of the VAP is to assign appropriate number of vehicles from the fleets to the municipalities so that the total evacuation time is minimal. But this leads to creating a nonlinear model [3]. In the iterative approach [1] it is not solved the VAP as big complex problem. Instead of that is iteratively solved the reduced vehicle assignment problem (RVAP) [4]. The target of the RVAP is to assign appropriate number of vehicles from the fleets to the municipalities so that the evacuation can be performed in the predetermined time. Let us denote this time with the symbol T^{max}. The information which we want to know in RVAP is only if we are able to perform the evacuation in the time T^{max} or not. If we are able to perform evacuation in the given time T^{max} so we decrease this time which represents upper bound of the time of the optimal solution and repeatedly solve the problem with the new value of T^{max}. If we are not able to perform evacuation in the given time then this time represents lower bound of the time of the optimal solution. By using such iterative approach we can reduce the interval given by lower and upper bound which contains the time of the optimal problem solution.

2 Solving RVAP by branch and bound method

There exists a set of municipalities J in the reduced vehicle assignment problem. Each municipality $j \in J$ is characterized by a number b_j of its population which must be evacuated to the predetermined refuge. The municipality is pre-assigned to the refuge in advance. The symbol I symbolizes the set of homogenous fleets. Each fleet $i \in I$

[1] University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Univerzitna 1, 01026 Žilina, Slovakia, lubomir.toman@fri.uniza.sk

has number N_i vehicles with the same capacity which is given by a number of people who can be transported in this vehicle simultaneously. The target of the RVAP is to assign appropriate number of vehicles from the homogenous fleets to the municipalities so that the evacuation can be performed in the predetermined time T^{max}.

Let us denote with the symbol q_{ij} the number of vehicles from the fleet i assigned for evacuation of the municipality j. The symbol a_{ij} symbolize the *evacuation capacity* that means the number of people who can be evacuated from the municipality j by one vehicle from the fleet i in the time T^{max}. The value of this symbol depends on a vehicle capacity and on a maximal number of visits of a vehicle in the municipality in the time T^{max} [5]. If this value is positive then the vehicle from the fleet i can be used for evacuation of the municipality j, in other words the municipality j is reachable from the fleet i in the time T^{max}. Let us denote with the symbol $I(j)$ the set of fleets $i \in I$ which the municipality j is reachable from. Analogically, the symbol $J(i)$ means the set of municipalities $j \in J$ which are reachable from the fleet i. Then the target of the RVAP is to find a feasible solution which satisfies the constraints (1)-(3) or to prove that such solution does not exist.

$$\sum_{j \in J(i)} q_{ij} \le N_i \qquad pre\ i \in I \tag{1}$$

$$\sum_{i \in I(j)} a_{ij} q_{ij} \ge b_j \qquad for\ \ j \in J \tag{2}$$

$$q_{ij} \in Z_0^+ \quad for\ \ i \in I,\ j \in J(i) \tag{3}$$

The constraints (1) ensure that we use for evacuation no more vehicles than the fleet contains. The constraints (2) ensure that every inhabitant of every municipality will be evacuated. The model (1)-(3) describes the decision problem because we want to just know if we are able to perform the evacuation in the time T^{max} or not. The graphic model of RVAP is on the **Figure 1**.

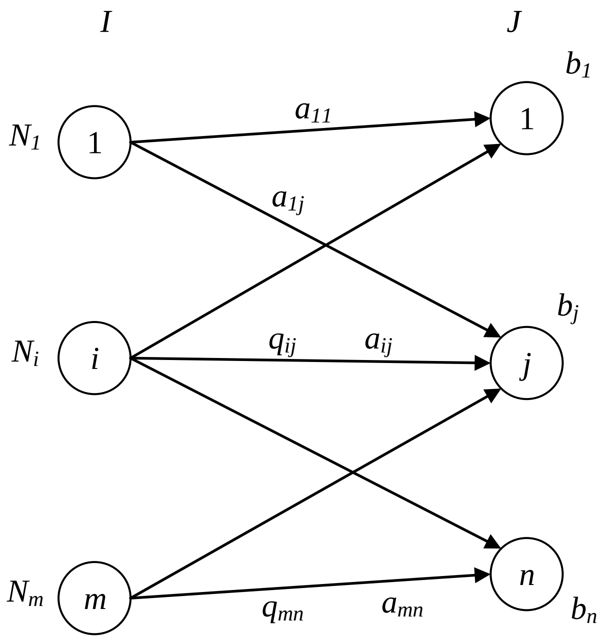


Figure 1 The graphic model of RVAP

The decision problem is transformed to the minimization one [5]. This minimization problem is solved by branch and bound method. In order to we can perform branching process we have added the constraints (4) and (5) to the model (1)-(3).

$$q_{ij} \le S_{ij} \quad for\ \ i \in I,\ j \in J(i) \tag{4}$$

$$q_{ij} \ge R_{ij} \quad for\ \ i \in I,\ j \in J(i) \tag{5}$$

The constraints (4) represent reasonable upper bound of the number of vehicles which are send out from the fleet i for evacuation of the municipality j. Analogically, the constraints (5) represent reasonable lower bound of this value. In the original approach the values of R_{ij} and S_{ij} are set to the following values for $i \in I$, $j \in J(i)$:

- $R_{ij} = 0$
- $S_{ij} = \min\{ N_i; \lceil b_j / a_{ij} \rceil \}$

There is no reason to send out less number of vehicles as zero and there is also no reason to send out more vehicles as the municipality j needs from the fleet i in the worst case ($\lceil b_j / a_{ij} \rceil$) or more as the fleet i contains (N_i). The minimization problem model with these coefficients creates the root of the searching tree.

2.1 Branch excluding

In the original approach the start values of the coefficients R_{ij} and S_{ij} are set to reasonable values which comprise the lower and upper bounds of q_{ij} values. In the alternative approach we try to set these lower and upper bounds more accurately. The alternative approach tries to accelerate branch and bound method convergence to the final solution with assistance of rapid excluding as many branches as possible. The branch excluding is used at the beginning of the solving process.

If we search for a feasible solution of LP relaxed minimization RVAP then the real values of q_{ij} will lie in the intervals which a solution exists for. Let us choose the particular couple vw (which is highlighted on the **Figure 2**) with the relevant coefficient q_{vw}, R_{vw} and S_{vw}.

If we decrease of one the value of S_{vw} stepwise to the value R_{vw} including, we can obtain the value $S`_{vw}$ which a feasible solution of LP relaxed problem does not exist for. If we do not allow to send out more than $S`_{vw}$ vehicles then a solution does not exist. So we must send out more than $S`_{vw}$ vehicles at least. Therefore the new and more accurate value of $R_{vw}$ will be $S`_{vw}+1$.

Analogically, if we increase of one the value of R_{vw} stepwise to the value S_{vw} including, we can obtain the value $R`_{vw}$ which a feasible solution of LP relaxed problem does not exist for. If we do not allow to send out less than $R`_{vw}$ vehicles then a solution does not exist. So we must send out less than $R`_{vw}$ vehicles at least. Therefore the new and more accurate value of $S_{vw}$ will be $R`_{vw}-1$.

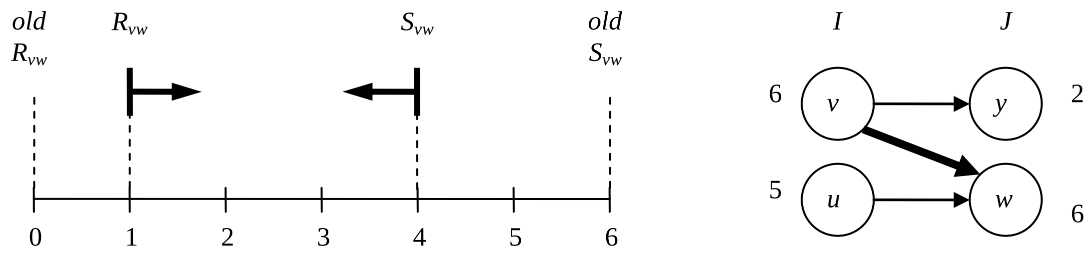


Figure 2 Example of more accurate setting of bounds

Let us give a small example (see the **Figure 2**). The set I contains the fleets v and u and the set J contains the municipalities y and w. The numbers on the right side of the figure have the following meaning. The numbers 6 and 5 which belong to the fleets v and u mean the number of vehicles which are at disposal in the particular fleets. The numbers 2 and 6 which belong to the municipalities y and w mean the number of vehicles which the particular municipalities need to get (let us assume that the coefficients a_{ij} are the same value for each fleet-municipality couple). The chosen couple is vw. In the original approach the values of R_{vw} and S_{vw} are set to values 0 and 6 respectively (old values). If we set the upper bound S_{vw} of q_{vw} stepwise to the values 5, 4, 3 etc. down to zero (actual value of R_{vw}) then there is no feasible solution for zero value of S_{vw}. This is because the municipality w can obtain only 5 vehicles from the fleet u and none vehicle from the fleet v but the municipality needs to obtain 6 vehicles. Therefore the value of q_{vw} must be 1 at least. Consequently the new value of R_{vw} will be 1 as well. Analogically, if we set the lower bound R_{vw} stepwise to the values 2, 3, 4 etc. up to value 6 (actual value of S_{vw}) then there is no feasible solution for the value 5 of R_{vw}. This is because the municipality y can obtain only 1 vehicle from the fleet v. But this number of vehicles is not sufficient for this municipality. Therefore the value of q_{vw} must be 4 at most. Consequently the new value of S_{vw} will be 4 as well. New values of R_{vw} and S_{vw} comprise more accurate lower and upper bound of the value q_{vw}. If the value of q_{vw} lied out of the interval $\langle R_{vw}; S_{vw} \rangle$ a feasible solution would not exist.

With using this approach it is possible to constrain the values of q_{ij} for $i \in I$, $j \in J(i)$ more accurately and also to exclude as many branches which do not include a feasible solution for (1)-(3) as possible from searching tree. It is likely that this approach makes the process of searching of a feasible solution more effective because there is a presumption that with using this approach we can obtain this solution by exploring fewer number of searching tree nodes.

3 Experiments

We have implemented and tested the suggested rapid excluding of branches from searching tree at the beginning of the branch and bound method. We have performed experiments on a personal computer equipped with Intel Pentium 4 with parameters 2.41 GHz and 1 GB RAM. We have used the specialized decision support system (DSS) [6] which is given for the evacuation problem solving. We have used for testing the suggested approach 20 instances of the problem which were created on transportation network of Slovak Republic. We have solved these instances in two series. The suggested excluding of branches have not been used in the first series and have been used in the second series. We have obtained the time of the best solution (row *Best*), the time of the potentially optimal solution (row *Pot*) and the computational time (*Comp*) which have been required for finding the solution. The computational time is given in format [HH:MM:SS]. The time of the potentially optimal solution is the time which a feasible solution can exist for but we have not been able to obtain it. We have also obtained this two times with using the general optimization solver Xpress IVE [7] [8] in order to we can determine efficiency of the suggested approach. The results of the experiments are shown in the **Table 1**. The symbol $|I|$ means the number of fleets and the symbol $|J|$ means the number of municipalities.

	Instance	**01**	**02**	**03**	**04**	**05**	**06**	**07**	**08**	**09**	**10**
	|I|	**10**	**7**	**11**	**7**	**9**	**9**	**9**	**10**	**10**	**28**
	|J|	**9**	**13**	**20**	**16**	**9**	**17**	**17**	**16**	**17**	**55**
Xpress	**Pot**	396	277	67	125	61	87	141	204	88	117
	Best	396	277	67	126	61	89	143	204	90	123
Without branch excluding	**Pot**	393	277	67	125	61	87	141	200	87	117
	Best	403	278	70	129	61	95	150	210	92	129
	Comp	1:10:10	0:27:39	0:44:16	0:44:09	0:09:30	1:20:02	1:40:02	1:40:01	0:23:04	1:26:07
With branch excluding	**Pot**	394	277	67	125	61	87	141	200	87	117
	Best	403	278	70	129	61	95	150	210	92	129
	Comp	1:09:59	0:21:56	0:44:28	0:44:35	0:09:30	1:00:07	1:40:03	1:40:02	0:23:08	1:34:57

	Instance	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
	|I|	**12**	**16**	**14**	**19**	**25**	**10**	**18**	**42**	**4**	**16**
	|J|	**22**	**33**	**20**	**24**	**37**	**24**	**14**	**32**	**15**	**26**
Xpress	**Pot**	119	155	117	192	430	196	409	259	60	81
	Best	121	158	120	195	430	200	409	265	60	81
Without branch excluding	**Pot**	119	154	117	188	424	191	409	259	60	81
	Best	127	170	120	205	463	218	409	289	61	81
	Comp	1:10:52	1:30:26	1:00:03	2:00:05	1:25:17	1:44:24	-	1:40:11	0:20:01	-
With branch excluding	**Pot**	119	154	117	188	424	193	409	259	60	81
	Best	127	170	120	205	463	218	409	289	61	81
	Comp	1:31:05	1:31:02	1:00:11	2:00:25	1:26:54	1:08:52	-	1:44:56	0:20:01	0:00:01

Table 1

As we can see in the **Table 1** the suggested approach has not brought an expected improvement because the solution which have been found with this approach are the same quality as the solution which have been found without this approach. This fact exists because the number of excluded branches has had the value zero for a lot of the values of the time T^{max}. It means that the suggested approach has not brought wished effect often. The branches have been excluded from searching tree mainly for the values of the time T^{max} which have been close to the time of the potentially optimal solution which have been obtained with using the specialized DSS. This fact has inspired us to test the behavior of the suggested algorithm for these times more carefully. The efficiency of the algorithm is given by the number of bounds of values q_{ij} which are set more accurately regardless if it is lower or upper bounds. We have performed new experiments. We have solved the instances for individual values of the time T^{max}. We have stepwise changed this time from the value of the time of the potentially optimal solution to the value of the time of the best found solution with specialized DSS. The results are shown in the **Table 2**. In the first row are the labels of the instances. *T* means the time T^{max} which has been the instance solved for. *NoB* means the number of bounds which have been set more accurately.

01		02		03		04		05		06		07		08		09		10	
T	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**
393	65	277	0	67	10	124	90	61	15	87	34	141	18	200	52	87	155	117	115
394	36	278	0	68	0	125	48	62	15	88	21	142	11	201	49	88	57	118	0
395	27			69	0	126	9	63	11	89	1	143	3	202	49	89	55	119	0
396	16			70	0	127	9	64	8	90	1	144	3	203	43	90	27	120	0
397	16					128	3	65	6	91	1	145	2	204	3	91	16	121	0
398	15					129	0	66	6	92	0	146	0	205	0	92	5	122	0
399	15							67	6	93	0	147	0	206	0			123	0
400	15							68	0	94	0	148	0	207	0			124	0
401	7							69	0	95	0	149	0	208	0			125	0

11		12		13		14		15		16		17		18		19	
T	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**	**T**	**NoB**
119	0	154	1	117	0	188	397	424	832	191	194	409	0	259	87	60	0
120	0	155	0	118	0	189	258	425	391	192	118			260	0	61	0
121	0	156	0	119	0	190	137	426	297	193	66			261	0		
122	0			120	0	191	113	427	289	194	6			262	0		
123	0					192	4	428	273	195	1			263	0		
124	0					193	0	429	120	196	0			264	0		
125	0					194	0	430	0	197	0						
126	0					195	0	431	0	198	0						
127	0					196	0	432	0								

Table 2

According to these results we can conclude that the suggested way of branch excluding is efficient mainly for the times T^{max} which are lower than the time of the optimal or the best obtained solution. This fact is valid for the instances 04, 06, 07, 08 and 15 and partially valid for instances 01, 09, 10, 14, 16 and 18. Based on these result we can say the hypothesis that the number of bounds which have been set more accurately is low for the time T^{max} which is equal to the time of the optimal or the best obtained solution. By using the suggested way of branch excluding we can give precision to estimation of the optimal solution time.

4 Conclusion

In this paper we have dealed with the evacuation problem solving particularly with the vehicle assignment problem solving. This problem comprises the key part of the evacuation problem from the operational research point of view. We have mentioned the iterative approach which is used for solving VAP. In this approach we solve RVAP for the particular times T^{max}. The branch and bound method is used for solving this problem. In the paper we have shown the way of excluding as many branches from searching tree as possible at the beginning of the solving process. We have experimentally verified the suggest way. The established results can be used for gaining more accurate estimation of evacuation time.

Acknowledgements

This contribution was supported by research grant VEGA 1/0361/10 „*Optimal design of public service systems under conditions of uncertainty*“.

References

[1] JANÁČEK, J. *Handling of a non-linear model of the evacuation plan design by IP-solver*. In *Proceedings of the 10th international Symposium on OPERATIONAL RESEARCH*. Nova Gorica, Slovenia : Slovenian Society Informatika, 2009. ISBN 978-961-6165-30-3, s. 279-287.

[2] JANÁČEK, J. *Tvorba evakuačních plánů pomocí na trhu dostupných IP-solverů*. In *Úlohy diskrétní optimalizace v dopravní praxi*. Pardubice : Univerzita Pardubice, 2009. ISBN 978-80-7395-193-1, s. 50-60.

[3] JANÁČEK, J., SIBILA, M. *Optimal evacuation plan design with IP-solver*. In COMMUNICATIONS. ISSN 1335-4205, 2009, vol. 11, č. 3, s. 29-35.

[4] TOMAN, Ľ. *Informatické nástroje a riešenie úloh evakuácie* : Písomná práca k dizertačnej skúške. Žilina : Žilinská univerzita v Žiline, 2011. 60 s.

[5] TOMAN, Ľ. *Min-Max Approach for Evacuation Problem Supported by BB-Search*. In Transcom : 9-th EUROPEAN CONFERENCE OF YOUNG RESEARCH AND SCIENCE WORKERS. Slovak republic, Žilina : University of Žilina, 2011. ISBN: 978-80-554-0369-4, p. 215-218.

[6] TOMAN, Ľ. *Object oriented design of decision support system for field of operational research* (in Slovak). In Zborník príspevkov konferencie Infotrans : Mezinárodní konference „Informační technologie v dopravě“. 2011.

[7] XPRESS-Mosel “User guide”. Dash Associates, Blisworth, 2005, UK, 99 s.

[8] XPRESS-MP“Getting Started”. Dash Associates, Blisworth, UK, 2005, 105 s.

[9] YUSOFF, M., ARIFFIN, J., MOHAMED, A. *An Improved Discrete Particle Swarm Optimization in Evacuation Planning*. In: International Conference of Soft Computing and Pattern Recognition. IEEE Computer Society, Malaysia. 2009. ISBN 978-0-7695-3879-2, s. 49-53.

Bayesian Estimation of a Comprehensive Small Open Economy Model on the Czech Data

Jaromír Tonner [1], Jiří Polanský [2], Osvald Vašíček[3]

Abstract. The contribution deals with the development and Bayesian estimation of the SOE model suitable for Czech data based analyses. The model follows recent development in monetary New Keynesian DSGE models and is extended with several technology processes to fit Czech stylized facts. The model is thus sufficiently rich and general for time-varying parameters estimations and non-linear filtrations. The second part of the contribution presents results of the Bayesian estimation of the model. Estimations of such complex models on Czech data are not usual in the literature, and we thus believe that such results might be informative.

Keywords: DSGE models, Bayesian methods

JEL classification: D58, E32, E47, C11, C13
AMS classification: 90C15

1 Introduction

DSGE models are today the most used (and suitable) tools for medium-term monetary policy analyses. Their structure provides a coherent framework that can capture main stylized facts of data. The core of various DSGE models is relatively similar. DSGE models consist of FOC equations from intra- and inter-temporal optimization problems of various agents and thus capture the behavioural characteristics of an economy and improve the Lucas critique problem [5]. With respect to model purposes, the FOC equations are subsequently suplemented with various features to fit the data. The most promiment model features are wedges (inserted into steady-state assumptions) and technology processes (see Tovar [10] and Andrle et al. [1]). These features play an important role for real-time forecasting or data filtrations and their "number" depends on the type of analyzed economy and objectives of the model.[1]

For our future work (Tonner [8] and Tonner et al. [9]), we need a sufficiently rich and general SOE DSGE model suitable for Czech stylized facts. The former stems from the complexness of the Czech small economy which is still undergoing through structural changes. The latter is directly designed for non-linear filtrations (e.g. there are no Phillips curves in the model). We believe that such a model framework is useful for (and in some respects necessary) for analyzing structural parameters drifting in the Czech economy model.

In this contribution we present the model framework and its Bayesian estimation. The model follows recent development in monetary New Keynesian DSGE models (Burriel et al. [2]) which is amended for the Czech data analysis (simplified fiscal sector, introducing reexports etc.) Subsequently, the model is extended with several technology processes to fit Czech stylized facts (Andrle et al. [1]). The overall model thus provides a framework for real-time Czech data forecasting and is suitable for various filtrations (linear or non-linear). The second part of the contribution presents the results of the Bayesian estimation of the model. Estimations of such complex models on Czech data are not usual in the literature, and we thus believe that such results might be (at least) informative for future work.

[1]Masaryk University/Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: jtonner@tiscali.cz
[2]Faculty of Economics and Administration, Masaryk University, Lipová 41a, 602 00 Brno and Macroeconomic Forecasting Division, Czech National Bank, Na Příkopě 28, 115 03 Praha 1., e-mail: Jiri.Polansky@cnb.cz
[3]Masaryk University/Faculty of Economics and Administration, Department of Economics, Lipová 41a, 602 00 Brno, e-mail: osvald@econ.muni.cz

[1]Adding various features into a DSGE model should not be ad hoc. As Andrle et al. [1] note for the case of regulated (administered) prices: "In a structural model regulated prices require structural interpretation". We believe that this might hold in general.

In our future work, we would like to investigate the stability of structural parameters of a rich DSGE model for the Czech economy. Our aim is to find out whether structural parameters are drifting in a framework without supplemented technologies and compare these results with parameter movements in a model that is equipped with technology processes. In both cases, we will set all the structural parameters as time-varying.[2] We will filter data using the Kalman filter on the first order approximated model with technologies (i) switched-on technologies and (ii) switched-off. In this contribution, we will compare the bayesian estimations on both versions of the model. This analysis shows how much sensitive are parameters estimates to the presence of technologies.

2 Model

The model is based on two current models. First, we use the model of Burriel et al. [2] designed for the Spanish economy as our starting framework. This model follows the current generation of DSGE models for the inflation targeting regime. It is sufficiently rich and general within its sectors' structure and contains many widely-used modelling features like real and nominal rigidities, technology growths, local currency pricing etc. Moreover, it is also described in literature in great detail.[3] To cope with the Czech data, we alter some sectors with the purpose to have a more simple fiscal policy treatment and introduce reexports. Subsequently, the model is extended with several features according to Andrle et al. [1] to fit medium-term Czech data characteristics.[4]

The model has a relatively standard and general structure with optimizing agents and rational expectations. It contains a set of real (internal habit formation, capital adjustment costs) and Calvo-type nominal rigidities with indexation parameters. The production structure with intermediate and final goods producing firms enables to capture the GDP accounts while the local currency pricing helps to incorporate the gradual exchange rate pass-through. The model is closed with a debt-elastic premium according to Schmitt-Groh and Uribe [7]. The overall structure of the model is described in Figure 1.

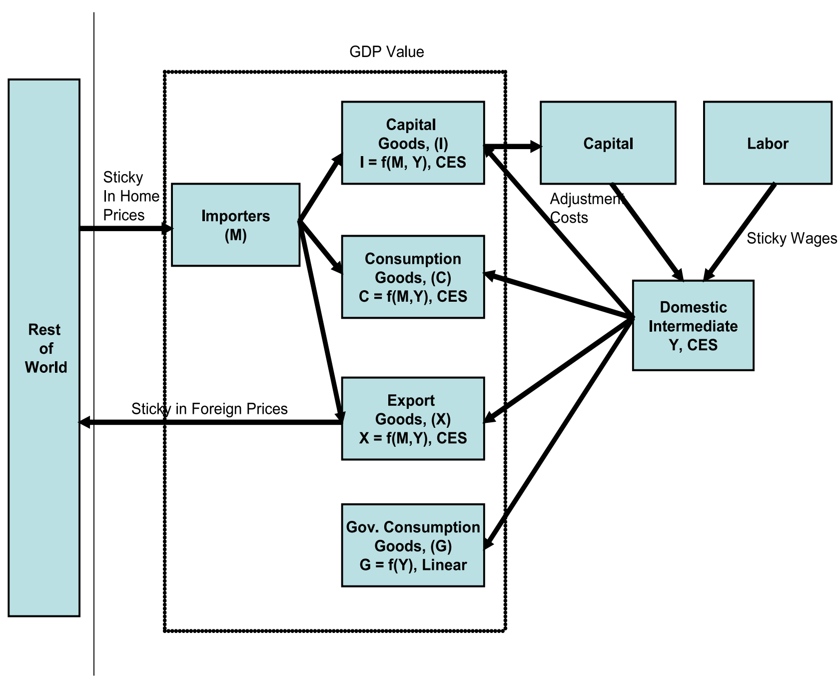


Figure 1: Structure of the Model

Households consume, save, and set their nominal wages subject to downward sloping demand and the

[2]We follow directly the methods proposed in Fernndez-Villaverde and Rubio-Ramrez [4].
[3]See also Fernndez-Villaverde and Rubio-Ramrez [4].
[4]Andrle et al. [1] describe the new Czech National Bank's core model and summarize the main stylized facts of the Czech economy Also, they discuss ways how to structurally incorporate them into the monetary DSGE framework.

Calvo specification. There are four final goods producing sectors in the model - consumption, investment, export and government. Consumption, investment and export firms purchase both (domestic and imported) intermediate composite inputs, the government sector uses the domestic intermediate goods only. The monopolistic competition is present in the intermediate sectors and in the export sector. The central bank operates under the inflation targeting regime. It sets its one-period nominal interest rate through open market operations according to a Taylor-type rule. The fiscal policy is Ricardian in the model. Besides our focus on the monetary analysis, we prefer a simple fiscal setting due to possible ambiguities and uncertainties in analyzing fiscal policy effects in a non-Ricardian setting.

As was noted, DSGE models are usually supplemented with technologies to get them closer to the data. Typical examples include sector technologies which capture important sector-specific features. These processes can be regarded as time-varying parameters. Following Andrle e al. [1], we incorporate three exogenous processes into the model. First, the export-specific technology captures the Harrod-Balassa-Samuelson effect, which implies real exchange rate appreciation in consumer prices in the steady state. Second, we aim to capture some aspects of the high openness of the Czech economy, especially the fact that exports are very import intensive. Thus, we assume the trade openness technology, which helps us to work with re-export effects in a model-consistent way. Third, since the Czech headline CPI inflation is still influenced by regulated prices, we incorporate regulated price technology into the model. This technology is a proxy for the regulated prices goods sector.

First, the export-specific technology captures the relative effectiveness of the tradable sector with respect to the non-tradable sector in the domestic economy (the Harrod-Balassa-Samuelson effect). Its growth increases the relative productivity in the tradable goods production sector resulting in a pressure for higher wages in the (less productive) non-tradable production sector. Preventing from higher wage gap leads to an inflation pressures. On the other hand, higher productivity in the tradable sector implies an exchange rate appreciation as higher production with lower costs moves balance of payment into surplus. The response of the monetary authority is thus ambiguous and depends on the relative strength of both effects. In our case, the effect of the stronger exchange rate prevails and the monetary authority decreases interest rates. Second, positive trade openness technology shock increases equally the growth rate of exports, imports and the foreign demand. The aim of this technology is primarily to remove the effect of re-exports from the observed time series. Other model variables are not affected by this shock. Third, a positive shock to the growth rate of regulated prices directly increases consumer price inflation. The consumer basket is composed of both non-regulated goods and goods whose prices are heavily regulated (e.g. rents and energy prices for the Czech case). The monetary authority must respond by raising interest rates. Since the income from increased regulated prices belongs to government, the government consumption increases. As government consumption goods are produced from domestic intermediate goods only, domestic intermediate goods are replaced by imported intermediate goods in other production sectors. The balance of payments moves to a deficit implying an exchange rate depreciation.[5]

3 Model Estimation

In this section, we present and discuss the results of the time-invariant Bayesian estimation on quarterly Czech and Eurozone data. The posterior distributions are constructed with the Metropolis-Hastings algorithm of the Dynare Toolbox [3] and [6].[6]

3.1 Data

The quarterly Czech data sample covers 60 observations from 1996Q1 to 2010Q4. We use 16 time series as observables for the estimations. Seasonally adjusted national accounts data stem from the Czech Statistical Office (CZSO). We use the real volumes of consumption, investment, government spending, export, import and deflators for investment, export and import. [7] The headline CPI inflation also comes from the CZSO. For the labour market data, we use the average nominal wage in the business sector (CZSO) and employed in the economy time series as an observable for the labour demand, the latter gained

[5] Regulated prices are approximated by an exogenous process in the model, i.e. there is no direct link to the observed regulated prices.

[6] One million of draws is used for the estimation.

[7] We use the CPI inflation instead of the consumption deflator. The government deflator is not necessary because of the simple fiscal policy treatment in the model.

from the Labour Force Sample Survey.[8] The exchange rate is the CZK/EUR while the domestic interest rate is the 3M PRIBOR. The foreign observables include the 3M EURIBOR, the effective Eurozone PPI from the Consensus Forecast (CF), and the foreign demand, acquired from the effective Eurozone GDP (CF).[9]

We allow for the measurement errors in the model to deal with high data uncertainty associated with frequent data revisions, methodology changes, or high volatility of quarter-on-quarter growth rates of several time series. Measurement errors are incorporated on levels via measurement equations where we let observables differ from measurements.

3.2 Priors

First of all, we set steady-state growth rates. The overall growth in the model is slightly above 4.5 % a year which is approximately consistent with the previous GDP growth of the Czech economy.[10] The steady-state population growth is set to zero as we assume its no role for determining the long-run growth in the model. We set the steady-state nominal appreciation rate to 2.4 % a year. This value corresponds approximately to the period until 2009.[11]

The steady state inflation corresponds to the 2 % inflation target set in annual terms.[12] The foreign inflation steady state is calibrated according to the 2 % inflation target of the ECB. The foreign demand growth for the domestic export is set at a pace of 9 % a year implying the long-run EU GDP growth at 2.25 % (dividing by a factor of four). The steady-state foreign nominal interest rate is calibrated to 4 % annually.

3.3 Posteriors

This subsection summarizes the results of the estimations. The table compares estimations of both versions of the model, i.e. with and without technologies. We comment only significant differences between posterior estimates.

The estimation without technologies has several different results. First, the risk premium parameter is higher. As this parameter affects the relationship between domestic and foreign nominal interest rates, the estimation has significant effects for the model behaviour. Second, there are significant differences of almost all price and wage setting parameters. The estimates of Calvo parameter for the export prices and the estimates of indexation parameter for domestic prices are the biggest differences what may completely modify pass-through of sector cost to final prices. Third, inactive technologies also modify the estimates of openness characteristics of the Czech economy. The investment sector becomes nearly 100 percent import-intensive. The share of domestic consumption goods in the total consumption basket (home bias in consumption) increases from approximately 37 percent to 62 percent. This is probably connected with switching-off of the regulated prices technology which represents a proxy for the regulated prices sector. And fourth, the estimation without technologies changes standard deviation of shocks, especially strenghten the investment specific technology at the expense of the neutral technology. Switching-off of the openness technology and export specific technology also increases the standard deviation of foreign demand shock, premium shock and wedge forex shock.

[8] Both time series are seasonally adjusted to get their trend-cyclical components.

[9] The weights used in the calculation of the effective variables are the shares of the individual euro area economies in the foreign trade turnover of the Czech Republic (see Inflation Reports of the CNB). The foreign demand is acquired by multiplying the foreign GDP by a factor four. See Andrle et al. [1].

[10] In the model, the overall steady-state growth is generated via the neutral technology.

[11] Adding the latest data would decrease the appreciation rate as there was a considerable depreciation. In this respect, we assume that financial crises might not affect long-run steady state of an economy and we thus assume that the Czech economy will return to the long-term appreciation. On the other hand, the period 1998-2008 would result in stronger appreciation but the ongoing convergence process would imply gradually lowering rates. Thus, the overall long-run growth might be relatively balanced.

[12] In quarterly terms relevant for the model, this target corresponds to (2/400+1=1.005).

Parameter		**Prior Mean**	**Dist**	**Posterior Mean**	**Posterior Mean NT**
Preferences					
Habits	h	0.900	beta	0.9286	0.9163
Labour supply coef.	ψ	8.832	gamma	8.8398	8.7053
Frisch elasticity	ϑ	1.250	gamma	1.2567	1.2677
Adjustment costs					
Investment	κ	20.000	gamma	20.0664	19.9685
Capital utilization	$\gamma 2$	0.280	gamma	0.2879	0.2650
Risk premium	ρ_{b_W}	0.018	beta	0.0009	0.0172
Elasticities of substitution					
Domestic intermediate goods	ϵ	5.000	gamma	4.9789	5.0645
Import goods	ϵ_M	9.000	gamma	9.0033	9.0927
Export goods	ϵ_x	9.400	gamma	9.4774	9.3152
World goods	ϵ_W	1.500	gamma	1.3049	1.2952
Consumption goods	ϵ_c	7.600	gamma	7.5346	7.4889
Investment goods	ϵ_i	7.600	gamma	7.5504	7.4865
Labour types	η	7.000	gamma	7.0007	6.9896
Price and wage setting					
Calvo dom. prices	θ_p	0.500	norm	0.5152	0.5982
Calvo exp. prices	θ_x	0.080	gamma	0.0816	0.3486
Calvo imp. prices	θ_M	0.750	norm	0.7645	0.6108
Calvo wages	θ_w	0.380	norm	0.5371	0.6021
Index. dom. prices	χ_p	0.750	gamma	0.6799	0.9709
Index. exp. prices	χ_x	0.350	gamma	0.2657	0.1130
Index. imp. prices	χ_M	0.500	gamma	0.6174	0.4864
Index. wages	χ_w	0.920	beta	0.8562	0.9550
Monetary policy					
Taylor rule (int. rates)	γ_R	0.750	beta	0.8300	0.8127
Taylor rule (output gap)	γ_y	0.220	gamma	0.2320	0.2225
Taylor rule (inflation)	γ_Π	1.150	gamma	1.1590	1.1523
Fiscal policy					
Public consumption	ρ_g	0.500	beta	0.4871	0.5258
Home bias					
Home bias in consump.	n_c	0.280	beta	0.3672	0.6237
Home bias in invest.	n_i	0.120	beta	0.0713	0.0165
Home bias in export	n_x	0.400	beta	0.4306	0.4123
Growth rates					
Invest. spec. tech.	Λ_μ	1.000	norm	1.0000	1.0000
General tech.	Λ_A	1.009	norm	1.0090	1.0090
Population	Λ_L	1.000	norm	1.0000	1.0000
ER appreciation	$\dot{ex}$	0.994	norm	0.9942	0.9942
Standard devs of shocks					
Invest. spec. tech.	σ_μ	0.032	invg	0.1517	0.2103
Neutral tech.	σ_A	0.032	invg	0.0084	0.0025
Intertemp. preferences	σ_d	0.044	invg	0.2164	0.2831
Hours preferences	σ_φ	0.001	invg	0.0008	0.0009
Monetary policy	σ_m	0.004	invg	0.0033	0.0029
Foreign prices	σ_{π_W}	0.010	invg	0.0097	0.0097
Foreign demand	σ_{y_W}	0.021	invg	0.0133	0.0771
World interest rate	σ_{R_W}	0.012	invg	0.0018	0.0017
Premium	σ_{prem}	0.004	invg	0.0030	0.0049
Openness	σ_{aO}	0.083	invg	0.0322	...
Regulated prices	σ_{aR}	0.010	invg	0.0084	...
Government specific	σ_{aG}	0.030	invg	0.0259	...
Target	σ_{target}	0.010	invg	0.0349	0.0494
Wedge forex	σ_{forex}	0.0100	invg	0.0610	0.2160
Wedge euler	σ_{euler}	0.0100	invg	0.0099	0.0069

4 Conclusion

Incorporating sector specific technologies has significant effects to the Bayesian estimation of the model. It strongly effects estimates of parameters describing the financial sector, pass-through of costs to final prices, openness characteristics of the economy as well as estimates of standard deviations of model shocks. As sector specific technologies play an important role in capturing changes in individual sectors, this result is intuitive.

