CZECH SOCIETY FOR OPERATIONS RESEARCH
CZECH ECONOMETRIC SOCIETY
FACULTY OF ECONOMICS, UNIVERSITY OF WEST BOHEMIA IN PILSEN

Proceedings of the 24th International Conference

Mathematical Methods in Economics

2006

13th – 15th September 2006

Pilsen
Czech Republic

Editor: Ladislav Lukáš
Technical editors: Ludmila Bokrošová , Kateřina Štruncová
Publisher: University of West Bohemia in Pilsen, Univerzitní 8, 306 14, Pilsen
Printing: TYPOS – Digital Print, spol. s r.o., Pilsen

Issue: 130 copies

ISBN 80-7043-480-5

Programme Committee

Mgr. Ing. Martin Dlouhý, Ph.D.
Faculty of Informatics and Statistics, University of Economics, Prague

Prof. RNDr. Ing. Petr Fiala, CSc.
Faculty of Informatics and Statistics, University of Economics, Prague

Doc. Ing. Jana Hančlová, CSc.
Faculty of Economics, Technical University of Ostrava

Prof. Ing. Josef Jablonský, CSc.
Faculty of Informatics and Statistics, University of Economics, Prague

Prof. RNDr. Ing. Jan Kodera, CSc.
Faculty of Finance and Accounting University of Economics, Prague

RNDr. Ing. Ladislav Lukáš, CSc.
Faculty of Economics, University of West Bohemia, Pilsen

Doc. Mgr. Jiří Málek, Ph.D.
Faculty of Finance and Accounting, University of Economics, Prague

Prof. RNDr. Jan Pelikán, CSc.
Faculty of Informatics and Statistics, University of Economics, Prague

Ing. Karel Sladký, CSc.
Institute of Information Theory and Automation of the Academy of Sciences of the Czech Republic, Prague

Prof. RNDr. Karel Zimmermann, CSc.
Faculty of Mathematics and Physics Charles University, Prague

Organisation Committee

RNDr. Ing. Ladislav Lukáš, CSc.
Faculty of Economics, University of West Bohemia, Pilsen

RNDr. Mikuláš Gangur, Ph.D.
Faculty of Economics, University of West Bohemia, Pilsen

Ing. Alexandr Kuchynka
Faculty of Economics, University of West Bohemia, Pilsen

Doc. Dr.Ing. Miroslav Plevný
Faculty of Economics, University of West Bohemia, Pilsen

RNDr. Jaroslav Potměšil, CSc.
Faculty of Economics, University of West Bohemia, Pilsen

Mgr. Blanka Šedivá
Faculty of Applied Sciences, University of West Bohemia, Pilsen

Table of Contents

Short Term Eqiulibrium in Press Distribution with Random Demand

Francisco J. Rodríguez Aragón
Dpt. of Distribution GELESA[1]. Franciscor@Gelesa.es

José M. Caridad y Ocerin
Dpt. of Statistics and Econometrics Córdoba
ETEA, Centro Andaluz de Prospectiva Spain. ccjm@uco.es

Abstract

Press distribution is done trough a network of selling points, and it can be compared to the commercialisation of non durable goods. Several actors participate in the distribution process: the editor, who produces the newspaper, a distributor, that estimate the demand to be supplied to each selling point, and collect the unsold copies, and finally, the retailer that could be specialized or not. The companies in charge of the distribution generally use heuristic methods, based on their market knowledge and of past experiences to estimate the demand at each selling point. Demand is a random variable, associated to each outlet; its distribution type is usually related to its mean sales, and the distribution company has a limited knowledge or it, trough the observation of a censored distribution, conditionally to the number of copies supplied to each point of the network (Caridad/Rodríguez/Ceular 2004). In this distribution process, conflicting goals arouse as the producer pays the distributor proportionally to the number of newspapers sended to the network plus the number of unsold copies. The retailer pays a fixed canon, so his interests of adjusting supply and demand are scarce.

A set of equations is proposed to represent the behaviour of retailers, distributors and producer, and their interrelations. The ideal economic optimization would be if the supply matched exactly the demand, at desagregated level, that is, in each outlet. This can not be achieved, due to the uncertainities involved. The producer would aim to avoid excessive oversupply, that increases his production, distribution cost and the collecting unsold copies. The distributor would try to estimate the demand, to minimize costs, and maintain the distribution contract. The retailer usually would desire to increase the supply, as his marginal cost for unsold copies is almost nil.

Using this model, and some proposed efficiency measures of the distribution process, the producer can control the process, forecast the demand and minimize costs. The possible conflicts between the different actors involved is analyzed, with the classical assumptions of rationality in the behaviour of producer, distributors and retailer. This efficiency function is also used to set distribution goals in the presence of different kinds of randomness in the demand, and to propose incentives for the maximization of profits.

Some real data about the applicability of the proposed methods is presented for a fully participated (by the producer) distributor of a mayor newspaper in Spain.

Keywords

Assignation of services, censored demand, stock rupture, demand estimation, press sales, press distribution, optimal distribution

JEL: C62

[1] GELESA company (www.Gelesa.es) has given support and data for this paper. Madrid. Spain.

1 SOME VARIABLES IN PRESS DISTRIBUTION ANALYSIS

Press distribution is a complex process that involve several sequential decisions by partially independent agents. The editor produce a newspaper, deciding upon the number of copies, N, produced. These are sended to one or several distributors, that have to determine how many copies would be supplied to each of the n outlet of the selling network. Finally these selling points offer the journals to the buyers, and return the unsold copies to the distributors.
The demand associated to the selling point i, on day t, is the random variable Dit that can be modelled by several usual distributions: of Poisson type for most selling points with average sales between 10 and 30 copies; in case of higher sales, a Normal distribution can be a fair approximation, while for low sales outlets, a geometric distribution is usually adequate. This demand variable con not be fully observed, as the corresponding sales is represented by the random variable

$$S_{it} = \begin{cases} D_{it} & \text{if} \quad D_{it} < s_{it} \\ s_{it} & \text{if} \quad D_{it} \geq s_{it} \end{cases}$$

being sit the number of copies or ‘service level’ provided to outlet i on day t. This amount is decided by the distributor on some exogenous base, as the historical data of sales, the number of additional copies then an outlet is asking for, the presence of some event, like a sport competition, the day of the week, and so on. Daily the truncated demand is observed as, the unsold copies, $u_{\mathrm{it}} = s_{\mathrm{it}} - d_{\mathrm{it}}$ if the demand does not exceed s_{it}, and $u_{\mathrm{it}} = 0$ otherwise. A theoretical model for the demand is necessary to evaluate the number of copies that could have been sold, in this last case. Theoretically, the demand could be known at each outlet, if its employees recorded data of unsold copies, when the last one has been sold. But this is not a realistic assumption, as there is no economic incentive to do so. The distributor could use a sampling method to estimate the demand at each selling point, but it is expensive, and can only be done during some short periods.

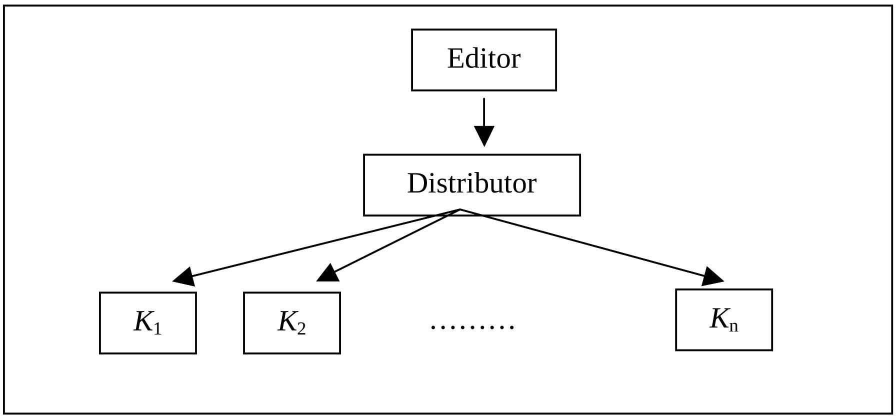
In the optimization process associated with press distribution there are three phases to be considered: goals associated with the distribution, the decisions about the number of copies to be supplied to each outlet, and finally, the control and evaluation process of the results achieved. The agents involved: producer or editor, distributor(s) and the selling points have some conflicting objectives. It is clear that each would want to maximize profits, but this does not necessary produce and absolute optimum, as inefficiencies arouse due to the organisation of all phases in the process. But, it is clear, that if the demand can be consistently estimated at every point of the network of selling points, an absolute optimum can be determined, and the whole process can be guided towards this optimum.

2 DISTRIBUTION ORGANIZATION

The distribution process of a newspaper can be summarized in Fig.1. The editor of producer send almost all the copies to a distributor, or to several distributors, each of them associated to a geographical area. The distributor is responsible for a network of n selling points, and daily, it has to decide how many copies are sended to each outlet. Usually a distributor will distribute several newspapers and journals from one or several editors, to minimize its distribution costs. To simplify the description a single editor with one distributor will be considered, although the procedures are general, and can be extended to more complex structures.

The editor or owner of the newspaper to be distributed has to control all the process, if his aim is to maximize profits. He has to look out not only at the production costs, but also to all the other actors structure, and decide about the level of the production, estimating the daily demand. His main sources of income are the sales procedures and the publicity included in the journal, M_e. Sales account for about 70% of the total, and usually cover the production costs, but do not produce benefits, and they will be proportional to total sales S: each paper sold produce a net income of P^I.

Figure 1: Distribution process.



Obviously the editor aim is to maximize sales, as his income from advertisements will be linked to the number of readers that could be have access to the publicity. He has also to bear a fixed amount, F_e, paid to distributor per selling point, beside his fixed costs per selling point, E_e, and the devolution costs for unsold copies, being P^D the corresponding unitary cost. Also, if an outlet runs out of papers, the additional demand non satisfied originate some unitary cost, P^A, associated to the future effect upon his non-served customers. The editor profit function will be desagregated in n parts, each of them associated to a selling point:

$$B_e(s_i) = \begin{cases} P_e^I d_i - P_e^D \left(s_i - d_i\right) - P_e^F s_i + M_e - E_e - F_e & \text{if} \quad d_i < s_i \\ P_e^I s_i - P_e^A \left(d_i - s_i\right) - P_e^F s_i + M_e - E_e - F_e & \text{if} \quad d_i \geq s_i \end{cases}$$

The subscript '*e*' is linked to variables and parameters related to the editor.