Acknowledgements

The financial support of specific research at ESF MU, project MUNI/A/0943/2009 and the MMT project Research Center 1M0524 are gratefully acknowledged. We are grateful to Jan Brůha for many helpful comments and advices.

References

[1] Andrle, M., Hlédik, T., Kameník, O., and Vlček, J. Implementing the New Structural Model of the Czech National Bank. CNB Working Paper Series 2/2009, 2009.

[2] Burriel, P., Fernández-Villaverde, J., and Rubio-Ramírez, J. F. MEDEA: a DSGE model for the Spanish economy. *Journal of the Spanish Economic Association*, 1 (2010), 175–243.

[3] Dynare. http://www.cepremap.cnrs.fr/dynare/.

[4] Fernández-Villaverde, J., and Rubio-Ramírez, J. F. How Structural Are Structural Parameters. NBER Working Paper 13166, 2007.

[5] Lucas, R. E. Econometric Policy Evaluation: A Critique *The Phillips Curve and Labor Markets*, 1 (1976), 19-46.

[6] Koop, G. Bayesian Econometrics. John Wiley and Sons, 2003.

[7] Schmitt-Grohé, S., and Uribe, M. Closing small open economy models. *Journal of International Economics 61* (2003), 163–185.

[8] Tonner, J. Expectations and Monetary Policy in Open Economy. PhD Thesis, Faculty of Economics and Administration, Masaryk University, forthcoming.

[9] Tonner, J., Polanský, J., and Vašíček, O. Parameters Drifting in an Estimated DSGE Model on the Czech Data. Mimeo.

[10] Tovar, C. E. DSGE Models and Central Banks. BIS Working Paper 258, 2008.

Fair voting majorities in proportional representation

František Turnovec[1]

Abstract. In parliaments elected by proportional systems the seats are allocated to the political parties roughly proportionally to the shares of votes for the party lists obtained in elections. Assuming that members of the parliament representing the same party are voting together, it has sense to require that distribution of the influence of the parties in parliamentary decision making is proportional to the distribution of seats. There exist measures (so called voting power indices) reflecting an ability of each party to influence outcome of voting. Power indices are functions of distribution of seats and voting quota (where voting quota means a minimal number of votes required to pass a proposal). By a fair voting rule we call such a quota that leads to proportionality of influence to relative representation. Usually simple majority is not a fair voting rule. That is the reason why so called qualified or constitutional majority is being used in voting about important issues requiring higher level of consensus. Qualified majority is usually fixed (60% or 66.67%) independently on the structure of political representation. In the paper we use game-theoretical model of voting to find a quota that defines the fair voting rule as a function of the structure of political representation. Such a quota we call a fair majority. Fair majorities can differ for different structures of the parliament. Concept of a fair majority is illustrated on the data for the Lower House of the Czech Parliament elected in 2010.

Keywords: fair majority, power indices, quota interval of stable power, simple weighted committee, voting power.

JEL Classification: C71, D72, H77
AMS Classification: 91A12, 91A40, 05C65

1 Fairness in voting

A qualified majority is a requirement for a proposal to gain a specified level or type of support which exceeds a simple majority (over 50%). In some jurisdictions, for example, parliamentary procedure requires that any action that may alter the rights of the minority has a qualified majority support. Particular designs of qualified majority (such as 60% or two-thirds majority) are selected "ad hoc", without quantitative justification. In this paper we try to provide such a justification, defining qualified majority by a "fair quota", providing each legislator with (approximately) the same influence, measured as an a priori voting power.

Let us consider a committee with n members. Each member has some voting weight (number of votes, shares etc.) and a voting rule is defined by a minimal number of weights required for passing a proposal. Given a voting rule, voting weights provide committee members with voting power. Voting power means an ability to influence the outcome of voting. Voting power indices are used to quantify the voting power.

The concept of fairness is being discussed related to the distribution of voting power among different actors of voting. This problem was clearly formulated by Nurmi [5], p. 204: "*If one aims at designing collective decision-making bodies which are democratic in the sense of reflecting the popular support in terms of the voting power, we need indices of the latter which enable us to calculate for any given distribution of support and for any decision rule the distribution of seats that is 'just'. Alternatively, we may want to design decision rules that – given the distribution of seats and support – lead to a distribution of voting power which is identical with the distribution of support.*"

Voting power is not directly observable: as a proxy for it voting weights are used. Therefore, fairness is usually defined in terms of voting weights (e.g. voting weights are proportional to the results of an election). Assuming that a principle of fair distribution of voting weights is selected, we are addressing the question of how to achieve equality of voting power (at least approximately) to relative voting weights. For evaluation of voting power we are using concepts of a priori power indices (a comprehensive survey of power indices theory see in Felsenthal and Machover [2]). The concepts of optimal quota, introduced by Słomczyński and Życzkowski [10], [11] for the EU Council of Ministers distribution of national voting weights (weights equal to square roots of

[1] Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00 Prague 1, Czech Republic, e-mail: turnovec@fsv.cuni.cz

.

population and quota that provides each citizen of the EU with the same indirect voting power measured by Penrose-Banzhaf index independently on her national affiliation), and of intervals of stable power (Turnovec [12]) are used to find, given voting weights, a fair voting rule minimizing the distance between actors' voting weights and their voting power.

In the second section basic definitions are introduced and the power indices methodology is shortly resumed. The third section introduces concepts of quota intervals of stable power and fair quota. The fourth section applies the concept of fair quota for the Lower House of the Czech Parliament elected in 2010..

2 Committees and voting power

A simple weighted committee is a pair [N, **w**], where N be a finite set of n committee members $i = 1,2,...,n$, and **w** = (w_1, w_2, …, w_n) be a nonnegative vector of committee members' voting weights (e.g. votes or shares). By 2^N we denote the power set of N (set of all subsets of N). By voting coalition we mean an element S $\in 2^N$, the subset of committee members voting uniformly (YES or NO), and $w(S) = \sum_{i \in S} w_i$ denotes the voting weight of coalition S. The voting rule is defined by quota q satisfying $0 < q \leq w(N)$, where q represents the minimal total weight necessary to approve the proposal. Triple [N, q, **w**] we call a *simple quota weighted committee*. The voting coalition S in committee [N, q, **w**] is called a winning one if $w(S) \geq q$ and a losing one in the opposite case. The winning voting coalition S is called critical if there exists at least one member k $\in$ S such that w(S\k) < q (we say that k is critical in S). The winning voting coalition S is called minimal if any of its members is critical in S.

A priori voting power analysis seeks an answer to the following question: Given a simple quota weighted committee $[N, q, \mathbf{w}]$, what is an influence of its members over the outcome of voting? The absolute voting power of a member i is defined as a probability $\Pi_i[N, q, \mathbf{w}]$ that i will be decisive in the sense that such a situation appears in which she would be able to decide the outcome of voting by her vote (Nurmi [6] and Turnovec [13]), and a relative voting power as $\pi_i[N, q, \mathbf{w}] = \frac{\Pi_i[N, q, \mathbf{w}]}{\sum_{k \in N} \Pi_k[N, q, \mathbf{w}]}$.

Two basic concepts of decisiveness are used: swing position and pivotal position. The s*wing position* is an ability of an individual voter to change the outcome of voting by a unilateral switch from YES to NO (if member j is critical with respect to a coalition S, we say that he has a swing in S). The *pivotal position is* such a position of an individual voter in a permutation of voters expressing a ranking of attitudes of members to the voted issue (from the most preferable to the least preferable) and the corresponding order of forming of the winning coalition, in which her vote YES means a YES outcome of voting and her vote NO means a NO outcome of voting (we say that j is pivotal in the permutation considered).

Let us denote by i the member of the simple quota weighted committee [N, q, **w**], W(N, q, **w**) the set of all winning coalitions and by W_i(N, q, **w**) the set of all winning coalitions with i, C(N, q, **w**) the set of all critical winning coalitions, and by C_i(N, q, **w**) the set of all critical winning coalitions i has the swing in, by P(N) the set of all permutations of N and P_i(N, q, **w**) the set of all permutations i is pivotal in. By card(S) we denote the cardinality of S, card(∅) = 0.

Assuming many voting acts and all coalitions equally likely, it makes sense to evaluate the a priori voting power of each member of the committee by the probability to have a swing, measured by the absolute Penrose-Banzhaf (PB) power index (Penrose [8], Banzhaf [1]) $\Pi_i^{PB}(N, q, \mathbf{w}) = \frac{card(C_i)}{2^{n-1}}$, where (card($C_i$) is the number of all winning coalitions the member i has the swing in and 2^{n-1} is the number of all possible coalitions with i). To compare the relative power of different committee members, the relative form of the PB power index $\pi_i^{PB}(N, q, \mathbf{w}) = \frac{card(C_i)}{\sum_{k \in N} card(C_k)}$ is used.

While the absolute PB is based on a well-established probability model (see e.g. Owen [5]), its normalization (relative PB index) destroys this probabilistic interpretation, the relative PB index simply answers the question of what is the voter i's share in all possible swings.

Assuming many voting acts and all possible preference orderings equally likely, it makes sense to evaluate an a priori voting power of each committee member by the probability of being in pivotal situation, measured by the

Shaply-Shubik power index (Shapley and Shubik [9]): $\Pi_i^{SS}(N,q,\mathbf{w}) = \dfrac{card(P_i)}{n!}$, where card($P_i$) is the number of all permutations in which the committee member i is pivotal, and n! is the number of all possible permutations of committee members. Since $\sum_{i\in N} card(P_i) = n!$ it holds that $\pi_i^{SS}(N,q,\mathbf{w}) = \dfrac{card(P_i)}{\sum_{k\in N} card(P_k)} = \dfrac{card(P_i)}{n!}$, i.e. the absolute and relative form of the SS-power index is the same.

It can be easily seen that for any $\alpha > 0$ and any power index based on swings, pivots or MWC positions it holds that $\Pi_i[N,\alpha q,\alpha\mathbf{w}] = \Pi_i[N,q,\mathbf{w}]$. Therefore, without the loss of generality, we shall assume throughout the text that $\sum_{i\in N} w_i = 1$ and $0 < q \le 1$, using only relative weights and relative quotas in the analysis.

3 Quota intervals of stable power and the fair quota

Let us formally define a few concepts we shall use later in this paper:

Definition 1. Let $\mathbf{w} = (w_1, w_2, ..., w_n)$ be a fair distribution of voting weights (with whatever principle is used to justify it) in a simple weighted committee [N, **w**], $\boldsymbol{\pi}$ is a relative power index, ($\boldsymbol{\pi}$[N, q, **w**] is a vector valued function of q), and d is a distance function, then the voting rule q_1 is said to be *at least as fair* as voting rule q_2 with respect to the selected $\boldsymbol{\pi}$ and distance d if d(**w**, $\boldsymbol{\pi}$(N, q_1, **w**)) $\le$ d(**w**, $\boldsymbol{\pi}$(N, q_2, **w**)).

Intuitively, given **w**, the voting rule q_1 is preferred to the voting rule q_2 if q_1 generates a distribution of power closer to the distribution of weights than q_2.

Definition 2. The voting rule q* that minimizes a distance d between $\boldsymbol{\pi}$[N, q, **w**] and **w** is called a fair voting rule (*fair quota*) for the power index π with respect to the distance d.

Proposition 1. Let [N, q, **w**] be a simple weighted quota committee and C_{is} be the set of critical winning coalitions of the size s in which i has a swing, then $card(P_i) = \sum_{s\in N} card(C_{is})(s-1)!(n-s)!$ is the number of permutations with the pivotal position of i in [N, q, **w**].[2]

From Proposition 1 it follows that the number of pivotal positions corresponds to the number and structure of swings. If in two different committees sets of swing coalitions are identical, then the sets of pivotal positions are also the same.

Proposition 2. *Let [N, q_1, **w**] and [N, q_2, **w**], $q_1 \neq q_2$, be two simple quota-weighted committees such that W(N, q_1, **w**) = W(N, q_2, **w**), then $C_i(N,q_1,\mathbf{w}) = C_i(N,q_2,\mathbf{w})$ and $P_i(N,q_1,\mathbf{w}) = P_i(N,q_2,\mathbf{w})$ for all i $\in$ N.*

From Proposition 2 it follows that in two different committees with the same set of members, the same weights and the same sets of winning coalitions, the PB-power indices and SS-power indices are the same in both committees, independently of quotas.

Proposition 3. *Let [N, q, **w**] be a simple quota weighted committee with a quota q, $\mu^+(q) = \min_{S\in W[N,q,w]}(w(S)-q)$ and $\mu^-(q) = \min_{S\in 2^N\setminus W(N,q,w)}(q-w(S))$. Then for any particular quota q we have W(N, q, **w**) =W(N, γ, **w**) for all $\gamma \in (q-\mu^-(q), q+\mu^+(q)]$.*

From Propositions 2 and 3 it follows that swing and pivot based power indices are the same for all quotas $\gamma \in (q-\mu^-(q), q+\mu^+(q)]$. Therefore the interval of quotas $(q-\mu^-(q), q+\mu^+(q)]$ we call an *interval of stable power* for quota q. Quota $\gamma^* \in (q-\mu^-(q), q+\mu^+(q)]$ is called the marginal quota for q if $\mu^+(\gamma^*) = 0$.

Now let us define a partition of the power set 2^N into equal weight classes $\Omega_0, \Omega_1, \ldots, \Omega_r$ (such that the weight of different coalitions from the same class is the same and the weights of different coalitions from different classes are different). For the completeness set $w(\varnothing) = 0$. Consider the weight-increasing ordering of equal weight classes $\Omega^{(0)}, \Omega(1), \ldots, \Omega^{(r)}$ such that for any t < k and $S \in \Omega^{(t)}$, $R \in \Omega^{(k)}$ it holds that w(S) < w(R). Denote q_t = w(S) for any $S \in \Omega^{(t)}$, t = 1, 2, …, r.

[2] The proofs of all propositions in this section see in Turnovec [10].

Proposition 4. *Let $\Omega^{(0)}$, $\Omega^{(1)}$, ..., $\Omega^{(r)}$ be the weight-increasing ordering of the equal weight partition of 2^N. Set $q_t = w(S)$ for any $S \in \Omega^{(t)}$, $t = 0, 1, 2, ..., r$. Then there is a finite number $r \leq 2^n\text{-}1$ of marginal quotas q_t and corresponding intervals of stable power $(q_{t-1}, q_t]$ such that $W(N, q_t, \mathbf{w}) \subset W(N, q_{t-1}, \mathbf{w})$.*

From Proposition 4 it follows that there exist at most r distinct voting situations generating r vectors of power indices.

Proposition 5. *Let [N, q, w] be a simple quota weighted committee and $(q_{t-1}, q_t]$ is the interval of stable power for quota q. Then $card(C_i(N,q,\mathbf{w})) = card(C_i(N,\gamma,\mathbf{w}))$ and $card(P_i(N,q,\mathbf{w})) = card(P_i(N,\gamma,\mathbf{w}))$ for any $\gamma = 1 - q_t + \varepsilon$, where $\varepsilon \in (0, q_t - q_{t-1}]$ and for all $i \in N$.*

While in [N, q, **w**] the quota q means the total weight necessary to pass a proposal (and therefore we can call it a *winning quota*), the *blocking quota* means the total weight necessary to block a proposal. If q is a winning quota and $(q_{t-1}, q_t]$ is a quota interval of stable power for q, then any voting quota $1 - q_{t-1} + \varepsilon$ (where $0 < \varepsilon \leq q_t - q_{t-1}$), is a blocking quota. From Proposition 5 it follows that the blocking power of the committee members, measured by swing and pivot-based power indices, is equal to their voting power. It is easy to show that voting power and blocking power might not be the same for power indices based on membership in minimal winning coalitions (HP and DP power indices). Let r be the number of marginal quotas, then from Proposition 4 it follows that for power indices based on swings and pivots the number of majority power indices does not exceed *int (r/2)+1*.

Proposition 6. *Let [N, q, w] be a simple quota-weighted committee, d be a distance function and $\pi_i(N, q_t, w)$ be relative power indices for marginal quotas q_t, and q_t* be the majority marginal quota minimizing the distance $d\left[\boldsymbol{\pi}(N,q_j,\mathbf{w}), w_i\right]$ where $j = 1, 2, ..., r$, r is the number of intervals of stable power such that q_j are marginal majority quotas, then the fair quota for a particular power index used with respect to distance d is any $\gamma \in (q_{t^*-1}, q_{t^*}]$ from the quota interval of stable power for q_t*.*

From Proposition 6 it follows that the voting rule based on quota q_t* minimizes selected distance between the vector of relative voting weights and the corresponding vector of relative voting power. The problem of fair quota has an exact solution via the finite number of majority marginal quotas

4 Fair quota in the Lower House of the Czech Parliament

The Lower House of the parliament has 200 seats. Members of the Lower House are elected in 14 electoral districts from party lists by proportional system with 5% threshold. Seats are allocated to the political parties that obtained not less than 5% of total valid votes roughly proportionally to fractions of obtained votes (votes for parties not achieving the required threshold are redistributed among the successful parties roughly proportionally to the shares of obtained votes). Five political parties qualified to the Lower House: left centre Czech Social Democratic Party (Česká strana sociálně demokratická, ČSSD), right centre Civic Democratic Party (Občanská demokratická strana, ODS), right TOP09 (Tradice, Odpovědnost, Prosperita – Traditions, Responsibility, Prosperity 2009), left Communist Party of Bohemia and Moravia (Komunistická strana Čech a Moravy, KSČM) and supposedly centre (but not very clearly located on left-right political dimension) Public Issues (Věci veřejné, VV).

In Table 1 we provide results of the 2010 Czech parliamentary election (by relative voting weights we mean fractions of seats of each political party, by relative electoral support fractions of votes for political parties that qualified to the Lower House, counted from votes that were considered in allocation of seats). Three parties, ODS, TOP09 and VV, formed right-centre government coalition with 118 seats in the Lower House.

	Seats	Votes in % of valid votes	Relative voting weight	Relative electoral support
ČSSD	56	22,08	0,28	0,273098
ODS	53	20,22	0,265	0,250093
TOP09	41	16,7	0,205	0,206555
KSČM	26	11,27	0,13	0,139394
VV	24	10,58	0,12	0,13086
$\sum$	200	80,85	1	1

Table 1 Results of 2010 election to the Lower House of the Czech Parliament
Source: http://www.volby.cz/pls/ps2010/ps?xjazyk=CZ

We assume that all Lower House members of the same party are voting together and all of them are participating in each voting act. Two voting rules are used: simple majority (more than 100 votes) and qualified majority (at least 120 votes). There exist 16 possible winning coalitions for simple majority voting (12 of them are winning coalitions for qualified majority), 16 marginal majority quotas and 16 majority quota intervals of stable power (see Table 2).

Parties of possible winning coalitions	Absolute marginal majority quota	Relative marginal majority quota	Intervals of stable power
ODS+KSČM+VV	103	0.515	(0.485, 0.515]
CSSD+KSČM+VV	106	0.53	(0.515, 0.53]
ČSSD+ODS	109	0.545	(0.53, 0.545]
ODS+TOP09+VV	118	0.59	(0.545, 0.59]
ODS+TOP09+KSČM	120	0.6	(0.59, 0.6]
ČSSD+TOP09+VV	121	0.605	(0.6, 0.605]
ČSSD+TOP09+KSČM	123	0.615	(0.605, 0.615]
ČSSD+ODS+VV	133	0.665	(0.615, 0.665]
ČSSD+ODS+KSCM	135	0.675	(0.665, 0.675]
ODS+TOP09+KSČM+VV	144	0.72	(0.675, 0.72]
ČSSD+TOP09+KSČM+VV	147	0.735	(0.72, 0.735]
ČSSD+ODS+TOP09	150	0.75	(0.735, 0.75]
ČSSD+ODS+KSČM+VV	159	0.795	(0.75, 0.795]
CSSD+ODS+TOP09+VV	174	0.87	(0.795, 0.87]
ČSSD+ODS+TOP09+KSČM	176	0.88	(0.87, 0.88]
ČSSD+ODS+TOP09+KSČM+VV	200	1	(0.88, 1]

Table 2 Possible winning coalitions in the Lower House of the Czech Parliament (own calculations)

For analysis of fair voting rule we selected Shapley-Shubik power index and Euclidean distance function. In Table 3 we provide Shapley-Shubik power indices (distribution of relative voting power) for all of marginal majority quotas. For any quota from each of intervals of stable power is Shapley-Shubik relative power identical with relative power in corresponding marginal majority quota. Entries in the row "distance" give Euclidean distance between vector of relative voting weights and relative power for each quota interval of stable power.

Party	Seats	Relative voting weight	SS power for q=0.515	SS power for q=0.53	SS power for q=0.545	SS power for q=0.59	SS power for q=0.6	SS power for q=0.605	SS power for q=0.615	SS power for q=0.665
ČSSD	56	0,28	0,3	0,35	0,3167	0,2667	0,3167	0,3667	0,3333	0,3
ODS	53	0,265	0,3	0,2667	0,3167	0,2667	0,2333	0,2	0,25	0,3
TOP09	41	0,205	0,1333	0,1833	0,2333	0,2667	0,2333	0,2	0,1667	0,1333
KSČM	26	0,13	0,1333	0,1	0,0667	0,1	0,15	0,1167	0,1667	0,1333
VV	24	0,12	0,1333	0,1	0,0667	0,1	0,0667	0,1167	0,0833	0,1333
Σ	200	1	0,9999	1	1,0001	1,0001	1	1,0001	1	0,9999
distance			0,08339	0,08169	0,10802	0,07271	0,07996	0,01195	0,08501	0,08339
Party	Seats	Relative voting weight	SS power for q=0.675	SS power for q=0.72	SS power for q=0.735	SS power for q=0.75	SS power for q=0.795	SS power for q=0.87	SS power for q=0.88	SS power for q=1
ČSSD	56	0,28	0,2667	0,2333	0,4333	0,3833	0,35	0,3	0,25	0,2
ODS	53	0,265	0,2667	0,2333	0,1833	0,3833	0,35	0,3	0,25	0,2
TOP09	41	0,205	0,1833	0,2333	0,1833	0,1333	0,1	0,3	0,25	0,2
KSČM	26	0,13	0,1833	0,15	0,1	0,05	0,1	0,05	0,25	0,2
VV	24	0,12	0,1	0,15	0,1	0,05	0,1	0,05	0	0,2
Σ	200	1	1	0,9999	0,9999	0,9999	1	1	1	1
distance			0,06238	0,07271	0,17874	0,20275	0,15637	0,14816	0,17875	0,14816

Table 3 Shapley-Shubik power of political parties for majority marginal quotas (own calculations)

The fair relative majority quota in our case is q = 0.675 (with respect to Euclidean distance between relative voting weights and relative voting power 0.06238), or any quota from interval of stable power (0.665, 0.675]. It means that minimal number of votes to approve a proposal is 135 (in contrast to 101 votes required by simple majority and 120 votes required by qualified majority). Voting rule defined by this quota minimizes Euclidean distance between relative voting weights and relative voting power (measured by Shapley-Shubik power index) and approximately equalizes the voting power (influence) of the members of the Lower House independently on their political affiliation.

The measure of fairness follows the same logic as measures of deviation from proportionality used in political science, evaluating the difference between results of an election and the composition of an elected body - e.g. Gallagher [3] based on the Euclidean distance, or Loosemore-Hanby [4] based on the absolute values distance. Using in our particular case the absolute values distance we shall get the same fair quota.

Acknowledgements

This research was supported by the Grant Agency of the Czech Republic, project No. 402/09/1066 “Political economy of voting behavior, rational voters’ theory and models of strategic voting”. The author is grateful to an anonymous referee for helpful comments and suggestions on an earlier draft of the paper.

References

[1] Banzhaf, J.F.: Weighted Voting Doesn't Work: A Mathematical Analysis, *Rutgers Law Review*, **19** (1965), 317-343.

[2] Felsenthal, D. S., and M. Machover : *The Measurement of Voting Power, Theory and Practice*, Edward Elgar, Cheltenham, 1998.

[3] Gallagher, M.: Proportionality, Disproportionality and Electoral Systems, *Electoral Studies*, **10** (1991), 33-51.

[4] Loosemore, J., and V.J.Hanby: The Theoretical Limits of Maximum Distortion: Some Analytical Expressions for Electoral Systems, *British Journal of Political Science*, **1** (1971), 467-477.

[5] Nurmi, H.: The Problem of the Right Distribution of Voting Power. *I*n: *Power, Voting and Voting Power* (Holler M.J., ed.), Physica Verlag, Würzburg, Wien, 1982, 203-212.

[6] Nurmi, H.: On Power Indices and Minimal Winning Coalitions, *Control and Cybernetics*, **26**, (1977), 609-611.

[7] Owen, G.: Multilinear Extensions of Games. *Management Science*, **18** (1972), 64-79.

[8] Penrose, L.S.: The Elementary Statistics of Majority Voting. *Journal of the Royal Statistical Society*, **109** (1946), 53-57.

[9] Shapley, L.S., and M. Shubik: A Method for Evaluation the Distribution of Power in a Committee System. *American Political Science Review*, **48** (1954), 787-792.

[10] Słomczyński, W., and K. Życzkowski: From a Toy Model to Double Square Root System. *Homo Oeconomicus*, **24** (2007), 381-400.

[11] Słomczyński, W., and K. Życzkowski: Penrose Voting System and Optimal Quota. *Acta Physica Polonica* **B37** (2006), 3133-3143.

[12] Turnovec F.: Fair Voting Rules in Committees, Strict Proportional Power and Optimal Quota. *Homo Oeconomicus*, **27** (2011), to appear.

[13] Turnovec, F.: Power, Power Indices and Intuition. *Control and Cybernetics*, **26** (1997), 613-615

Examining the relationship between economic performance and unemployment: the case of Visegrad countries

Michal Tvrdoň[1], Tomáš Verner[2]

Abstract. In this paper, macroeconomic links between economic performance and labour market performance are analysed. The paper is focused on the relationship between the change in GDP and the change in the number of the unemployed labour force. Economic growth is considered to be a pathway to decrease the level of unemployment. On contrary, the economic crisis can be a potential source of growth in unemployment which could have a serious social impact. Even though the economy seems to be growing again, it may be a while before the unemployment rate begins to decline, and it may even continue rising for some time after the resumption of sustained economic growth. The study is aimed to test long-term relationship between GDP growth and unemployment. Quarterly OECD data were used for the analysis of both GDP and labour market performance. The Johansen test was applied on 1995-2010 data to examine cointegration between the number of the unemployed labour force and real gross domestic product. On the basis of the unit root test, we found that in all countries, both variables are stationary except for their first differences. Cointegration was not proved in the case of all Visegrad Group countries.

Keywords: cointegration, economic performance, unemployment, Visegrad countries

JEL Classification: J64
AMS Classification: 62P20

1 Introduction

After 1989, all transition economies had to cope with a new phenomenon – unemployment. The situation is described best in [9]: "the most important change that occurred in the labor market after 1989 was the change of conditions, when from the long-term persisting "lack" of labor force we almost immediately meet a phenomenon, by that time non-registered, latently existing, however extraneous into the vocabulary of a centrally planned economy, with unemployment."

We can also assume that the transitive economies started to be confronted with a number of resulting tasks: i) how to take care of those who lose a job and simultaneously ii) do not create inadequate fiscal costs and iii) to minimize a reluctance to work related with this protection [8]. The labor markets in the Visegrad group countries (V-4) recorded several alike and several different experience in the period of transition. Whilst in Slovakia, Poland and Hungary had been recorded a sharp increase of the unemployment rate (up to double figure numbers), in the Czech Republic the low level of the unemployment rate persisted in comparison with other V-4 countries in the early 90s. The high unemployment rate in these countries is explained generally as a result of (i) macroeconomic policies or main external shocks; (ii) problems associated with an economic structure in these countries or (iii) unfinished transition from a centrally planned economy to the market one [6].

Unemployment fluctuates consistent with phases of the business cycle in the most countries. In other words, the number of unemployed during a recession is increasing, while the number of unemployed is decreasing in the phase of economic growth. However, while the increase in unemployment occurs very quickly after the first signs of an economic recession, the dynamics of reducing the number of unemployed is significantly lower. In addition, Abraham and Shimer [1] mention that at the most of proceeded economic cycles it was proved rather strong correlation between the unemployment level and average duration of unemployment.

We set the year 1995 as a starting year for the time series because the V-4 economies already passed through fundamental economic reforms and previous data had been strongly influenced by the systematic transition of the economy.

[1] Silesian University/School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, tvrdon@opf.slu.cz.
[2] Silesian University/School of Business Administration in Karviná, Department of Economics, Univerzitní nám. 1934/3, Karviná 733 40, verner@opf.slu.cz.

The aim of this paper is to estimate, based on econometric approach, relationship between economic performance and unemployment in Visegrad Group countries, especially in the context of long-term time period. For this purpose the paper is divided into several parts. The first part is devoted to description of analytical tools. The second part contains empirical results of cointegration analysis and the final section summarizes all the key findings

2 Econometric methodology

Cointegration is an econometric technique for testing the relationship between non-stationary time series variables. This technique is often used because of many macroeconomic time series are not stationary in their levels. If two or more series each have a unit root, that is I(1), but a linear combination of them is stationary, I(0), then the series are said to be cointegrated. Thus cointegration analysis is an extension of the simple correlation based analysis. The objective of this article is to analyze the effects of economic growth on unemployment in the Visegrád group countries.

The problem then is to find a way to work with two possibly non-stationary series in a fashion that allows us to capture both short run and long run effects. In more technical parlance, cointegration is the link between integrated processes and steady state equilibrium. If the time series are stationary in first differences than it is fulfilled requirements for the implementation of cointegration. Although we have two non-stationary time series, their common cointegration long-term shift in time moves towards some equilibrium.

We used Phillips-Perron (PP) test as the unit root test. We used this approach to test the null hypothesis that a time series in integrated of order 1. The PP method estimates the non-augmented DF test equation, and modifies the *t*-ratio of the α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic. The PP test makes a non-parametric correction to the t-test statistic.

The PP test is based on the statistic [7]:

$$\bar{t}_\alpha = t_\alpha \left(\frac{\gamma_0}{f_0}\right)^{1/2} - \frac{T(f_0-\gamma_0)(se(\hat{\alpha}))}{2 f_0^{1/2} s} \tag{1}$$

where $\hat{\alpha}$ is the estimate, and t_α the *t*-ratio of α, $se(\hat{\alpha})$ is coefficient standard error, and *s* is the standard error of the test regression. In addition, γ_0 is a consistent estimate of the error variance. The remaining term, f_0, is an estimator of the residual spectrum at frequency zero.

Cointegration test is based on the determination of r cointegration relations in the VAR model. Cointegration is confirmed, if true, that r > 0. For testing purposes, we used Johansen cointegration test.

It is necessary to obtain an indication of optimal time delay before the implementation of Johansen cointegration test, which was in our case according to the Schwarz information criterion (SC) applied to estimate the VAR model of differentiation two periods. The SC criterion is defined as [4]:

$$SC = n^{k/n} \frac{\sum \hat{u}^2}{n} = n^{k/n} \frac{RSS}{n} \tag{2}$$

where RSS means the residual sum of squares, k/n is the penalty factor.

We used two tests for determining the number of cointegration vectors: (i) the Trace test; and (ii) the Maximal Eigenvalue test.

The Trace test for the number of cointegrating vectors determines the number of cointegrating equations r:

$$(\mathrm{r}) = -N * \sum_{i=r+1}^{m} \ln(1-\hat{\lambda}_i) \tag{3}$$

We tested hypothesis by the Trace test for H_0 r=0 (there are no cointegration vectors) and H_1 r≤1 (there is cointegration equation). We did not reject the H_0 hypothesis if the Trace statistics is no larger than the 5% critical value).

Another test is the Maximal Eigenvalue test:

$$(\mathrm{r}, \mathrm{r}+1) = -N * \ln(1-\hat{\lambda}_{r+1}) \tag{4}$$

We tested hypothesis by the Maximal Eigenvalue test for the same H_0 and H_1 like the Trace test. We also do not reject the H_0 hypothesis if the Maximal Eigenvalue statistics is no larger than the 5 % critical value.

As the following step the Error Correction Term (ECT) should be estimated and test for stationarity. The result of the PP test for the unit root should confirm integration in order I(0). It means that the Y_t and X_t are cointegrated or that the regression of equation in no longer spurious, and we can also find the linear combination that connects Y_t and X_t in the long run [3] or we can say that there is a long-run equilibrium relationship between X and Y:

$$\hat{\mu}_t = Y_t - \hat{\beta}_1 + \hat{\beta}_2 X_t \qquad (5)$$

Finally the Error Correction Model (ECM) should be estimated (if Y_t and X_t are cointagrated). Thus, we can express the relation between Y_t and X_t with an ECM specification as [3]:

$$\Delta Y_t = \alpha_0 - \alpha_1 (Y_{t-1} - \beta_1 X_{t-1}) + \beta_0 X_t + \varepsilon_t \qquad (6)$$

where current changes in *Y* are a function of current changes in *X* (the first difference of *X*) and the degree to which the two series are outside of their equilibrium in the previous time period. Specifically, β_0 captures any immediate effect that *X* has on *Y* , described as a contemporaneous effect or short-term effect. The coefficient, β_1 reflects the equilibrium effect of *X* on *Y*. It is the causal effect that occurs over future time periods, often referred to as the long-term effect that *X* has on *Y* . Finally, the long-term effect occurs at a rate dictated by the value of α_1.

Arlt [1] shows that it should be stressed that the importance of the ECM lies in the fact that it allows us to combine statistical and econometric approach to modelling economic time series.

3 Empirical results

The data used in this study are real Gross Domestic Product (GDP) and a number of the unemployed labor force (U). We used quarterly OECD data between the first quarter of 1995 and the fourth quarter of 2010. Both GDP and U data were seasonally adjusted (see Figure 1 and 2). The first step was to transform these variables into logs (ln GDP and ln U) and then establish that every variable is integrated of order one or I(1), i.e. stationary at the first difference.

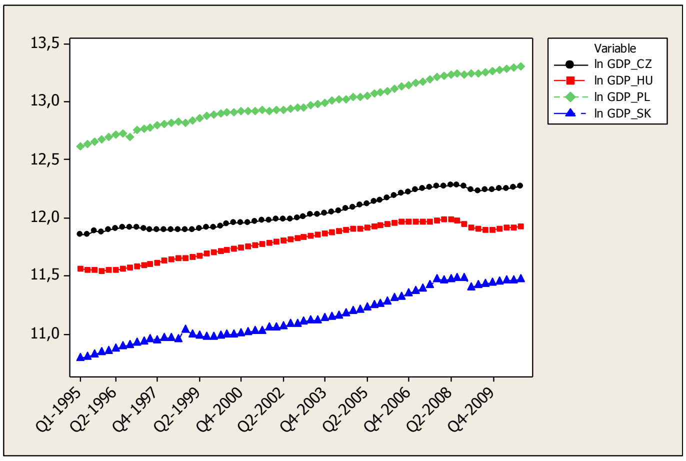


Figure 1 Real GDP time series (ln)

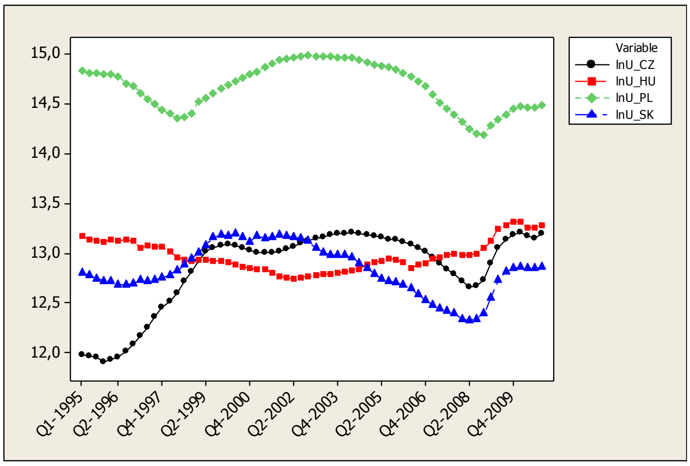


Figure 2 Unemployed labour force time series (ln)

Table 1 shows that the statistics for all the variables (GDP, U) in all Visegrad Group countries are greater than the critical values at 5% levels from Phillip- Perron test (PP test). Thus, the results show that the null unit roots cannot be rejected, suggesting that all the variables are non-stationary in their level forms. The results of the first differenced variables show that the PP test statistics for all the variables are less than the critical values at 5% levels. That results show that all the variables are stationary after differencing once, suggesting that all the variables are integrated of order I(1).

		Level form		**First difference**	
	variable	Test statistic	5% level	Test statistic	5% level
Czech Rep.	GDP	3.2363	-1.9461	-3.8735	-1.9462
	U	1.3893	-1.9461	-2.3711	-1.9462
Hungary	GDP	2.7965	-1.9461	-2.4569	-1.9462
	U	0.2805	-1.9461	-4.2242	-1.9462
Poland	GDP	8.1390	-1.9461	-5.3094	-1.9462
	U	-0.5298	-1.9461	-2.5556	-1.9462
Slovakia	GDP	4.5517	-1.9461	-7.2717	-1.9462
	U	0.0503	-1.9461	-3.0802	-1.9462

Table 1 Unit root test (PP)

Next step was to obtain an indication of optimal time delay before the implementation of Johansen cointegration test, which was in the case of the Visegrad Group countries according to the Schwarz information criterion (SC) applied to estimate the VAR model of differentiation two periods.

	$\mathbf{H_0}$	**Eigenvalue**	**Trace Statistic**	**0.05 Critical Value**	**Probability**
Czech Rep.	r = 0	0.2097	23.0510	20.2618	0.0201
	r = 1	0.1329	8.6969	9.1645	0.0613
Hungary	r = 0	0.2557	27.0436	20.2618	0.0050
	r = 1	0.1375	9.0256	9.1645	0.0531
Poland	r = 0	0.3807	35.6451	20.2618	0.0002
	r = 1	0.0999	6.4183	9.1645	0.1608
Slovakia	r = 0	0.4611	41.7871	20.2618	0.0000
	r = 1	0.0647	4.0774	9.1645	0.4007

Table 2 Unrestricted Cointegration Rank Test (Trace)

In determining the cointegration it is important to define a number of r (rank) cointegration relations (relations) in the VAR model. Cointegration is confirmed in the case of $r > 0$. The widespread test for detecting cointegration is Johansen cointegration test (see Johansen [5]). Two statistics were applied for the number of relations – Trace statistic and Maximum eigenvalue statistic. Results of the unrestricted cointegration rank test can be seen in Table 2 and Table 3.

	H_0	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Probability
Czech Rep.	r = 0	0.2097	14.3540	15.8921	0.0858
	r = 1	0.1329	8.6969	9.1645	0.0613
Hungary	r = 0	0.2557	18.0180	15.8921	0.0229
	r = 1	0.1375	9.0256	9.1645	0.0531
Poland	r = 0	0.3807	29.2268	15.8921	0.0002
	r = 1	0.0999	6.4183	9.1645	0.1608
Slovakia	r = 0	0.4611	37.7097	15.8921	0.0000
	r = 1	0.0647	4.0774	9.1645	0.4007

Table 3 Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

In determining the cointegration relationship the null hypothesis and the alternative hypothesis were established: H_0: r = 0 and H_1: r> 0. H_0 was in the case of the Trace Statistic for a five percent significance level rejected in all cases, which means that the number of cointegration relations is not equal to zero. A similar result we get from the other test criterion (Max-eigenvalue statistics) with the exception of the Czech Republic, where the cointegration relationship was not confirmed.

The next step was to determine the null hypothesis: H_0: r = 1 and the alternative: $H_{1:}$ r > 1, as shown in the tables, the null hypothesis is accepted and the number of cointegration relations, r is equal to 1 (with the exception of the Czech Republic by the Max-eigenvalue, where r = 0, in which case we followed the Trace statistic, where the relationship was confirmed). Both tests indicate one cointegrating equation at the 5 % significance level in the case of other Visegrad Group countries.

Based on performed tests of cointegration, we obtained the following cointegration equations for the period 1995-2010:

$$LnU\_CZ = 3.438918*LnGDP\_CZ – 30.80355 \tag{7}$$

$$LnU\_HU = 0.872977*LnGDP\_HU + 2.968167 \tag{8}$$

$$LnU\_PL = -0.988913*LnGDP\_PL + 28.80581 \tag{9}$$

$$LnU\_SK = -0.718846*LnGDP\_SK + 21.44964 \tag{10}$$

If we look at equations (7) and (8) we can see that the relationship between two variables is not consistent with economic theory. We expected negative relationship as seen in equation (9) and (10) which means, in other words, that the number of unemployed will decrease with the economic growth and vice versa.

Time series of residues obtained from equations above were not stationary at levels which means that they are not integrated of order I(0). This was confirmed by the PP test, so the time series are not cointegrated. Although cointegration between economic performance and the number of unemployed was not confirmed, we can interpret this relationship as a regression based on the VAR model in the case of equations (9) and (10) (we refused equations (7) and (8) because they are in inconsistent with elementary patterns of economics). The Equation (9) shows that the GDP growth by 1% reduces the number of unemployed also almost by 1% in the case Poland. In the case of Slovakia equation (10) suggests that GDP growth by 1% reduces the number of unemployed by 0.72%.

4 Conclusion

The aim this paper was to perform an empirical analysis of long-term relationship between economic growth and a number of unemployed. In other words, we try to find possible cointegration between changes of real gross domestic product and a number of unemployed. On the basis of the unit root test, we found that in all countries, both variables are integrated of order I(1). This result allowed us to continue and after establishing a period of lag (two periods), we implemented the Johansen cointegration test. This test showed that although one cointegration relationship, ECT, however, showed that residues are not stationary at I (0). Cointegration relationship thus has not been demonstrated and we can interpret relationship as regression based on VAR model only.

Acknowledgements

The research behind this paper was supported by the Czech Science Foundation within the project GAČR 402/09/P142 “Institutional labour market framework in the context of economic convergence and adopting single currency (application on Visegrád group)”.

References

[1] Abraham, K. G., and Shimer, R.: *Changes in Unemployment Duration and Labor Force Attachment*. NBER Working Paper No. 8513. NBER, Cambridge (MA), 2001.
[2] Arlt, J.: Kointegrace v jednorovnicových modelech. *Politická ekonomie* **45** (1997), 733-746.
[3] Asteriou D., and Hall S.G.: *Applied Econometrics*. Palgrave MacMillan, New York, 2007.
[4] Gujarati, D., and Porter, D. C.: *Basic Econometrics*, 5th ed. McGraw-Hill, Boston, 2009.
[5] Johansen, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models, Oxford University Press.
[6] Münich, D., and Švejnar, J.: *Unemployment in East and West Europe*. IZA Discussion Paper No. 2798, 2000.
[7] Phillips, P., and Perron, P.: Testing for a Unit Root in Time Series Regression. *Biometrika* 75 (1988), 335-346.
[8] Raju, D., Vodopivec, M., and Wörgötter, A.: Unemployment Benefit Systems in Central and Eastern Europe: A Review of the 1990s. [Online]. *World Bank Social Protection Discussion Paper No. 0310.* Available: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2003/07/26/000094946_03071512065264/Rendered/PDF/multi0page.pdf
[9] Spěváček, V. et al.: *Transition of the Czech Economy: political, economic and social aspects*. Linde, Praha, 2002.

Labour market frictions in global economic crisis: RBC model of Czech economy

Stanislav Tvrz[1], Osvald Vašíček[2]

Abstract. This paper deals with analysis of labour market frictions of Czech economy based on RBC model framework. Labour market frictions are modelled as aggregate costs that firms and workers face in the process of forming a match. The question of interest is how do the labour market frictions behave during business cycle and what are the implications for development of unemployment rate. The role of labour frictions during recent global financial and subsequent economic crisis is investigated. Possible structural changes of economy caused by global crisis are considered as well. Explaining potential of labour market frictions concept is evaluated by comparison of marginal likelihood of different specifications of the model. To capture the changing structure of economy a recursive estimation method is employed. Sensitivity and identification analysis of estimated parameters is performed as well. Computations are carried out using Dynare toolbox for Matlab. Results of the estimations confirm the importance of labour market frictions and suggest their procyclical nature.

Keywords: RBC, Bayesian methods, recursive estimation, labour market frictions, economic crisis.

JEL classification: J64
AMS classification: 91B64

1 Introduction

As reported in Galí [2], the structural VAR model with long-run restrictions shows contractionary effect of productivity shock on employment. López-Salido and Michelacci in [6] find evidence for negative response of employment to technology shocks in a structural VAR model with job flows. Classical RBC model in accordance with Kydland and Prescott [5] on the other hand produces positive response of employment to the technology shocks. Since other shocks play substantially less important role in explaining aggregate fluctuations, it is the persistent technology shock that is the true driving force of RBC models. The discrepancy between empirical evidence and theoretical framework of RBC models thus calls for explanation.

Basic Hansen's RBC model assumes perfectly competitive markets with no frictions. In reality we often observe different wage bargaining power of employers and employees or long-term labour contracts, which causes suboptimal allocation of labour. This fact suggests that there are additional costs that firms and workers face in forming a match, i.e. hiring costs. In theory, it could be the presence of real labour market frictions, that cause the failure of RBC model.

Hiring costs represent the costs incurred at all stages of recruitment. They include costs of advertising the job vacancy and screening of potential candidates as well as the costs of training. The hiring costs can behave either pro-cyclically or counter-cyclically. Low opportunity costs during recessions imply more restructuring of the workforce and more hiring. This means that the firms spend more resources on screening and training and hiring costs rise. High availability of unemployed workers during recessions, on the other hand, lowers the cost of advertising and hence lowers the hiring costs. Domination of the former effect would cause the hiring costs to behave counter-cyclically and vice versa.

The reaction of employment to productivity shock then depends on the response of hiring costs. With

[1]Masaryk University, Brno, e-mail: 207212@mail.muni.cz
[2]Masaryk University, Department of Economics, Brno, e-mail: Osvald.Vasicek@econ.muni.cz

pro-cyclical hiring costs the productivity shock would increase the marginal product of labour as well as the hiring costs. Domination of the latter effect would allow for negative reaction of employment to a productivity shock.

2 The Model

The simple framework of RBC model can be easily enhanced with labour search and matching frictions, as shown by Blanchard and Galí in [1]. In this paper we will be working with RBC model framework proposed by Mandelman and Zanetti in [7]. This model specification allows the hiring costs to react directly to productivity and leaves the data to decide on the type of their cyclical behavior. Our original research consists of adding the production factor of capital to the model in accordance with [3].

The representative household maximizes the expected value of its utility function, which takes following form:

$$E_0 \sum_{t=0}^{\infty} \beta^t \varepsilon_t^b \left(\ln C_t - \varepsilon_t^l \frac{N_t^{1+\phi}}{1+\phi} \right). \tag{1}$$

C_t and N_t stand for consumption in period t and amount of supplied labour in period t respectively. β is the discount factor. ϕ stands for inverse Frisch labour supply elasticity. ε_t^b stands for preference shock in discount rate and ε_t^l represents shock in labour supply.

Cobb-Douglas production function with production factors of labour L_t and capital K_t determines the aggregate product Y_t. ε_t^a stands for the level of technology or the total factor productivity. Parameter α represents elasticity of production function with respect to capital.

$$Y_t = \varepsilon_t^a K_t^{\alpha} N_t^{1-\alpha} \tag{2}$$

A fraction of previous period jobs given by separation rate ϱ is destroyed at the beginning of each period. Some jobs are created and filled by newly hired workers H_t.

$$N_t = (1-\varrho)N_{t-1} + H_t \tag{3}$$

Pool of unemployed and available for hiring U_t is complementary to the number of employed after the job destruction.

$$U_t = 1 - (1-\varrho)N_{t-1} \tag{4}$$

Job-finding rate x_t is defined as fraction of new hires and number of unemployed.

$$x_t = \frac{H_t}{U_t} \tag{5}$$

Unitary hiring costs G_t are defined as a product of job-finding rate, level of technology and a scale parameter B. γ represents elasticity of hiring costs with respect to productivity and ω stands for hiring costs elasticity with respect to job-finding rate.

$$G_t = A_t^{\gamma} B x_t^{\omega} \tag{6}$$

Unemployment rate u_t is defined as complement of the workforce to one.

$$u_t = 1 - N_t \tag{7}$$

Aggregate product is utilized as consumption and investment and part of the product is also spent to cover aggregate hiring costs $G_t H_t$.

$$Y_t = C_t + I_t + G_t H_t \tag{8}$$

Capital is accumulated in accordance with following law of motion. Parameter δ represents depreciation rate. ε^i stands for exogenous shock in marginal efficiency of investment.

$$K_{t+1} = (1-\delta)K_t + I_t \varepsilon_t^i \tag{9}$$

All the exogenous shocks are modelled as AR(1) processes. ρ_x is the autoregressive coefficient of the process and $\eta_{x,t+1} \sim N(0, \sigma_x)$ is shock innovation.

$$\varepsilon_{t+1}^x = (\varepsilon_t^x)^{\rho_x} \exp(\eta_{x,t+1}) \tag{10}$$

3 Methodology

The model is estimated with the use of Bayesian methods based on Random walk Metropolis-Hastings algorithm and Kalman filter. Computations are carried out using Dynare toolbox for Matlab. Dynare version 4.1.3. and Matlab version R2010a are used. The estimates are based on two Metropolis-Hastings runs of 1 200 000 draws each. First 50 % of each sequence is discarded as burn-in before the computation of the estimates.

A method of recursive estimation is used in the paper. It is a tool which enables us to see a changing structure of the model as the time flows. Using the same techniques as above we estimate a sequence of models with changing historical data input. Let T be the total number of observations. We choose number of observations $t < T$ and estimate models with t, $t+1,\ldots$, T first observations of historical data. We can then use the output of this recursive estimation to see the effects of new information in added observations on estimates of structural parameters or impulse response functions. In the paper we used recursive estimation with $t = 18$. In each estimation we used two Metropolis-Hastings runs with 100 000 samples each. Again 50 % samples of each sequence were discarded as burn-in.

4 Data

For estimation we used quarterly data of the Czech economy in period between the first quarter of 1999 and last quarter of 2010 from publicly available Czech National Bank database. Four observed variables are used – aggregate product, real investment, job-finding rate and unemployment rate. Time series of unemployment rate correspond with the methodology of Ministry of Labour and Social Affairs of the Czech republic. Original time series were seasonally adjusted.

To be able to use the data in the RBC model we needed time series representing gaps of each variable with respect to its stedy-state. Since the steady state values are in general unobservable we approximated the steady state values of aggregate product and real investment by global linear trend and the steady state of unemployment rate and job-finding rate by constant level of variable's mean. Using this approximated steady state values we calculated logarithmic deviations[1] from steady state for each variable. These deviations are presented in figure 1.

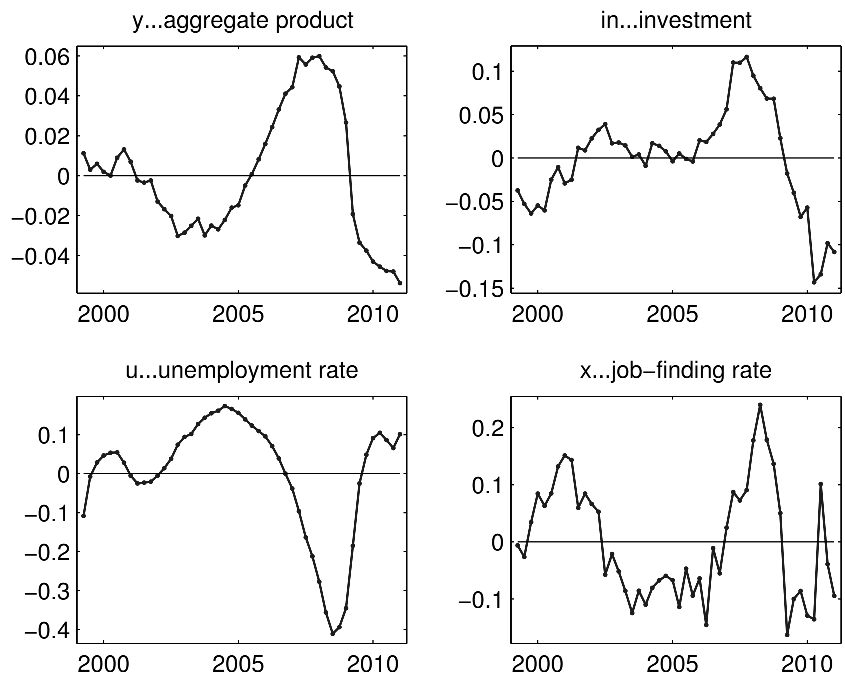


Figure 1: Input data

[1]Logarithmic deviation from steady state is $\ln\left(\frac{X_t}{\overline{X}_t}\right)$, where X_t is observed value of given variable at time t and $\overline{X}_t$ is approximated steady state value of given variable at time t.

5 Estimation results

Several structural parameters of the model were calibrated. The job destruction rate ϱ can be expressed from steady-state relations of the model. Using the equation $\varrho = ux/[(1-u)\cdot(1-x)]$ and the means of observed variables we calibrated this parameter to the value of 0.008. Scale factor of labour frictions B was set to the value of 10. Quarterly discount parameter is calibrated at the value of 0.995. Depreciation rate of 2.5 % per quarter is assumed. Capital share of national income is set at the value of 0.33.

I present the results of parameters estimation in the table 1. Parameter of labour friction costs elasticity with respect to productivity γ is statistically insignificant, which is probably caused by the choice of strictly uninformative prior and lack of strong evidence in the data.

The estimate of inverse labour supply elasticity parameter ϕ imply labour supply elasticity around 3. HPD or highest posterior density interval is rather wide implying labour supply elasticity between 2.14 and 5.2.

Parameter of labour friction costs elasticity with respect to job-finding rate ω is estimated at the value of 0.7741 which suggests weaker than proportional reaction of hiring costs to changes in job-finding rate. HPD interval also suggests that values of ω higher than 1 are rather unlikely.

Estimates of persistences and volatilities of exogenous shocks are in line with our expectations. They show high persistence of productivity shock and rather high volatility of labour supply shock.

	Prior			Posterior	
Parameter	Distribution	Mean	St.dev.	Mean	95 % HPD interval
γ	normal	0.000	3.00	-3.7547	(-8.6651, 1.1558)
ϕ	gamma	0.350	0.10	0.3253	(0.1924 , 0.4663)
ω	gamma	1.000	0.30	0.7741	(0.4771 , 1.0538)
ρ_a	beta	0.700	0.15	0.9747	(0.9504 , 0.9977)
ρ_b	beta	0.500	0.20	0.5330	(0.2094 , 0.8961)
ρ_l	beta	0.500	0.20	0.7434	(0.6063 , 0.8864)
ρ_i	beta	0.500	0.20	0.8360	(0.7531 , 0.9524)
σ_a	inv.gamma	0.007	∞	0.0078	(0.0065 , 0.0091)
σ_b	inv.gamma	0.002	∞	0.0022	(0.0004 , 0.0042)
σ_l	inv.gamma	0.010	∞	0.1092	(0.0570 , 0.1612)
σ_i	inv.gamma	0.005	∞	0.0092	(0.0071 , 0.0124)

Table 1: Results of estimation

According to [4] we analysed possible identification issues caused by multicollinearity or general misspecification. The analysis showed that the parameter of elasticity of labour supply ϕ is rather difficult to estimate. Also parameters of preference shock and investment efficiency shock are hard to estimate due to the structure of the model. The rest of the parameters should be estimated correctly.

5.1 Recursive estimation

Recursive estimates of most interesting model parameters are presented in figure 2. Black line represents the mean of given estimate while gray lines represent the bounds of 95 % HPD interval. The period of global economic crisis is highlighted by light grey shading.[2]

The estimates of parameter γ are at first estimated around -1. At the beginning of 2009 the estimates fall to values around -4. The economic crisis, therefore, caused a growth of labour frictions sensitivity to the changes in productivity. The estimates of parameter γ are however statistically insignificant.

Parameter ω, representing labour frictions elasticity with respect to job-finding rate, is initially estimated at the values around 1.05 with slight downward trend. In the end of 2008 the estimates sharply

[2]In this paper we consider the period between the beginning of 2008 and the second quarter of 2010 to represent the period of global economic recession.

decline to the values around 0.75. Due to the economic crisis the labour market frictions became significantly less sensitive to the changes in job-finding rate.

Rather stable estimates of productivity shock volatility, σ_a, gradually grew since 2005 and reached the values around 0.6 % in 2008. In the end of 2008 and the beginning of 2009 the estimates sharply rose above 0.8 %. Similarly, the estimates of labour supply shock volatility, σ_l, grew significantly during the economic contraction of 2008–2009. From values around 5 % the values rose above 10 %. The period of economic instability in Czech republic beginning in 2008 is, therefore, well identified by the recursive estimation.

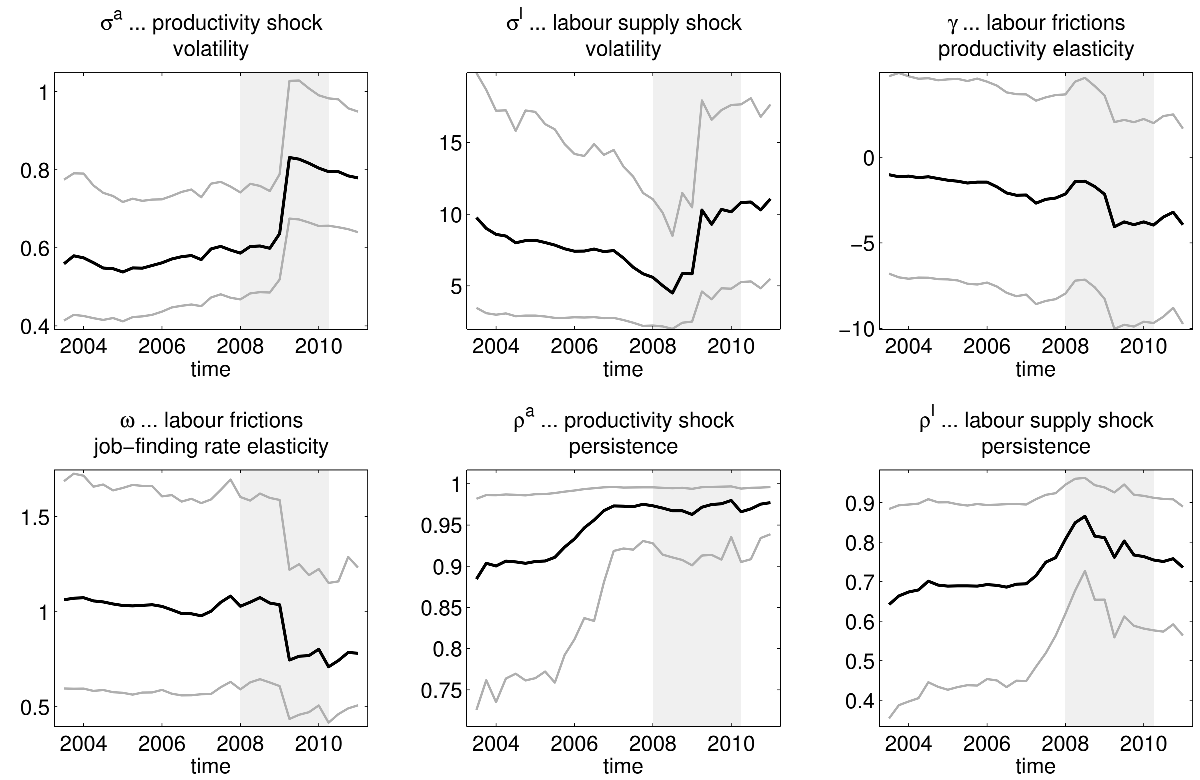


Figure 2: Recursive estimates

5.2 Cyclical behavior

Hiring costs, which represent the labour market frictions, are unobserved endogenous variable and we can, therefore, use the smoothed values of this variable to analyse its cyclical behavior. We can calculate the correlations between leading or lagging hiring costs and aggregate product in order to quantify the quality of given relation. Obtained values are shown in table 2. We find the highest value of correlation with output of 0.63 for lead of two quarters. Corresponding values of correlation with output are presented for the observed variables of job-finding rate and unemployment rate. Job-finding rate was pro-cyclical and leading by one quarter ahead of the output. The unemployment rate was counter-cyclical and lagging behind the aggregate product by three quarters.

	Lag of given variable								
Variables	−4	−3	−2	−1	0	1	2	3	4
Y	0.43	0.74	0.87	0.96	1.00	0.96	0.87	0.74	0.58
G	0.43	0.56	0.63	0.62	0.60	0.48	0.34	0.20	0.07
x	0.51	0.64	0.72	0.72	0.71	0.58	0.43	0.27	0.13
u	0.14	-0.08	-0.30	-0.50	-0.69	-0.82	-0.88	-0.89	-0.86

Table 2: Correlation with output

6 Conclusion

In this paper we tried to explain the behavior of labour market frictions modelled as aggregate hiring costs using RBC model framework. Results of the estimation show that the direct link between unitary hiring costs and productivity represented by parameter γ is statistically insignificant. This, however, doesn't mean that the hiring costs would show no distinct cyclical behavior. On the contrary, smoothed unitary hiring costs are clearly pro-cyclical. This is accomplished through the definition of hiring costs, that includes a link to job-finding rate. The job-finding rate is pro-cyclical, as can be seen in observed data, and the parameter ω, which represents hiring costs elasticity with respect to this variable, is estimated to be near 0.77. According to the model, the nature of unitary hiring costs is therefore sufficiently defined with the use of job-finding rate.

Model comparison based on logged marginal density approximations supports previous statement, because it showed no persuasive evidence in favour of model specification with direct link of labour market frictions to productivity through parameter γ.

Pro-cyclical labour market frictions act as a stabilising element of labour market. During recessions the effect of high availability of unemployed workers lowers the costs of advertising and dominates the opposing effect of increased screening and training costs. Unitary hiring costs, therefore, decrease, which ceteris paribus motivates a growth of employment. Since the unemployment rate is obviously a counter-cyclical variable, the effect of lower hiring costs during recessions is dominated by the effect of general decline of productivity.

Recursive estimation of the model captured the increase of shock volatilities during the period of global crisis. Persistence of exogenous shock increased as well. Some structural parameters changed in this period too. Decline of parameter γ suggests slightly growing importance of productivity in explaining unitary hiring costs. Decline of parameter ω estimate shows, on the other hand, slightly declining influence of job-finding rate.

For further research, extension of the capital market should be considered. In this paper the capital market framework is rather simplified. Adjustment costs of capital stock could improve the explaining potential of the model. Addition of human capital to the model could be considered as well.

Acknowledgements

This work was supported by funding of specific research at ESF MU, project MUNI/A/0943/2009 and by MŠMT project Research centers 1M0524.

References

[1] Blanchard, O. - Galí, J.: *A new Keynesian model with unemployment.* Research Series 200610-4. National Bank of Belgium. 2006.

[2] Galí, J.: *Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?* American Economic Review, vol 89(1) (March), 1999, pp. 249-271.

[3] Greenwood, J. - Hercowitz, Z. - Huffman, G.: *Investment, capacity utilization, and the real business cycle.* The American Economic Review. American Economic Association. 1988.

[4] Iskrev, N. - Ratto, M.: *Analysing identification issues in DSGE models.* Bank of Portugal, Lisbon, Portugal. European Comission, Joint Research Centre, Ispra, Italy. 2010.

[5] Kydland, F. E. - Prescott E. C.: *Time to build and aggregate fluctuations.* Econometrica, vol. 50(6) (November), 1982, pp. 1345-1370.

[6] López-Salido, D. - Michelacci, C.: *Technology shocks and job flows.* Review of Economic Studies, vol. 74(4) (October), 2007, pp. 1195-1227.

[7] Mandelman, F. - Zanetti, F.: *Technical Handbook No. 1 - Estimating general equilibrium models: an application with labour market frictions.* Centre for Central Banking Studies, Bank of England. 2008. ISSN 1756-7297. Available from: http://www.bankofengland.co.uk/education/ccbs/technical_handbooks/pdf/techbook1.pdf

Analyzing relations of insured accident and selected risk factors with interactions

Jiří Valecký[1]

Abstract. The paper is devoted to the analyzing impact of selected risk factors on insured accident and forecasting probability of loss occurrence. On the basis of these factors and measurement of their relation to the insured accident, it is possible to determine the probability of insured accident for each policyholder and also to differentiate the paid insurance premiums in compliance with risk factors. The insurance premium is proposed in this way that it corresponds to undertaken risk, i.e. occurrence probability. In the purpose of identifying and modelling this relation, the logistic regression model is primarily employed, but potential interactions among factors during the proposing and estimating are often neglected. It results in the inaccurate parameter estimates and also in misinterpreting and incorrect assessing relation to the outcome generally. In the paper, the model with and without interaction is proposed and estimated and then, in compliance with both models, the relation of selected risk factor and the insured accident are interpreted and assessed. The differences are explained including the recommendation concerning the rules for differentiation of paid insurance premiums.

Keywords: logit model, interaction term, motor hull insurance, insurance premium, insured accident, incidence of loss.

JEL Classification: C31, C58, G22

1 Introduction

Insurance premium (regardless to various insured amount) is determined in accordance with undertaken risk. In other words, the premium is determined in accordance with risk behavior of given policyholder and the more risky client should pay higher premium. This trend has been already observed for many years already. It is very common that the insurers set the premium in motor hull insurance in compliance with the volume of an engine. Nowadays, some of them also distinguish the size of district where the client lives and some insurers respect even client´s age. Determining the relation of particular risk factors which are characteristic for the policyholder and the incidence of loss and also their precise quantification are crucial for differentiating premiums.

For this purpose, one can employ several models regarding the issue which is needed to be examined. Count and Duration models can be applied there, see [5], [7], [9], or [11], but for analyzing the effect of risk factor on loss occurrence, the logistic regression is primarily used. The estimation of these models is associated with many problems which lead to misestimation and misinterpretation. For example, neglecting of confounders results in incorrect evaluation of impact of risk factor under interest within various groups. For example, the probability of car accident can be higher for 40 year old man than for woman at the same age. Also assuming linearity of logit can lead to excessive distortion of impacts of risk factors on the loss occurrence. In addition, existing interactions between risk factors are the most difficult problem in spite of the fact that it is generally easy to identify them. In complex full models, there are many interactions between risk factors and moreover some of them are related to more covariates than only two, for example [2]. Some researchers neglect them, for example [4]; [8] and others, but if the variables are interacted, one independent variable modifies the relation between the outcome and the other independent variable. It results of course in inaccurate or incorrect assessing of the impact of risk factors on the outcome.

Thus, when we suspect the presence of interaction between some variables, we regress the outcome on both variable and also on their interaction term. In linear models, the marginal effect of the change in both interacted variable equals to the marginal effect of changing the interaction terms. On the other hand, according to [1], many econometricians and statisticians misinterpret the coefficients of the interaction term in non-linear models because the marginal effect of changing the interaction terms is dependent on interacted variables. Moreover, also assessing statistical significance via traditional z-test cannot be conducted for the same reason, i.e. the dependency on the level of interacted variables. On the contrary, Greene [3] argues that evaluating marginal effect is not necessary informative and is also difficult to interpret it in terms of the relationships among the variables.

[1] VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská tř. 33, 701 21, jiri.valecky@vsb.cz.

He also argues that graphical devices can be much more informative than the test statistics suggested by Ai and Norton [1].

The aim of this paper is to determine interactions between risk factor under interest and assess their relation to the probability of incidence of loss. We focus on the interpreting effect of risk factors on the outcome correctly including the statistical verification of interaction and we also propose the rules and recommendation for determining insurance premiums in compliance with these risk factors. The paper is organized as follows. The general logistic regression and interaction terms are described in Section 2 and Section 3 consist of the evaluating effects of selected risk factors on the insured accident. Section 4 concludes the paper.

2 Identifying effect of risk factor on insured accident

Next, for the purpose of quantifying relation of risk factors to the insured accident, we focus on the logistic regression and interaction term.

2.1 Logistic regression model

Consider a binary variable Y_i characterizing the occurrence of insured accident for given policyholder, thus

$$Y_i = \begin{cases} 1 & \text{if insured accident occurs,} \\ 0 & \text{otherwise, for } i = 1, \dots, N, \end{cases} \tag{1}$$

where N is the number of policyholders and each client is characterized by the vector of individual K risk factors $\mathbf{x}_i = \left(1, x_{1i}, x_{2i}, \dots, x_{Ki}\right)$.