The distributor has three sources of income: the variable part proporcional to the number of copies distributed, P^I per unit (usually around 10% of the price), a commission from each selling point *i*, G_i, that usually is fixed, but can depend also of the level of copies distributed to this outlet, and a fixed amount, F_e, from the editor. The distributor also associate a unitary cost, P^A, for lost sales due to an unsatisfied demand at a selling point, as, he is responsible of this distribution. He has to bear his fixed costs E_d. The subscript '*d*' is linked to variables and parameters related to the distributor, and his profit function is

$$B_d(s_i) = \begin{cases} P_d^I d_i - E_d + G_i + F_e & \text{if} \quad d_i < s_i \\ P_d^I s_i - P_d^A (d_i - s_i) - E_d + G_i + F_e & \text{if} \quad d_i \geq s_i \end{cases}$$

Finally, at each selling point, the income corresponding to a newspaper is a commission, PI, usually 20% of the price charged to the final buyer. There is also an implicit unitary cost, P^A, due to lost sales derived from having run out of copies, beside the fixed costs E_i, and the charges G_i of the distributor. Unsold copies are returned to the distributor without any charges. The profit function of the selling point is, thus,

$$B_i(s_i) = \begin{cases} P_i^I d_i - E_i - G_i & \text{if} \quad d_i < s_i \\ P_i^I s_i - P_i^A \left(d_i - s_i\right) - E_i - G_i & \text{if} \quad d_i \geq s_i \end{cases}$$

The subscript '*i*' is linked to variables and parameters related to the *i* selling point.
It is clear that the implicit costs, P^A, attributed to lost sales due to lack of newspapers at a selling point, is not an objective data, but derived from business policies.

3 SHORT TERM EQUILIBRIUM WITH PERFECT INFORMATION

All the agents involved in the distribution process are assumed to be rational, and to maximize profits, they will try to increase sales as much as possible. Nowadays, the importance attributed to the sales maximisation is different for each of them. And this is the cause of inefficiencies in all the distribution process. The main interest of a selling point is to maximize sales; the amount of papers, si, provided to the point i, is decided by the distributor, but the outlet can ask for additional copies. As there is no marginal cost for additional or for unsold copies, selling points tend to demand excessive number of newspapers, originating additional costs to distributor and editor, as the former has to collect the unsold papers, and for the producer it is a waste in production cost, although the paper is recycled. The distributor would also want to maximize sales, but his main interest is to keep his customer, the editor, happy, as his sales commission is quite low. Also, as he has to collect unsold copies, he has no special interest in oversupplying the points of sales. The editor tries to increase sales, to maximize his income both from sales and publicity, but taking into account that distribution costs can overshot if the distribution is not well carried off, originating lost sales for undersupply, or production and distribution costs for recalling unsold copies.

In theory, in a perfect information and competitive environment, global profits would tend towards cero, if the demand was known at each selling point, and there would not be cost for unsold copies, leading to this trivial relationship

$$\left[\left(P_i^I + P_d^I + P_e^I\right) - P_e^F\right]d = E_i + E_d + E_e - M_e$$

Increasing income due to more demand, would enable better service and investments in fixed cost and structure, and also, would permit lower publicity inserted in the newspaper, as a base for future growth. But in the profits relations, demand is not deterministic, so unsold copies can not be avoided as the undersupply of some outlets, originating additional costs for the agents involved. Thus the last trivial relationship does not hold, and the profit functions for each of the three types of agents have to be considered, as they have partially conflicting interests. At the short term, some of the cost structure and sales, publicity income and distributions commissions can be assumed constant, and, thus, reformulated in a simpler version, and including the cost and income functions

$$B_i(s) = I_i(s) - C_i^A(s) = \begin{cases} P_i^I d & \text{if} \quad d < s \\ P_i^I s - P_i^A\left(d - s\right) & \text{if} \quad d \geq s \end{cases}$$

$$B_d(s) = I_d(s) - C_d^A(s) = \begin{cases} P_d^I d & \text{if} \quad d < s \\ P_d^I s - P_d^A\left(d - s\right) & \text{if} \quad d \geq s \end{cases}$$

$$B_e(s) = I_e(s) - \left[C_e^D(s) + C_e^A(s) + C_e^F(s)\right] = \begin{cases} P_e^I d - P_e^D\left(s - d\right) - P_e^F s & \text{if} \quad d < s \\ P_e^I s - P_e^A\left(d - s\right) - P_e^F s & \text{if} \quad d \geq s \end{cases}$$

The expected values of these profit functions should be maximized by the selling points, the distributors and the editor. This would no lead to a global equilibrium state, as the conflicting interest can be easily stated: the distributor benefit function increases with s, so his main interest would lie within augmenting the number of copies supplied; something similar can be said about the retailers. On the other side, the editor would want to increase supply to each outlet up to a certain point, but over this limit, the devolution cost will decrease its global benefit in the short run. Also, the demand is of random nature, and its distribution has to be estimated for each selling point, and this information translated to all intervening agents. The valued attributed to unsold copies when an outlet run out of papers, by the editor, distributor and retailers are quite different, and could also change within the network. The only agent that can be able to control the whole process is the editor, but it would need information gathered by the distributor about the results of daily sales and unsold copies, as well as the supply awarded to each outlet.

4 SHORT TERM DECISIONS WITH RANDOM DEMAND AT EACH OUTLET

To optimize the expected profits for each agent, it is necessary to consider some realistic model for the demand at each point of the sales network. Several usual distributions have been considered with real data from a leading distributor in Spain (Caridad, Rodríguez y Ceular 2004). For retailers with medium mean daily sales, a Poisson distribution provides a reasonable probabilistic model. Its mean can be estimated using likelihood methods with simple data from the truncated Poisson variable corresponding to daily sales. In case of low sales points, other discrete distributions have to be considered.

The Poisson type demand will be assumed form now on. In a network of 613 points of sales studied between 2001 and 2003, (Caridad y Rodriguez, 2004b), more than 20% of these had a Poisson demand. Also, different demand distribution can be used with the same methodology. Obviously, at each point of the sale network, the lambda parameter is estimated and in the following expressions, $\lambda = \lambda_i$, is the estimated mean value of the demand at the i outlet. Its expected income function is

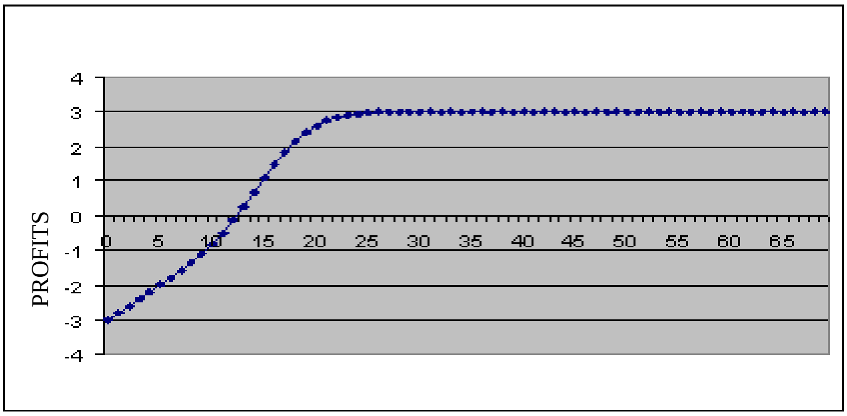
$$I_i(s) = P_i^I e^{-\lambda} \sum_{j=0}^{s} j \frac{\lambda^j}{j!} = P_i^I e^{-\lambda} F_\lambda(s-1)$$

being F_λ the distribution function of the demand. The expected cost function is

$$C_i^A(s) = P_i^A e^{-\lambda} \sum_{j=s}^{A} (j-s) \frac{\lambda^j}{j!}$$

The profit function increases with the number of copies supplied; the retailer does not have any incentive to limit the desired supply, even if this would increase the editor and the distributors costs, and the number of unsold copies. The following figure shows this situation for an outlet with expected daily sales of 15 copies, and using as unit costs $P_i^I = P_i^A = P_i^D = 0.2$.

Figure 2: Retailer profit function



In this case, a supply of s = 24 copies should be enough to allow a 'near-maximun' expected profit for the retailer, as the marginal increase is almost nil. But the retailer does not known this profit function, and even if it did, there no incentive for him not to demand additional copies, as the unsold papers are returned with no cost for him. With a supply of s = 20 copies he should notice that his optimum is not attained, as some days he would run out of stock.

The influence of the relative values of P_i^I, P_i^A, P_i^D affects the profit curve, and it can be different from one retailer to another. This last case is not considered here. If $P_i^I = P_i^A$, a supply s = 0, the expected profit (loss) for the selling point is

$$B_i(0) = I_i(0) - C_i^A(0) = - P_i^A \lambda$$

Over-restricting supply to a selling point will originate additional local demand, as the retailer depends of a minimum level of sales to compensate his fixed costs.

For the distributor, the expected income is similar, including his sale commission (around 10%, P_d^I =0.1), and his commission per point of sale (P_d^I = 0.2)

$$I_d(s) = P_d^I \lambda e^{-\lambda} F_\lambda(s - 1)$$

He has also to consider the cost of collecting unsold copies, as this can leads to an oversupply of the points of sales. As this cost is lower for the editor than to the editor, this can be a real problem in the distribution process.

The editor has an expected income function similar so the distributor

$$I_e^I(s) = P_e^I \lambda e^{-\lambda} F_\lambda(s - 1)$$

but he has to take into account the expected costs associated with undersupply

$$C_i^A(s) = P_i^A e^{-\lambda} \sum_{j=s}^{A} (j - s) \frac{\lambda^j}{j!}$$

for unsold copies, returned by the distributor

$$C_i^D(s) = P_e^D e^{-\lambda} \sum_{j=0}^{s} (s - j) \frac{\lambda^j}{j!}$$

and the production costs

$$C_e^F(s) = P_e^F s$$

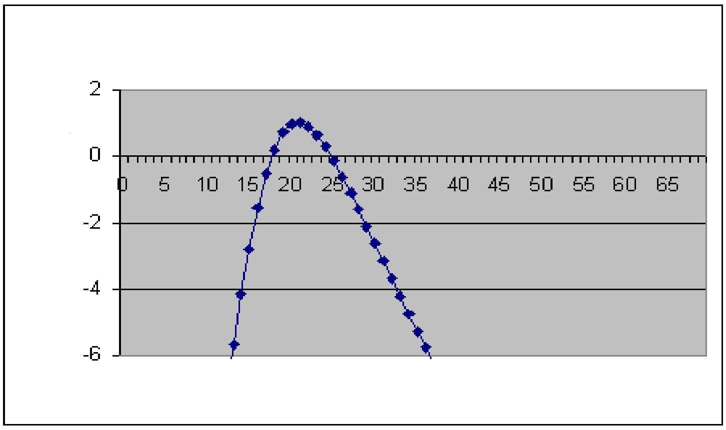
all referred to a particular selling point.