The probability of an insured accident for a given policyholder, $P_i = P\left(Y_i = 1\right)$, is possible to express, on the basis of its characteristic vector $\mathbf{x}_i$, as a function $F\left(\boldsymbol{\beta};\mathbf{x}_i\right)$ which is monotonically increasing $F'\left(\boldsymbol{\beta};\mathbf{x}_i\right) \geq 0$ and has a domain of definition $\left(-\infty, +\infty\right)$ and a range $\left(0,1\right)$. Thus, it holds that $F\left(-\infty\right) = 0$ and $F\left(+\infty\right) = 1$, the probability function can be written in the form of

$$P_i = F\left(\boldsymbol{\beta};\mathbf{x}_i\right), \tag{2}$$

where $\boldsymbol{\beta}$ is vector of parameters $\left(\beta_0, \beta_1, \dots, \beta_K\right)$.

These properties are satisfied by the cumulative distribution function of the logistic distribution

$$P_i = P\left(Y_i = 1\right) = F\left(\boldsymbol{\beta};\mathbf{x}_i\right) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_i}}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_i}} = \frac{1}{1 + e^{-\left(\boldsymbol{\beta}'\mathbf{x}_i\right)}} \tag{3}$$

which is also a function of the probability that insured accident occurs. The probability that the accident does not occur can be written as

$$1 - P_i = P\left(Y_i = 0\right) = 1 - F\left(\boldsymbol{\beta};\mathbf{x}_i\right) = \frac{1}{1 + e^{\boldsymbol{\beta}'\mathbf{x}_i}}. \tag{4}$$

The ratio of probabilities (3) and (4) is referred to as odds and it takes the form of

$$\frac{\pi}{1-\pi} = \frac{P\left(Y_i = 1\right)}{P\left(Y_i = 0\right)} = e^{\boldsymbol{\beta}'\mathbf{x}_i} \tag{5}$$

and the logarithm of (5) is termed logit or log-odds, thus

$$\ln\left[\frac{\pi}{1-\pi}\right] = \boldsymbol{\beta}'\mathbf{x}_i = g\left(\mathbf{x}_i\right). \tag{6}$$

The odds ratio for any covariates x_j (considering binary, categorical and continuous variable) is determined by using the following equation

$$OR\left(a,b\right) = \frac{\pi\left(x_j = a\right) \big/ \left[1 - \pi\left(x_j = a\right)\right]}{\pi\left(x_j = b\right) \big/ \left[1 - \pi\left(x_j = b\right)\right]} = e^{\beta_j\left(a-b\right)}. \tag{7}$$

2.2 Interaction terms

Dependent variable under interest is the probability that $Y_i = 1$. In other words, we are interested in probability that the insured accident occurs. In literature, there are two definitions and interpretations of the interaction effect. Firstly, Hosmer and Lemeshow [6], Vittingoff et al. [10] and others define the interaction effect as changing the slope parameter of one variable depending on the level of the other variable. On the other hand, Ai and Norton [1] interpret the interaction effect as marginal effect of interacted x_1 and x_2 via cross-derivative of probability function $F(\boldsymbol{\beta};\mathbf{x}_i)$.

Thus, let us assume the interaction between two continuous variables x_1 and x_2 and let variable x_1 be a risk factor under interest. The changing slope parameter of x_2 depending on the level of x_1 is

$$\beta_{x_1 \times x_2} = \beta_1 x_1 + \beta_{12} x_1 x_2 \text{ for all } x_2 . \tag{8}$$

It is obvious that the coefficient $\beta_{x_1 \times x_2}$ varies dependently on the level of x_2.

The other interpretation of the interaction is proposed by Ai and Norton [1] who define the interaction effect on the estimated cross-derivative of the terms in the interaction, not on the coefficient of the interaction term, thus

$$\frac{\partial^2 F(\boldsymbol{\beta};\mathbf{x}_i)}{\partial x_1 \partial x_2} = \beta_{12} F'(\boldsymbol{\beta};\mathbf{x}_i) + (\beta_1 + \beta_{12} x_2)(\beta_2 + \beta_{12} x_1) F''(\boldsymbol{\beta};\mathbf{x}_i). \tag{9}$$

We can see from Equations (8) and (9) that the interaction term can be non-zero in both cases even if $\beta_{12} = 0$. Except for that, the interaction effect can have different signs for different values of covariates, thus the sign of β_{12} is ambiguous. Moreover, the statistical significance of the interaction effect cannot be tested via z-test on the coefficient β_{12} because of changing standard error. Thus, the standard error of the estimated changing parameter $\beta_{x_1 \times x_2}$ is computed as

$$SE_{x_1 \times x_2} = \sqrt{x_1^{\,2} \operatorname{var}(\beta_1) + (x_1 x_2)^2 \operatorname{var}(\beta_{12}) + 2 x_1^{\,2} x_2 \operatorname{cov}(\beta_1;\beta_{12})} \tag{10}$$

and the standard error of estimated interaction effect on $F(\boldsymbol{\beta};\mathbf{x}_i)$ according to [1] is derived via Delta method as follows

$$SE_{x_1 \times x_2} = \sqrt{\frac{\partial^2}{\partial \beta^T} \left\{ \frac{\partial^2 F(\boldsymbol{\beta};\mathbf{x}_i)}{\partial x_1 \partial x_2} \right\} \hat{\Sigma}_\beta \frac{\partial^2}{\partial \beta} \left\{ \frac{\partial^2 F(\boldsymbol{\beta};\mathbf{x}_i)}{\partial x_1 \partial x_2} \right\}}, \tag{11}$$

where $\hat{\Sigma}_\beta$ is covariance matrix of estimated β and *T* denotes transposition. From the computation of standard errors (10) and (11) follow that z-statistics and the significance are dependent on the level of x_2.

3 Evaluating effect of selected risk factors

In this section, we focus on the volume (*Volume*) and performance (*kW*) of a car engine in relation to the probability of insured accident and we verify their statistical dependencies. We also examine the probabilities for various levels of volume and performance of a car engine. For this purpose, we work with data encompassing characteristics of policyholders during the year 2009 (82 132 of insurance policies) within a motor hull insurance.

Firstly, we simply analyze the effect of these risk factors on the outcome separately and then we evaluate the effect when both risk factors are included in the model. We reveal that there is an interaction between them. To highlight this issue, we evaluate the effect of both variables on the probability with and without an interaction. Further, we make some remarks about model building and statistical significance and we conduct the statistical verification of interaction effect significance. Finally, we compare the results.

3.1 Model estimation

Interaction terms are commonly identified in this way that firstly the univariate logistic regression is run and then other variables are appended into the model. If the estimated coefficients of the previous regression change substantially, the interaction is possible. Thus, we performe univariate logistic regression. The next table records obtained results.

Variable	Coef.	LL	Chi2	Prob
Volume	.0002218	-3.90e+04	337.801	0.0000
kW	.0005199	-3.91e+04	39.086	0.0000

Table 1 Univariate logistic regression

As we can see, both variables are statistically significant. When we run regression on both variables simultaneously, the coefficients changed, see the next table.

Claim	Coef.	Std. Err.	z	P>z	95% Conf. Interval	
Volume	.0002208	.0000127	17.40	0.000	.0001959	.0002457
kW	.0000961	.0000885	1.09	0.278	-.0000774	.0002695
cons	-1.882157	.0231398	-81.34	0.000	-1.92751	-1.836804

Model significance: LR-statistics 343.41 (0.000)

Table 2 Logistic regression without interaction

We can see that the change of *Volume* coefficient is not significant (not statistically evaluated), but the change of coefficient of *kW* is substantially different. Moreover, the *kW* is now statistically insignificant (according to the p-value), therefore we can assume that there is an interaction between them. Thus, we should create the interaction term as the product of both variables and we run the logistic regression again, see results in the next table.

Claim	Coef.	Std. Err.	z	P>z	95% Conf. Interval	
Volume	.0004235	.0000279	15.18	0.000	.0003688	.0004782
kW	.001516	.0001927	7.87	0.000	-.0011384	.0018936
kvol	-9.74e-07	1.18e-07	-8.23	0.000	-1.21e-06	-7.42e-07
cons	-2.192789	.0440378	-49.79	0.000	-2.279102	-2.106476

Model significance: LR-statistics 416.53 (0.000)

Table 3 Logistic regression with interaction

Both interaction term and also the *kW* are significant. As it was mentioned above, we cannot evaluate the statistical significance on the basis of traditional z-test because the z-statistics is dependent on the level of other variables. Thus, we evaluate the significance on the basis of Delta method only, see the next figure.

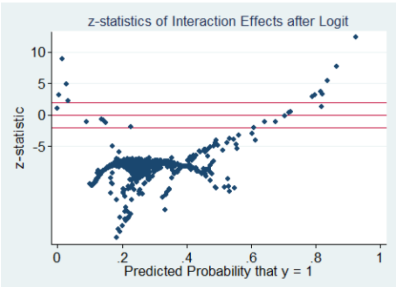
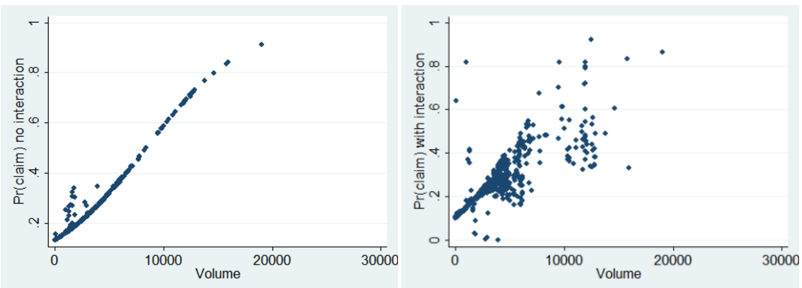


Figure 1 z-statistics of interaction term

The points between lines indicate the statistical insignificance. As we can see, the interaction is statistically significant for the most of policyholders, especially for such policyholders whose probability of accident is not higher than 0.5.

3.2 Results comparison and discussion

Now, we evaluate the relation of engine´s volume and performance to the occurrence of probability. In the next figure we can see the relation of probability and volume/performance in both models (without interaction on the left and with interaction on the right).



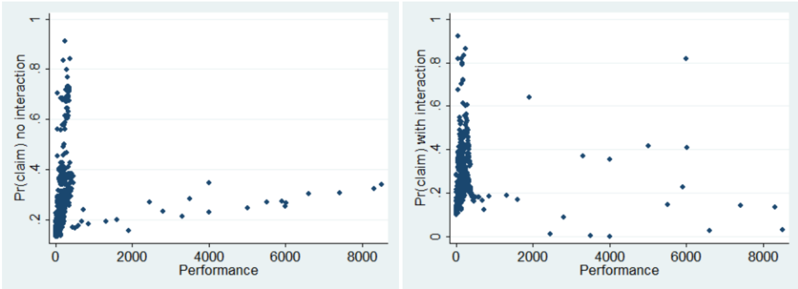


Figure 2 Probability and volume of engine

According to the figures above, we can say that policyholders should pay higher insurance premium with increasing engine`s volume but this conclusion is not so apparent and unique in the case of the model with interaction. It appears to be more chaotic and the model without interaction appears to represent the relation better. But this conclusion is incorrect due to the neglecting interaction. This is explained later. In the case of the second risk factor, it is obvious that any conclusion about the relation of performance to probability is almost impossible to formulate in both models. Let´s focus on some cars with similar or same performance of engine. We have chosen the most frequent values and in the next figure we depict the relation between volume and probability given by the performance.

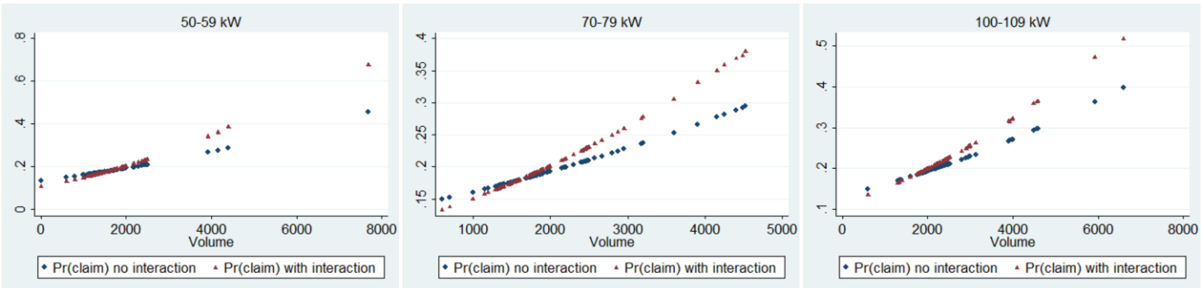


Figure 3 Probability and volume of engine for given performance

The probability is increasing in both models, but when the interaction is considered, the growth of probability is higher than in the model without interaction. We can also see and we prove it later that the growth decelerates with increasing performance. Further, we depict the relation between performance and probability given by the volume.

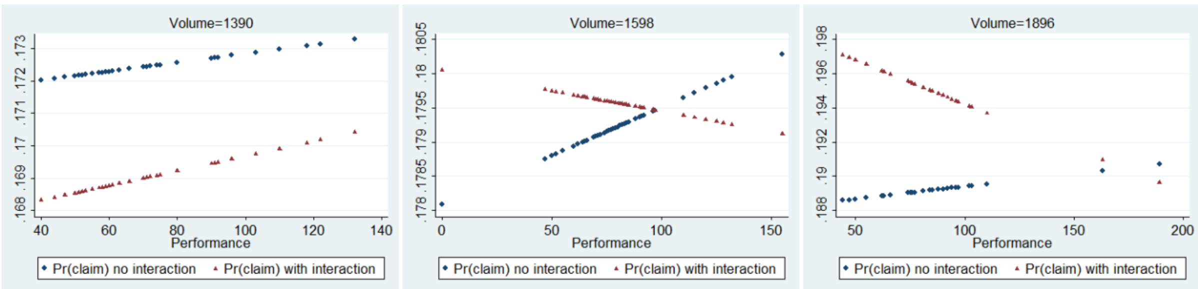


Figure 4 Probability and performance of engine for given volume

Here, the change of slopes is more apparent. The probability given by volume of 1390 cm^3 is increasing with growth of performance in case of both models. On the contrary, the slope has changed from positive sign to negative if volume equals 1598 cm^3 and is lower in case of 1896 cm^3. Because the change of slope parameter is not so apparent in the case of the first analyzed relation, we show both risk factors against the linear predictor of logit.

The slope of volume in logit is decreasing when performance increases. Moreover, the slope of all selected performance classes is higher than the slope of all cars when the interaction is not considered. Also the positive slope of performance in logit is decreasing but in addition, the slope changes from positive sign to the negative around the volume 1598 cm^3 and decreases more depending on increasing volume.

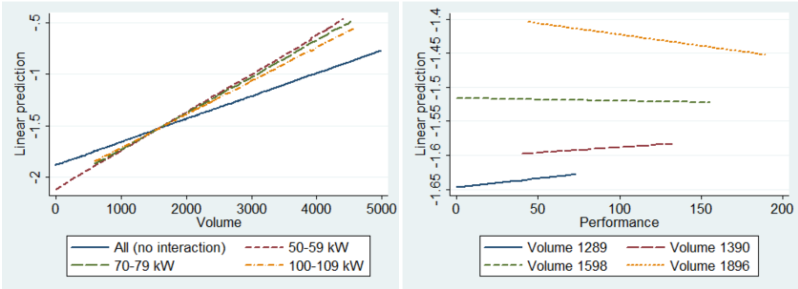


Figure 5 Comparison of changing probability given volume/performance

4 Conclusion

The paper was devoted to analyzing impact of selected risk factors on insured accident and forecasting probability of loss occurrence within motor hull insurance. We identified the interaction between performance and volume of car engine and we revealed the substantial change of relation between risk factors and occurrence probability due to the interaction. We also analyzed these relations and compared the results with the case when no interaction was considered.

Thus, based on the obtained results, we revealed statistically significant interaction which changes effect of risk factors on the outcome. The slope of performance and volume were changed by interaction effect. The slope parameter of volume was decreasing when performance decreased. Also the slope of performance changed due to the level of volume and moreover, the slope parameter changed the sign around volume at 1598 cm^3. It means that the accident is more likely for given volume above 1598 cm^3 for lower performance and less likely for volume below this level.

On the basis of these analyzed relations, we can say that the policyholders should pay higher premium with increasing volume generally but the raise of the insurance price should decelerate when the performance increases. On the other hand, the policyholder should pay more with increasing performance only when the volume is not higher 1598 cm^3, otherwise he should pay less.

Acknowledgements

This paper was solved within the project SP2011/38 Estimating and forecasting individual risk in financial institutions.

References

[1] Ai, C., and Norton, E.: Interaction terms in logit and probit models. *Economics Letters* **80** (2003), 123-129.

[2] Cornelißen, T. and Sonderhof, K.: Partial effects in probit and logit models with a triple dummy-variable interaction term. *The Stata Journal* **4** (2009), 571-583

[3] Greene, W.: Testing hypotheses about interaction terms in nonlinear models. *Economics Letters* **107** (2010), 291-296.

[4] Gurny, P. and Gurny, M.: Comparison of the credit scoring models on PD estimation of US banks. In: *28th International Conference on Mathematical Methods in Economics 2010*. University of West Bohemia, Ceske Budejovice, 2010, 188-194.

[5] Hausman, J., Hall, B. and Griliches, Z.: Econometric models for count data with an application to the parents-R&D relationship. *Econometrica* **52** (1984), 909-938.

[6] Hosmer, D.W. and Lemeshow, S.: *Applied Logistic Regression*. John Wiley & Sons Inc., New York, 2000.

[7] Lawless, J.: Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15** (1987), 209-225.

[8] Valeckova, J.: An Analysis of the Influences of Demographic Factors on Customer Satisfaction with Using Logistic Regression. *ECON11* **1**, 2011, 123-131.

[9] Van den Berg, G.: Duration models: specification, identification and multiple durations. In: *Handbook of Econometrics* (Heckman, J. and Leamer, E. eds.). North-Holland, Amsterdam, 2001, 3381-3480.

[10] Vittingoff, E., Glidden, D.V., Shiboski, S.C. and McCulloch, C.E.: *Regression Methods in Biostatistics: Linear, Logistic, Survival and Repeated Measures Models*. Springer Verlag, New York.

[11] Winkelmann, R.: Duration dependence and dispersion in count data models. *Journal of Business and Economic Statistics* **13** (1995), 467-474.

Analysis of occurrence of extremes in a time series with a trend

Petr Volf[1]

Abstract. We consider a random series of values and are interested in the analysis and modeling the occurrence of extremes. There exist several possible approaches. One of them is the analysis of sequence of block maxima. As we assume that the series has a trend, we first select a proper regression model for the block maxima development. From it, a Markov chain of the sequence of extremes is derived. As the transition probabilities of the chain are not tractable analytically, we use the Monte Carlo generation of the chain behavior. Then, from the sample representing the series of block maxima development we obtain a representation of corresponding predictive distribution. Finally, we shall apply such a method to real data.

Keywords: extreme, regression model, random walk, prediction.

JEL classification: C41, J64
AMS classification: 62N02, 62P25

1 Introduction

Let us consider a sequence of maximal values in a series of random variables $X_1, X_2, \ldots$. New maximum (record) is established when $X_{t+1} > max\{X_1, ..., X_t\}$. The case of i.i.d. (independent, identically distributed) random variables X_t has been analyzed by many authors (cf. Embrechts et al, 1997). It has been shown that the probability of new record at time t is proportional to $1/t$, and that the sequence new record values $R_1 < R_2 < \ldots$ behaves as a random point process with intensity $h_x(r)$ equal to the intensity of distribution of random variables X_t.

However, quite frequently the assumption of i.i.d. variables is not adequate. Especially in cases when the series of variables is dependent and changes along certain trend. This contradiction between reality and the i.i.d. scheme led to the construction of models describing the sequence of extremes (i. e. values, increments, times) with the aid of convenient functional models for intensity, regression, or time-series (though we shall speak mostly on maxima, the same concerns the minimal values).

In the paper we are interested in the following questions: In Section 2 we deal with the problem of trend model fitted to the data. We use the approach of block maxima, i.e. the data are reduced to maximal values over certain periods. It has an advantage that the dependence in a series is reduced, on the other hand some local extremes are lost. The statistical tools for the model fit diagnostics are recalled, too. Then the method of the prediction of further series development is proposed. It is based on the simulation, as the numerical computations are hardly tractable in this case. The simulation of future trajectories enables us to estimate prediction bands, i.e. curves which are crossed with given probability (see "Peaks over threshold", POT method, for instance in Beirlant et al, 2004). Notice that when we deal with block maxima, the POT approach yields a piece-vise constant threshold curve, which is also one of possible choices of thresholds of POT analysis for the whole data series.

In Section 3 we recall the attempts to model the occurrence of extremes as a random point process. We actually link up both approaches, by the formulation of random walk process of new extremes based on the analysis of trend of block maxima. Finally, the last part deals with a brief illustration of presented methods.

[1]Institute of Information Theory and Automation, Prague 8, e-mail: volf@utia.cas.cz

2 A model of trend

Let $X(t)$ be a series of maximal values in periods t, so that t is the discrete time, periods $t = 1, 2, \ldots$ given by our selection of data. We shall consider the following model form:

$$X(t) = m(t) + r(t),$$

where $m(t)$ stands for a trend function, $r(t)$ are the residuals, random errors. We allow for their autoregressive structure, so that

$$r(t) = \sum_{j=1}^{K} a_j \cdot r(t-j) + \sigma \cdot \varepsilon_t,$$

where ε_t are already i.i.d. $N(0,1)$ (standard normal) variables, $\sigma > 0$ is constant. Naturally, during the analysis those assumptions have to be checked. An alternative can consider time-dependent variance, either given by a function of time (as in Volf, 2011) or for instance by an ARCH model. As regards the selection of trend function, we consider two possibilities:

1. A parametrized function corresponding to the shape of trend. Frequent choices are S-shape Gompertz function or exponential decay curve (see Kuper and Sterken, 2003).
2. Trend constructed from a linear combination of functional units (eg. polynomials, goniometric functions, polynomial splines or others). Optimal selection of units and a degree of model is achieved with the aid of tests of parameters significance, the choice may be supported by the use of some penalized criterion, for instance the BIC. However, as we are interested in a global trend and also in the possibility to extrapolate (predict) it, the choice of basic functions is rather limited.

Estimation procedure is then based on the method of least squares, even in the case of autoregression among errors $r(t)$. It can seem that, because we deal with values of block maxima, the use of GEV (generalized extremal values) distribution should be preferred to normal errors. In such a case, method of maximal likelihood should be employed. However, the difference is negligible, as we show also in our example.

Diagnostics of model fit: The goodness-of-fit of selected model, namely the correspondence of errors ε_t to the standard normal distribution, can be tested both graphically (by the Q-Q plot) and numerically, e.g. with the aid of the Kolmogorov-Smirnov test. The selection of degree of eventual autoregression is standardly based on the maximum likelihood estimation (i.e. the mean squares in the Gauss distribution case) of autoregression parameters, on the tests of their significance, and also may be supported by the BIC criterion.

The constantness of σ, i.e. the homoskedasticity of remaining term, can be tested by the White test (White, 1980). It is based on the coefficient of determination in the linear regression of squared residuals (i.e. estimated $\sigma \cdot \varepsilon_t$) on regressors contained in $m(t)$. The test statistics is the coefficient of determination multiplied by the sample size, its critical value is given by corresponding chi-square quantile with $p - 1$ degrees of freedom, where p is the number of parameters in $m(t)$.

Finally, the independence of ε_t can be tested for instance by simple nonparametric tests (”series above and below median”, ”series up and down”).

Generally, the models with smooth trend do not consider any change of conditions, though such changes are quite frequent in practice. Then, the analysis can be amended by a method searching for potential changes, as well as by the detection of outlied values.

3 Models of random point process

Such models are based on the notion of intensity of new extreme occurrence (cf. Embrechts at al, 1997, Beirlant et al, 2004, and references there). The methodology is borrowed from the survival analysis. Models also allow to incorporate the dependence of intensity on influencing factors, for instance in the framework of Cox's regression model. In order to enlarge the point process scheme to the description of both new extremes times and values, we can use a model of compound point process. Compound process means the process of random increments at random times, formally

$$C(t) = \int_0^t Z(s)\,\mathrm{d}N(s),$$

where $Z(s)$ are (nonnegative) random variables and $N(s)$ is a counting process. If $N(s)$ has intensity $\lambda(s)$ and the mean and variance of $Z(s)$ are $\mu(s)$, $\sigma^2(s)$, respectively, then the mean development of $C(t)$ is given as

$$EC(t) = \int_0^t \lambda(s)\,\mu(s)\,\mathrm{d}s \quad \text{and} \quad \operatorname{var} C(t) = \int_0^t \lambda(s)\,\left(\mu^2(s) + \sigma^2(s)\right)\mathrm{d}s.$$

Now, both components of compound process can depend on explaining factors, via conveniently selected regression model (as is for instance the model presented in Volf, 2005). In the discrete-time case, i.e. also in the block maxima approach, we register just whether the new maximum was achieved or not in certain period. Then the compound process changes to a random walk model. It is described by probabilities $p(t)$ of new extreme occurrence in period t and random variables $Z(t)$ of its increase.

3.1 New maximum occurrence and value

Let us assume that up to time 0 the maximal value of the series was R. Further, let the block maxima in following periods be described as (continuous type) random variables $X(t)$, $t = 1, 2, \ldots$, with probability densities, distribution functions, survival functions $f_t,\ F_t,\ , S_t = 1 - F_t$, respectively. Then the probability that a new maximum will occur in period k is

$$p(k,R) = P\left\{X(j) < R, j = 1, 2, .., k-1,\ X(k) > R\right\} = \tag{1}$$

$$= \left\{\prod_{j=1}^{k-1} P(X(j) < R)\right\} \cdot P(X(k) > R) = \left\{\prod_{j=1}^{k-1} F_j(R)\right\} \cdot S_k(R)$$

when $X(t)$ are independent (conditionally, given the trend function). Further, the new maximum value is given by the density

$$g_k(r,R) = P(X(k) = r | X(k) > R) = \frac{f_k(r)}{S_k(R)}, \quad \text{for } r > R, \tag{2}$$

provided k is the period of new maximum occurrence.

Therefore, when the joint distribution of $X(t)$ is known (estimated, in the present context), the distributions of random variables T_R - the period of new record occurrence, and Z_R of new maximum improvement, can be derived easily. Namely,

$$P(T_R = k) = p(k,R), \quad k = 1, 2, \ldots$$

and the distribution of Z_R has the density

$$g(z,R) = \sum_{k=1}^{\infty} p(k,R) \cdot g_k(R+z,R). \quad \text{for } z > 0. \tag{3}$$

In most instances, including the case of normal distribution, these formulas can be evaluated just numerically.

3.2 A Markov chain of extremes, its prediction

The process of growth of the maximal value can also be treated as a Markov chain, with discrete time and continuous state space. Again, assume that at time t the actual maximum value is R_t. Then, at time $t+1$, no change will occur with probability $P(R_{t+1} = R_t) = P(X(t+1) \le R_t) = F_{t+1}(R_t)$, while the transition probability to a higher value $r > R_t$ is given as in (2), namely by density $g_{t+1}(r, R_t) = f_{t+1}(r)/S_{t+1}(R_t)$.

Such a Markov scheme is convenient for random generation of future process paths. Namely, assume that the data up to period T, $X(1), \ldots, X(T)$, are available and that the parameters of trend model are estimated from them. The objective is to predict the process for next periods. First, the trend of $X(t)$ is extrapolated to $t > T$. On this basis, we generate random trajectories of Markov process of extremes described above, starting from value R_T at T. From a large number of such random trajectories, the development of certain sample characteristics of future process course can be computed. For instance the mean values, variances as well as different quantiles.

3.3 Prediction bands and POT view

Random generation of future trajectories of analyzed process yields a sample of them, say $f_m(t), t \in (T, T_1), m = 1, .., M$. Then the point-wise (at each t) 'prediction' intervals for $X(t)$ are obtained immediately from sample quantiles of f_m at t. Methods for construction of prediction bands on the whole interval (T, T_1) could be, theoretically, connected with the concept of 'depth of data' (see for instance Zuo and Serfling, 2000). Practically, the approach corresponds to the construction of multivariate quantiles, for instance in the following way: Let us consider a sample of functions $f_m(t)$ given empirically by values at the same set of points $t_j; j = 1, .., J$. For each $k < M/2$, point-wise k/M or $(M-k)/M$ sample quantiles (i.e. at each t_j) can be constructed. If we join them to a band, we can try to find such k that, approximately, a given proportion (95%, say) of functions lies below it. As an additional finer criterion we can compare numbers of points at which the quantiles are crossed. In other words, in such a way we construct an empirical version of the threshold which is crossed (on the whole interval) just with probability 5%.

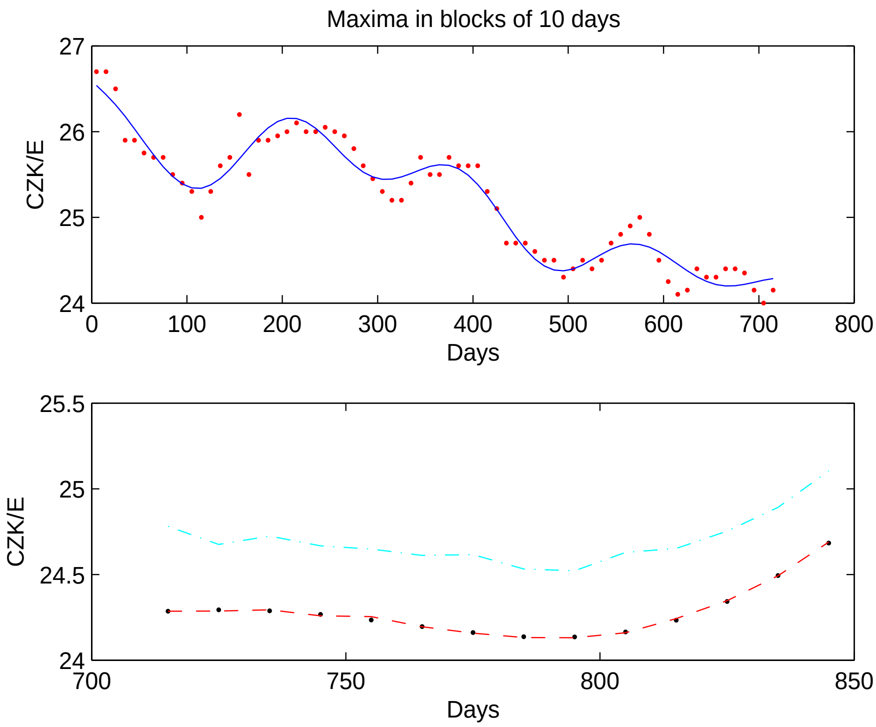


Figure 1: Block maxima and trend function (above), extrapolation of trend and 90% prediction band

4 Application

Figure 1, in its upper subplot, displays, by points, the development of exchange rate of CZK to Euro, namely 10 days maxima during approximately last 2 years (May 2009 to April 2011). As we know, the trend was decreasing, moreover, the data show certain seasonal components and also other non-regular disturbances. The plot contains also fitted trend curve. It was created as a linear combination of certain basic function, their selection was optimized as described in Part 2, following the approach 2. Namely,

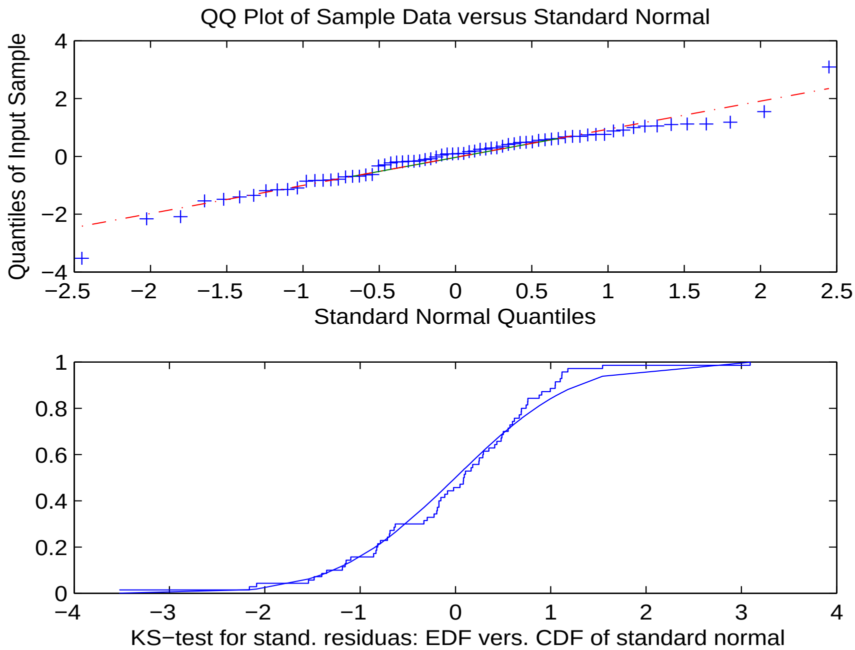


Figure 2: Q-Q plot (above) and K-S test (below) comparing empirical distribution of standardized errors ε_t with standard normal distribution

we obtained

$$m(t) = b_1+b_2\cdot t+b_3\cdot t^2+b4\cdot cos(2\pi/T)+b_5\cdot sin(4\pi/T)+b_6\cdot sin(6\pi/T)+b_7\cdot sin(8\pi/T)+b_8\cdot cos(8\pi/T). \quad (4)$$

As T is actually the range of times t, $sin(2k\pi/T)$ is a function with k periods during T. Hence, the trend function has also several periodic components, with a period of 1 year and also with another period half-year long.

Further, it was detected that the residual values behaved as AR(1) series,

$$r(t) = a \cdot r(t-1) + \sigma \cdot \varepsilon_t.$$

Parameters were estimated by the least squares method, with results (half–widths of 95% confidence intervals are in parentheses):

$b_1 = 27.05412$ (18.01561), $b_2 = -0.00896$ (0.00271),

$b_3 = 0.000008$ (0.000004), $b_4 = -0.58707$ (0.20364),

$b_5 = -0.25745$ (0.06260), $b_6 = -0.12884$ (0.05968),

$b_7 = 0.21106$ (0.05863), $b_8 = 0.11985$ (0.05860),

$a = 0.42870$ (0.21395).

Residual standard deviation σ was estimated as $s = 0.15344$.

Figure 2 shows the Q-Q plot (above) and graphical version of the Kolmogorov-Smirnov test (below), both comparing empirical distribution of realizations of ε_t with $N(0,1)$ distribution. Neither graph contradict to the good fit. Numerically, in K-S test the maximal departure of empirical and hypothetical distribution function was 0.0903 while the critical value for $n = 72$ is larger, 0.1601.

The homoskedasticity of residual term was tested with the White test (described in Part 2). The test statistics yielded 12.27, while the 95% quantile of the chi-square distribution with 7 degrees of freedom was 14.07, so that the hypothesis of constant σ was not rejected.

The assumption of independence of errors ε_t was checked, too. Two nonparametric tests ("series above and below median", "series up and down") were employed. The independence was not formally rejected by any test, P-values were 0.0576 and 0.1406, respectively (though these P-values are quite close to their critical border).