The shape of the expected profit function, when the assumed demand is of $P(\lambda = 15)$ and the prices included in the former income and cost functions are

$$P_e^I = 0.7 \quad P_e^A = 2 \quad P_e^I = 1.4 \quad P_e^D = 0.5 \quad P_e^I = 0.35 \quad P_e^F = 0.5 \quad P_e^D = 0.175$$

is in Fig.3.

Figure 3: Expected profit function for the distributor for a selling point

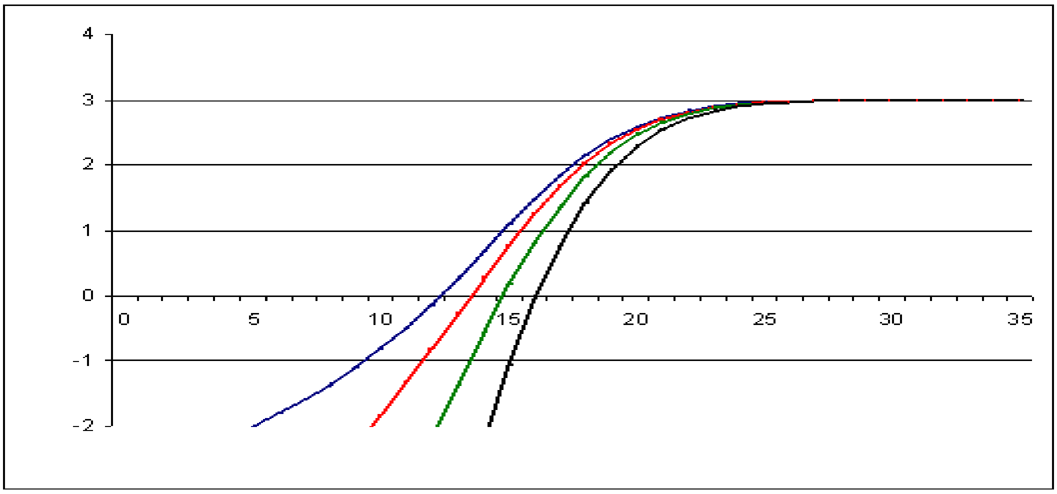


An undersupply or an oversupply, will result into a loss, so the number of copies, s, distributed to a selling point has to be kept within strict limits, if the editor would expect a profit in it. But these decisions are taken by the distributor, so it is clear that the editor can not ignore this fact, and should receive daily information about the results of the distribution process, and of the implicit prices assumed for undersupply. Trust would be better served with a flow of information about the results of the distribution process: full information to the editor from his own papers, and clear statistical information about the results of the distribution results for other editor titles distributed by the same distributor.. Also, it is clear that the outlets can have different perception about losses associated to undersupply, and this would affect the overall distribution process, but average values should be used by both the editor and distributor. Conflicts of interest between different agents are clear comparing the profit functions of the editor, the distributor and the selling points. The distributor can analyse the market, as he usually control the network of selling points, and he can increase profits using this information not available to the editor, who, on the other hand, has monopolist power over his product, and will use it to obtain the distribution information, and to maximize his expected profit.

5 DISTRIBUTION WITH FULL INFORMATION

The selling point has little influence over his supply, and his fear of running out of copies will induce him to assign a higher implicit value to P_i^A, and thus, it will try to augment s; the distributor has to take this into account when deciding the number of copies for each outlet. The influence of this value over the selling point profit function is presented in Fig.4, also with demand of Poisson type, with mean 15.

Figure 4: Expected profits for an outlet with $P_i^A = 1, 2, 3, 8$ times P_i^I

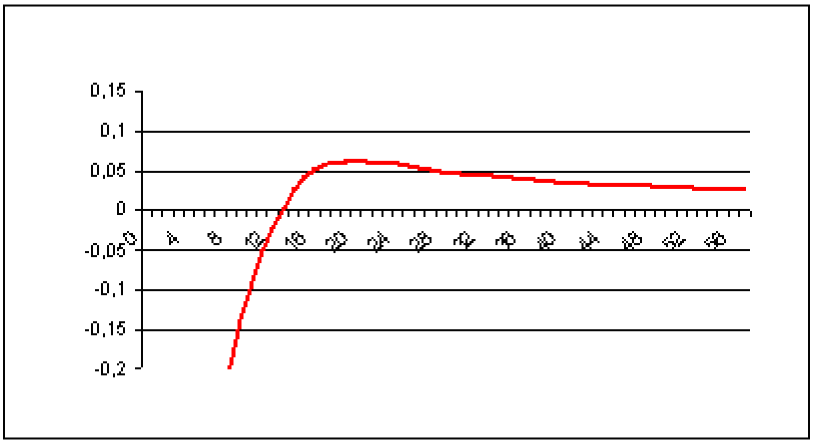


As the distributor does not bear the production costs, there is no incentive for him to control strictly *s*. His commission over sales without cost for unsold copies, would induce him to increase *s*, as his interests are, in this aspect, similar to those of the selling point, colluding with the editor. The latter should ask for a full daily account of the distribution process, as the distributors policies could harm him, in the short term.
Nowadays, the distributor marginal benefit will decrease, tending to cero

$$\frac{B_d(s)}{s} = \frac{I_d(s) - C_d^A(s)}{s}$$

In Fig.5, the expected distributor marginal profit has a maximum, but this happens for a value of *s* considered too high by the editor.

Figure 5: Distributor marginal expected profits.

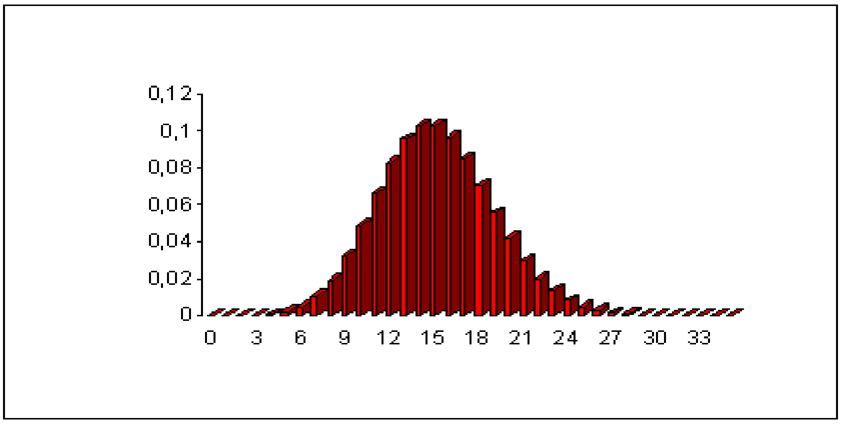


If the distributor is able to impose his value for the cost for undersupply of the selling point, the distribution process can be simplified and reduced to maximize the expected profit functions $B_d(s) = I_d(s) - C_d^A(s)$ and $B_e(s) = I_e(s) - \left[C_e^D(s) + C_e^A(s) + C_e^F(s) \right]$. The distributor can expect to distribute a higher number of copies, if the editor is not aware of the situation, originating also a larger amount of oversupply. If the editor has full information, he can maximize his expected profits at a lower value of the distributor wishes (Fig.5).

The distributor has a clear interest in concealing information to the editor, who needs this information, and would be willing to invest in a better estimation of the demand at each selling point (Caridad/Rodríguez/Ceular 2004). An alternative is that the distributor company is at least partially controlled by the editor. This is a usual situation in Spain, where these are participated by one or several editors.

For example, for a selling point with average sales λ = 15, and having previously established that the demand is of Poisson type, that is, its probability function is show in Fig.6

Figure 5: Demand distribution at a selling point.



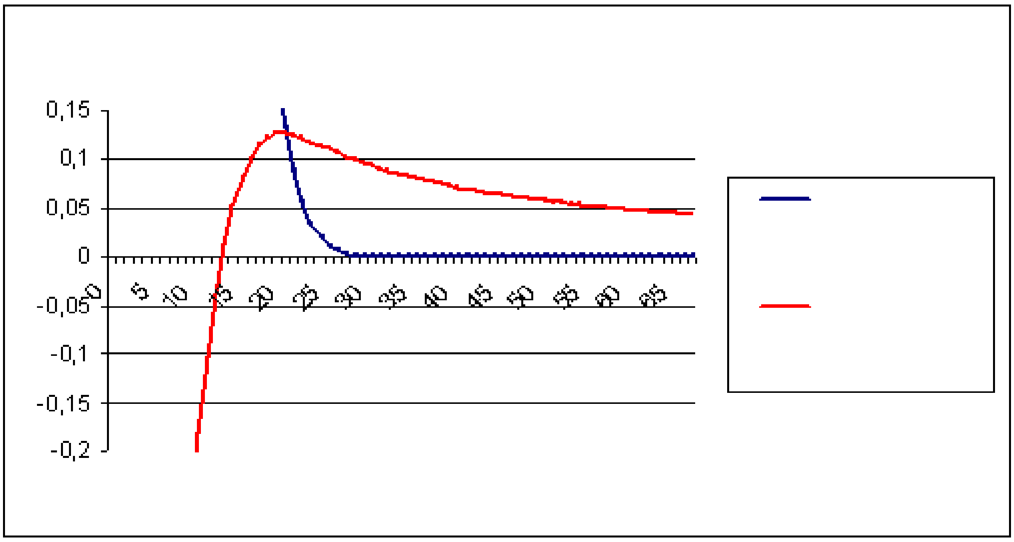
For s = 22, the demand hill be fully served 95% of the time. The selling point profit will depend upon the relative value of P_i^A against P_d^I, as can be shown on Table 1.

Table 1: Expected profits for a selling point

P_d^A / P_d^I	**Profit for *s* = 22**
1	2.8254
4	2.7797
8	2.7187

The expected profit function is bounded, but $\nabla B_d(s) = B_d(s) - B_d(s-1) > 0$, so its marginal value will decrease from a certain point, as can be show on figure 7

Figure 7: Expected and marginal profits for $P_i^A / P_i^I = 4$



The selling points are individual companies, with no influence over prices of the product they sell, that is provided to them with no marginal cost (excluding the limited space available on the outlet). Sales for each of them can be considered independent, so they will ask for a minimum number of copies, not below the amount that maximize their marginal profits. In table 2 are presented this minimum number of copies that a selling point (again with P(15) demand).