Finally, we also tried to predict the behavior of series $X(t)$ for next 13 periods 10 days long. The result is displayed in Figure 1, lower subplot. The points correspond to extrapolated trend curve $m(t)$, at $t = 715, 725, ..., 845$. We then generated 1000 realizations of series of future block maxima, following the approach described in subsection 3.2. Dashed curve close to $m(t)$ is connecting point-wise sample medians obtained from them. Remaining curve, dot-dashed, is the threshold which was crossed just by 100 (i.e 10%) of trajectories, in other words, it is an empirical (so that approximate) 90% prediction band for maxima in the series of exchange rate development.

5 Conclusion

We have proposed a model for the occurrence of extremal values, taking into account the series of block maxima. Its development is represented by a nonlinear regression (trend) model. From it, the Markov chain of new extremes occurrence and values has been derived. While, explicitly, the model depends just on time, an implicit dependence of new maximum on the past maximum duration and value is involved, too. An application to real data has shown usefulness and good performance of the model. A future improvement should concentrate to the problems of detection (and prediction) of changes in analyzed time series. Except the use of statistical methods for changes and outliers detection, one can think also on selection of informative factors indicating the changes of conditions, and on methods of pattern recognition for the analysis of those factors.

Acknowledgements

The research is supported by the project of GA CR No 402/10/0956.

References

[1] Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J.: *Statistics of Extremes: Theory and Applications.* Wiley, New York, 2004.

[2] Embrechts, P., Kluppelberg, K., and Mikosch, T.: *Modelling Extremal Events.* Springer, Berlin, 1997.

[3] Kuper, G.H. and E. Sterken (2003): Endurance in speed skating: the development of world records, *European Journal of Operational Research* **148** (2003), 293–301.

[4] Volf, P.: Hazard rate model and statistical analysis of a compound point process, *Kybernetika* **41** (2005), 773–786.

[5] Volf, P.: A stochastic model of progression of athletic records, *IMA Journal of Management Mathematics* **22** (2011), 157–169.

[6] White, H.: A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* **48** (1980), 817-838.

[7] Zuo, Y. and Serfling, R.: General notion of statistical depth function. *Annals Statist.* **28** (2000), 461–482.

Proposal of determination of the rules for design of fuzzy sets of qualitative indicators used in the decision-making processes in Lean Company

Petr Wolf[1], Pavel Machula[2]

Abstract. This paper promotes the establishment of the rules for the construction of a fuzzy set of quality indicators for use in the decision-making process of Lean Company. Its aim is to provide an example of specific fuzzy economic indicators of the quality of the company's management through the definition and establishment of the general rules applicable to the use of this fuzzy set which will be accomplished with the simplest possible processing whilst meeting the necessary conditions associated with it. The decision-making is based on the use of a fuzzy controller, and fuzzy sets are included in the base rules and data resources.

Keywords: fuzzyfication, fuzzy controller, fuzzy sets, Lean Company

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

Information and the knowledge in the information society are the source of the power and the fortune [2], [11]. Who is owner of the information sources, who is able to transfer information into knowledge is in the advantage, he has power. The change is now truly global. Never has this statement been more true as today. The global market is real battle field and business is about war [15]. The management strategic decision making tasks are classified according to a large number of different criteria [1], [3]. Using the fuzzy sets theory device is unjustly neglected while using them, although similar methods and principles implemented and verified in practice in technical disciplines can be used by analogy.

1.1 Lean Company

With conception of Lean Company began Henry Ford in his car factories in 1903-1923. Conception based on reduction in waste was expanded in car factory Toyota in Japan in the 50's – 60's of 20th century. From a number of characteristics we assessed as best for our project the following characteristic: „Leanness of company means to do only such activities that are necessary, to do it correct on the first occasion and to do it faster than others and therewithal spending less money“ [17]. The conception of lean company can be characterized by the following five basic principles [5], [13], [17]: value identification, mapping the value chain, creation of the flow, creation of the move, permanent improvement. There are next advantages: higher quality, higher profit, higher flexibility, bigger strategic orientation. The Lean Company concept represents the approach to company management which influences company's duality, flexibility, bottom line and production [12]. Conception of Lean Company is a complex of effective managerial methods with association of ICT, which are incorporated into a single philosophy - "lean" that must be continuously applied as a whole, in order to achieve the desired long term objective - the Lean Company.

1.2 Using of fuzzy sets

Fuzzy sets are generalizations of classical two-valued crisp set, which limits of examining whether the elment belongs to a given set or not. The result of the characteristic function $\mu_A(x)$ also called the membership functions are the only values $\mu_A : U \to \{0,1\}$. Generalization in the form of fuzzy sets lies in changing the approach to the examination propriety of elements to a specific set. The result of membership functions is the value from the interval $\mu_A : U \to \langle 0,1 \rangle$. Each function from U of uniquely identifies some fuzzy set. In some

[1] Slezská univerzita v Opavě, Obchodně podnikatelská fakulta, Univerzitní nám. 76, Karviná, p.wolf@seznam.cz
[2] MWare CZ, s.r.o, B. Martinů 733, Trutnov, pavel.machula@mware.cz

textbooks the fuzzy sets are differentiated by their membership functions, it will be further verified, simplify the registration and there is no misunderstanding [10], [16].

To solve these problems, it can be used expert system based on principle of fuzzy control (fuzzy controller). Expert system appears from the identification of quality indicators used in the decision-making processes their subsequent fuzzyfication and making rules IF THEN. The values of linguistic variables can be interpreted as fuzzy sets. A set of linguistic values is a set of terms, which are defined on the universe.

In the case of decision task by manager, the natural human understanding and expression of quantitative traits plays an important role [6]. Rules for the construction of fuzzy sets are as follows: identification of appropriate indicator - linguistic variables, the definition of the universe, setting the values of variables (fuzzy sets), definition of characteristic function (membership function) for particular fuzzy sets, i.e., determination the degree of membership with which the element (defined on the universe) belongs to set.

The principle of the identification of qualitative indicators is to find such indicators, in the company which are interrelated and their relationship can be expressed by (mostly two-dimensional) functional dependence of linguistic variables and simultaneously to find operational variable, which is able to affect the value of terms at least one of the linguistic variables. As an example see Table 1.

Linguistic variable #1	**Linguistic variable #2**	**Action value**
Added value	Level of process necessity	Internal processes/arrangement change
Level of investment	Usage efficiency of means of production	Change in production program
Cost of share of material costs of foreign suppliers for one product	Presumption of long-term changes in the exchange rate	Change of inventories
Total stocks	Presumption of monthly production development	Change of inventories
Market share	Usage level of production factors	Change in advertising investment

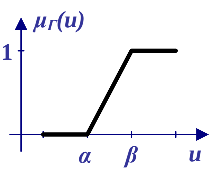
Table 1 Example of linguistics variables and action value. Source: Authors‘ table

To each linguistic variable, it is necessary to define its universe and a set of terms. The universe is a study of all acceptable values of linguistic variables, terms represent particular fuzzy sets. From perspective of other usage, it is appropriate to specify terms in the entirety of the universe.

Because of the need for further mathematical processing, it is required to convert linguistic variable to a particular fuzzy sets. An important factor is the choice of membership function conformation, which may be different. Its conformation usually depends on the location of fuzzy sets within its universe. The conformation can be expressed as a curve, the trapezoid, or its parts. The conformations of the membership function that are often used appertain Γ - function, L - function, Λ - function, Π - function, see Table 2.

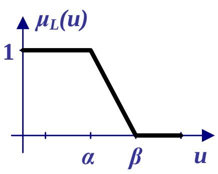
In the first step of linguistic variable fuzzyfication, it is necessary to make normalization of universe at the selected interval. The following is assignment by the degree of membership to fuzzy sets of each value in the universe so that the universe would be fully covered by of all fuzzy supports. The overlap of fuzzy supports in the universe is desirable. The last step is to choose a particular form of membership function of particular fuzzy sets. The form of membership functions is recommended to choose, so as to avoid to unnecessary computational complexity during the next defuzzyfication. Usually, for practical reasons, use a linear function with unilateral constraints (L, Γ), a linear function with both-sided restriction (Λ, Π) function or gradual function, see Table 2. Probability density function of normal distribution may serve as an example of using smooth nonlinear membership function, there is a need to realize that a relationship need not apply: $\int_{-\infty}^{\infty} f(x;\mu,\sigma^2)dx = 1$.

Process and definition of Γ -function:



$$\Gamma(u,\alpha,\beta) = \begin{cases} 0 & u < \alpha \\ (u-\alpha)/(\beta-\alpha) & \alpha \le u \le \beta \\ 1 & u > \beta \end{cases}$$

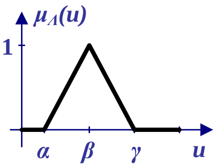
Process and definition of L-function:



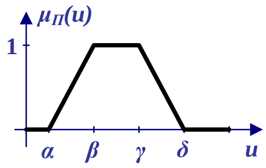
$$L(u,\alpha,\beta) = \begin{cases} 0 & u > \beta \\ (\beta-u)/(\beta-\alpha) & \alpha \le u \le \beta \\ 1 & u < \alpha \end{cases}$$

Process and definition Λ -function:

Process and definition Π - function:



$$\Lambda(u,\alpha,\beta,\gamma)=\begin{cases}0 & u<\alpha\\ (u-\alpha)/(\beta-\alpha) & \alpha\le u\le\beta\\ (\gamma-\alpha)/(\gamma-\beta) & \beta\le u\le\gamma\\ 0 & u>\gamma\end{cases}$$



$$\Pi(u,\alpha,\beta,\gamma,\delta)=\begin{cases}0 & u<\alpha\\ (u-\alpha)/(\beta-\alpha) & \alpha\le u\le\beta\\ 1 & \beta\le u\le\gamma\\ (\delta-u)/(\gamma-\delta) & \lambda\le u\le\delta\\ 0 & u>\delta\end{cases}$$

Table 2 Conformations of the membership function that are often used appertain Γ - function, L - function, Λ - function, Π - function. Source: Derived from [7]

Partial decision-making process is based on the implication of two-dimensional dependability with one rule in the form of IF <antecedent> THEN <consequent>. Antecedent contains fuzzy operation includes two fuzzy sets. The result of the fuzzy operation is level α value for the consequent fuzzy set (a measure of action incidence). Fuzzy operation is usually fuzzy conjunction, fuzzy disjunction or fuzzy negation.

The base of rules can be created either on the basis of empirical knowledge of service or on the basis of applicable meta-rules [8], [14]. From the view of decision-making process, it is necessary to fulfill the above condition - coverage of the universe of particular fuzzy sets. In case of defaulting of condition by fuzzy sets in antecedent could occur to stage, to which would not suit any inference rule. Failure to condition of fuzzy set of consequent is in terms of reality less likely, because non-existence of fuzzy set consequent would be arguable during the construction of rule. But this case can not be completely ruled out especially in combination with the absent fuzzy set of antecedent.

The use of Expert System based on fuzzy regulation in the Lean Company is different in method of processing and interpretation of results from the standard educational example. In reality it is an n-dimensional dependability with m-dimensional vector of actuating variables. Method how to find a solution, i.e. the opportune vector of actuating (manipulated) variables, consists in the decomposition of task to depending on the aforementioned two-dimensional dependability and its processing. It should be noted that the processing is done sequentially. The results of the processing of some tasks can provide quite a different value for particular actuating variable of m-dimensional vector. The resulting value can be influenced by the expert method, e.g. method of determining weights of criteria.

2 Model example

As a simplified example, we present the general management recommendations for stock producers. The model is based on the evaluation of variables of share in the costs of material of foreign suppliers to a one product with the expected long-term development of the exchange rate and the total amount of material resources with foresight monthly production. The example based on the recommendations and methodology used. Fuzzy sets have a linear shape of membership functions fuzzy supports overlap each other. Antecedent of decision rules includes standard fuzzy conjunction. Term definitions and graphics processing is based on [4], [7][3].

2.1 Definition of linguistic variables

The share of material costs of foreign suppliers for one product: $< I, L_I, U_I, M_I >$ where

I	is denotation of the linguistic variable " Cost of share of material costs of foreign suppliers for one product",
L_I	= { N_I – almost zero, S_I – low, M_I – medium, H_I – high}, set of function values,
U_I	= [0,100] – set in view of all variants, below universe, i.e. physical extent of the values in units [%],
M_I	is function that charts verbal values in the universe:

$$N_I \rightarrow \int_0^{100} L(I,10,30)/I \qquad S_I \rightarrow \int_0^{100} \Pi(I,0,20,30,50)/I \qquad M_I \rightarrow \int_0^{100} \Pi(I,30,50,60,80)/I$$

[3] The sign $\int_{min}^{max}$ is used for definition of appropriate fuzzy set (with emphasis on the range of values and continuity of universe), not for definite integral.

$$H_I \to \int_0^{100} \Gamma(I,60,80)/I$$

Presumption of long-term changes in the exchange rate: $< J, L_J, U_J, M_J >$ where

J is denotation of the linguistic variable " Presumption of long-term changes in the exchange rate",
L_J = { N_J – negative, C_J – stabile, P_J – positive}, set of function values,
U_J = [-30, 30], universe, i.e. physical extent of the values in units [%],
M_J is function that charts verbal values in the universe:

$$N_J \to \int_{-30}^{30} L(J,-12,0)/J \qquad C_J \to \int_{-30}^{30} \Pi(J,-18,-6,6,18)/J \qquad P_J \to \int_{-30}^{30} \Gamma(J,0,12)/J$$

Total stocks: $< T, L_T, U_T, M_T >$ where

T is denotation of the linguistic variable „Total stocks"
L_T = { C_T – critical, S_T – small, M_T – optimum, H_T – high, E_T – too high }, set of function values,
U_T = [0, 20] – universe, i.e. physical extent of the values in units [mil. Kč],
M_T is function that charts verbal values in the universe:

$$C_T \to \int_0^{20} L(T,5,8)/T \qquad S_T \to \int_0^{20} \Pi(T,5,7,9,11)/T \qquad M_T \to \int_0^{20} \Pi(T,8,10,12,14)/T$$

$$H_T \to \int_0^{20} \Pi(T,11,13,15,17)/T \qquad E_T \to \int_0^{20} \Gamma(T,14,17)/T$$

Presumption of monthly production development: $< S, L_S, U_S, M_S >$ where

S is denotation of the linguistic variable „Presumption of monthly production development"
L_S = { S_S – decreasing, M_S – stabile, H_S – increasing}, set of function values,
U_S = [0, 100] – universe, i.e. physical extent of the values in units [%],
M_S is function that charts verbal values in the universe:

$$S_S \to \int_0^{100} L(S,60,70)/S \qquad M_S \to \int_0^{100} \Pi(S,60,70,80,90)/S \qquad H_I \to \int_0^{100} \Gamma(S,80,90)/S$$

Change of inventories: $< K, L_K, U_K, M_K >$ where

K is denotation of the linguistic variable „Change of inventories"
L_K = { *NB* – negative big, *NM* – negative middle, *NS* – negative small, *Z0* – almost zero, *PS* – positive small, *PM* – positive middle, *PB* – positive big }, set of function values,
U_K = [-10, 10] – universe, i.e. physical extent of the values in units [mil. Kč],
M_K is function that charts verbal values in the universe:

$$NB \to \int_{-10}^{10} L(K,-6,-4)/K \qquad NM \to \int_{-10}^{10} \Lambda(K,-6,-4,-2)/K \qquad NS \to \int_{-10}^{10} \Lambda(K,-4,-2,0)/K$$

$$Z0 \to \int_{-10}^{10} \Lambda(K,-2,0,2)/K \qquad PS \to \int_{-10}^{10} \Lambda(K,0,2,4)/K \quad PM \to \int_{-10}^{10} \Lambda(K,2,4,6)/K \quad PB \to \int_{-10}^{10} \Gamma(K,4,6)/K$$

2.2 Definition of decision-making rules 1

ID	Decision-making rule	ID	Decision-making rule
R$^{(1)}$	$(I = N_I) \wedge (J = N_J) \Rightarrow (K = PS)$	R$^{(7)}$	$(I = M_I) \wedge (J = C_J) \Rightarrow (K = Z0)$

R$^{(2)}$	$(I=S_I)\wedge(J=N_J)\Rightarrow(K=PM)$	R$^{(8)}$	$(I=H_I)\wedge(J=C_J)\Rightarrow(K=Z0)$
R$^{(3)}$	$(I=M_I)\wedge(J=N_J)\Rightarrow(K=PB)$	R$^{(9)}$	$(I=N_I)\wedge(J=P_J)\Rightarrow(K=NS)$
R$^{(4)}$	$(I=H_I)\wedge(J=N_J)\Rightarrow(K=PB)$	R$^{(10)}$	$(I=S_I)\wedge(J=P_J)\Rightarrow(K=NM)$
R$^{(5)}$	$(I=N_I)\wedge(J=C_J)\Rightarrow(K=Z0)$	R$^{(11)}$	$(I=M_I)\wedge(J=P_J)\Rightarrow(K=NM)$
R$^{(6)}$	$(I=S_I)\wedge(J=C_J)\Rightarrow(K=Z0)$	R$^{(12)}$	$(I=H_I)\wedge(J=P_J)\Rightarrow(K=NB)$

Table 3 The decision-making rules, dependence the share of material costs from foreign suppliers for one product vs. expectation of long term change of exchange rate

2.3 Definition of decision-making rules 2

ID	Decision-making rule	ID	Decision-making rule
R$^{(1)}$	$(R=C_R)\wedge(S=S_S)\Rightarrow(K=Z0)$	R$^{(9)}$	$(R=H_R)\wedge(S=M_S)\Rightarrow(K=NS)$
R$^{(2)}$	$(R=S_R)\wedge(S=S_S)\Rightarrow(K=Z0)$	R$^{(10)}$	$(R=E_R)\wedge(S=M_S)\Rightarrow(K=NM)$
R$^{(3)}$	$(R=M_R)\wedge(S=S_S)\Rightarrow(K=NS)$	R$^{(11)}$	$(R=C_R)\wedge(S=H_S)\Rightarrow(K=PB)$
R$^{(4)}$	$(R=H_R)\wedge(S=S_S)\Rightarrow(K=NM)$	R$^{(12)}$	$(R=S_R)\wedge(S=H_S)\Rightarrow(K=PM)$
R$^{(5)}$	$(R=E_R)\wedge(S=S_S)\Rightarrow(K=NB)$	R$^{(13)}$	$(R=M_R)\wedge(S=H_S)\Rightarrow(K=PS)$
R$^{(6)}$	$(R=C_R)\wedge(S=M_S)\Rightarrow(K=PM)$	R$^{(14)}$	$(R=H_R)\wedge(S=H_S)\Rightarrow(K=Z0)$
R$^{(7)}$	$(R=S_R)\wedge(S=M_S)\Rightarrow(K=PS)$	R$^{(15)}$	$(R=E_R)\wedge(S=H_S)\Rightarrow(K=Z0)$
R$^{(8)}$	$(R=M_R)\wedge(S=M_S)\Rightarrow(K=Z0)$		

Table 4 Decision-making rules, dependence of the total number of stocks vs. expectation of monthly production

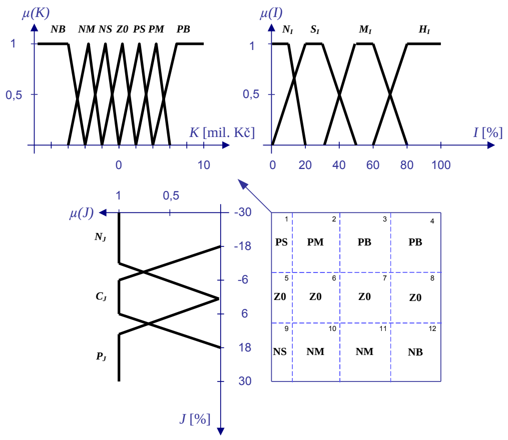


Figure 1 Rules for Table 3

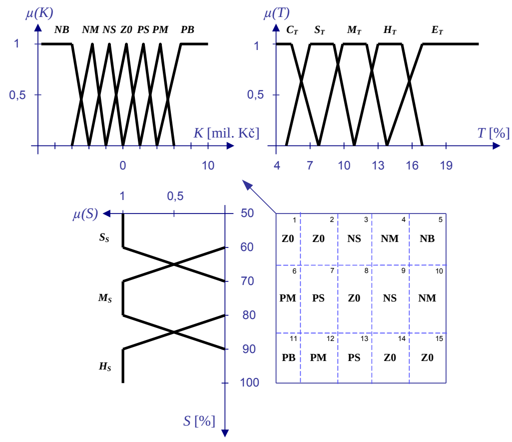


Figure 2 Rules for Table 4

2.4 Analysis of dependence

Based on the comparison table values and graphic solutions of both cases, it can be noted that various combinations of functional values of linguistic variables imply the same action hit. The solution also shows that actually there may be some situations which have contradictory effect, in view of setup of actuating variable. A concrete example is, when the value of the share of material costs from foreign suppliers for one product is low or larger than the small, which can happen due to seasonal ingredients appreciation of the domestic currency. This case implies the action hits of decision-making rules number 10, 11, 12, see Fig 1, namely {NM, NM, NB}.

Situation, when the current amount of material inventories is critically low and at the same time small and there is an intention to increase the monthly output, conversely it implies the action hits of decision-making rules number11 and number 12, see Fig 2, i.e. {PP, PM}. By reciprocal combination comparing of functional values of linguistic variables (implying the same action hit of both resolutions), there can be detected other relations between pairs of function values of linguistic variables, or otherwise hidden ones!

3 Conclusion

The methodology of using fuzzy sets for representation linguistic variables in the strategic decision-making enables (besides option the possibility to use not very accurate values and non-numeric expression values close to human understanding) composition assembly simply and quickly of understandable schematic procedures for dealing with all possible combinations. Visual solution helps to detect potential conflict situations.

The main benefit of using the above methodology is the ability to analyze appearance of combination of terms of linguistic variables in relation to the same value of the action variables. It is a generally applicable method for analysis of specific relationship between different linguistic variables terms, however, it is valid only for solving specific task. Limitation of validity is determined purely by setting individual decision rules. Using this fuzzy oriented approach in the Lean Company can be effected by analysis of internal structures and relationships on whose basis we can proceed to optimize them.

References

[1] Bellman, R., Zadeh, L. A.: *Decision Making in Fuzzy Environment*. Management Science 179(4), 1970.

[2] Drucker, P. F.: *Managing in Turbulent Times*. Harper and Row, New York, 1980. (in Czech - *Řízení v turbulentní době*. Praha: Managemant Press, 1994.).

[3] Harrison, E. F.: *The Managerial Decision-Making Process*. South-Western Coollege Pub. 576 p. Interprise Solution. System Manager, Cincinnati, Retrieved February 18th, 2010.

[4] Jura, P.: *Basics of fuzzy logic control and modeling*. Vitium, Brno, 2003, ISBN 80-214-2261-0. (in Czech - *Základy fuzzy logiky pro řízení a modelování*).

[5] Košturiak, J., Frolík, Z.: *Slim and innovative company*. Alfa Publishing, Praha, 2006, ISBN 80-86851-38-9. (in Czech: *Štíhlý a inovativní podnik*).

[6] Klir, G. J., Yan, B.: *Fuzzy Sets and Applications.* Prentice Hall, Upper Saddle river, NJ, 1995.

[7] Machula, P.: *Application of fuzzy sets in management decisions*. TU VŠB Ostrava, 2005, ISBN 80-248-0952-4. (in Czech - *Aplikace fuzzy množin při rozhodování managementu*).

[8] Modrlák, O.: *Automation Control Theory II.* Liberec, 1992. (in Czech - *Teorie Automatického řízení*).

[9] Navara, M., Olšák, P.: *Basics of Fuzzy Sets*. ČVUT, Praha, 2002, ISBN 80-01-02585-3. (in Czech – *Základy fuzzy množin*).

[10] Ramík, J., Vlach, M.: *Generalized Concavity in Fuzzy Optimization and Decision Analysis*. Kluwer Academic Publishers, Boston/Dordrecht/London. USA, 2002, ISBN 0-7923-7495-9.

[11] Tofler, A.: *The Third Wave*. William Morrow, New York, 1980.

[12] Wolf, P.: *Successful business company in the global market*. CS Profi – Public, Bratislava, 2006, ISBN: 80-969546-5-2. (in Czech - *Úspěšný podnik na globálním trhu*)

[13] Wolf, P., Besedová, H., Wolf, J.: . Proc. 10th International Carpathian Control Konference ICCC'2009, May 24-27, Zakopane, Poland, 2009, ISBN 83-89772-51-5.

[14] Wolf, P. 2007: Fuzzy Logic in Decission – Making Process. *In Proceedings of WSEAS PROCEEDINGS 10/2007, Conference ID: 571-319. The 3rd WSEAS International Conference on DYNAMIC SYSTEMS and CONTROL (CONTROL´07),* Arcachon, France, October 2007, ISSN: 1790-5117, ISBN: 978-960-6766-08-4, 13-15.

[15] Wolf, P., Machula, P.: Information Society and Management. *Acta Academica karviniensia, 1/2005. Karviná: Slezská univerzita v Opavě, Obchodně podnikatelská fakulta v Karviné,* ISSN 1212-415X, 270.

[16] Zadeh, L. A.: Outline of a new approach to the analysis of complex systems and decision processes. *IEEE trans. SMC*, 1973, 3, 1, 28 - 44.

[17] Electronic Publication: V. Kotelnikov: *Lean Production – Doing More With Less*. 2008 [online] URL: <http://www.1000ventures.com/business_guide/lean_production_main.html>.

Evaluating Value at Risk an Expected Shortfall of Individual Insurance Claims

Kateřina Zelinková[1], Jiří Valecký[2]

Abstract. Value at Risk can be considered as a basic measure for quantifying market, insurance or credit risk. It can be also referred to as methodology which is used especially to determine capital requirement in banks or in insurance companies. This paper deals with estimating Value at Risk and conditional Value at Risk of motor hull insurance portfolio within the Solvency II. Therefore, the VaR is estimated at 99.5% confidence level over one year risk horizon. We evaluate both risk measures analytically under assumption of traditional distributions, i.e. exponential, gamma and Weibull´s, and also using Extreme Value Theory to respect the fat tail of empirical distribution. We give the evidence that the both risk measures are highly underestimated when the traditional distributions are assumed. First and foremost, we describe VaR and CVaR and then we focus on Extreme Value Theory. Subsequently, we estimate both risk measures under all mentioned probability distribution conditions. In the end, we compare all estimates with each other.

Keywords: Value at Risk, Extreme Value Theory, Conditional Value at Risk, generalized Pareto distribution, Solvency II, motor hull insurance.

JEL Classification: C16, G22, G32

1 Introduction

The Value at Risk (VaR) is a risk measure representing a value of loss which will not be exceeded over given risk horizon at some significance level. There is also a possibility to refer to VaR as managing risk methodology which is applied widely to modelling credit, operational, market and also insurance risk. A good introduction to Value at Risk methodology is provided by the technical document from [9] or by many follow-up books such as [5], [6], [7], and also in insurance area [8] and others.

Nevertheless, we can find a lot of VaR criticism, for instance [7]. [2] dealt with the features of good risk measure (called coherent) which is defined by four assumption imposed on the ideal risk measure, i.e. monotonicity, sub-additivity, homogeneity and translational invariance. Value at Risk satisfies all these features only in specific case. Specifically, the sub-additivity is violated as far as the portfolio´s profit/loss or portfolio´s return cannot be characterized by some elliptical probability distribution; see [3] for more details. In addition, the VaR says nothing about the loss behind the VaR. Therefore, other risk measures are preferable such as conditional Value at Risk (CVaR) which represents the average of losses exceeding the VaR.

Application of VaR to insurance risk consists in the evaluation of maximum individual claim or maximum total claim at given significance level. Within Solvency II the VaR is needed to estimate at 99.5% confidence level over one year risk horizon. Moreover, if the VaR and CVaR are estimated analytically, the distribution assumption is needed. Traditional distributions such as exponential, gamma or Weibull can be supposed but this assumption results in underestimation of VaR and CVaR due to the existence of fat tails. In facts, to solve this problem, only two approaches seem to be applicable. One can consider mixture distribution to fit the empirical distribution the most, the other can apply Extreme Value Theory (EVT) focused on fitting the tail distribution only which is approximated mostly via general Pareto distribution.

Thus, the aim of our paper is analytical evaluation of VaR and CVaR on individual claims from motor hull insurance portfolio during the year 2009. To respect the fat tail, we apply EVT assuming the general Pareto distribution and we compare our estimates with values obtained under assumption of traditional probability distributions. We highlight in this paper that the VaR and CVaR are highly underestimated in that case.

[1] VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská tř. 33, 701 21, katerina.zelinkova.@vsb.cz.
[2] VŠB-TU Ostrava, Faculty of Economics, Department of Finance, Sokolská tř. 33, 701 21, jiri.valecky@vsb.cz.

The paper is organized as follows. Section 2 is devoted to the description of general Value at Risk methodology and Extreme Value Theory. The VaR and CVaR estimates under traditional probability distribution condition and within EVT are determined in Section 3 and Section 4 concludes the paper.

2 Value at Risk methodology and Extreme Value Theory

In this part, we firstly characterize general formula of Value at Risk and then we describe Conditional Value at Risk. Explanation of general principle of Extreme Value Theory follows.

2.1 Value at Risk

VaR measures were developed in response to the financial disasters in the 1990s and have obtained an increasingly important role in risk management. The advantage of this criterion is that it provides a single quantity that summarizes the total risk in a portfolio of financial assets. Therefore, it gained popularity among managers and regulators. The VaR measure is used in setting capital requirements for banks and insurance companies throughout the world.

Value at Risk is defined as the smallest loss for the predicted level of probability for a given time interval. It is a function of two parameters, i.e. the risk horizon and the confidence level. Within Solvency II, Value at Risk is defined at 99.5% confidence level over one year risk horizon. We can also characterize the Value at Risk as a one-sided confidence interval of potential loss of portfolio value for a given holding period, which can be written:

$$F(x) = P\left(X \le -VaR_{\alpha,\Delta t}(x)\right) = \alpha\,, \tag{1}$$

where $F(x)$ is cumulative distribution function, α is confidence level and Δt is holding period or risk horizon. From (1) follows, that

$$-VaR_{\alpha,\Delta t}(x) = \inf\left\{x \in R : \alpha \le F(x)\right\}. \tag{2}$$

2.2 Conditional Value at Risk

Conditional Value at Risk (*CVaR*) is also sometimes referred to as expected shortfall or expected tail loss. Measurement of risk by the CvaR is better than the VaR because we respect the extreme values at the tail of the distribution. VaR represents level of unexpected loss only but CVaR informs what the losses would exceed this level. This method fulfils the conditions for coherent measure of risk. The characteristics of coherence risk measure were analyzed by [3] and it should satisfy these following conditions:

1. monotonicity: if $X \le Y$, then $\rho X \le \rho Y$;
2. sub-additivity for random variable X and Y: $\rho(X+Y) = \rho(X) + \rho(Y)$;
3. positively homogeneity for any constant $h > 0$: $\rho(hX) = hp(X)$;
4. translational invariance for any constant $h > 0$: $\rho(h+X) = h + \rho(X)$.

Conditional VaR can be generally defined in the form of

$$CVaR_{\alpha,\Delta t} = -\alpha^{-1}\int_{-\infty}^{VaR_{\alpha,\Delta t}} x\, f(x)\,dx, \tag{3}$$

where $f(x)$ is density function.

2.3 Extreme Value Theory

Extreme value theory (EVT) provides a way of smoothing and extrapolating the tails and thus it improves Value at Risk estimates at very high confidence level when the fat tail occurs in empirical distribution. There are several ways how to determine VaR and CVaR within EVT. One can apply Block of Maxima method (BM) when the maximum variables in successive periods are considered; the other one can use Peak over Threshold (POT) method focusing on variables above a given threshold. In accordance with Picklands-Dalkema-de Hann theorem, the limiting distribution of $F(x)$ in (1) is a generalized Pareto distribution (GPD) with cumulative distribution function given by

$$F_{\xi,\beta}(x)=\begin{cases}1-\left(1+\xi\dfrac{x}{\beta}\right)^{-1/\xi} & ;\xi\neq 0,\\ 1-\exp\left(-\dfrac{x}{\beta}\right) & ;\xi=0,\end{cases}\tag{4}$$

where ξ is a shape parameter determining the heaviness of the tail of the distribution and β is a scale parameter.

The VaR within EVT is given by

$$VaR_{\alpha,\Delta t}=u+\frac{\beta}{\xi}\left[\left(\frac{n}{n_u}(1-\alpha)\right)^{-\xi}-1\right],\tag{5}$$

where α is confidence level, n is number of all observations and n_u is number of exceedance over threshold value u. The CVaR estimate is calculated as follows

$$CVaR_{\alpha,\Delta t}=VaR_{\alpha,\Delta t}+\frac{\beta+\xi(VaR_{\alpha,\Delta t}-\mu)}{1-\xi}=\frac{VaR_{\alpha,\Delta t}}{1-\xi}+\frac{\beta-\xi\mu}{1-\xi}\,.\tag{6}$$

3 VaR and CVaR estimates of individual insurance claims

In this part, Value at Risk and conditional Value at Risk are determined under assumption of traditional distributions, i.e. exponential, gamma and Weibull´s distributions, and both risk measure are also estimated within Extreme Value Theory via generalized Pareto distribution. All VaR estimates are obtained in compliance with Solvency II, thus both risk measure are calculated at 99,5% confidence level over 1 year risk horizon. For our purpose, we worked with data containing the claims of individual policyholders within motor hull insurance during the year 2009. Some statistical characteristics of our data sample, namely mean, standard deviation, skewness, kurtosis and the total number of insurance claims, are shown in Table 1 and Figure 1 records the histogram.

Mean	56 441 CZK
Standard deviation	96 534 CZK
Kurtosis	1 403
Skewness	29
The total number of insurance claims	5 164

Table 1 Basic statistical characteristic

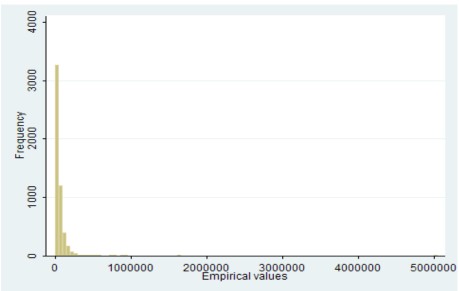


Figure 1 Histogram of empirical values

We can see from the Figure above that the most of the losses belong to the interval 1 - 350 000 CZK. We can also mention that there are also a few observations above the level of about 1.3 million CZK. Subsequently, we verify whether the application of EVT is reasonable. Thus, we estimate via maximum likelihood method the parameters of traditional distributions (exponential, gamma and Weibull´s) on our data sample and we plot the quantiles of values from theoretical distributions against quantiles of empirical values in the following QQ plots. Calculation these probability distributions we find in [4]. The parameter estimates of all probability distributions are recorded in Table 2.

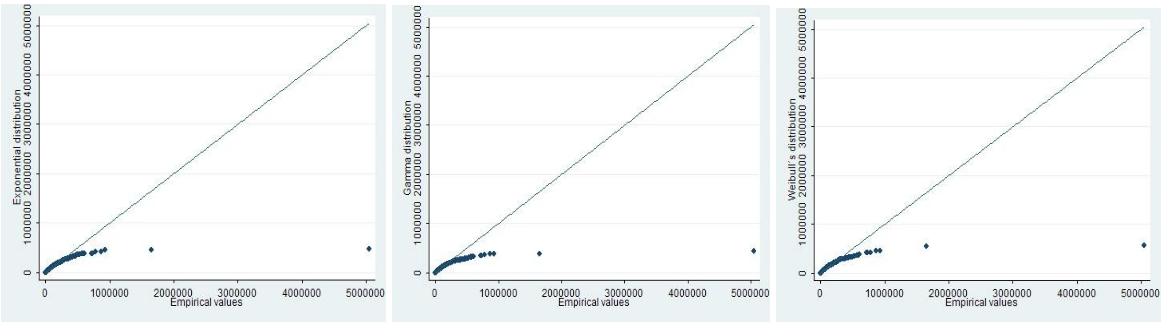


Figure 2 The results of graphics test via QQ plots

Exponential	λ	
	56 440.971	
Gamma	α	β
	1.112	50767.864
Weibull	α	β
	56 234.865	0.993

Table 2 Parameter estimates of traditional distributions

On the basis of Figure 2, we can conclude that data sample are characteristic by heavier fat tail than the traditional distributions and therefore it is appropriate to determine VaR and CVaR within EVT.

Next, we applied EVT. Firstly, we determined the threshold value via mean excess loss function depicted in the Figure 3. We should note here that it is necessary to select such threshold value from which the mean excess function can be approximated by straight line. We decided to select the threshold value at the level of 150 000 CZK.

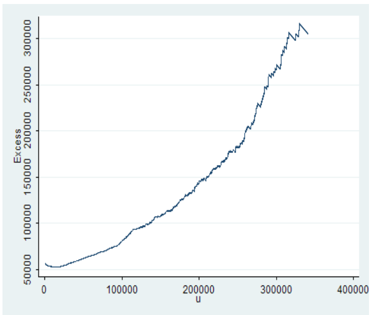


Figure 3 Mean excess loss function

To determine the value of VaR and CVaR under EVT, it is necessary to estimate the parameters of generalized Pareto distribution on the data exceeding the threshold value. To estimate parameters, we employed maximum likelihood method again. Obtained results are recorded in next table.

Generalized Pareto distribution	**Value**
ξ	0,0198
β	252 769,06
Threshold	150 000
Number of empirical data over the threshold	316

Table 2 Parameter estimates of generalized Pareto distribution

Finally, VaR and CVaR were determined under assumption of traditional distributions and within EVT. All results are in next table.

	Distribution			
	Pareto	exponential	gamma	Weibull
Extreme value theory ($VaR_{99,5\%}$)	799 062	299 042,18	282 385,96	301 618,71
Conditional Value at Risk ($CVaR_{99,5\%}$)	918 928	355 483,15	333 936,19	344 310,37

Table 3 The amount of risk for difficult distribution

The amount of capital to cover unexpected losses at 99.5% confidence level is 799 062 CZK in the case of VaR within EVT. We can see that the difference from the VaR estimates under traditional distribution assumption is much lower. The same results we conclude according to the CVaR results. Also in this case, it is obvious that the CVaR estimates are highly underestimated. Thus, the importance of applying EVT to quantify the risk measure in the form of VaR or CVaR is obvious and we can highly recommend it.

4 Conclusion

The paper deals with quantification of risk measure using Value at Risk methodology on individual insurance claims given by portfolio of motor hull insurance. We presumed the fat tail of probability distribution and therefore we applied Extreme Value Theory. We estimated VaR and CVaR at 99.5% confidence level over one year risk horizon in compliance with Solvency II and we compared estimates of both risk measures in the case of presuming exponential, gamma and Weibull´s distribution with estimates under generalized Pareto distribution assumption within Extreme Value Theory.

On the basis of obtained results, we confirmed the fat tail in empirical probability distribution of our data sample and we also verified it via QQ plots. We revealed that no one presumed theoretical distributions fits the empirical sufficiently. Thus, we computed the VaR and CVaR under assumption of generalized Pareto distribution. We also highlighted the advantage of using EVT over the traditional approach. Thus, neglecting fat tail can lead to the very imprecise and very different Value at Risk estimates resulting in insufficient capital which should cover the loss from holding our insurance portfolio.

Acknowledgements

This paper was solved within the project SP2011/38 Estimating and forecasting individual risk in financial institutions.

References

[1] Alexander, C.: *Value at Risk models*. John Wiley& Sons Inc , Chichester, 2008.
[2] Artzner, P., Delbaen, F., Eber, J.–M., and Heath, D.: *Coherent Measures of Risk. Mathematical Finance* 9, 1999, 203–228.
[3] Artzner, P., Delbaen, F., Eber, J.–M., and Heath, D.: *Thinking Coherently. Extremes and Integrated Risk Management*, P. Embrechts, London, 2000.
[4] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. *Statistics of Extremes: Theory and Applications.* John Wiley& Sons Inc, Chichester, 2004.
[5] Hull, J. C. *Risk management and Financial Institutions*. Pearson Education, New Jersey, 2007.
[6] Jorion, P. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, New York, 2007.
[7] Knobloch, A.P. *Value at Risk: Regulatory and Other Applications, Methods, and Criticism. Risk management*. Springer Verlag, Berlin, 2005, 99-124.
[8] Mikosch, T.: *Non-Life Insurance Mathematics*. Springer Verlag, Berlin, 2009.
[9] Morgan, J.P.: *RiskMetricsTM Technical Document*, New York, 1996.

Fuzzy Model for Determining the Type of Worker

Blanka Zemková[1], Jana Talašová[2], Pavel Holeček[3]

Abstract. In the human resources management, it is necessary to treat each employee individually with regard to his/her abilities and personality traits. Being an effective manager who is able to work with people, one should know the work type of the person and, accordingly, choose an appropriate management strategy. The model described within this is to support classification of various types of workers. It extends the model of employees' evaluation of an IT company. The outputs of the model are evaluations of workers in the three following areas: input (knowledge, skills), process (work behavior) and output (working results). Evaluations in these areas are expressed by values of linguistic fuzzy scales.

The presented model is based on a fuzzy rules base. On the left sides of the rules, there are combinations of scale values representing evaluation of workers in these three areas, on the right sides there are corresponding job types. For each worker and each rule there is calculated the degree to which his/her assessment matches the rule. The aim of the model is to determine the most appropriate work type of the given employee.

Keywords: evaluation, fuzzy sets, human resources, fuzzy rule base.

JEL Classification: C44
AMS Classification: 90C15

1 Introduction

The aim of this paper is to demonstrate possible application of fuzzy sets in human resources management (see also [10] and for comparison e.g. [1]). Other applications of fuzzy sets in behavioural and social science can be found in [7] or [9]. The first chapter summarizes mathematical notions and methods used in the project. The second chapter shows the competency model, the base for the employees' evaluation. Next part describes the process of evaluation, whose output is the evaluation of workers in the particular areas. In the last chapter there is presented the way, how the type of worker is determined on the basis of evaluations calculated before.

2 Applied Notions of the Fuzzy Sets Theory

A **fuzzy set** A in a universal set U is uniquely determined by its membership function $A(.)$, $A: U \rightarrow [0,1]$. Sets $A_\alpha = \{x \in U, A(x) \geq \alpha\}$, $\alpha \in [0,1]$, are called α-cuts of the fuzzy set A, a set $Supp\,A = \{x \in U, A(x) > 0\}$ is a support of the fuzzy set A, and a set $Ker\,A = \{x \in U, A(x) = 1\}$ is called a kernel of the fuzzy set A. The fuzzy set A is called normal if $Ker\,A \neq \emptyset$. The symbol *hgt A* denotes the height of the fuzzy set *A*, and $hgt\ A = \sup\{A(x) \mid x \in X\}$ (see [2]).

A **fuzzy number** C is a fuzzy set defined on the set of real numbers $\Re$ that possesses the following properties: $Ker\,C \neq \emptyset$, $Supp\,C$ is bounded, and α-cuts C_α represent for all $\alpha \in (0, 1]$ closed intervals. Real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$ are called significant values of the fuzzy number C if the following holds: $Cl(Supp\,C) = [c^1, c^4]$, $Ker\,C = [c^2, c^3]$, where $Cl(Supp\,C)$ denotes a closure of $Supp\,C$. A fuzzy number C is said to be defined in $[a,b]$ if $Supp\,C \subseteq [a,b]$.

Fuzzy numbers $T_1, \ldots T_s$ that are defined in the interval $[a,b]$ form a fuzzy partition on the interval $[a,b]$ if $\sum_{i=1}^{s} T_i(x) = 1$ for any $x \in [a,b]$. A **fuzzy scale** on $[a,b]$ is defined as a set of fuzzy numbers $T_1, \ldots T_s$ that are

[1] AXIOM PROVIS Int. s.r.o., Pod Vrškem 5360, 760 01 Zlín, blanka.zemkova@seznam.cz.
[2] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, jana.talasova@upol.cz
[3] Palacký University/Faculty of Science, Dept. of Mathematical Analysis and Applications of Mathematics, 17. listopadu 1192/121, 77146 Olomouc, Czech Republic, pavel.holecek@upol.cz.

defined in the interval $[a,b]$, form a fuzzy partition on this interval, and are numbered according to their order (see [8]).

A **linguistic variable** defined in $[a,b]$ is a quintuple $(V, T(V), [a,b], G, M)$, where V is a name of the variable, $T(V)$ is a set of its linguistic values, $[a,b]$ is a closed interval of real numbers, G is a syntactic rule for generating linguistic values from $T(V)$, and M is a semantic rule that assigns to each linguistic value $A \in T(V)$ its mathematical meaning, a fuzzy number A defined in $[a,b]$.

Let us suppose that $(V, T(V), [a,b], G, M)$, $T(V) = \{T_1, \ldots T_s\}$, is a linguistic variable, and that the meanings of its linguistic values form a fuzzy scale in $[a,b]$. Then we say that the linguistic variable represents a **linguistic fuzzy scale** in $[a,b]$.

The instruments of linguistic fuzzy modelling can be applied for solving a classification problem. Let C = {1,…,k}, k ∈ *N*, be a set of numeric identifiers of the classes of interest. A fuzzy classification system can then be described by means of a fuzzy rule base in the following form:

If F_1 is A_{11} and ... and F_m is A_{1m}, then class D_1.
If F_1 is A_{21} and ... and F_m is A_{2m}, then class D_2.
....
If F_1 is A_{n1} and ... and F_m is A_{nm}, then class D_n,

where for $i = 1,\ldots, n$, $j = 1,\ldots, m$:

- $(F_j, T(F_j), [c_j, d_j], M_j, G_j)$ are linguistic scales for the individual features,
- $A_{ij} \in T(F_j)$ are their linguistic values and $A_{ij} = M_j(A_{ij})$ are fuzzy numbers representing their meanings,
- $D_i \in \{1, \ldots, k\}$ are the class identifiers.

In case of **Voting by Multiple Rules Method,** the classification of an object can be described as follows. Let us suppose that an object is described by the fuzzy numbers $A'_1, \ldots, A'_m$ representing the uncertain values of the characteristics $F_1, \ldots, F_m$. The degrees of correspondence, h_i, $i=1,\ldots,n$, between the inputs and the left-hand sides of the rules are calculated as follows: $h_i = hgt\left((A'_1 \times \ldots \times A'_m) \cap (A_{i1} \times \ldots \times A_{im})\right)$, for $i = 1,\ldots,n$. The number of votes v_T for each of the classes is calculated as the sum of correspondences degrees of rules which voted for this class: $v_T = \sum_{i \in \{1,\ldots,n\}: D_i = T} h_i, \quad T = 1, \ldots, k$. The resulting class is the one with the maximum value of v_T.

3 Competency Model

Although periodic evaluation of employees is an important part of human resources management, in many companies it is perceived rather as an "inevitable evil", by both bosses and subordinates. The evaluation is often based only on measurable characteristics (the number of contracts, the number of invoiced hours etc.). Characteristics that are not directly measurable are often ignored.

In this project, the assessment of an employee is based on multiple-criteria evaluations. The criteria are derived from typical competencies of the employee. For any given working role, a competency model has been created with different weights assigned to different competencies. The weights are normalized, their sum is equal to 1.

Competencies are the summary of knowledge, skills, abilities, and attitudes necessary for personal development and self-assertion of each member of the society. The competency model in terms of human resources management is always created for a given working role. It reflects the combination of competencies that are necessary for carrying out particular type of work. More about the competency models is shown in [4], [5]. The example of the competency model is displayed in Figure 1.

		Senior Executive	Project manager	Analyst	Consultant	Developer	Dealer	Marketing Agent
Input		**0,33**	**0,33**	**0,33**	**0,33**	**0,33**	**0,33**	**0,33**
	Eduaction	0,12	0,18	0,18	0,16	0,11	0,1	0,12
	Experience	0,12	0,18	0,16	0,2	0,12	0,12	0,1
	Certification	0,06	0,08	0,08	0,07	0,08	0,06	0,06
	Language Skills	0,08	0,08	0,04	0,04	0,08	0,18	0,16
	Stress Resistance	0,08	0,12	0,06	0,13	0,08	0,12	0
	Creative Thinking	0,06	0	0,14	0,05	0,04	0,06	0,18
	Customer Orientation	0,08	0,1	0	0,12	0,08	0,16	0,12
	Openness toward Changes	0,08	0,08	0	0	0,08	0,06	0,12
	Learning Skills	0,06	0	0,12	0,08	0,11	0	0
	Willingness to learn	0,08	0	0	0,08	0,11	0	0
	Analytical Thinking	0,04	0	0,16	0,07	0,11	0	0
	Vision and Strategy	0,14	0,18	0,06	0	0	0,14	0,14
	TOTAL	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Output		**0,33**	**0,33**	**0,33**	**0,33**	**0,33**	**0,33**	**0,33**
	Fullfilment of given goals from previous period	0,2	0,4	0,2	0,06	0,25	0,45	0,45
	Quality of Work	0,4	0,3	0,4	0,5	0,4	0,25	0,3
	Perminf One's Tasks inTime	0,4	0,3	0,4	0,44	0,35	0,3	0,25
	TOTAL	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Proces		**0,34**	**0,34**	**0,34**	**0,34**	**0,34**	**0,34**	**0,34**
	Centrality of Work	0,1	0	0,2	0,18	0,2	0,13	0,16
	Interpersonal Sensitivity	0,12	0,28	0,08	0,1	0	0,13	0,08
	Team Work	0,08	0	0,34	0,27	0,4	0	0
	Communication and Influence	0,2	0,2	0,22	0,31	0,3	0,14	0,3
	Integrity	0,2	0,22	0,16	0,14	0,1	0,2	0,26
	Organisational Behaviour, Leadership	0,3	0,3	0	0	0	0,4	0,2
	TOTAL	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Figure 1 Competency model

The competencies are divided into three groups: INPUT (knowledge, skills), OUTPUT (results), and PROCESS (behaviour). The weights of competencies for a particular working role are mathematically represented by the structure of normalized weights.

4 Evaluation Based on Competencies

For the evaluation against the expertly evaluated criteria, the evaluators have at their disposal an uneven six-element fuzzy scale. Using the scale values, the evaluator expresses to what extent, in his/her view, the evaluated employee meets the required level of the given competence.

In the next stage, the partial evaluations against particular competencies are aggregated to form the overall evaluation within the whole competency model. First, the aggregation in groups INPUT, OUTPUT, PROCESS is carried out, followed by the aggregation of these groups. The evaluation tree is defined in accordance with the specified competency model. The evaluation is calculated as a weighted average of the partial fuzzy evaluations against criteria [6]. Evaluation based on the competencies is described in detail in [10]. The similar evaluation problem is solved also in [1].

5 Determining the Type of Worker

For the human resources management, the evaluations in groups INPUT, OUTPUT, and PROCESS are particularly important, since interrelations among these evaluations determine the type of employee, which allows the management to choose the appropriate control strategy.

For determining the worker type a fuzzy rule base was constructed (Figure 2). The fuzzy rule base describes the relation between the aggregated evaluations in the areas of INPUT, OUTPUT and PROCESS on one side, and the worker type on the other side (see Table 1).

For the aggregated evaluations (INPUT, OUTPUT and PROCESS) on the left-hand sides of the fuzzy rules, six-elements linguistic fuzzy scales were used (see Figure 3). On the right sides, there are integers denoting the classes of worker types. To each of the classes, an appropriate strategy for managing the worker is assigned. The classes and the appropriate strategies are described in the Table 1.The structure of the worker types is illustrated by the Figure 5.

For a given worker, the fuzzy evaluations of INPUT, OUTPUT and PROCESS enter the fuzzy classification system (see Figure 4). The degrees of correspondence between input values and the left-hand sides of rules are calculated. For each of the classes, the "number of votes" is calculated as the sum of degrees of correspondence of the fuzzy rules which voted for this class, i.e. the *Voting by Multiple Rules Method* is used. The class with the maximum number of votes is labeled as the winner. If the number of votes for the winner is low, or if victory of the winner is not distinct enough, then the worker is not classified (none of the strategies is applicable for the given worker). For calculation purpose, the FuzzME software was used (see [3]).

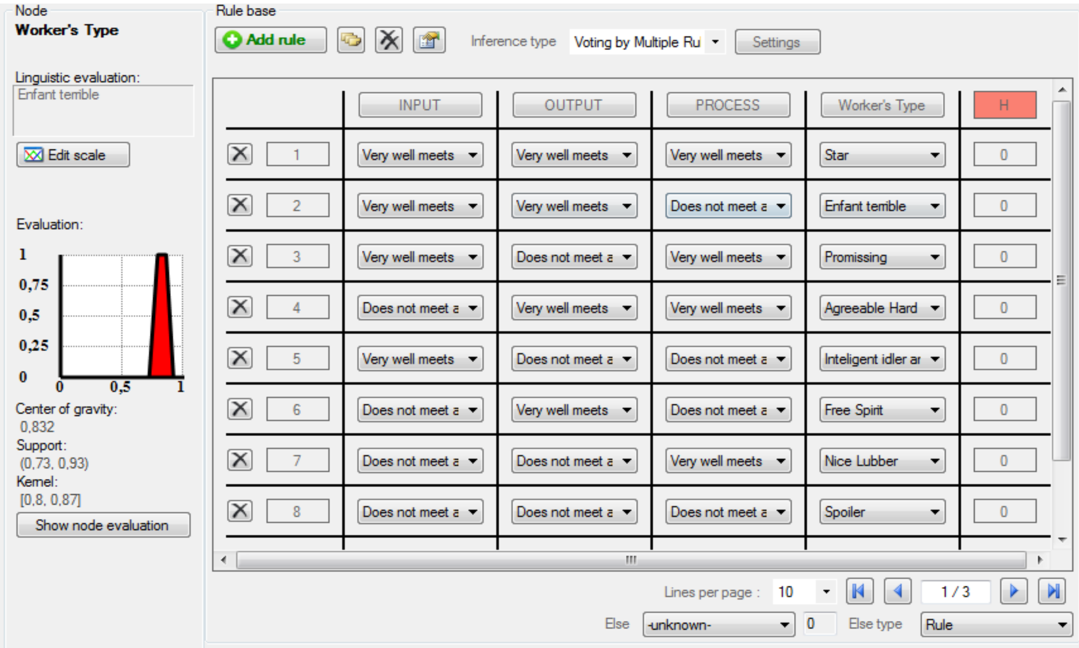


Figure 2 Part of the used fuzzy rule base

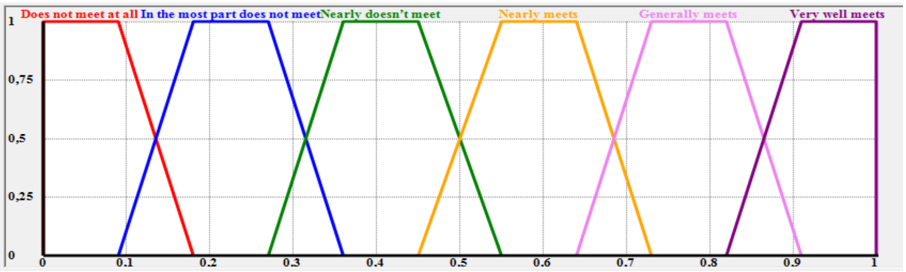


Figure 3 Evaluation scale used for the inputs

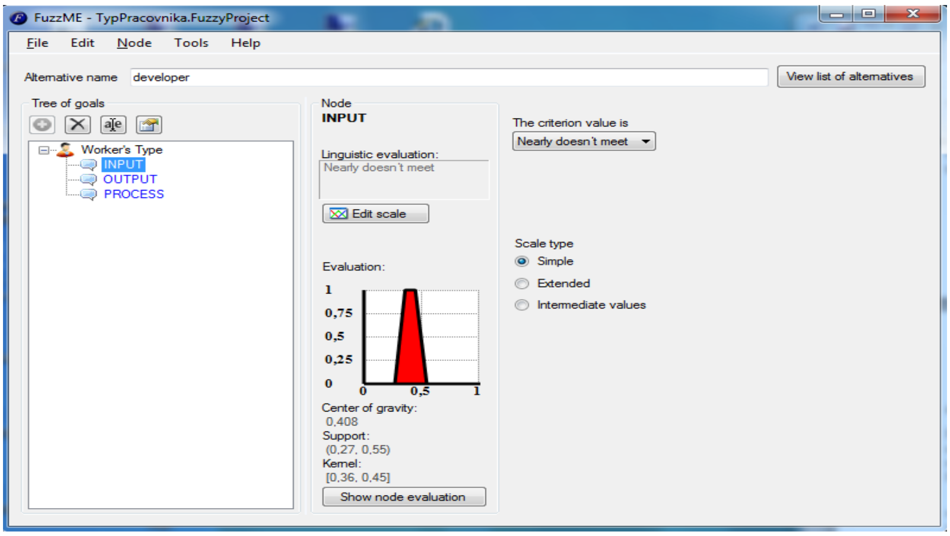


Figure 4 Evaluation in groups

6 Human Resources Management Based on the Type of Worker

The type of worker is given by interrelation among evaluation in groups INPUT, OUTPUT and PROCESS. The figure 5 shows the graphical view.

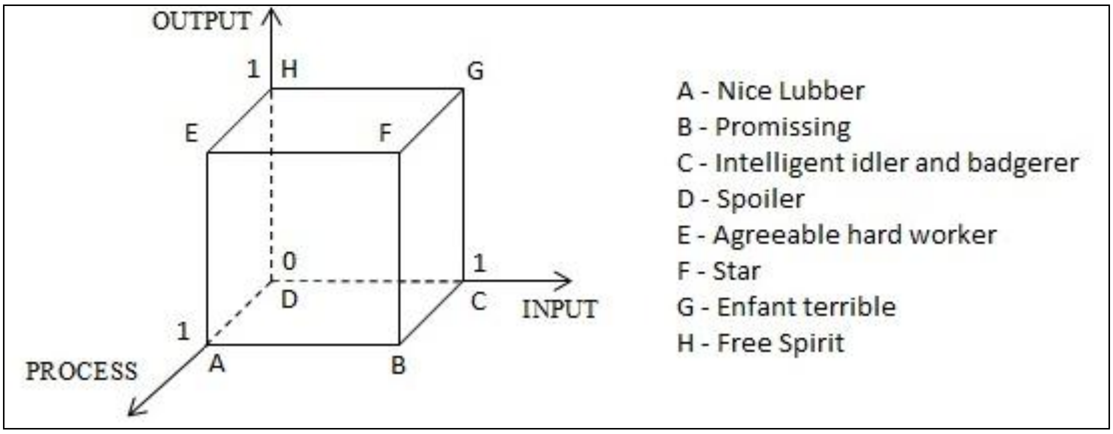


Figure 5 Worker types - graphical view

Management strategy suitable for the job types is shown below, in the table 1 (Hroník, [4]).

OUTPUT	INPUT	PRO-CESS	WORK TYPE	MOTIVATION STRATEGY
1	1	1	Star	He/She should be given more ambitious tasks, his/her informal authority should be encouraged, should be given as an example, should be delegated, promoted
1	1	0	Enfant terrible	should be more involved in group tasks or on the contrary trusted with independent tasks, depending on the personality type; requires consistent, unforgiving approach, not-ignoring any assets; needs acceptance by the others, without being preachy;
1	0	1	Agreeable Hard Worker	to instruct and create conditions for self-education;
0	1	1	Promising	more support, stimulate courage, resilience and self-confidence;
1	0	0	Free Spirit	shape up;
0	1	0	Intelligent idler and a badgerer	limits need to be set as well as deadline for change with clear implications in case of non-compliance; as much feed-back as possible;
0	0	1	Nice Lubber	if tutoring by others (of the *infant terrible* type) fails, consider redeployment;
0	0	0	Spoiler	problematic choice; it is necessary to speak expressly on the possibility of departure; the initiative of improvement is to be left solely on the person in question; deadline for visible improvements needs to be nailed down.

Table 1 Management strategy based on the type of worker

7 Conclusion

The aim of this paper was to demonstrate application of fuzzy sets in human resources management, in this case for determining the type of worker. The advantage of the linguistic fuzzy approach is, on the one hand, mathematical data processing excluding subjective bias, and, on the other hand, natural process of evaluation and natural expression of evaluation results in a language natural for the evaluator.

Acknowledgements

The research was supported by the grant *PrF_2011_022 Mathematical Models and Structures* of Internal Grant Agency of the Palacky University.

References

[1] de Andres R., Garcia-Lapresta J., Martinez L.: *A multi-granular linguistic model for management decision-making in performance appraisal.* Soft Computing - A Fusion of Foundations, Methodologies and Applications. Springer Verlag, Heidelberg 2010, 21-34

[2] Dubois D., Prade H.: *Fundamentals of Fuzzy Sets*. Dordrecht: Kluwer Academic Publishers Group, 2000

[3] *Fuzzy Models of Multiple-Criteria Evaluation*. http://fuzzme.wz.cz (27.11.2010)

[4] Hroník F.: *Hodnocení pracovníků*. Praha, Grada Publishing, a. s., 2006

[5] Kessler R.: *Competency-based Performance Reviews: How to Perform Employeee Evaluations the Fortune 500 Way.* Career Pr Inc., 2008

[6] Pavlačka O., Talašová J.: *Application of the Fuzzy Weighted Average of Fuzzy Numbers in Decision Making Models*. Proceedings of the 5th Conference of the European-Society-for-Fuzzy-Logic-and-Technology, Ostrava 2007, 455 – 462

[7] Smithson M.: *Fuzzy Set Analysis for Behavioral and Social Science*. New York: Springer - Verlag, 1987

[8] Talašová J.: *Fuzzy metody vícekriteriálního hodnocení a rozhodování.* Olomouc: Vydavatelství UP, 2003

[9] Talašová J., Stoklasa J.: *Fuzzy approach to academic staff performance evaluation*. Proceedings of the 28th International Conference on Mathematical Methods in Economics, České Budějovice 2010, 621-626

[10] Zemková B., Talašová J.:, *Application of Fuzzy Sets in Human Resources Management*, Proceedings of the 28th International Conference Mathematical Methods in Economics, České Budějovice, 8th – 10th September 2010, 682 – 687

Investigation of the Real Switch Option Value Sensitivity

Zdeněk Zmeškal[1]

Abstract

Switch options are the generalised methodology of the multi-choice real options. The generalised approach of the switch option value on switch cost is based on fuzzy-stochastic methodology application. The paper includes the switch options value sensitivity on switch cost. The switch option valuation model is based on Bellman stochastic dynamic programming approach and discrete binomial model. Scenario sensitivity approach and fuzzy-stochastic model are applied including illustrative example.

Keywords: Real options; Switch option; Discrete Binomial Model; Stochastic Dynamic Programming; Sensitivity analysis; Fuzzy-stochastic model; Fuzzy number

JEL Classification: C6, C 44, C53, F2, F21, G1, G11, G15, G2, G21
AMS Classification: 91B25, 91G20, 91G50, 91G60, 91G80

1 Introduction

The real options methodology could be considered to be a generalised approach encompassing risk and flexibility aspects simultaneously in a valuation. We can present publications concerning real options, see e. g. Dixit and Pindyck (1994), Sick (1995), Smith and Nau (1995), Trigeorgis (1998), Zmeškal (2001), Brandao and Dyer (2005), Zmeškal (2008), Guthrie (2009), Čulík (2010). Relatively new topic of option valuation are fuzzy-stochastic models, examples of papers are e.g. in Zmeškal (1999), Zmeškal (2001), Simonelli (2001), Yoshida (2002), Carlsson and Fuller (2003), Yoshida (2003), Muzzioli and Torricelli (2004), Wu (2005), Cheng et al. (2005), Liyan Han and Wenli Chen (2006), Zhang et al. (2006), Wu (2007), Muzzioli and Reynaerts (2007) and Thiagarajaha et al. (2007), Guerra et al. (2007), Chrysafis and Papadopoulos (2009), Zmeškal (2010).

Owing to the economic assets types, random processes complexity and decisions variables and functions, the real options are largely of the American options type, discrete binomial (multinomial) models, with multinomial options to switch. Simultaneously, fundamental approach of valuation under complete market is the replication strategy and no-arbitrage principle using the risk-neutral probability and general principle is the martingale approach. Switching options modelling is a generalised approach of real options modelling including other real option models as a special type, e. g. sequential options, learning options, growing options, see e.g. Trigeorgis (1998), Guthrie (2009), Zmeškal (2010). The switching cost is influential input parameter of switch option valuation. A sensitivity of the value and results are the important aspect of the decision-making and analysis.

The intention of the paper is to analyse the possibilities of the sensitivity of the switch options value on the switching cost. The traditional scenario sensitivity analysis approach and fuzzy-stochastic approach is investigated. Methodology and notation is derived and linked to the papers of Zmeškal (2008) and Zmeškal (2010).

2 Valuation procedure of an American option with switching cost

The switch options methodology is described in several papers. Applying notation in Zmeškal (2008) the crucial recurrent valuation formulas are following,

$$V_N^m = \max_{q \in S}\left(x_0^q - C_{m,q} + \beta \cdot \widehat{E}\left(V_{N-1}^q\right)\right), \tag{1}$$

$$V_{N-k}^m = \max_{q \in S}\left(x_k^q - C_{m,q} + \beta \cdot \widehat{E}\left(V_{N-1-k}^q\right)\right), \tag{2}$$

$$V_1^m = \max_{q \in S}\left(x_{N-1}^q - C_{m,q} + \beta \cdot V_0^q\right). \tag{3}$$

Here V_{N-k}^m is value for *N-k* periods to final period and x_k^q is cash flow in the particular period k and mode *q*, $C_{m,q}$ are switch cost from mode *m* to *q*, $\beta_t = (1+R)^{-t}$ is discount factor, $\widehat{E}\left(V_{N-1-k}^q\right)$ is risk-neutral expected value of subsequent mode and *N-k-1*is time to final period.

[1] VSB – Technical University of Ostrava, Faculty of Economics, Department of Finance, Sokolská 33, Ostrava, Zdenek.Zmeskal@vsb.cz.

2.1 Switch option valuation procedure description

Valuation procedure of multinomial options with non-symmetrical switch options reflecting the stochastic dynamic programming on the Bellman's principle expressed by recurrent equations, under the discrete binomial model and risk-neutral probability and geometric Brown motion is possible to describe in following steps.

a) The determination of the risk-neutral growth parameter $\widehat{g}$.
b) Cash flow modelling so as an underlying asset. Subjective approach by virtue of expert estimation and forecast. Objective approach on the basis of statistical estimation and forecasting of random process. In the case of Brown's geometrical process due to Cox, Ross, Rubinstein (1979) calibration,
   $x_{t+1,s+u}^{u} = x_{t,s} \cdot U; \quad x_{t+1,s+d}^{d} = x_t \cdot D$. Here $U = e^{\sigma \cdot \sqrt{dt}}$, $D = e^{-\sigma \cdot \sqrt{dt}}$.
c) At the beginning of the second phase the value for the second phase is $V_{0,s}^{q}$, here s is state and q is mode.
d) By backward recurrent procedure from the end of binomial tree to the beginning for states s and modes q of particular period in accordance with the generalised recurrent Bellman's stochastic equations (1), (2), (3) correspondingly values are calculated. Here $\widehat{p}$ is risk-neutral probability of up movement and $\widehat{q} = 1 - \widehat{p}$ is risk-neutral probability of down movement. Valuation formula for one period to the end of the first phase,
   $$V_{1,s}^{m} = \max_{q \in S}\left(x_{N-1,s}^{q} - C_{m,q} + \beta \cdot V_{0,s}^{q}\right).$$
   Valuation formula for other periods by virtue of the recurrent procedure,
   $$V_{N-k,s}^{m} = \max_{q \in S}\left(x_{k,s}^{q} - C_{m,q} + \beta \cdot \left(\widehat{p} \cdot V_{N-1-k,s+u}^{q} + \widehat{q} \cdot V_{N-1-k,s-d}^{q}\right)\right).$$
   Valuation formula at the beginning of the whole first phase (the first period),
   $$V_{N,s}^{m} = \max_{q \in S}\left(x_{0,s}^{q} - C_{m,q} + \beta \cdot \left(\widehat{p} \cdot V_{N-1,s+u}^{q} + \widehat{q} \cdot V_{N-1,s-d}^{q}\right)\right).$$
e) Identification of the decision variant for state s and time t, $Q_{t,s}$,
   $$Q_{t,s} = \arg\max_{q \in S}\left(x_{k,s}^{q} - C_{m,q} + \beta \cdot \left(\widehat{p} \cdot V_{N-1-k,s+u}^{q} + \widehat{q} \cdot V_{N-1-k,s-d}^{q}\right)\right).$$
f) The implementation of the sensitivity analysis concerning the input data.

3 Fuzzy-stochastic elements

For application of fuzzy-stochastic methodology the crucial terms are fuzzy number, fuzzy operations and decomposition principle. A fuzzy set meeting preconditions of normality, convexity, continuity with the upper semi-continuous membership function and closeness and being defined as the quadruple $\tilde{s} = \left(s^{L},\, s^{U}, s^{\alpha}, s^{\beta}\right)$, where $\phi(\mathrm{x})$ is a non-decreasing function and $\psi(\mathrm{x})$ is a non-increasing function, is called the *T-number*. Let us denote the set of T-numbers on n-dimensional Euclidean space E by $F_T(E)$. *T-number* is defined as follows,

$$\tilde{s} \equiv \mu_{\tilde{s}}(x) = \left\{ \begin{matrix} 0 & for & x \le \mathrm{s}^{\mathrm{L}} - s^{\alpha}; & \phi(\mathrm{x}) & for & \mathrm{s}^{\mathrm{L}} - s^{\alpha} < \mathrm{x} < \mathrm{s}^{L}; \\ 1 & for & \mathrm{s}^{\mathrm{L}} \le x \le s^{U}; & \psi(\mathrm{x}) & for & \mathrm{s}^{\mathrm{U}} < \mathrm{x} < \mathrm{s}^{\mathrm{U}} + s^{\beta}; \\ 0 & for & x \ge \mathrm{s}^{\mathrm{U}} + s^{\beta} & & & \end{matrix} \right\}.$$

The ε -cut of the fuzzy set $\tilde{s}$, depicted $\tilde{s}^{\varepsilon}$, is defined as follows. $\tilde{s}^{\varepsilon} = \{x \in E; \mu_{\tilde{s}}(x) \ge \varepsilon\} = [{}^{-}s^{\varepsilon},\, {}^{+}s^{\varepsilon}]$ where ${}^{-}s^{\varepsilon} = inf\{x \in E; \mu_{\tilde{s}}(x) \ge \varepsilon\}$, ${}^{+}s^{\varepsilon} = sup\{x \in E; \mu_{\tilde{s}}(x) \ge \varepsilon\}$.