Table 2: Minimum number of copies expect by an outlet related to P_d^A / P_d^I

P_d^A / P_d^I	$\max_s \left(\frac{B_i(s)}{s} \right)$
1	21
4	22
8	22

The distributor can use the same information to establish his own expected profit function $B_d(s) = I_d(s) - C_d^A(s)$, tending to oversupply, if the distributor has not access to the full information of supply and sales at each outlet, and his profit function will be aggregated over all the network.

Considering the aggregated profit function

$$B_e(s) = I_e(v) - C_e^D(s - v) - C_e^S(s) - C_e^A(a)$$

where $s = \sum_{i=1}^{n} s_i$ is global supply and $v = \sum_{i=1}^{n} v_i$ the overall sales, and *a* the number of selling points that outsold their copies. For example, using $I_e(v) = P_e^I v$, $P_e^I = 0.7$, $C_e^D(s - v) = P_e^D(s - v)$ with $P_e^D = 0.35$, $C_e^S(s) = P_e^S s$ with $P_e^S = 0.175$, and $C_e^A(a) = 0$

Thus, the profit function is $B_e(s) = 1.5v - 0.525s$, and it will produce benefits for $s \leq 2v$. But, although margins are quite high, for a constant value of sales, ν , it is possible to reduce *s*, increasing the editor profits, lowering the distributor, and maintaining the selling points income. In this example, with selling point with average sales of 15 copies, the distributor will try to almost double the supply, and the editor profits will be .

$B_e(s) = 1.05 \times 15 - 0.525s$

If $s = 25$, $B_e = 3.25$, while for $s = 30$, $B_e = 0.75$, that is, quite a difference, although that the supply 25 will cover the demand 99.38% of the time, and increasing it by five additional copies, this ratio only augment to 99.98%. These figures show the importance for the editor of controlling the distribution process, as even with a lower supply, demand could be reasonably attended, and the editor profits witll grow significantly, without harm for the distributor and selling points.

6 EDITOR'S CONTROL WITH FULL INFORMATION

The editor objectives are associated with his knowledge of the results of the distribution process and sales at the selling point level, being his expected profit function (for each selling point)

$$B_e(s) = I_e(s) - \left[C_e^D(s) + C_e^A(s) + C_e^F(s) \right]$$

including, now, the returned copy cost, $C_e^D(s)$ (usually, 50% of its retail price), and the production cost $C_e^F(s)$, being $P_e^D = 0.35$ and $P_e^F = 0.175$. In this situation, the editor can assign value to unsold copies and for undersupply, mostly overriding the opinions of the other agents. Usual values employed in Spain are $P_e^A = 2P_e^I = 1.4$. In this case, the optimum supply is $s = 21$, and the editor profit $B_e(21) = 1.0251$. The distributor profit is affected, as can be seen in Table 3.

Table 3: Changes in the distributor expected profits.

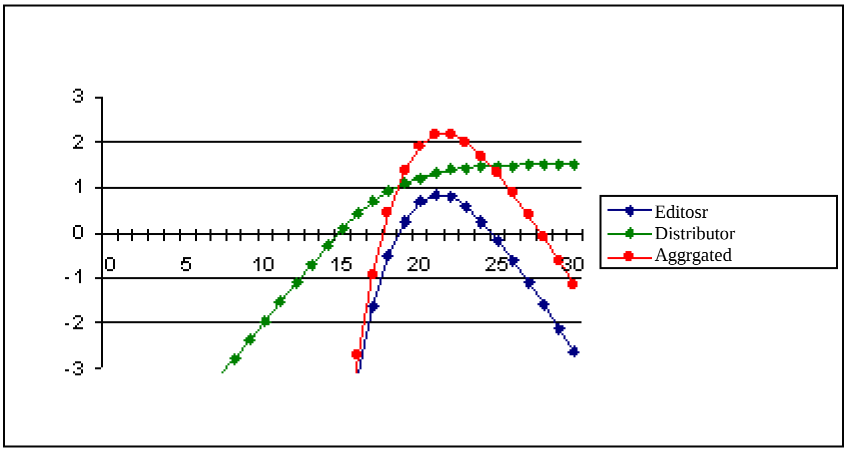
P_d^A / P_d^I	$B_d(21) - B_d(25)$	$B_d(21) - B_d(26)$	$B_d(21) - B_d(27)$
1	-0.1193	-0.1274	-0.1321
4	-0.1543	-0.1642	-0.1699
8	-0.2009	-0.2133	-0.2203

The distributor expectations will be diminished, and should be aware of this circumstances. An alternative would be to try to maximize the aggregated profit function editor-distributor.

$$B(s) = B_e(s) + B_d(s) = I_d(s) + I_e(s) - \left[C_e^D(s) + C_e^A(s) + C_d^A(s) + C_e^F(s) \right]$$

which is shown on Fig.8.

Figure 8: Editor, distributor and aggregated profit functions



This leads to a supply higher that the 21 copies desired by the editor, but not reaching the levels that would be desired by the distributor and the retailers:

$$\max_s B(s) = 22$$

In case of asymmetries, with the distributor with full information and the editor with aggregated data, the supply decided by the distributor will increase, with no incentive from the retailers in limiting it, while the distributor profits do not fall sharply. In case of full information over the distribution process, the editor will send a lower number of copies to the distributor, while satisfying the demand almost with no change, and reducing sharply the returned unsold copies, and lowering slightly the distributor profits. A negotiated settlement between both agents will increase the supply, in benefit of the distributor. The implicit prices associated for running out of stock in a selling point are important in the relations between distributor and editor, although they have no direct impact on the cash-flow. Finally, one can wonder about the position of the retailer. In fact, they do not have individual bargaining power, unless they get associated, which, in practice, can be opposed successfully by the distributors. Their hypothetical refusal to sell a particular newspaper is not realistic, as some additional points of sales could be added to the network2. In some regions, there are distributors wich own a significant number of point of sales, limiting the hypothetical power of these associations. Some commercial centres with high level of sales, as Carrefour group do have this bargaining power.
In summary, with this simple approach, the market forces can be explained, leading to forecast about the results of policies changes for both the distributors and the editors. The logistics aspects of press distribution (and some products with similar characteristics, as pharmaceuticals) is of increasing importance. as well as the information situations of a particular market. With full information, the editor should have the leverage power upon the global results, and he needs to control, at least in part the distribution companies. Otherwise, if the information about the supply and sales in all the network does not reach the editor, he will be in a weak commercial position when establishing the distribution contracts. The trend towards vertical integration in press distribution, as done by the PRISA group in Spain, and a more precise demand estimation at the outlets, will accelerate in the near future.

The cross section analysis presented can be extended in several lines: the use of panel data, combining temporal information with spatial distributions is clear, and can be combined with other approaches, (Caridad/Rodriguez, 2004d), using clusters of selling points as an intermediate aggregation level.

[2] In the biggest Spanish cities, there are some associations with these objectives: in Madrid, A.V.V.P.M.

References

Caridad y Ocerin, J.M. (1998). *Econometría: modelos econométricos y series temporales*.Ed. Reverté. Barcelona.

Caridad y Ocerin, J.M. (2002). Hybrid and classical forecasting methods: the demand of water in urban areas. MME’02. pp. 33-45. Praga. ISBN 80-248-0153-1.

Caridad y Ocerin J. M.; Rodríguez Aragón F. J.; Ceular Villamandos N. (2005). Lost Sales and Press Distribution. XV Jornadas Hispano-Lusas de Gestión Científica. Sevilla. ISBN 84-96378-14-4

Caridad y Ocerin J. M.; Rodríguez Aragón F. J.; Ceular Villamandos N. (2004). Demand estimation in press distribution. MME’04. .42-55 Brno. ISBN:80-210-3496-3.

Heskes T.(2002).Innovatie distributiesoftware verhoogt verkoop: Just-Enough-Delivery interessant voor supers. Levenmiddelenkrant, 13.

Heskes T.; Spanjers J. J.; Bakker B.; Wiegerinck, Wim. (2003). Optimizing newspaper sales using neural-Bayesian. Neural Computing & Applications, 12(3), 212-219.

Maron, M. J. (1987). Numerical Analysis. A practical Approach. Mc Millan Publishing Company.

Pyndick, RS y Rubinfeld, D. L. (2001). Microeconomía. Ed. Prentice Hall. Madrid.

Rodríguez Aragón, F. J. y Caridad y Ocerin J. M. (2004a). Asignación de servicios de prensa en un *punto de demanda media*. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba. ISBN. 84-95723-17-4.

Rodríguez Aragón, F. J. y Caridad y Ocerin, J. M. (2004b). Demanda insatisfecha en una red de puntos de venta de prensa. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba. ISBN. 84-95723-18-2.

Rodríguez Aragón, F. J. y Caridad y Ocerin, J. M. (2004c). Sistemas de difusión de prensa en puntos de venta con demanda aleatoria de distribución geométrica. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba. ISBN. 84-95723-19-0.

Rodríguez Aragón, F. J. y Caridad y Ocerin, J. M. (2004d). Equilibrio de mercado en un punto de venta con demanda Normal. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba. ISBN. 84-95723-22-0.

Van Kranenburg, H.L. (2001). Economic Effects of Consolidation of Publisher and Newspaper in the Netherlands. Journal of Media Economic, 14(2), 61-76.

Rent Extraction by Large Shareholders:

Evidence Using Dividend Policy in the Czech Republic *

Jan Bena (j.bena@lse.ac.uk)†

Department of Accounting and Finance, LSE, London and CERGE-EI, Prague‡

Jan Hanousek (jan.hanousek@cerge-ei.cz)§

CERGE-EI, Prague

Abstract. Using cross-sectional analysis of corporate dividend policy we show that large shareholders extract rents from firms and expropriate minority shareholders in the weak corporate governance environment of an emerging economy. By comparing dividends paid across varying corporate ownership structures—concentration, type, and domicile of ownership—we quantify these effects and reveal that they are substantial. We find that the target payout ratio for firms with majority ownership is low but that the presence of a significant minority shareholder increases the target payout ratio and hence precludes a majority owner from extracting rent. In contrast to other studies from developed markets, our unique dataset from the Czech Republic for the period 1996-2003 permits us to take account of the endogeneity of ownership.

Keywords: Rent extraction, Large shareholders, Corporate governance, Dividend policy

JEL codes: D21, G32, G35

---

* We are grateful to Ron Anderson, Andrew Ellul, Štěpán Jurajda, David Webb, Randall K. Filer, Joachim Inkmann, Michela Verardo, Oriana Bandiera, Antoine Faure-Grimaud, and seminar participants at LSE, London, CERGE-EI, Prague, and the Czech Economic Society meeting in Prague for useful comments and suggestions.

† Please address correspondence to: Jan Bena, The Financial Markets Group, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, Email: j.bena@lse.ac.uk, Tel: +44 (0) 20 7955 6270, Fax: +44 (0) 20 7852 3580.

‡ A joint workplace of the Center for Economic Research and Graduate Education, Charles University, Prague, and the Economics Institute of the Academy of Sciences of the Czech Republic.

§ While preparing this paper Jan Hanousek benefited from GACR grant No. 402/06/1293.

1. Introduction

Theoretical papers suggest that large shareholders have a dual impact on firms. On the one hand, significant owners have a strong incentive to monitor management to ensure that a firm's value is maximized, while on the other hand, their behaviour is motivated by the possibility to extract rents and enjoy the private benefits of control.[1] Hence, as argued in Shleifer and Vishny (1997), the overall effect of large shareholders on firms is ambiguous and has to be tested empirically.

In this paper we provide evidence that large shareholders extract rents from firms and expropriate minority shareholders, by showing that some corporate ownership patterns are consistently associated with higher/lower target dividend payout ratios and different levels of dividend smoothing in the cross-section. Moreover, by comparing dividends paid across various ownership structures we quantify the rent extraction associated with the presence of large shareholders and show that it is substantial. We consider several levels of ownership concentration, several types of the single largest owner, and investigate the difference between domestic and foreign owners.

We find that presence of a significant minority shareholder prevents majority owners from extracting rent by increasing the target payout ratio. This finding is much stronger for domestic owners than for foreigners. Our results are consistent with the hypothesis that strong minority owners play a crucial role in dividend policy, especially in the weak corporate governance environment of an emerging economy.

We use data from the Czech Republic for the period 1996-2003. This dataset allows us, first, to account for the endogeneity of ownership and, second, to separate the effect of ownership from a broader institutional corporate governance framework. The unique modern economic history of the Czech Republic helps to explain the ownership endogeneity problem, as the initial ownership structure of companies was set exogenously by government bureaucrats during privatization in 1991-1994. The dataset we use in this

[1] Shleifer and Vishny (1986) were the first to formally investigate the role of large investors in firms, and Shleifer and Vishny (1997) provide a systematic survey of costs and benefits associated with the presence of large shareholders in firms. More recently, Bolton and von Thadden (1998) model the trade-off between costs and benefits of concentrated versus dispersed ownership and Burkart et al. (2000) show how large shareholders and the private benefits they enjoy influence takeovers.

study includes detailed variables from the privatization process as well as variables capturing pre-market firm-level conditions, which we employ as instruments for ownership. After privatization, ownership rights were fully honoured which helped early corporate development,[2] but the evolution of institutional structures was considerably slower; corporate governance was virtually nonexistent, and corporate law was only weakly enforced. As a result, corporate governance mechanisms which are present in developed economies and which play a key role in the relationship between corporate insiders and outsiders, including dividend policy, were missing.[3] These conditions forced shareholders to act based on fundamental rights derived from ownership only, and hence the environment of the Czech Republic fits closely our model's assumptions of large shareholders' behaviour. In this way, privatization and the fact that corporate law and governance developed from scratch in the Czech Republic help focus our analysis on the effect of ownership only.

This paper is the first empirical study of dividends from a transition country in Central and Eastern Europe. Since many CEE countries underwent a similarly quick transition from a state-directed to a market economy, our findings based on data from the Czech Republic may to a large extent be valid for them as well.

The structure of the paper is as follows: In the next section we survey the literature; in section 3 we provide an institutional outline and explain in detail how private ownership developed in the Czech Republic over the 1990s; in section 4 we define ownership variables, describe our model, and present our econometric technique; section 5 contains a description of our data and summary statistics; in section 6 we present our results; section 7 contains some robustness checks; section 8 summarizes the paper and concludes.

---

[2] Using data from transition countries Johnson et al. (2002) find that property rights are the most important determinant of investment by entrepreneurs. Weak property rights discourage firms from reinvesting their profits, even when bank loans are available.

[3] In their international study Laporta et al. (2000) offer evidence that countries with laws protecting the rights of minority shareholders are associated with higher dividend payout ratios and show that companies pay out a smaller proportion of earnings in those countries where laws are more relaxed about overinvestment and empire building. Other economic institutions are important determinants of dividend policy as well. Dewenter and Warther (1998) compare dividend policies of U.S. and Japanese corporations and link them to institutional differences in the structure of corporate ownership. Japanese firms face fewer agency conflicts and information asymmetries than do U.S. firms. Consistent with the agency theory of dividends, Japanese firms experience smaller stock price reactions to dividend omissions and initiations, they are less reluctant to omit and cut dividends, and their dividends are more responsive to earnings changes.

2. Literature

The existing empirical evidence on rent extraction by large shareholders deals with developed economies only and gives mixed results. Demsetz and Lehn (1985) show that private benefits of control affect ownership structure in the U.S. and Zingales (1994) argues that expropriation by large shareholders is significant in Italy. On the contrary, Bergström and Rydqvist (1990) and Barclay and Holderness (1989, 1992) do not find evidence of substantial expropriation in Sweden or the United States, respectively. In the paper closest to our own, Gugler and Yurtoglu (2003) suggest that this problem is present in Germany. The authors show that announcements in dividend change provide new information about conflicts between a controlling owner and small outside shareholders in Germany, and document how small shareholders use dividends to limit rent extraction by controlling owners.[4] Faccio et al. (2001) find evidence of systematic expropriation of the outside shareholders in Western Europe and East Asia at the base of extensive corporate pyramids. They show that corporations in Europe pay significantly higher dividends than in Asia and that in Europe other large shareholders contain the controlling shareholder's expropriation of minority shareholders whereas in Asia they collude in that expropriation.

Our paper is novel since by working in the Czech transition environment we can fully account for ownership endogeneity and focus on fundamental rights derived from ownership. We also benefit from a large sample that covers a majority of the country's economic activity.

Our work is also linked to a rich empirical literature on corporate dividend policy. According to free cash flow theory[5] dividends are a control mechanism used by shareholders to divert free funds, which managers have power over within corporations, away from them. The shareholders' goal is to prevent managers from perk consumption, empire building/overinvestment, or management entrenchment[6]. In support of the free cash

[4] Similarly, Gugler (2003) estimates the effect of ownership on dividend policy using data from Austria. He finds that ownership and control structure of a firm are significant determinants of its dividend policy.

[5] First mentioned by Easterbrook (1984), reinvented by Jensen (1986), and modelled in a dynamic setting in Zwiebel (1996).

[6] Shleifer and Vishny (1989) model management entrenchment as one possible driving force behind inefficient investments undertaken by managers with free cash flows at hand.

flow theory Lang and Litzenberger (1989) find that the market reacts favourably to dividend announcements made by firms with characteristics suggesting that they might otherwise overinvest their funds. Brook et al. (1998) show that firms poised to experience large, permanent cash flow increases after four years of flat cash flow tend to boost their dividends before cash flow jumps, but are hesitant to adjust them afterwards.

The competing argument to free cash flow is based on the idea that management uses dividend policy to communicate to investors the level and growth of income or future prospects of the company because ordinary accounting reports are insufficient or inadequate to convey this information.[7] In their test of signalling hypothesis versus other agency models Bernheim and Wantz (1995) find support for signalling theory. Similarly, Offer and Siegel (1987) show that equity analysts revise their earnings forecasts following the announcement of an unexpected dividend change. Also, in their event study of stock price reactions to dividend change announcements Amihud and Murgia (1997) find some dividend-signalling patterns in Germany. On the other hand, DeAngelo et al. (1996) argue that dividend changes lag behind earnings changes and conclude that managers do not signal their negative information with dividends. An even stronger argument appears in a study by Benartzi et al. (1997). They find no evidence that changes in dividends carry information about future earnings changes.

Both signalling and free cash flow theory were developed for firms with dispersed ownership structures and hence with managerial control. Similar to other continental European countries, the ownership of Czech firms is rather concentrated in the period we analyse.[8] For a firm with concentrated ownership, the free cash flow and signalling rationale for paying dividends still applies but, in this case, dividends are used to solve the agency issues and/or the asymmetry of information between a dominant shareholder who colludes with management (appoints the management) and remaining shareholders. Therefore, corporate dividend policy in a firm with concentrated ownership is predomi-

[7] This literature was started by Bhattacharya (1979) and Miller and Rock (1985), and was extended by John and Williams (1985) and Bernheim (1991).

[8] High ownership concentration is present in most Continental European countries. See La Porta et al. (1999) for a description of prevailing ownership structures in Europe. Additional relevant descriptions are in Gugler (2003) for Austria, Gugler and Yurtoglu (2003) for Germany, and Kočenda (1999) for the Czech Republic.

nantly determined by how the conflict among the firm's shareholders about distribution of profits (benefits) is resolved. Legally, all shareholders have the same cash flow rights in the Czech Republic. Paying dividends follows this principle as cash reaches all shareholders proportionally, but a dominant shareholder exerting effort to seek private benefits associated with ownership does not. In other words, in contrast to the case of dispersed ownership where the main corporate governance issue is to solve moral hazard between management and shareholders, good governance in concentrated ownership structures predominantly means equal treatment (per unit of stake in the firm) of all shareholders. From the minority shareholders point of view, dividend payments alleviate the free cash flow problem or serve as a signal.

3. Institutional Environment

3.1. Privatization

Since the ownership structure of companies is a key explanatory variable in our study we describe in detail how these structures developed. Since 1989 the Czech Republic has undergone overwhelming economic changes that have resulted in the quick introduction of a modern market economy. At the beginning of the transition process, almost all productive assets were state-owned, the separation of ownership and control did not exist, there was no modern corporate law and financial markets, and corporate governance structures were only about to start evolving.