Application of the Decomposition principle for a function of fuzzy numbers allows expressing the selected fuzzy operations $\tilde{*}$ among fuzzy numbers directly, as follows: $\tilde{w} = \tilde{s} \,\tilde{*}\, \tilde{r} = \bigcup_{\varepsilon} \varepsilon\left(w^{\varepsilon}\right) = \bigcup_{\varepsilon} \varepsilon\left(s^{\varepsilon} * r^{\varepsilon}\right)$.

Fuzzy addition $\quad s^{\varepsilon} + r^{\varepsilon} = \left\lfloor {}^{-}s^{\varepsilon} + {}^{-}r^{\varepsilon}; {}^{+}s^{\varepsilon} + {}^{+}r^{\varepsilon} \right\rfloor$.

Fuzzy subtract $\quad s^{\varepsilon} - r^{\varepsilon} = \left\lfloor {}^{-}s^{\varepsilon} - {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} - {}^{-}r^{\varepsilon} \right\rfloor$.

Fuzzy scalar product $\quad k \cdot s^{\varepsilon} = \left\lfloor k \cdot {}^{-}s^{\varepsilon}; k \cdot {}^{+}s^{\varepsilon} \right\rfloor$ for $k \ge 0$, $k \cdot s^{\varepsilon} = \left\lfloor k \cdot {}^{+}s^{\varepsilon}; k \cdot {}^{-}s^{\varepsilon} \right\rfloor$ for $k < 0$.

Fuzzy multiplication $\quad s^{\varepsilon} \cdot r^{\varepsilon} = \left\lfloor {}^{-}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon}; {}^{+}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon} \right\rfloor$ for $\tilde{s} > 0,\ \tilde{r} > 0$,
$s^{\varepsilon} \cdot r^{\varepsilon} = \left\lfloor {}^{-}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon} \right\rfloor$ for $\tilde{s} < 0,\ \tilde{r} > 0$, $s^{\varepsilon} \cdot r^{\varepsilon} = \left\lfloor {}^{+}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon}; {}^{-}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon} \right\rfloor$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy division $\quad s^{\varepsilon} : r^{\varepsilon} = \left\lfloor {}^{-}s^{\varepsilon} : {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} : {}^{-}r^{\varepsilon} \right\rfloor$ for $\tilde{s} > 0,\ \tilde{r} > 0$, $s^{\varepsilon} : r^{\varepsilon} = \left\lfloor {}^{+}s^{\varepsilon} : {}^{+}r^{\varepsilon}; {}^{-}s^{\varepsilon} : {}^{-}r^{\varepsilon} \right\rfloor$ for $\tilde{s} < 0,\ \tilde{r} > 0$,
$s^{\varepsilon} : r^{\varepsilon} = \left\lfloor {}^{+}s^{\varepsilon} : {}^{-}r^{\varepsilon}; {}^{-}s^{\varepsilon} : {}^{+}r^{\varepsilon} \right\rfloor$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy max, $\quad \max\left(s^{\varepsilon}\right) = \left\lfloor \max {}^{-}s^{\varepsilon}; \max {}^{+}s^{\varepsilon} \right\rfloor$.

Here $\tilde{s} > 0$ is positive fuzzy number, positive $\tilde{s} : \{x; \text{for which } \mu_{\tilde{s}} \geq 0\}$ and simultaneously $x \in E^{+}$ (set of positive numbers), negative $\tilde{s} : \{x; \text{for which } \mu_{\tilde{s}} \geq 0\}$ and simultaneously $x \in E^{-}$ (set of negative numbers). *Decomposition principle* (Resolution identity) is defined as follows, $\mu_{\tilde{s}}(y) = \sup_{\varepsilon}\{\varepsilon \cdot I_{\tilde{s}^{\varepsilon}}; y \in \tilde{s}^{\varepsilon}\}$ for any $y \in E$ *and* $\varepsilon \in [0;1]$, where $\tilde{s}^{\varepsilon} = [{}^{-}s^{\varepsilon}, {}^{+}s^{\varepsilon}]$ is ε-cut, $\mathrm{y} = \mathrm{f(x)}$, ${}^{-}s^{\varepsilon}(x) = \min_{x \in \tilde{x}^{\varepsilon} \subset E} f(x)$, ${}^{+}s^{\varepsilon}(\mathrm{x}) = \max_{x \in \tilde{x}^{\varepsilon} \subset E^{n}} \mathrm{f(x)}$. Here $I_{\tilde{s}^{\varepsilon}}$ is the characterisation function, $I_{\tilde{s}^{\varepsilon}} = \{1\, if\ \mathrm{y} \in [{}^{-}s^{\varepsilon}, {}^{+}s^{\varepsilon}]; 0\, if\ \mathrm{y} \notin [{}^{-}s^{\varepsilon}, {}^{+}s^{\varepsilon}]\}$.

4 Sensitivity analysis value on switch cost

According to methods and complexity of the sensitivity and influence of switch cost on value of the switch options we can distinguish two basic approaches. The first is traditional *scenario approach* and the second one is *fuzzy-stochastic approach* based on fuzzy numbers. In the scenario approach could be investigated influence of the one factor or more factors simultaneously. In fuzzy-stochastic approach one variable should be given as fuzzy number or more variables could be introduced in the form of fuzzy number.

4.1 One factor scenario sensitivity analysis example

We can express the sensitivity (change) of switch option value on a switching cost, deriving from (1), (2) and (3) as follows. Let the parameter $\alpha \in [-1;+1]$ expresses the level and direction of the sensitivity.

$$\Delta_{\alpha} V_{N}^{m} = \max\left(x_{0}^{m-1} - \alpha \cdot C_{m,m-1} + \beta \cdot \hat{E}\left(V_{N-1}^{m=1}\right); x_{0}^{m} + \beta \cdot \hat{E}\left(V_{N-1}^{m}\right)\right), \tag{4}$$

$$\Delta_{\alpha} V_{N-k}^{m} = \max\left(x_{k}^{m-1} - \alpha \cdot C_{m,m-1} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m-1}\right); x_{k}^{m} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m}\right)\right), \tag{5}$$

$$V_{1}^{m} = \max\left(x_{N-1}^{m} + \beta \cdot V_{0}^{m}\right) \tag{6}$$

4.2 Fuzzy-stochastic sensitivity analysis example

We can generally express, applying the decomposition principle, ε-cut of fuzzy number $\tilde{f}_{t}$, composed from two fuzzy numbers $\tilde{G}_{t}$ and $\tilde{H}_{t}$ by relation fuzzy maximization, as follows,

$\left[f_{t}^{\varepsilon}\right] = \max\left[G_{t}^{\varepsilon}; H_{t}^{\varepsilon}\right] = \max\left[\left({}^{-}G_{t}^{\varepsilon}; {}^{-}H_{t}^{\varepsilon}\right); \left({}^{+}G_{t}^{\varepsilon}; {}^{+}H_{t}^{\varepsilon}\right)\right] = \left[\max\left({}^{-}G_{t}^{\varepsilon}; {}^{-}H_{t}^{\varepsilon}\right); \max\left({}^{+}G_{t}^{\varepsilon}; {}^{+}H_{t}^{\varepsilon}\right)\right] = \left[{}^{-}f_{t}^{\varepsilon}; {}^{+}f_{t}^{\varepsilon}\right]$.

Here ${}^{-}f_{t}^{\varepsilon} = \max\left({}^{-}G_{t}^{\varepsilon}; {}^{-}H_{t}^{\varepsilon}\right)$, ${}^{+}f_{t}^{\varepsilon} = \max\left({}^{+}G_{t}^{\varepsilon}; {}^{+}H_{t}^{\varepsilon}\right)$. And ${}^{-}G_{t}^{\varepsilon} = \min G_{t}^{\varepsilon}$, ${}^{-}H_{t}^{\varepsilon} = \min H_{t}^{\varepsilon}$, furthermore ${}^{+}G_{t}^{\varepsilon} = \max G_{t}^{\varepsilon}$, ${}^{+}H_{t}^{\varepsilon} = \max H_{t}^{\varepsilon}$.

Applying previous common fuzzy maximization relation with fuzzy switch cost, we can formulate following recurrent formulas,

$$\left(V_{N}^{m}\right)^{\varepsilon} = \left[{}^{-}\left(V_{N}^{m}\right)^{\varepsilon}; {}^{+}\left(V_{N}^{m}\right)^{\varepsilon}\right], \tag{7}$$

where ${}^{-}\left(V_{N}^{m}\right)^{\varepsilon} = \max\left[x_{0}^{m-1} - {}^{+}\left(C_{m,m-1}\right)^{\varepsilon} + \beta \cdot \hat{E}\left(V_{N-1}^{m=1}\right); x_{0}^{m} + \beta \cdot \hat{E}\left(V_{N-1}^{m}\right)\right]$ and

$${}^{+}\left(V_{N}^{m}\right)^{\varepsilon} = \max\left[x_{0}^{m-1} - {}^{-}\left(C_{m,m-1}\right)^{\varepsilon} + \beta \cdot \hat{E}\left(V_{N-1}^{m=1}\right); x_{0}^{m} + \beta \cdot \hat{E}\left(V_{N-1}^{m}\right)\right];$$

$$\left(V_{N-k}^{m}\right)^{\varepsilon} = \left[{}^{-}\left(V_{N-k}^{m}\right)^{\varepsilon}; {}^{+}\left(V_{N-k}^{m}\right)^{\varepsilon}\right], \tag{8}$$

where ${}^{-}\left(V_{N-k}^{m}\right)^{\varepsilon} = \left[x_{k}^{m-1} - {}^{+}\left(C_{m,m-1}\right)^{\varepsilon} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m-1}\right); x_{k}^{m} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m}\right)\right]$ and

$${}^{+}\left(V_{N-k}^{m}\right)^{\varepsilon} = \max\left[x_{k}^{m-1} - {}^{-}\left(C_{m,m-1}\right)^{\varepsilon} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m-1}\right); x_{k}^{m} + \beta \cdot \hat{E}\left(V_{N-1-k}^{m}\right)\right];$$

$$\left(V_{1}^{m}\right)^{\varepsilon} = \left[{}^{-}\left(V_{1}^{m}\right)^{\varepsilon}; {}^{+}\left(V_{1}^{m}\right)^{\varepsilon}\right] = \left[\left(x_{N-1}^{m} + \beta \cdot V_{0}^{m}\right); \left(x_{N-1}^{m} + \beta \cdot V_{0}^{m}\right)\right] \tag{9}$$

5 Illustrative example of sensitivity value calculation

The switch option value is to be calculated for three variants: Variant 1 – Mode A (normal production), Variant 2 – Mode B (expansion production), Variant 3 – Mode C (contraction production). The binomial model, American option; non-symmetrical switching cost; replication value strategy, risk-neutral approach; expected present value objective function will be employed. The applied model is of two-phase one type, the first phase with the random cash flow takes 4 years, and the second one a non-random phase is of perpetuity version. We suppose that random cash flow (underlying asset) follows geometric Brown process. We analyse two variants, scenario one and fuzzy-stochastic one.

Input data of the model are following: risk-free rate r = 10%; up-movement index U = 1,2; $D = (1{,}2)^{-1}$; the value for the beginning of the second phase $V_{0,s}^{q}$ for the states s and modes q, see Fig. 1. The price of underlying asset S is 10, rate of contraction or expansion is 15 %. Risk-neutral probability of up movement is $\widehat{p} = 72{,}73\%$ and down movement is $\widehat{q} = 1 - \widehat{p} = 27{,}27\%$. Tab. 1 shows switch costs C_{ij} connected with switching among particular modes. The keeping the same mode is linked with no switch cost, of course. Positive values means cash outflow and negative ones cash inflow. Non-possibility of switching between modes is expressed by penalty value, ∞. It is apparent that switching costs are non-symmetrical.

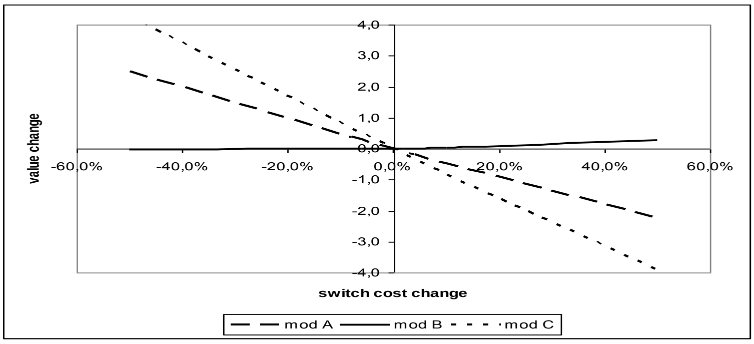
Switch cost $C_{i,j}$		**Subsequent mode *j***		
		A	**B**	**C**
Initial mode *i*	**A**	0	5	-4
	B	-3	0	∞
	C	4	∞	0

Tab. 1 Switch costs between modus

Computation procedure in coincidence with the methodology described above is on Fig. 1. There are three columns showing procedure for three initial states: Mode A (normal production), Mode B (expansion production) and Mode C (contraction production). In the rows are four steps of procedure: (I) cash flow tree, (II) pre-calculation (present expected value), (III) maximal value, (IV) assigned mode. Resulting value of mode A is 73,424 c. u, mode B 78, 424 c. u. and mode C 68,379 c. u..

5.1 One factor scenario sensitivity analysis

Switch cost scenario sensitivity results due to Chapter 4.1 are on Graph 1. The most influenced is the mode C, subsequent is the mode A, and the mode B does not significantly influence the value.



Graph 1 Value sensitivity on switch cost quantity for particular mode

5.2 Fuzzy-stochastic sensitivity analysis

Fuzzy-stochastic sensitivity analysis according to Chapter 4.2 is used. The fuzzy switch costs are given in the Tab. 2. in the form of the T-number $\widetilde{s} = \left(s^{L}, s^{U}, s^{\alpha}, s^{\beta}\right)$.

Switch cost $C_{i,j}$		**Subsequent mode *j***		
		A	**B**	**C**
Initial mode *i*	**A**	0	4;6;2;1	-5;-3;1;2
	B	-4;-2;1;1	0	∞
	C	3;5;2;1	∞	0

Tab. 2 Fuzzy switch costs between mode (T-numbers)

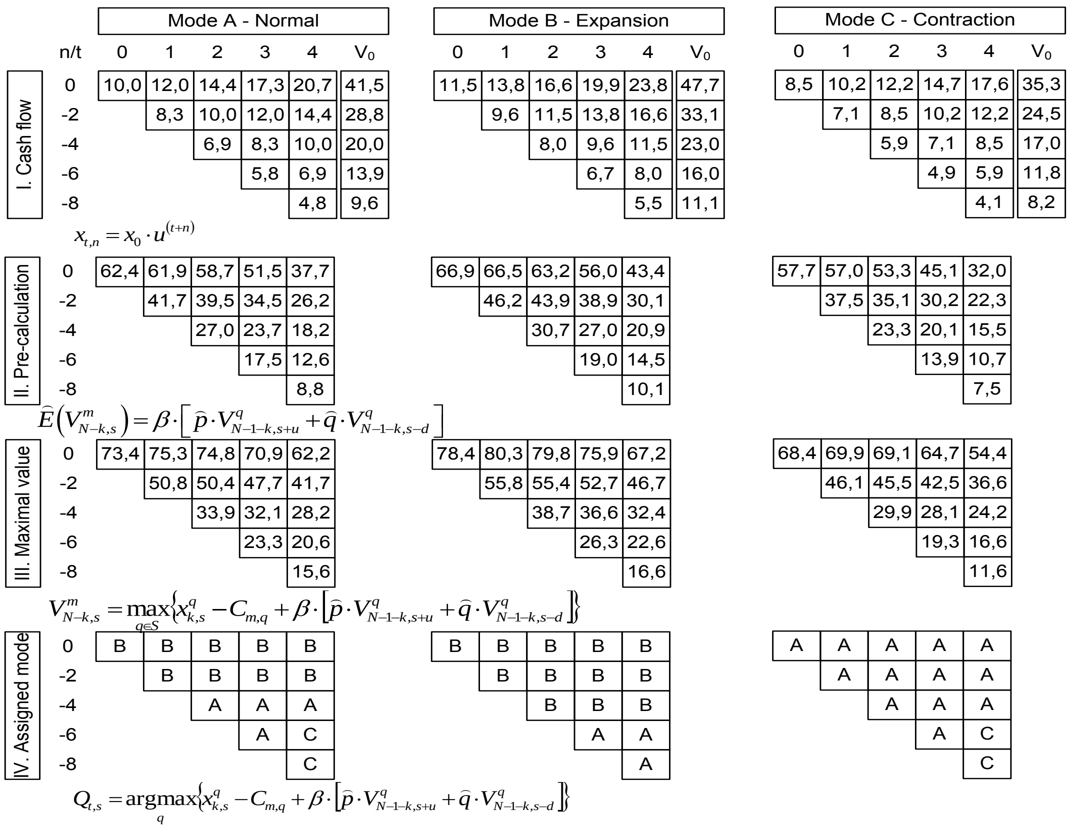


Fig. 1 Valuation procedure of dynamic flexible multinomial switch options methodology

epsilon	**A - mode**		**B - mode**		**C - mode**	
	min	**max**	**min**	**max**	**min**	**max**
1	72,409	74,484	78,409	78,484	66,455	70,347
0,75	71,909	74,830	78,409	78,580	65,500	70,921
0,5	71,409	75,279	78,409	78,779	65,045	71,608
0,25	70,909	76,350	78,409	79,389	63,648	72,389
0	70,409	77,653	78,409	80,252	63,817	73,252

Tab. 3 Fuzzy value of switch option

6 Conclusions

Two approaches of the sensitivity analysis of switch options on switch cost were investigated: the traditional scenario one and innovative fuzzy-stochastic one. The ranking of the sensitivity of modes A, B and C were in the example similar. However the fuzzy-stochastic methodology could more deeply show the sensitivity dimensions and better catch the non-linearity, non-symmetry of switch cost and impreciseness as well. Therefore, we can conclude, that the fuzzy-stochastic approach could be generalised sensitivity methodology.

References

[1] Brandao, L. E., Dyer, J. S.: Decision Analysis and Real Options: A Discrete Time Approach to Real Option Valuation. *Annals of Operations Research* **135** (2005), p. 21.

[2] Carlsson, C., Fuller, R.: A fuzzy approach to real option valuation. *Fuzzy Sets and Systems* **139** (2003), pp. 297-312.

[3] Cheng Few-Lee, Tzeng Gwo-Hsiung, Wang Shin-Yun,: A Fuzzy Set Approach for Generalized CRR Model: An Empirical Analysis of S&P 500 Index Options. *Review of Quantitative Finance and Accounting*, **25** (2005) , pp. 255–275.

[4] Chrysafis, K. A., Papadopoulos, B. K. : On theoretical pricing of options with fuzzy estimators. Journal of Computational and Applied Mathematics**. 223/ 2**(2009) , pp. 552-566

[5] Čulík, M.: Valuing of real options portfolio and its influence on project value. In: *22nd International Conference on Mathematical Methods in Economics*. 2004, pp. 56-62.

[6] Čulík, M.: Real option application for modular project valuation. In: *24th International Conference on Mathematical Methods in Economics*. 2006, pp. 123-130.
[7] Čulík, M.: Emissions trading: investment decision-making and real option analysis (simulation approach). 4th International Scientific Conference on Managing and Modelling of Financial Risk, Ostrava. 2008, pp 11-18
[8] Čulík, M.: Flexibility and project value: interactions and multiple real options. In. 3rd Global Conference on Power Control and Optimization, Gold Coast, AUSTRALIA. POWER CONTROL AND OPTIMIZATION Book Series: AIP Conference Proceedings . 2010 Volume: 1239, pp 326-334
[9] Dixit, A. K., Pindyck, R.S.: *Investment under Uncertainty*. Princeton University Press, 1994.
[10] Dluhošová, D.: An analysis of financial performance using the EVA method. *Finance a úvěr- Czech Journal of Economics and Finance* **11-12** (2004).
[11] Guerra, M. L., Sorini, L., Stefanini, L.: Parametrized fuzzy numbers for option pricing. IEEE International Conference on Fuzzy Systems, London, 2007, VOLS 1-4, pp 727-732.
[12] Guthrie, G.: *Real Options in Theory and Practice*. Oxford University Press, 2009.
[13] Liyan Han, Wenli Chen: The Generalization of λ-Fuzzy Measures with Application to the Fuzzy Option. In: Wang, L. et al. (Eds.) Proceedings: Fuzzy Systems and Knowledge Discovery, 2006, LNAI 4223, pp. 762–765, Springer-Verlag Berlin Heidelberg.
[14] Muzzioli, S., Reynaerts, H.: The solution of fuzzy linear systems by non-linear programming: a financial application. *European Journal of Operational Research* **177** (2007), pp 1218–1231
[15] Muzzioli, S., Torricelli, C.: A multi-period binomial model for pricing options in a vague world. *Journal of Economic Dynamics and Control* **28** (2004), pp. 861-887.
[16] Smith, J. E., Nau, R. F.: Valuing risky projects: Option pricing theory and decision analysis. *Management Science* **14** (1995), p. 795.
[17] Thiagarajaha, K., Appadoob, S.S., Thavaneswaranc, A.: Option valuation model with adaptive fuzzy numbers, *Computers and Mathematics with Applications* **53** (2007), pp. 831–841.
[18] Trigeorgis, L.: *Real Options - Managerial Flexibility and Strategy in Resource Allocation*. Harvard University, 1998.
[19] Wu, H.-C.: Using fuzzy sets theory and Black–Scholes formula to generate pricing boundaries of European options. Applied Mathematics and Computation, **185/ 1** (2007) , pp. 136-146.
[20] Yoshida, Y.: A discrete-time model of American put option in an uncertain environment. *European Journal of Operational Research* **151** (2002), pp 153-166.
[21] Yoshida, Y.: The valuation of European options in uncertain environment. *European Journal of Operational Research* **145** (2003), pp. 221-229
[22] Zhang Jinliang, Du Huibin, Tang Wansheng: Pricing R&D option with combining randomness and fuzziness. Computational intelligence, PT 2, Proceedings Book Series: Lecture notes in artificial intelligence, 2006, Volume: 4114, pp 798-808
[23] Zmeškal, Z.: Fuzzy-stochastic estimation of a firm value as a call option. *Finance a úvěr - Czech Journal of Economics and Finance* **3** (1999), pp. 168-175 .
[24] Zmeškal, Z.: Application of the fuzzy - stochastic Methodology to Appraising the Firm Value as a European Call Option. *European Journal of Operational Research* **135/2** (2001), pp. 303-310.
[25] Zmeškal, Z.: Hedging strategies and financial risks. *Finance a Úvěr - Czech Journal of Economics and Finance* **54** (2004), pp. 50-63.
[26] Zmeškal, Z.: Approach to Real Option Model Application on Soft Binomial Basis Fuzzy - stochastic approach. In: *23rd International Conference on Mathematical Methods in Economics*. 2005, pp. 433-439.
[27] Zmeškal, Z.: Real option applications based on the generalised multinomial flexible switch options methodology. In: *24th International Conference on Mathematical Methods in Economics*. 2006, pp. 545-553.
[28] Zmeškal, Z.: Application of the American Real Flexible Switch Options Methodology A Generalized Approach. *Finance a Úvěr - Czech Journal of Economics and Finance* **58** (2008), pp. 261-275.
[29] Zmeškal, Z.: Generalised soft binomial American real option pricing model (fuzzy–stochastic approach). *European Journal of Operational Research*. **207/ 2** (2010) , pp 1096-1103. doi:10.1016/j.ejor.2010.05.045.
[30] Zmeškal, Z.: Real compound sequential and learning options modelling. In: *28th International Conference on Mathematical Methods in Economics*. 2010, pp. 688-693.

Simultaneous Scheduling of Material Handling and Production Processes: Classification and Heuristics for Parallel Machine Problems

Jan Zouhar[1]

Abstract. The paper deals with optimal scheduling of operations in a production system. Roughly speaking, these operations can be divided into manufacturing and logistical ones. The traditional OR approaches analyze the two types separately: the wide class of vehicle routing problems deals with the logistical issues, while the scheduling theory focuses on the tasks carried out by manufacturing machines. However, in many today's manufacturing systems, especially those using automatically guided vehicles (AGVs) to transport work-in-progress to manufacturing machines, the logistical and manufacturing processes are tightly interconnected, and thus have to be coordinated simultaneously.

In the paper, we focus on simultaneous scheduling of material handling and manufacturing tasks in production systems with multiple AGVs and parallel manufacturing machines. First, we present a classification of these systems and provide a measure of their complexity in terms of the number of feasible schedules. Next, we provide a mixed-integer program to minimize the total makespan, and suggest a computationally tractable heuristic.

Keywords: scheduling theory, parallel machine problems, transportation, automatically guided vehicles.

JEL classification: M11, L23
AMS classification: 90B35, 90B06

1 Introduction

Since the 1950s, scheduling theory has been devoted to finding optimal schedules of machines in various types of manufacturing systems. The traditional modeling frameworks, such as the various shop problems (flow shops, job shops etc.) focus solely on the operation of manufacturing machines and neglect the material transportation issues. As Abdelmaguid and Nassef [1] argue, this "is unjustified, since the capacities of the material handling equipment [...] are limited, and the transportation times of parts are dependent on their routings which differ considerably."

It is not that the traditional OR techniques are not well-suited to optimize material handling processes: various types of VRP and VRPTW heuristics can be applied to find an efficient routing of transportation vehicles, once the schedule of the manufacturing machines is known (see [5], [6], [8]). However, for a manufacturing system to be efficient, the scheduling of material handling and production processes needs to be done *simultaneously.* Ganesharajah et al [4] identify the inefficiencies that result from addressing both issues in isolation, suggesting the need for integration.

Since the late 1990s, integrated scheduling of manufacturing and material handling machines has attracted several researchers' attention. One of the reasons was a technology change: as noted in [3], in real manufacturing systems, automatically guided vehicles (AGVs) came to a widespread use for work-in-progress transportation. Obviously, the use AGVs required that explicit schedules were set up, which in turn created a natural demand for heuristics that devise these schedules in an efficient way. Most of the extant heuristics are either a direct extension of the solutions designed for shop problems (e.g. [1]) or an application of meta-heuristic techniques such as genetic algorithms (e.g. [7]).

[1]University of Economics, Prague, Dept. of Econometrics, Nám. W. Churchilla 4, Praha 3, 130 67, e-mail: zouharj@vse.cz

Apart from this line of research, there's another classical OR problem that is connected with material handling in manufacturing systems: the quadratic assignment problem (QAP). Rather than scheduling the machines in a manufacturing system with a fixed layout, the goal in QAP is to find an efficient facility layout that would help cut down on the amount of time needed for material handling. In this paper, we aim at adding the facility-layout element to the task of simultaneous scheduling of material handling and manufacturing processes. In the resulting optimization problem, the following decisions are made:

- where to store the materials, components or unfinished parts (henceforward collectively termed *work-in-progress*) needed to perform the individual manufacturing tasks,
- how to schedule the manufacturing tasks,
- how to route the material handling machines (henceforward referred to as *AGVs*).

For simplicity, we restrict ourselves only to the parallel-machine mode of operation in the manufacturing part, and to single-load AGVs (i.e., at a given time, an AGV can only carry the work-in-progress of a single manufacturing task).

2 Problem description and classification

The main elements of the manufacturing system can be organized into the following sets:

Tasks The set of all manufacturing tasks that need to be carried out. With each task, a specific work-in-progress item is connected.

Locations The set of locations where work-in-progress can be stored. For simplicity, we assume that each location can accomodate a single work-in-progress item, and therefore, there are as many locations as tasks.

Processors The set of manufacturing machines. Machines are identical and in a parallel mode of operation: a task can be processed by any processor, with an invariant processing time.

AGVs The set of all AGVs that transfer the work-in-progress from locations to processors. Like processors, the AGVs are treated as identical (all can handle any task's work-in-progress and operate with equal transport speeds). As we noted before, we consider single-load AGVs only. Therefore, the route of an AGV has to be an alternating sequence of locations and processors – apart from its beginning and ending: all AGVs are assumed to start and finish in a common depot. Figure 1 shows a possible route of an AGV in the manufacturing system (L's are locations, P's are processors); the dashed arrow that goes back to the depot indicates the fact that this path is not counted in the total makespan of the system in our model (see below).

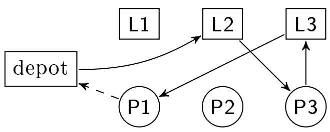


Figure 1: A possible route of an AGV.

The objective of the optimization problem is to find a feasible schedule that minimizes the total makespan, i.e. the time elapsed before all tasks are completed. A feasible schedule must make sure that all tasks are processed, no tasks overlap on a processor (and no preemption of the tasks is possible), and before the processing of a task starts, its work-in-progress has to be delivered an unloaded at the particular processor; moreover, the processor must finish the preceding task before the unloading starts.

We need the following data to describe the system:

- *task durations* (i.e., the processor time needed to complete each task),
- *task loading/unloading times* (we assume that for each task, the loading time equals the unloading time),
- *travel times* between locations, processors, and the depot.

Furthermore, we distinguish two different modes of operations of the AGVs: they can be either *independent* or *coupled* with processors. In the latter case, there as many AGVs as processors, and each AGV serves a particular processor (i.e., delivers the work-in-progress to a single processor only). Figure 2 depicts the various ways a task can pass through a system with 3 processors and 3 AGVs that are either coupled (on the left) or independent (right); A's stand for AGVs, P's for processors.

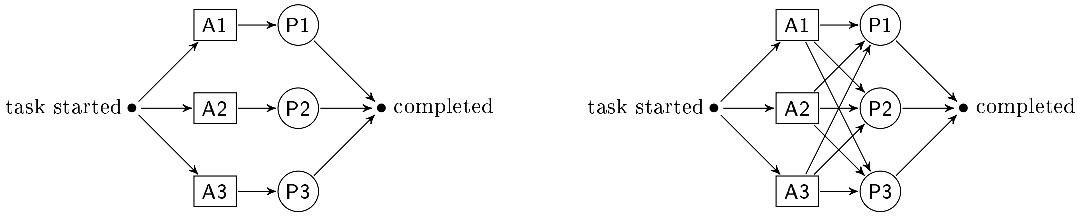


Figure 2: Task scheduling options with *coupled* (left) and *independent* (right) AGVs.

For the sake of comparison, we also consider manufacturing systems with a fixed layout, i.e. systems where the tasks have already been assigned to individual locations; these systems then become a special case of the systems analyzed in [1] and [4]. We'll refer to the two alternatives as systems with *fixed* and *assignable* locations, respectively.

The two modes of AGV operation and two location modes give us four different classes of the system. To distiguish between them, we'll use the notation AGV_mode | location_mode, where the modes are indicated with their first letters: e.g., C|F denotes a system with coupled AGVs and fixed locations.

It is worth noting that the scheduling problem in any of these classes is NP-hard; this comes immediately from the fact that the classical scheduling problem with parallel machines $P||C_{max}$, which is itself NP-hard (see [2]), is easily reduced to the scheduling problem in any of the four classes by setting the travel and loading/unloading times to zero. (Similarly, we can reduce TSP to any of the four problems.)

Still, it is obvious that some classes are easier to schedule than others: for instance, scheduling a problem with fixed locations (say, C|F) is evidently less complicated than scheduling its counterpart with assignable locations (C|A). A measure of system complexity that recognizes this is the *number of feasible schedules*. Let P denote the number of processors, A the number of AGVs, and T the number of tasks; then, the number of feasible schedules for each of the four classes can be expressed as

$$\begin{aligned}
\mathsf{C|F}: &\quad \textstyle\prod_{i=1}^{P}(P+T-i),\\
\mathsf{C|A}: &\quad P!\cdot\textstyle\prod_{i=1}^{P}(P+T-i),\\
\mathsf{I|F}: &\quad \textstyle\prod_{i=1}^{P}(P+T-i)\cdot\textstyle\prod_{j=1}^{A}(A+T-j),\\
\mathsf{I|A}: &\quad P!\cdot\textstyle\prod_{i=1}^{P}(P+T-i)\cdot\textstyle\prod_{i=j}^{A}(A+T-j).
\end{aligned} \tag{1}$$

Table 1 shows the dramatic difference in the number of schedules even for small-scale problems [the P, A, T notation is the same as in (1)].

P, A, T	C|F	C|A	I|F	I|A
2,2,2	6	12	36	72
2,2,3	24	144	576	3,456
3,3,4	360	8,640	129,600	3,110,400
3,3,5	2,520	302,400	6,350,400	762,048,000

Table 1: # of feasible schedules (examples)

3 MIP formulation

In this section, we formulate a linear mixed-integer program that finds the optimal schedule for the I|A class (the remaining classes can be viewed as special cases of the I|A class, and their MIP formulations are easily derived from the I|A's one). For the sake of the mathematical model, we introduce a set of fictitious tasks that represent the start of each processor: $\mathsf{S_P} = \{\textsc{start}_k : k \in \mathsf{Processors}\}$, a set of fictitious tasks that represent the start of each AGV: $\mathsf{S_A} = \{\textsc{start}_k : k \in \mathsf{AGVs}\}$, and a set containing another fictitious task that represents the end of operation of all machines (processors and AGVs alike): $\mathsf{F} = \{\textsc{finish}\}$. By

D, we denote the depot. Table 2 shows the description of all input data and variables used in the MIP formulation, and the respective notation. All data regarding the fictitious tasks are treated as zeros. The MIP formulation that minimizes the total makespan in the I|A system is given below.