The ownership structure of most Czech companies was set during the mass privatization of medium and large enterprises in the first half of the 1990s.[9] The majority of shares of these companies were offered through the voucher scheme to the general public. All citizens 18 years and over could buy, for a tiny nominal fee, a package of vouchers worth 1,000 points. With these points they could bid for the shares on offer or they could place (part of) their points in investment privatization funds, which could then bid for

[9] This section is based on Gupta et al. (2001) and Hanousek et al. (2004). The Czech privatization process has been described in detail in Švejnar and Singer (1994), Kotrba (1995), and Coffee (1996).

shares. After bidding was completed, points were exchanged for shares and secondary market trading started at the Prague Stock Exchange.[10] A large number of investment privatization funds emerged on a voluntary basis. Although funds were started by various sponsors (domestic and foreign banks, corporations, or individuals), most funds were sponsored by domestic banks, with several banks starting more than one fund. Funds ended up with about 70 % of all points. Bank-sponsored funds acquired most of the points, with the ten largest bank-sponsored funds holding 67 % of all points acquired by all funds (or about 44 % of all points initially bought by individuals). Control of the largest privatization funds by majority state-owned banks was an unexpected outcome for the government and had a major impact on the emerging corporate governance structure in the middle of the 1990s.[11]

The privatization process was designed to find private owners of firms very quickly rather than to look for optimal ownership structures. The decision-making of the Ministry of Privatization was rapid and rule-based, and the initial ownership structures emerging from privatization in 1994 can be considered exogenous with respect to future performance, capital structures, and dividend policies of firms. The suboptimality of the first ownership structures was confirmed by the rapid reallocation of shares across new owners in 1995-1996.[12] The 1995-1996 ownership changes were massive, unregulated, and frequently unobservable to outsiders. Investors—especially privatization funds—engaged in direct swaps of large blocks of shares and off-market share trading was common. The first ownership patterns that were consistent with market economy principles emerged in 1996 and hence we chose this year as the beginning of our analysis.

[10] Before privatization, firms were transformed into joint stock companies. After incorporatization the firms' current management had to submit privatization proposals and other individuals and institutions submitted competing proposals. The privatization proposal was a business plan, which determined the equity share offered in the voucher scheme to the public and the stake that remained in state hands in the form of temporary or permanent holdings. The Ministry of Privatization picked and approved the winning proposal. If a direct domestic or foreign investor had been identified who was willing to buy (part of) the firm, the required stake in the firm was sold to the investor and the rest was offered in the voucher scheme. The level of managerial and employee ownership was low. In the first wave, only a limited number of firms ended up with managerial or employee ownership; in the second wave, more firms did, but the ownership stakes were low. Also, only very limited restructuring happened prior to privatization.

[11] See Kočenda (1999) for a detailed description of how chains of ownership linked banks, investment privatization funds, and industrial companies.

[12] Cull et al. (2001) document how quickly post-privatization dispersed ownership structure became increasingly concentrated in 1995-1996.

In 2003, the last year of our analysis, the Czech Republic was characterized by private ownership, competitive product markets with unregulated prices, business law to a large extent compliant with EU rules, a private banking sector, stock market, and an economy with links to all major developed countries in the world. In May 2004 the country was integrated into the EU.

3.2. Legal Framework

A new corporate law which reflected market economy principles was introduced in 1993. Since lawmakers were well behind the economic activity, Czech law was incomplete and kept changing literally every year.[13] As a result, only very fundamental and robust ownership rights were effectively enforced. High legal uncertainty and weak/slow law enforcement[14] suggest that, in the period we analyse, shareholders acted based on fundamental rights derived from ownership. More subtle rights, e.g., rights protecting minority shareholders, were either nonexistent or very poorly enforced. The ownership structures that were evolving in this environment reflected its specific conditions, and large shareholding was quite naturally the most important control device. Only highly concentrated owners are able to control managers effectively and, on the other hand, because of the underdeveloped legal system and financial market, dispersed ownership structures cannot enjoy benefits from greater market liquidity and better risk diversification.[15] Overall, Czech corporate ownership structures are very different from those of

[13] To illustrate the situation we describe the evolution of the income tax law in detail. The modern tax system implemented from 1993 was completely novel for most of the citizenry as well as for the public administration. Regulatory institutions and enforcement procedures developed gradually and the tax law was amended many times. During 1993–2002, there were 43 amendments —approximately one modification every quarter. Not only did the income tax law change substantially in character, it also became extensive. The first version of the law contained fewer than 14 thousands words, whereas the one in 2002 was composed of nearly 57 thousand words. Income tax law modifications were typically introduced to correct previous mistakes or to launch new policies, though sometimes they emerged in reaction to lobbying. Even tax advisors complain that the law is too difficult for them to follow, so that the ordinary public has little chance of grasping it.

[14] To settle business disputes at court takes a lot of time: for example, lawsuits related to purchase agreements took on average 452, 594, and 655 days to settle at court in 1998, 1999, and 2000 respectively (from statistics of Ministry of Justice of the Czech Republic).

[15] See the survey by Shleifer and Vishny (1997).

large publicly-traded firms from developed countries for which there exists a vast majority of dividend empirical research.

3.3. Taxes

Taxation is one of the key determinants of corporate dividend policy and the different treatment of various types of owners might explain varying dividend policies across ownership structures.[16] We argue that this cannot be the case in the Czech Republic since the marginal tax rate on cash dividends is the same for all types of shareholders and stock repurchases are not used at all. Czech companies distribute dividends from after-tax profits. In the period of our analysis the same dividend tax treatment applied to individuals and corporations. In the case of individuals, income from dividends was taxed at the source separately from all other income using the flat tax rate.[17] The same treatment and rate applied to corporations (including financial institutions). If the receiver was foreign the taxation of dividends was governed by the treaty between the Czech Republic and the country of the receiver. These treaties prevented double taxation of dividends and existed with all major developed countries.[18] Overall, tax considerations or tax clientele effects cannot drive cross-sectional differences in dividend policies.

During 1996-2003 individuals were exempted from the capital gains tax if they held shares for at least 6 months. On the other hand, corporations paid standard income tax on capital gains; the corporate income tax rate was on average close to 30 percent and decreased gradually. Pension, mutual, and investment funds had a preferential lower income tax rate. The described taxation applied to capital gains realized by trading on the stock market, whereas share repurchases were taxed in the same way as cash dividends independent of shareholder type. As expected, we do not observe any share repurchases in the period of our analysis in the Czech Republic.

---

[16] See Allen et al. (2000) and Dhaliwal et al. (1998), for example.

[17] In 1996-1998 the income dividend tax rate was 25 percent and from 1999 it was lowered to 15 percent.

[18] Foreign owners in our sample are mainly from the EU and we have very few foreign owners incorporated in offshore centres or low-income-tax countries.

4. Model

4.1. Ownership Structures

Our data allows us to track ownership in line with how Czech corporate law assigns control rights to different ownership levels. We distinguish three ownership categories: majority ownership (more than 50 percent of shares)[19], blocking minority ownership (more than 33.3 but not more than 50 percent of shares), and legal minority ownership (at least 10 but not more than 33.3 percent of shares).[20] A majority owner has the right to select management and a supervisory board, to decide whether the company distributes profits as dividends or reinvests them, and to adopt almost all decisions at general shareholders' meetings. Blocking minority ownership gives the right to block some decisions at general shareholders' meetings, mainly those related to implementing major changes in business activities and changing the firm's capital structure.[21] Finally, legal minority ownership can be considered a form of dispersed ownership since its direct impact on business decisions is limited. On the other hand, the corporate law entitles minority shareholders to call a general shareholders' meeting to decide on issues put on the meeting's agenda by a minority shareholder.[22] The ability to identify owners according to these categories is a key to understanding corporate control in the Czech Republic.

Based on these ownership levels we define the following concentration of ownership dummy variables: Majority: The company is controlled by a single majority owner and the next largest owner holds less than 10 percent of equity. Monitored majority: The majority owner is checked by the presence of at least one significant minority owner (either blocking minority or legal minority owner). Minority: The largest owner is only a

[19] We define the majority as holding more than 50 percent of shares or alternatively as holding more than 66.6 percent of shares.

[20] Czech law does not require reporting of stakes of less than 10 percent. This does not restrict our analysis since by having data on all owners with 10 percent and more we are able to estimate the effect of the most relevant degrees of concentration and dispersion of ownership, ranging from a single owner having majority ownership, to no single owner having legal minority ownership.

[21] A blocking minority owner may block a decision to change the articles of incorporation, liquidate the company, issue priority or convertible bonds, issue equity, and increase or decrease equity capital in some other way.

[22] There were some cases in which minority shareholders obstructed a company's operations by delaying implementation of stronger shareholders' decisions through lengthy court proceedings.

blocking minority owner. Dispersed: All shareholders have less than 10 percent of equity. In addition to concentration we are able to identify types of owners: industrial firm, private individual, financial institution, and state. Domicile of the owners is either Czech or foreign.[23]

4.2. Hypotheses

The motives of owners regarding the distribution of profits might vary across ownership stake sizes. Majority owners may maximize shareholder value[24] but they can also loot firms at the expense of small shareholders.[25] After controlling for capital structure and investment opportunities, shareholder value maximization is associated with high dividend payouts. In contrast, if the majority shareholders goal is to loot the firm, dividends are paid less often and the target payout ratio is low.

These predictions are altered if the behaviour of the majority owner is monitored by the presence of a significant minority shareholder. Bargaining between majority and powerful minority shareholder(s) induces the majority shareholder to pay dividends and not to misappropriate profits.[26] Hence we expect the monitored majority ownership structure to be associated with a higher probability to pay dividends and with a higher target payout ratio relative to the majority ownership structure. This pattern is difficult to explain by an alternative story. For example, there is neither theory nor empirical evidence arguing

[23] Type and domicile ownership structure is identified by the type and domicile of the single largest owner (SLO).

[24] Majority owners are expected to have access to more information about the firm and to be able to use more efficient control mechanisms, most importantly a credible threat to dismiss management. In the context of the Czech Republic it was documented that a firm's value and profitability increase with ownership concentration. See Hanousek et al. (2004); Claessens (1997); Claessens and Djankov (1999); or Claessens et al. (1997). This contrasts with a finding by Demsetz and Lehn (1985) from the U.S., that no significant relationship between ownership concentration and profit rates exists.

[25] In the Czech Republic, this behaviour was extensively documented by Cull et al. (2001).

[26] This result is documented by Gugler and Yurtoglu (2003) for Germany. They show that dividend change announcements provide new information about the conflict between a controlling owner and small outside shareholders. "Majority-controlled and unchecked" firms have the smallest target payout ratio, "majority-controlled and checked" firms have the largest target payout ratio, and minority-controlled firms lie in between. This implies that minority shareholders with large stakes press successfully for dividends to be paid out, consistent with the rent extraction hypothesis.

that the size of ownership stake is systematically linked to varying rates of time preference or different evaluation of investment opportunities.

Firms with dispersed ownership structures might not suffer from misappropriating efforts of the majority shareholder but dispersed owners might be weak in exercising their power against management. On the other hand, since in dispersed ownership private benefits of control are diluted among large number of shareholders, dividend payments are the only effective way to disseminate profits and we expect these firms to have a high target payout ratio. We also expect some dividend smoothing as free cash flow theory predicts for cases when asymmetric information is high.

For many reasons we expect foreign owners to behave differently from Czech owners. Foreign owners have better business, managerial, and corporate governance expertise than do Czech owners. On the other hand, foreign owners are less familiar with local corporate, employment, and other laws relevant to the operations of the firms they own, and they have to overcome some additional, e.g., language or cultural, barriers. Therefore, agency conflicts and asymmetric information between foreign owners and management/other domestic owners are different than those between management and Czech owners. With better business know-how and knowledge of technology, foreign investors can assess the profitability of firms[27] and collect these profits as dividends to prevent managers from misappropriating them.[28] Due to ability to tap more developed capital markets foreign owners have easier access to external finance sources relative to Czech owners. At the same time, we expect foreign owners to loot firms less than would Czech owners since foreign owners have a bigger reputation at stake and are subject to more stringent corporate governance (discipline imposed by more developed capital markets) in their home countries. Also, foreign owners in our sample are predominantly industrial firms and financial institutions, while we have many individuals and state institutions among Czech

[27] In the context of the Czech Republic, this argument is supported by Claessens and Djankov (1999) or Hanousek et al. (2004), who show that foreign ownership is associated with improved performance.

[28] Hines (1996) finds that U.S. corporations pay dividends out of their foreign profits at roughly three times the rate they do out of their domestic profits. In a related paper, Desai et al. (2002) analyze dividend remittances by foreign affiliates of U.S. multinational firms. The fact that parent firms are willing to incur tax penalties by simultaneously investing funds while receiving dividends from foreign affiliates allows Desai et al. to argue that payout policies are largely driven by the need to control managers of foreign affiliates by diverting funds.

owners as well. Overall, we expect firms with foreign ownership to have a higher target payout ratio and to pay dividends more often relative to Czech owners and we provide key results for ownership concentration separately for domestic and foreign owners.

In our sample majority owners from the financial sector are banks, bank-sponsored funds, and insurance companies. Banks are usually described in the literature as good monitors, and a combination of equity ownership and debt claims can reduce the shareholder-debtholder conflict. In the Czech Republic, banks seem to serve an especially positive role in corporate governance since the profitability and value of firms under bank ownership is high.[29] Despite increasing profitability, however, the effect on dividend policy has to be qualified by the fact that paying high dividends could endanger banks' loans. After controlling for this effect we expect financial institutions with large shareholding to impose financial discipline and aim at high dividend payout ratios. We expect no looting from banks as they are subject to much stricter regulation and care more about their reputation than do industrial firms and individuals. We also expect low dividend smoothing since information asymmetry in the case of bank ownership is small.

Finally, the most common owners among state-controlled firms are municipalities and especially the National Property Fund.[30] This suggests that dividends paid within this category will be determined by the political process without aiming for a specific target payout ratio or the level of dividend smoothing.

4.3. Estimation

Our specification of dividend payoffs builds upon the seminal model by Lintner (1956):[31]

$$D_{i,t} = \beta_i + \alpha_i \tau_i \pi_{i,t} + (1 - \alpha_i) D_{i,t-1} + \varepsilon_{i,t}, \tag{1}$$

---

[29] See Claessens and Djankov (1999) and Claessens et al. (1997).

[30] The National Property Fund manages shareholdings of the Czech state and sells these ownership stakes over time by direct sales or auctions mainly to foreign investors.

[31] As noted by Benartzi et al. (1997): "... the conclusion we draw from [our] analysis is that Lintner's model of dividends remains the best description of the dividend setting process available."

where $D_{i,t}$ is dividend per share company i pays in year t, $\pi_{i,t}$ denotes earnings per share company i reports in year t, τ_i is the target payout ratio of company i, and $\epsilon_{i,t}$ is the error term. Parameters α_i and $1-\alpha_i$ correspond to the weight placed on current earnings and lag dividends, respectively. In order to test our hypothesis that dividend payments vary with ownership in our sample we augment specification (1) by ownership:

$$D_{i,t} = \sum_j \left[\beta_j + \alpha_j \tau_j \pi_{i,t} + (1-\alpha_j) D_{i,t-1}\right] OWN\,(j)_{i,t} + \zeta_{i,t}, \tag{2}$$

where $OWN\,(j)_{i,t}$ is a dummy variable equal to 1 if company i belongs to ownership structure j in year t and is zero otherwise. With respect to chosen ownership structure $OWN\,(j)_{i,t}$, parameter τ_j of model (2) reflects the target payout ratio of ownership structure j, and parameters α_j and $1-\alpha_j$ correspond to the weight placed on current earnings and lag dividends, respectively. Ownership structure as entered in (2) can be easily specified to account for majority/monitored majority/minority/dispersed concentration level as well as its interaction with domicile and type of owner.

A direct application of Lintner's model suffers on several fronts in an emerging market environment. First, we do not observe a majority of firms paying dividends (less than ten percent of our sample) and hence a direct application of Lintner's model leads to biased results due to sample selection (see Heckman, 1979). Second, due to weak market supervision and regulation enforcement we have to address the problem of missing financial data for firms that do not pay dividends (in the case of the Czech Republic it reduces the original data panel to less than half of a fully defined data point). Third, we study dividend payments shortly after privatization, when ownership is potentially endogenous with respect to corporate performance (e.g., state versus private, domestic versus foreign). Since profit influences dividends we therefore expect a bi-directional link between ownership structure and the decision to pay dividends.

To address sample selection biases (missing data and a relatively low frequency of observed dividends) and ownership endogeneity we model dividend payments as a two stage process. In the first stage, firms decide whether a dividend will be paid or not, while in the second stage the size of a dividend payment is decided. Technically, this approach is a Heckit regression, in which we model separately the decision to pay dividends as a 0-1 variable (the first stage) and, in the second stage, we estimate specification (2) for those firms paying dividends. Based on a thorough discussion provided by Angrist and Krueger (2001) we use a linear probability model instead of probit in the first stage. The linear probability model allows us to instrument ownership and provides consistent estimates under standard assumptions, while probit regression with plugged predicted values of ownership "do not generate consistent estimates unless the nonlinear model happens to be exactly right, a result which makes the dangers of misspecification high" (ibid). Also, the linear probability model can be corrected for sample selection. We redo the first stage using probit as a robustness check.

Besides its easy implementation, each estimation stage sheds light on the dividend decision process: 1) linear probability regression (2SLS/IV) used as the first step provides a clear-cut decision if the company pays dividends in a given year; 2) the ordinary least square method, which we run on a subset of companies that decided to pay dividends, estimates what influences the size of dividends in a Lintner-type specification augmented by various ownership structures. Formally, the whole estimation logistics is described in the next section.

4.3.1. *Two Stage Process for Dividend Payout*

STAGE 1: We estimate the decision to pay dividends (0-1 variable) as a linear probability regression model:

$$\begin{aligned}\mathbf{I}[D_{i,t} > 0] \;=\; & \sum_j p(j) \cdot OWN\,(j)_{i,t} + CONTROLS_{i,t} + \\ & + EFFICIENCY_{i,t} + t \cdot TAX_{96-98} + \\ & + d \cdot DIV_{i,t-1} + \lambda_1 \cdot M1_{i,t} + \eta_{i,t}, \qquad (3)\end{aligned}$$

where $OWN\,(j)_{i,t}$ is a dummy variable equal to 1 if company i belongs to ownership structure j in year t and coefficient $p(j)$ is the probability with which the ownership structure j pays dividends. As controls ($CONTROLS_{i,t}$) we use financial variables: total assets, total liabilities to total assets, bank loans to total liabilities, cash holdings to total assets, and the growth rate of average sales in the industry the firm is part of, excluding the firm itself. After controlling for capital structure and investment opportunities, the only variables that might drive the decision to pay dividends from outside the shareholders' perspective are efficiency measures: profit (or total sales) to total assets and total sales to total labour costs. We include these variables in model (3) as $EFFICIENCY_{i,t}$. To account for a change in dividend taxation in the period of our analysis we include a dummy variable TAX_{96-98} which is equal to 1 for the time period with a higher dividend income tax rate (1996-1998). We also include dummy variable $DIV_{i,t-1}$ that is equal to 1 if the firm paid dividends in the last year. We estimate model (3) using the instrumental variable approach (the set of instruments for ownership variables is described and discussed in detail in the next subsection).

Variable $M1_{i,t}$ in (3) stands for an inverse Mills ratio which we use to address the issue of missing financial data. Mills ratio comes from the following probit regression (which we run as a "0 stage") with missing financial data in our sample as a binary response:

$$\begin{aligned}\mathbf{I}[MissF] &= f(const, TNS_i, NSVP_i, MissF\_91/93_i, \\ &\quad AP_i, IPF_i, II_i) + \vartheta_{i,t}, \end{aligned} \tag{4}$$

where TNS_i denotes the original total number of shares[32] in the voucher privatization scheme (in 1992); $NSVP_i$ denotes the number of shares offered under the voucher privatization scheme; $MissF\_91/93_i$ stands for a set of 0/1 indicators of missing financial data (profit, sales, debt, and the number of employees) prior to privatization (in 1991-1993); AP_i is the average price for which the shares were sold in the voucher scheme; IPF_i and II_i denote total holdings (in percent) of the investment privatization funds after the voucher scheme (here we consider also disaggregation to the five largest owners) and individual investors, respectively.