		Subscript range	Value	Description
Data:	δ_i	$i \in \mathsf{Tasks} \cup \mathsf{S_P}$	Non-negative	Task duration
	λ_i	$i \in \mathsf{Tasks} \cup \mathsf{F}$	Non-negative	Loading/unloading time
	τ_{kl}	$k \in \mathsf{Processors} \cup \{D\},\ l \in \mathsf{Locations}$	Non-negative	Travel time $k \leftrightarrow l$
Variables:	s_i	$i \in \mathsf{Tasks} \cup \mathsf{S_P} \cup \mathsf{S_A} \cup \mathsf{F}$	Non-negative	Processing start time
	u_{ik}	$i \in \mathsf{Tasks},\ k \in \mathsf{Processors}$	Binary	$=1$ if i processed by k
	v_{jl}	$j \in \mathsf{Tasks},\ l \in \mathsf{Locations}$	Binary	$=1$ if j stored at l
	x_{ij}	$i \in \mathsf{Tasks} \cup \mathsf{S_P},\ i \neq j \in \mathsf{Tasks} \cup \mathsf{F}$	Binary	$=1$ if i followed by j on a processor
	y_{ij}	$i \in \mathsf{Tasks} \cup \mathsf{S_A},\ i \neq j \in \mathsf{Tasks} \cup \mathsf{F}$	Binary	$=1$ if i followed by j on an AGV

Table 2: Data and variables

$$\begin{aligned}
&\text{minimize } s_{\text{FINISH}} \\
&\text{subject to} \quad \textstyle\sum_{k\in\mathsf{Processors}} u_{ik} = 1, && i \in \mathsf{Tasks}, && \text{(2a)}\\
&\textstyle\sum_{l\in\mathsf{Locations}} v_{il} = 1, && i \in \mathsf{Tasks}, && \text{(2b)}\\
&\textstyle\sum_{i\in\mathsf{Tasks}} v_{il} = 1, && l \in \mathsf{Locations}, && \text{(2c)}\\
&\textstyle\sum_{i\in\mathsf{Tasks}\cup\mathsf{S_P},\, i\neq j} x_{ij} = 1, && j \in \mathsf{Tasks}, && \text{(2d)}\\
&\textstyle\sum_{j\in\mathsf{Tasks}\cup\mathsf{F},\, j\neq i} x_{ij} = 1, && i \in \mathsf{Tasks}\cup\mathsf{S_P}, && \text{(2e)}\\
&\textstyle\sum_{i\in\mathsf{Tasks}\cup\mathsf{S_A},\, i\neq j} y_{ij} = 1, && j \in \mathsf{Tasks}, && \text{(2f)}\\
&\textstyle\sum_{j\in\mathsf{Tasks}\cup\mathsf{F},\, j\neq i} y_{ij} = 1, && i \in \mathsf{Tasks}\cup\mathsf{S_A}, && \text{(2g)}\\
&s_i + \delta_i + \lambda_j \le s_j + M(1 - x_{ij}), && i \in \mathsf{Tasks}\cup\mathsf{S_P},\ j \in \mathsf{Tasks}\cup\mathsf{F} && \text{(2h)}\\
&s_i + 2\lambda_j + \textstyle\sum_{k\in\mathsf{Processors}} \tau_{kl}(u_{ik} + u_{jk}) \le s_j + M(2 - v_{jl} - y_{ij}), && i, j \in \mathsf{Tasks},\ l \in \mathsf{Locations}, && \text{(2i)}\\
&2\lambda_j + \tau_{Dl} + \textstyle\sum_{k\in\mathsf{Processors}} \tau_{kl} u_{jk} \le s_j + M(2 - v_{jl} - \textstyle\sum_{i\in\mathsf{S_A}} y_{ij}), && j \in \mathsf{Tasks},\ l \in \mathsf{Locations}, && \text{(2j)}\\
&u_{ik} \le u_{jk} + 1 - x_{ij}, && i \in \mathsf{Tasks}\cup\mathsf{S_P},\ j \in \mathsf{Tasks},\ k \in \mathsf{Processors}. && \text{(2k)}
\end{aligned}$$

We omitted the variable specification part from the mathematical model, as it is contained in Table 2; M denotes a high enough (prohibitive) number – the sum of all processing, loading and travel times will suffice. Let us briefly comment on the individual constraints. (2a) through (2g) are the obvious assignment constraints that follow immediately from the description of binary variables in Table 2. (2h) requires that if task i is followed by task j on a processor, the start of j must come after i is finished and j is unloaded from an AGV. If j is stored at l and preceded by i on an AGV, then (2i) makes sure that j is transported to its processor and unloaded before its processing starts; (2j) is the same kind of constraint, only for a task j that comes first on an AGV (having no predecessors, meaning that the AGV starts in the depot, D). (2k) requires that if task i is processed by processor k and followed by task j, then j has to be processed by k as well.

4 A heuristic for the C|A system

It is worth noting that for any but the C|F class, the MIP solution turns out to be computationally intractable even for small-scale problems – hence the need for heuristics. As the caption of this section suggests, we only deal with the C|A version of the problem here. According to our preliminary experiments, the heuristic seems to perform fairly well; however, furhter experimentation is needed to provide more conclusive results regarding its properties. The basic idea of the heuristic is to decompose the scheduling problem into three steps:

Step 1: Distribute tasks among the processors (and AGVs).

Step 2: Determine the order of tasks on each processor (and AGV).

Step 3: Assign the work-in-progress to individual locations.

Even though these three scheduling operations are carried out successively, the aim of the heuristic design in to "think ahead" in the first two steps. In the following discussion, we use the same notation for data and variables as in the previous section.

Step 1. If we forget about work-in-progress for a while, we're left with a standard $P||C_{max}$ scheduling problem: tasks with duration $\delta_i + \lambda_i$ are assigned to parallel processors, minimizing the total makespan. Note that the eventual order of tasks on each processor is irrelevant here; the makespan is fully determined by the assignment variables u_{ik}. The problem is a generalized version of the bottleneck assignment problem, and is NP-hard; for smaller-scale problems, however, its mixed-integer program is still computationally tractable. For medium- to large-scale problems, efficient heuristics have been developed, see [2]. Any of these will do, yielding the values of u_{ik}. For further discussion, let us denote by Tasks_k the set of tasks assigned to processor k.

Step 2. When we neglected the material handling processes in step 1, the ordering of tasks on a processor didn't matter as far as the makespan on the particular processor was concerned. In step 2, the aim is to establish the ordering of tasks on each processor in such a way that the chances of the processor having to wait for work-in-progress are minimized. In order to leave as much time for material handling as possible, we apply the following heuristic rules:

(i) The task with the lowest value of $\delta_i + \lambda_i$ comes last, and the task with the lowest λ_i comes first.

(ii) The remaining tasks are scheduled in such a way that the minimum of $t_i - v_j$ for successive tasks i and j is maximized.

Note that in (ii), the problem resembles a bottleneck TSP; the solution can either be found using MIP or heuristics. As the number of tasks assigned to a particular processor is not expected to be large in practical applications, we opted for the MIP solution. The procedure that finds an ordering on processor k according to (i) and (ii) can be formalized as follows:

Procedure *OrderProc*(k):

$\text{FIRST} \leftarrow \arg\min_{i \in \mathsf{Tasks}_k} (\delta_i + \lambda_i)$;

$\text{LAST} \leftarrow \arg\min_{i \in \mathsf{Tasks}_k \setminus \{\text{FIRST}\}} \lambda_i$;

$x_{\text{START}_k,\,\text{FIRST}} \leftarrow 1$;

$x_{\text{LAST},\,\text{FINISH}} \leftarrow 1$;

solve maximize T

$$\begin{aligned}
\text{subject to } \textstyle\sum_{i \in \mathsf{Tasks} \setminus \{\text{LAST}, j\}} x_{ij} &= 1, && j \in \mathsf{Tasks}_k \setminus \{\text{FIRST}\},\\
\textstyle\sum_{j \in \mathsf{Tasks} \setminus \{\text{FIRST}, i\}} x_{ij} &= 1, && i \in \mathsf{Tasks}_k \setminus \{\text{LAST}\},\\
\textstyle\sum_{i \in \mathsf{Tasks}_k \setminus \{\text{FIRST}\}} x_{i,\text{FIRST}} &= 0, &&\\
\textstyle\sum_{j \in \mathsf{Tasks}_k \setminus \{\text{LAST}\}} x_{\text{LAST},j} &= 0, &&\\
t_i - v_j + M(1 - x_{ij}) &\geq T, && i \in \mathsf{Tasks}_k \setminus \{\text{LAST}\}, j \in \mathsf{Tasks}_k \setminus \{\text{FIRST}\},\\
z_i + 1 - M(1 - x_{ij}) &\leq z_j, && i \in \mathsf{Tasks}_k \setminus \{\text{LAST}\}, j \in \mathsf{Tasks}_k \setminus \{\text{FIRST}\},\\
z_i &\geq 0, && i \in \mathsf{Tasks}_k,\\
T &\geq 0;
\end{aligned}$$

Step 3. After steps 1 and 2, the shedule is nearly complete: the only remaining decision regards the assignment of the tasks' work-in-progress to individual locations. For this, we need to express the impact of work-in-progress assignment on the makespan at the individual processors. If task i is followed by task j on a processor, two operations are running simultaneously: the delivery of j's work-in-progress and the processing of i. It is the longer of the two which influences the processor's makespan; therefore, we define

$$\gamma_{jl} = \max\left\{2\lambda_i + \textstyle\sum_{k \in \mathsf{Processors}} u_{jk}\tau_{kl}\,,\ \lambda_j + \textstyle\sum_{i \in \mathsf{Tasks}} x_{ij}\delta_i\right\}, \quad j \in \mathsf{Tasks},\ l \in \mathsf{Locations}, \tag{3}$$

and γ_{jl} measures the impact of assigning task j to location l on the makespan of j's processor. If we denote by LAST_k the last task scheduled on processor k, we can express the total makespan at k as:

$$\delta_{\text{LAST}_k} + \textstyle\sum_{j \in \mathsf{Tasks}} \sum_{l \in \mathsf{Locations}} u_{jk} c_{jl} v_{jl},$$

where u_{jk} and c_{jl} are now given parameters, and v_{jl} are variables determining the assignment of tasks to locations (as in Table 2). Therefore, in order to minimize the total makespan, we solve:

$$\begin{aligned} &\text{minimize } T \\ &\text{subject to} \quad \delta_{\text{LAST}_k} + \textstyle\sum_{j\in\mathsf{Tasks}} \sum_{l\in\mathsf{Locations}} u_{jk} c_{jl} v_{jl} \le T, && k \in \mathsf{Processors}, \\ &\textstyle\sum_{j\in\mathsf{Tasks}} v_{jl} = 1, && l \in \mathsf{Locations}, \\ &\textstyle\sum_{l\in\mathsf{Locations}} v_{jl} = 1, && j \in \mathsf{Tasks}, \\ &T \ge 0. \end{aligned} \tag{4}$$

To sum up, the three steps outlined in the beginning of the section are carried out as follows:

Step 1: Solve $P||C_{max}$ with $\delta_i + \lambda_i$ task durations to find the values of u_{ik}.
Step 2: For each processor k, run *OrderProc*(k), which gives the values of x_{ij} (and $y_{ij} = x_{ij}$).
Step 3: Calculate (3) and solve (4) to find the values of v_{jl}.

5 Conclusion and future research

In this paper, we presented a scheduling problem that integrates the decisions about material handling and production processes. In comparison to the extant studies, we added the option to adjust the facility layout. We formulated the problem of makespan minimization in the form of a linear mixed-integer program. As the MIP solution is intractable even for small-scale problems, we provided a heuristic; although the efficiency of the heuristic is still being tested, the preliminary results seem to be quite satisfactory. A proper testing of the heuristic and its possible refinements are the subject of an ongoing research. Another line of research aims at extending the results to a wider range of manufacturing systems. In our study, we restricted ourselves to single-load AGVs and parallel machine problems only; in future, we would like to provide heuristics that cover the flow shop and job shop sheduling problems with multiple-load AGVs.

Acknowledgements

This research was supported by the Internal Grant Agency of the University of Economics, Prague, project no. F4/18/2011.

References

[1] Abdelmaguid, T., Nassef, A.: A constructive heuristic for the integrated scheduling of machines and multiple-load material handling equipment in job shops, *The International Journal of Advanced Manufacturing Technology*, **46** (2010), 1239–1251.

[2] Brucker, P.: *Scheduling Algorithms.* Springer, New York, 2007.

[3] Chung, J., Tanchoco, J. M A.: Material handling automation in production and warehouse systems, *Springer Handbook of Automation* (S. Y. Nof, ed.), Springer Berlin Heidelberg, 2009, 961–979.

[4] Ganesharajah, T., Hall, N. G., Sriskandarajah, C.: Design and operational issues in AGV-served manufacturing systems, *Annals of Operations Research*, **76** (1998), 109–154.

[5] Ioannou, G., Kritikos, M.: Optimization of material handling in production and warehousing facilities, *Operational Research*, **4** (2004), 317–331.

[6] Pelikán, J.: A Time Limited Vehicle Routing Problem, *Communications*, **3** (2006), 9–11.

[7] Reddy, B., Rao, C.: A hybrid multi-objective GA for simultaneous scheduling of machines and AGVs in FMS, *The International Journal of Advanced Manufacturing Technology*, **31** (2006), 602–613.

[8] Zanjirani Farahani, R., Laporte, G., Miandoabchi, E., Bina, S.: Designing efficient methods for the tandem AGV network design problem using tabu search and genetic algorithm, *The International Journal of Advanced Manufacturing Technology*, **36** (2008), 996–1009.

Model for Assessment of the Social Economic Level of Municipalities

Miroslav Žižka[1]

Abstract. The purpose of this contribution is to propose a methodology for assessment of social economic disparities on the level of municipalities. The data base for analysis of municipalities is consequently derived from official data by the Czech Statistical Office and represents 33 indicators for the basic group of 6,249 municipalities within the Czech Republic. Explanatory factor analysis was used to reduce the number of variables. The suitability of factor analysis was first assessed on the basis of a correlation analysis, a Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity. Factors were extracted using various methods. A suitable method of extraction was selected on the basis of analysis of communality. Seven factors characterizing the social economic level of the municipality were defined following comparison of the extraction results. From the aspect of factor loadings these factors were interpreted as follows: civic amenities, unemployment, economic activity, domicile attractiveness, population settlement, age structure of inhabitants and branch structure.

Keywords: Explanatory factor analysis, factor loadings, municipality, civic amenities, unemployment, economic activity, domicile attractiveness, population settlement, age structure, branch structure

JEL Classification: C38, O18, R15
AMS Classification: 62H25, 62J15

Introduction

The issue of assessment of the performance of regions, or more precisely establishment of regional disparities, is very current. The differences between the social and economic level in countries and regions can be measured using various indicators. The gross domestic product per capita indicator is used most frequently. However, this indicator is only available on the level of larger territorial units. It is not available for assessing smaller territorial units (micro-regions, administrative districts of municipalities) and so different indicators must be used.

Various indicators used for quantification of social economic disparities within regions can be found in literature. Labudová et al. [3, p. 6] gives the unemployment rate, average income per person, the aforementioned GDP level or the poverty indicator for example. Martinčík [4, p. 16] recommends a set of 18 indicators divided into three groups – macro-economic performance, growth potential and quality of life – for assessment of the economic-social level of regions. According to Uramová and Kožiak [9, p. 8] among the most important and most frequently monitored indicators are gross domestic product per capita, employment and unemployment, dynamics of inflow of foreign direct investment and the level and development of wages.

The common denominator of all the aforementioned approaches is the focus on measuring the social economic level of regions and higher territorial units. The goal of this paper is to propose a model for assessing municipalities, as a basic building block for all regions. For this reason it was necessary to find indicators that are monitored on the level of municipalities. Because, in general, it applies that the volume of available statistical data and the periodicity of its monitoring is falls along with the size of territorial unit. The first analysis on this subject was performed in 2007. This analysis was based on 26 indicators, which were subsequently reduced to eight factors. The methodology of this research is described in sections [2], [5] and [6]. The intent was to create a methodology that would be sustainable in the long-term, and easily updated, i.e. would be based as much as possible on official data by the Czech Statistical Office (CZSO). Unfortunately this intention was only partially fulfilled. The reason for this was changes to the methodology and the periodicity of monitoring data about municipalities. For this reason, following a consultation with the CZSO, changes were made and only those indicators, monitoring of which it is assumed is sustainable in the long-term, were newly selected. At the same time indicators from the census in 2001, which are no longer current, were eliminated. This paper represents the results of the reviewed analysis of municipal data and compares them to the results achieved in 2007.

[1] Technical University of Liberec, Faculty of Economics, Department of Business Administration, Voroněžská 13, CZ 461 17 Liberec, e-mail: miroslav.zizka@tul.cz

1 Research Methodology

The goal of the analysis was to execute a repeated investigation of the social economic level of municipalities within the Czech Republic. With regard to the possibility of comparing results with those from 2007, which was the date of the preceding analysis, the intention was to use the greatest possible number of identical indicators from the preceding investigation. The list of indicators used in 2007 is given in the work [6]. During the first phase original indicators were checked. On the basis of a consultation with the CZSO, indicators based on the census in 2001 were eliminated first of all, because their predicative ability is now quite low with regard to their age. At the same time these indicators were replaced with others, which are regularly monitored by the CZSO and enable implication of the originally monitored phenomenon. For instance, the original indicators “proportion of economically active inhabitants to the total number of inhabitants” was replaced with the indicator titled “proportion of employees to the total number of inhabitants.“ The unemployment rate indicator in the primary, secondary and tertiary sector was replaced with the “proportion of active subjects in agriculture, industry and services to the total number of active economic subjects“, etc. Indicators that the CZSO no longer monitors or that cannot be considered sufficiently reliable (see [2, pp. 102-104] for more detail) were also eliminated. This chiefly concerns availability of civic amenities in municipalities such as technical infrastructure (connection to the water mains, canalization with wastewater treatment plant, gas mains). An additional local investigation is essential for these indicators.

To prevent random deviation, particularly with regard to small municipalities, sum totals for the last five years (natural and migration increase/decrease in numbers of inhabitants, divorce rate, numbers of completed apartments) were used for selected indicators. Data on pre-primary and primary educational facilities and data on voter turnout in parliamentary elections has newly been included in the analysis. Most of the data dates from 2009, with the exception of numbers of employees, where the last available data is from 2008, numbers of secondary school facilities dating from 2005 and data on voter turnout, which dates from 2010. Following the update the initial set of indicators numbers 33 (the original set from 2007 numbered 26 indicators, and 10 of these indicators were used in the new analysis), see Table 1.

The basic set of 33 indicators for all 6,249 municipalities in the Czech Republic was analyzed. Steps of the analysis:

1. Preparation of the data set for the explanatory factor analysis. The data of all indicators were pasted into the MS Excel 2007 file.
2. The data were imported into the statistical programmes STATGRAPHICS and OpenStat.
3. The suitability of use of factor analysis was verified in several phases. During the first phase Pearson’s and partial correlation coefficients were determined and relations between individual variables were assessed. In the second phase the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) and Bartlett’s test of sphericity were calculated. The variables with the low KMO values (KMO < 0.5) were eliminated from the analysis.
4. Making decisions on the number of factors. Several methods were used to extract factors (see Table 2). A suitable extraction method was selected on the basis of communality analysis. The number of factors was determined on the basis of Kaiser’s rule taking into account the percentage of explained variance and the progress of the curve of eigenvalues.
5. Factor rotation was performed using the Varimax method.
6. Evaluation of factor loadings. If a factor loading is high (above 0.3), then the relevant variable helps to describe that factor quite well.
7. Identification and interpretation of factors resulting from the factor loadings.

2 Analysis results

The correlation analysis indicated that 59% of pair correlations are statistically important on a significance level of 95%. Weak dependences ($r < 0.3$) were established in 14 variables: OHUZA09, HMROZ09, MRNEZ09, AMNEZ09, DANPO09, PZAMO08, ESPOP09, SUZPR09, KEKST09, POMSP09, OPRVU09, PDBRD09, POBPL09 and HOPEN09.

The suitability of use of factory analysis was initially assessed using the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO). According [1, pp. 224-225], KMO is based on an index which compares correlation and partial correlation coefficients. KMO statistics take values between 0 and 1. High values are obtained if the sum of the correlation coefficients are relatively large compared to the sum of the partial correlation coefficients. In this case there are likely to be patterns of correlation in the data indicating that a factor analysis might be an appropriate technique to use The Kaiser-Meyer-Olkin measure for the examined correlation matrix achieves the value of 0.70. According to Kaiser (see for instance [1, p. 225]) the suitability of use of factor analysis in this case can be rated as middling. For individual variables the KMO ranges between 0.17 and 0.87, during which time a value of less than 0.5 is considered unsatisfactory. The low KMO value concerned the following variables: AMNEZ09 (0.29), DANPO09 (0.17), SUZPR09 (0.32), KEKST09 (0.44), POMSP09 (0.37) and PDBRD09 (0.46). Therefore these 6 variables were eliminated from further analysis. The recalculated KMO value for the whole matrix following this elimination was 0.75 and all KMO values for individual variables were satisfactory (higher than 0.5).

Bartlett's test of sphericity was also performed (see Stewart [8, p. 57] for more detail). The hypothesis tested is that the correlation matrix came from a population of variables that are independent. Rejection of the hypothesis is an indication that the data are appropriate for factor analysis. The calculated significance level (0.001) is lower than the assigned (0.05) therefore we reject the zero hypothesis. The condition for use of factor analysis is consequently fulfilled.

Code/year	**Description of indicator**
OHUZA09	general population density
SHUZA09	specific population density
KEKST09	proportion of ecologically favourable areas to areas burdening the environment
PVEKC09	average age of living inhabitants in total (men and women)
ISTAR09	age index
IPMPR09	intensity of natural and migration increment of inhabitants during the last 5 years (2005-2009)
IEZOB09	economic burden on inhabitants index according to the number of inhabitants in pre-productive (0-14 years) and post-productive (65+) age to the average number of inhabitants of productive age
HMROZ09	gross divorce rate; number of divorces 2005-2009 per average number of inhabitants in 2005-2009
MRNEZ09	registered unemployment rate in job applicants total on 31/12/2009 (%)
MDNEZ09	long-term unemployment rate on 31/12/2009 (%)
MTLPM09	job pressure rate on 31/12/2009 (%)
AMNEZ09	proportion of graduate and youthful job applicants compared to the total number of job applicants
DANPO09	tax yields per 1 inhabitant
PZAMO08	proportion of employees to the total number of inhabitants (average number on 31/12/2009)
ESPOP09	proportion of active economic subjects to population aged between 15-64 years
SPODN09	proportion of private entrepreneurs to population aged 15-64 years
SUZEM09	proportion of subjects active in agriculture/forestry to the total number of active economic subjects
SUZPR09	proportion of active subjects in industry to the total number of active economic subjects
SUSLU09	proportion of active subjects in services to total number of active economic subjects
POMSP09	proportion of small and medium enterprises to total number of enterprises
OPRVU09	proportion of self-employed individuals (without employees) to total number of economic subjects
DBYTY09	number of completed apartments during the period of 2005-2009 per 1,000 inhabitants
PDBRD09	number of completed apartments in family houses during period of 2005-2009 to total number of completed apartments during period of 2005-2009
POBPL09	average floor space in one apartment in family or apartment houses during period of 2005-2009
KAHUZ09	capacity of mass accommodation facilities (number of beds) per one thousand inhabitants
HOPEN09	proportion of beds in hotels and guesthouses to total mass accommodation facilities capacity
PRLEK09	number of inhabitants per general practitioner's office for adults, children and youth
AMBUL09	number of inhabitants per one outpatient medical facility
PFARM09	number of chemists per 1,000 inhabitants
MATSK09	number of inhabitants aged between 3-5 years per one nursery school
ZAKSK09	number of inhabitants aged 6-14 years per one primary school
STRSK05	number of inhabitants aged 15-19 years per one secondary school
VOLBY10	voter turnout (%) during elections to the Chamber of Deputies of the CZ Parliament

Table 1 List of indicators used for evaluation of municipalities

Table 2 gives communality values after extraction of factors according to individual methods. The number of factors was determined on the basis of Kaiser's rule (the factor eigenvalue must be higher than one) taking into account the percentage of explained variance (requirement of at least 60%). Table 2 indicates that variable variability is best explained during use of the principal component method (approximately two thirds). The best result was achieved for variable MRNEZ09 (nearly 91% explained variability), variability is explained most poorly for variable ISTAR09 (approx. 40%). According to Kaiser's rule, 9 factors should be extracted using the principal component method, which explain a total of approx. 69% of variance. However, the differences between eigenvalues of factors with sequence numbers 6 to 10 are very small. According to Rummel [7, pp. 363-364] the eigenvalue-one criterion may discriminate between factors that have little difference in eigenvalues. This small variance difference between the factors appears hardly meaningful, yet one factor is retained and the other dropped. In such cases it is recommended that the number of factors is determined on the basis of the progress of the curve of eigenvalues. When the last meaningful or substantively important factor is extracted, the eigenvalues will show a discontinuity – a sharper drop than for adjacent factor. Factors subsequent to the discontinuity will then have a fairly constant slope for their eigenvalues, implying that they mainly extract random error.

This discontinuity occurred between the fifth and sixth factor. This means that it would be sufficient to extract only five factors, which would only explain 52% of the variance. Numbers of extracted factors using other methods, including the percentage of explained variance, are given in Table 2. The final decision on the number of factors will be made after assessment of factor loadings.

Factor rotation was performed using the Varimax method. Table 3 gives the results for extraction of rotated factors during use of the abovementioned methods. The table only gives variables for which the absolute value of factor loadings was higher than 0.3.

It is clear from Table 3 that the results of extraction using various methods are similar to a certain degree. The first factor is chiefly saturated by indicators that are related to the civic amenities available in the municipality, including medical, educational and accommodation facilities. This factor is also indirectly proportionately affected by the specific population density, which can be considered logical because the occurrence of all these facilities is dependent on the size, or number of inhabitants in the municipality. The factor can be named **civic amenities.**

Type of Analysis	**Principal Components**	**Partial Image**	**Guttman Image**	**Harris Scaled Image**	**Maximum Likelihood**	**Principal Axis Factors**
PVEKC09	79.36	67.83	40.51	51.95	66.41	80.24
ISTAR09	40.01	3.17	2.62	2.62	3.10	3.03
IPMPR09	78.05	48.97	36.51	32.24	54.55	60.20
IEZOB09	51.54	23.32	18.28	18.73	24.41	23.64
OHUZA09	86.61	7.43	2.14	6.60	30.47	9.73
SHUZA09	81.73	54.30	33.05	43.19	65.95	68.33
HMROZ09	63.21	2.93	2.44	2.40	2.95	2.75
MRNEZ09	90.97	89.63	78.59	80.65	87.89	96.56
MDNEZ09	70.39	51.02	48.34	47.45	51.69	51.61
MTLPP09	85.53	82.59	78.82	77.43	84.35	84.35
PZAMO08	53.52	8.33	7.08	6.69	7.26	8.71
ESPOP09	85.38	83.48	68.04	78.22	86.06	86.58
SPODN09	85.54	78.92	66.68	76.78	84.61	82.15
SUSLU09	68.36	39.46	20.10	30.18	38.21	42.84
OPRVU09	68.03	41.70	21.44	33.90	44.21	46.23
DBYTY09	75.68	46.20	32.90	31.94	51.56	56.47
POBPL09	55.10	20.17	16.49	16.53	19.58	20.92
KAHUZ09	66.64	16.44	9.38	10.69	12.08	21.02
HOPEN09	55.03	28.61	21.63	21.19	23.94	32.49
SUZEM09	61.11	35.55	22.04	27.43	33.78	38.47
PRLEK09	82.01	79.61	76.76	80.43	87.04	79.26
AMBUL09	80.29	77.51	76.02	79.34	86.14	77.10
PFARM09	43.74	26.90	24.48	23.79	25.63	27.31
VOLBY10	62.52	44.21	28.85	35.81	45.95	47.17
MATSK09	56.26	36.54	33.28	32.01	35.61	39.07
ZAKSK09	74.51	66.01	62.30	63.36	67.61	66.49
STRSK09	49.82	32.90	28.30	26.61	33.90	35.60
Number of factors	9	4	4	5	6	5
Cumulative % of variance	68.55	39.19	34.96	36.50	44.08	45.64

Table 2 Communality Estimates as Percentages and Number of Extracted Factors

The second factor is affected by indicators related to unemployment. A significant negative correlation between this factor and voter turnout was also established for five methods of extraction. This may indicate that voter turnout has a specific link to unemployment in the municipality. Municipalities with a high unemployment rate usually have a lower voter turnout than municipalities with a low unemployment rate. The second factor was identified as **unemployment.**

The third factor was defined similarly using all methods. It is fulfilled by indicators that reflect the proportion of private entrepreneurs and active economic subjects to the population in the municipality. The capacity of mass accommodation facilities in the municipality is also reflected here, which is also related to economic activity within the municipality. This is why this factor was called **economic activity.**

The fourth factor is fulfilled using all methods by the indicators of intensity of natural and migration increment in inhabitants, the average age of inhabitants and the number of complete apartments per 1,000 inhabitants. Apart from one method the indicator of average floor area per apartment is also important. Consequently this concerns the characteristics of residential development and demographic indicators. To a measure these characteristics are related to how attractive the environment of the municipality is for the life of local inhabitants. For this reason this factor was called **domicile attractiveness.**

Factors with subsequent sequence numbers are more difficult to interpret. At the same time greater differences occurred here in the individual methods of extraction used. The greatest accord occurred in three methods in the case of the general and specific population density indicators. Therefore the fifth factor was called **population settlement.**

Extract. M. Factor	**Principal Components**	**Partial Image**	**Guttman Image**	**Harris Scaled Image**	**Maximum Likelihood**	**Principal Axis Factors**
F1	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 POBPL09 *SHUZA09*	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 SUZEM09 *SUSLU09* *SHUZA09*	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 SUZEM09 *SHUZA09*	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 *SHUZA09*	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 *SHUZA09*	HOPEN09 PRLEK09 AMBUL09 *PFARM09* MATSK09 ZAKSK09 STRSK09 SUZEM09 *SHUZA09*
F2	MDNEZ09 MTLPP09 MRNEZ09 *VOLBY10*	MDNEZ09 MTLPP09 MRNEZ09 *VOLBY10*	MDNEZ09 MTLPP09 MRNEZ09 *VOLBY10*	MDNEZ09 MTLPP09 MRNEZ09 *VOLBY10*	MDNEZ09 MTLPP09 MRNEZ09	MDNEZ09 MTLPP09 MRNEZ09 *VOLBY10*
F3	SPODN09 KAHUZ09 ESPOP09	SPODN09 KAHUZ09 ESPOP09	SPODN09 KAHUZ09 ESPOP09	SPODN09 KAHUZ09 ESPOP09	SPODN09 KAHUZ09 ESPOP09	SPODN09 KAHUZ09 ESPOP09
F4	IPMPR09 *PVEKC09* DBYTY09 *POBPL09*	*IPMPR09* PVEKC09 *DBYTY09* POBPL09 SUZEM09 *SUSLU09* *SHUZA09*	IPMPR09 *PVEKC09* DBYTY09 *POBPL09* *SUZEM09* SUSLU09 SHUZA09	*IPMPR09* PVEKC09 *DBYTY09* POBPL09 SUZEM09 *SUSLU09*	IPMPR09 *PVEKC09* DBYTY09 SUSLU09	*IPMPR09* PVEKC09 *DBYTY09* POBPL09 IEZOB09 SUZEM09 *SUSLU09*
F5	MATSK09 PVEKC09 POBPL09 ISTAR09 IEZOB09			OHUZA09 SHUZA09	MATSK09 PVEK09 POBPL09 IEZOB09 VOLBY10 SUZEM09	OPRVU09 PZAM08
F6	*SUZEM09* SUSLU09				OHUZA09 SHUZA09	
F7	OHUZA09 SHUZA09					
F8	*KAHUZ09* *VOLBY10* HMROZ09 OPRVU09					
F9	PZAM08					

Comment: Negative factor loadings are written in italics.

Table 3 Summary of Extracted Factors

In the case of the Principal Components and Maximum Likelihood methods a sixth factor can be defined, which is saturated identically to indicators MATSK09, PVEK09, POBPL09 and IEZOB09. These indicators are particularly affected by the **age structure** of the inhabitants of the municipality. For this reason this factor was also interpreted as such.

The Principal Components and Principal Axis Factors methods were used to define the factor that reflects the proportion of employees to the total number of inhabitants and the proportion of small entrepreneurs without employees to the total number of economic subjects. This concerns a factor that characterizes the economic activity of inhabitants. With regard to the fact that similar phenomena were already defined in the third factor, it does not seem rational to define another individual factor separately.

A similar decision was made in the case of factor F8 defined using the Principal Components method. This factor is made up of divorce rate, minor business activities and indirectly proportionately by voter turnout and capacity of accommodation facilities indicators. Interpretation of such a factor is considerably difficult, in general it can be surmised that it is linked to the relationship between inhabitants and their site of residence.

The last factor extracted using the Principal Components method is fulfilled using indicators that characterize the branch structure of economic subjects within the municipality (the proportion of subjects in agriculture and in services to the total number of active economic subjects). This can be called the **branch structure** factor.

Conclusion

On the basis of assessment of factor loadings it seems optimum to define seven factors, which describe the social economic level of the municipality. The extracted factors explain approx. 61% variance in original data (during use of the Principal Components method). If we compare the results of the previous analysis realised in 2007, we discover that the number of factors is one less. Factors can be interpreted identically, even though the database used in the new analysis was more extensive and with regard to changes in the methodology of monitoring some indicators by the CZSO, it was not wholly identical. From the methodological aspect only 10 indicators were identical. In the newly performed analysis the factor of sustainable development was not confirmed as being statistically significant. However, this factor was also problematic in the previous survey and was difficult to interpret in practice (see article [10]), therefore its elimination can be accepted. Repeated investigation also confirmed that the proposed model for assessment of the social economic level of a municipality is also sustainable in the long-term from the aspect of updating the database. This demonstrates the fact that the monitored phenomena were selected correctly. At the same time the results of the analysis must be considered support for assessment of individual municipalities. The results of analysis of official statistical data provide a basic picture of the state of the municipality, but some phenomena are difficult to establish or cannot be established at all using “hard“statistical data. Consequently the results of the factor analysis must be supplemented by local additional investigation of some phenomena.

This paper is supported by a project of the Ministry for Regional Development – WD-30-07-1 Innovation approach to analysis of disparities on regional level.

References

[1] Hutcheson, G.D., and Sofroniou, N.: *The Multivariate Social Scientist.* SAGE Publications, London, 1999.
[2] Jáč, I. et al.: *The Uniqueness of a Municipality in the Region.* Professional Publishing, Prague, 2010.
[3] Labudová, V., Linda, B., and Vojtková, M.: Application of Multidimensional Methods to Measure Poverty. *E & M Ekonomie a Management* **13** (2010), 6-22.
[4] Martinčík, D.: Regional socio-economic level - Comprehensive comparative analysis. *E & M Ekonomie a Management* **11** (2008), 14-25.
[5] Prskavcová, M. and Řehořová, P.: Methodology of Questionnaire Survey of Economic Growth of Municipalities in the Czech Republic. *E & M Ekonomie a Management* **11** (2008), 77-83.
[6] Rydvalová, P. and Žižka, M.: *Competitiveness and Uniqueness of a Municipality.* Technical University of Liberec, 2008.
[7] Rummel, R.J.: *Applied Factor Analysis.* Northwestern University Press, Evanston, 1988.
[8] Stewart, D.W.: The Application and Misapplication of Factor Analysis in Marketing Research. *Journal of Marketing Research* **18** (1981), 51-62.
[9] Uramová, M., and Kožiak, R.: Regional Disparities in Slovakia from the Aspect of Average Nominal Wage. *E & M Ekonomie a Management* **11** (2008), 6-18.
[10] Žižka, M., Hovorková Valentová, V., and Rydvalová, P.: Statistical Comparison of Economic Situation Evaluation in Municipalities. In: *Mathematical Methods in Economics 2009* (Brožová, H., and Kvasnička, R., eds.). Czech University of Life Sciences, Prague, 2009, 366-371.