STAGE 2: We estimate the decision about the size of dividends paid on a subset of firms paying dividends (i.e., $D_{i,t} > 0$). The final specification we use is an extension of (2):

$$D_{i,t} = \sum_j \left[\beta_j + \alpha_j \tau_j \pi_{i,t} + (1-\alpha_j) D_{i,t-1}\right] OWN(j)_{i,t} +$$

[32] This is equal to the book value (or subscribed capital), since original shares were issued in the nominal value of 1,000 CZK per share.

$$+CONTROLS_{i,t} + \lambda_2 \cdot M2_{i,t} + \nu_{i,t}. \tag{5}$$

We follow the established dividend literature (e.g., Fama and French, 2001), and use the following control variables ($CONTROLS_{i,t}$) to isolate corporate dividend policy from firms' capital budgeting and borrowing decisions: Firm Size (Total assets, $TA_{i,t}$; we expect a positive relationship), Leverage (Total liabilities as a fraction of total assets, $\frac{TL}{TA}_{i,t}$; we expect a negative relationship), Bank Power (Bank loans as a fraction of total liabilities, $\frac{BL}{TL}_{i,t}$; we expect a negative relationship but this effect might interact with the aggregate leverage measure), Cash Holdings (Cash as a fraction of total assets, $\frac{CH}{TA}_{i,t}$; we expect a positive relationship), and Investment Opportunities (Growth rate between the current year and the following year of average sales in the industry the firm is part of, excluding the firm itself, $grSA_{i,\frac{t+1}{t}}$; we expect a negative relationship).[33] We also include dummy variables for every year. Since less than ten percent of firms in our sample pay dividends, we add the inverse Mills ratio, $M2_{i,t}$, computed from regression (3) to remove the sample selection bias.

While estimating (5) we test for ownership endogeneity by employing a Hausman-type test for specification. In contrast to the first stage, ownership endogeneity is rejected in all second stage specifications and hence we employ simple OLS regression.

4.3.2. *Instruments Used for Endogeneity of Ownership in Dividend Payment Process*

As instruments for ownership variables we use pre-privatization data coming from detailed information on all proposed privatization projects that were submitted to the government before privatization, and data related to voucher privatization (voucher privatization bids) available at the Ministry of Finance. We have available all existing pre-privatization financial data, together with the ownership structure specified in the winning privatization proposal. Despite the fact that all our IVs are strictly pre-determined through time,

[33] Accounting variables: Earnings, total assets, total liabilities, bank loans, cash holdings, and sales come from audited accounting statements as published by companies in their filings to the Prague Stock Exchange. We use consolidated statements if available. All accounting statements are based on Czech accounting law and standards. Cash is defined as the sum of two items in Czech accounting statements: "Cash in hand" and "Cash in transit". Sales are named as "Sales of own production, services, and goods bought for resale" in the Czech accounting statements. We include Bank Power to control for the possibility that a commercial bank is a shareholder and a debtholder at the same time. This is quite common in our sample.

we employ the Sargan test of overidentifying restrictions and use only a subset of variables that do not interfere with the formal test at the 10% significance level or stricter.[34]

The full set of available instruments consists of a set of regional (REG_i) and industrial (IND_i) dummies; basic accounting variables (sales, profit, and debt) from 1991-1993 (FIN_i); TNS_i, the total number of shares (the share of each company was set at the same nominal value before large-scale privatization); the set of variables collected from the database of privatization projects: NP_i, which refers to the number of privatization projects submitted to the government in 1991; $VPOWN_i$, which stands for the ownership structure proposed by the government in 1991 in the winning privatization project—expressed in percentage intended for certain ownership types (state, municipalities, foreign and domestic owners, intermediaries, etc.); and the information coming from the voucher privatization scheme: AP_i, the average price per share of a company in the voucher privatization scheme (this reflects the demand for a particular firm in the privatization process). In addition, since we have a relatively unique dataset on privatization outcomes, we also have information on the proportion of company shares allocated to investment privatization funds IPF_i (in the estimation we consider five additional variables containing the holdings of the five largest investment funds) and individual investors II_i, respectively, during large-scale privatization in 1992-1994.[35]

5. Data and Summary Statistics

Our analysis is based on data from 1996 to 2003 on the complete population of 1,664 medium and large firms privatized in 1991-1994 and consequently traded on the Prague Stock Exchange, which constituted most of the country's economic activity in the late

[34] Some of the pre-determined variables do not pass the test of being strictly exogenous and hence we do not use them in certain equations. For example, percentage of the firm's shares to be sold to foreign owners (as proposed in a winning project) typically does not pass the Sargan test.

[35] The effects of variables such as the firm's total number of shares and shares allocated to the institutional and individual investors may be nonlinear, so we use a Taylor series expansion of the third order to obtain a specification that can take into account potential nonlinearities.

1990s. Financial and ownership data come from the private database ASPEKT.[36] Data for the privatization period come from the Ministry of Privatization of the Czech Republic. To estimate dividend equations we use data from 1996-2003 (post-privatization market economy period). We use data from 1991-1994 (privatization period) as instrumental variables that allow us to control for the endogeneity of ownership.

Companies with dispersed ownership seem to be big, not profitable, and dividend-paying. The most effective firms are those with monitored majority ownership, but they seem to pay the lowest dividends among the concentration structures we consider. Majority controlled firms are the smallest and seem to pay the largest dividends (see Table I). Czech controlled companies seem to be on average smaller, more leveraged, and seem to pay lower dividends relative to companies controlled by foreigners. Czech controlled firms are also not profitable (see Table II). State-controlled companies are on average the largest and, surprisingly, seem to pay the highest dividends among all ownership types; they are profitable and less levered than firms from other control groups, but have relatively low sales relative to assets. Companies controlled by individuals are on average small and have low profitability, yet still seem to pay some dividends. Companies controlled by financial institutions seem to be just profitable, have the highest leverage, and pay very low dividends. Companies controlled by industrial firms seem to pay no dividends at all and to have the highest sales relative to staff costs (see Table III).

The total number of dividends paid is evenly spread over the whole period we analyse (see Table IV). In the category Foreign and Financial we observe just a few dividend payments. In the category Czech (or Foreign) and Industrial, SLOs seem to be well spread across many industries. We observe very few dividends paid by firms in which SLO is an individual (Czech or Foreign).

[36] ASPEKT collets data mainly from the Prague Stock Exchange and the Czech Statistical Office. This database is the Czech source for AMADEUS, a pan-European database containing financial statements data.

6. Results

Table V reports estimates from the stage one regression describing the decision to pay dividends for the entire sample of 1,664 firms over the period 1996-2003, and Table VI reports estimates from the stage two regression describing the conditional decision about the size of dividends paid over the same period. All regressions contain the full set of ownership structure dummies; the residual group of firms not assigned to any ownership category is denoted as "Other". We present three specifications which differ based on how we cut the sample according to ownership: domicile, concentration combined with domicile, and type.

The Czech largest owner has a positive but small effect on the probability to pay a dividend, 0.11 significant at the 1% level (column "Domicile" in Table V). If the largest owner is foreign, the probability to pay a dividend is positive and the effect is very large: 0.35 significant at the 1% level. In line with this, the target dividend payout ratio (column "Domicile" in Table VI) for foreign-owned firms of 0.46 (significant at the 1% level) is substantially higher than Czech-owned firms at 0.12 (significant at the 5% level). These results are consistent with the hypothesis that foreigners use dividends to distribute profits more often and aim at a higher target payout ratio than Czechs (the difference in the target payout ratios is significant at the 1% level).

The main results are reported in the column "Concentration" in Tables V and VI. The probability that a firm with a Czech majority owner pays a dividend is 0.09 (significant at the 5% level). If the Czech majority owner is accompanied by a significant minority shareholder the probability increases to 0.16 (significant at the 1% level). The same pattern holds for foreigners. The probability that a firm with a foreign majority owner pays a dividend of 0.26 (significant at the 1% level) is a lot lower than the same probability if the majority owner is accompanied by a significant minority shareholder 0.58 (significant at the 1% level). The associated target payout ratios for these ownership structures ("Concentration" column in Table VI) are as follows: positive but not significant for the Czech majority ownership structure; 0.82 (significant at the 1% level) for the Czech monitored majority ownership structure; 0.61 (significant at the 1% level) for the foreign majority ownership structure; and 0.86 (significant at the 1% level) for the foreign monitored majority ownership structure. The difference in target payout ratios for Czech majority controlled and Czech monitored majority controlled firms is significant at the 10% level, but the same test of difference of target payout ratios for firms with a foreign largest owner is significant only at the 34% level. This set of results supports our hypothesis that significant minority shareholders limit rent extraction by increasing the probability that a dividend is paid and increasing the target payout ratio. This holds both for Czech as well as for foreign largest owners after controlling for firm size, performance, investment opportunities, leverage, and bank influence on the firm. Rent extraction and dilution of minority shareholders seems to be associated predominantly with Czech owners.

Ownership by financial institutions (column "Type" in Table VI) is associated with a high target payout ratio of 0.54 (significant at the 1% level) and no dividend smoothing since the weight put on current earnings is 1.0 (significant at the 1% level). In line with predictions of the free cash flow theory this result confirms that financial institutions act as sophisticated monitors that do not rely on dividend smoothing as a controlling mechanism and collect about half of the profits as dividends every year. If the largest owner is a financial institution, the effect on the probability to pay dividends depends on the domicile (column "Type" in Table V). A Czech financial institution has a positive effect on the probability to pay dividends (coefficient 0.24 significant at the 1% level). In contrast, the coefficient associated with a foreign financial institution is 1.22 (significant at the 1% level).

If the largest owner is an industrial firm the target payout ratio is 0.56 (significant at the 1% level) and we observe significant dividend smoothing; the weight associated with current earnings is 0.47 (significant at the 1% level). Industrial owners smooth dividends considerably more than do owners from the financial sector; the difference in weights placed on current earnings is significant at the 1% level. Ownership by private individuals has no effect on the probability to pay dividends (coefficient 0.06 is not significant) and the target payout ratio is not significantly different from zero either. This seems to suggest that private individuals as largest owners do not pay dividends and extract rents instead. The state as an owner is associated with a positive probability that dividends are paid, 0.26 (significant at the 1% level) but decisions about dividend payments do not seem to be consistent with Lintner's model, as neither the weight coefficient nor the target payout ratio coefficient are significant. We believe this is because dividends are paid according to the fiscal needs of the government or municipalities with no aim to establish a target payout ratio.

In Tables V and VI, the ownership category "Dispersed or unknown" contains firms of two types that we cannot distinguish: Firms with dispersed ownership without legal obligation to disclose their owners, and firms that do not report their ownership structure. This makes interpretation of the results difficult since, e.g., firms with both Czech and foreign ownership might have reasons not to disclose their ownership structures. For the "Dispersed or unknown" ownership structure the probability to pay dividends is on average 0.18 (significant at the 1% level in all specifications) and the target payout ratio is large, on average 0.94 across all three specifications (significant at the 1% level). This suggests that dividends are used to distribute profits if there is no large shareholder with a strong incentive to extract rents or to dilute, but our data do not allow us to draw any strong conclusion.

The coefficients in front of the control variables have similar signs as found in the previous literature in both regressions: Firm size has a positive and significant effect on the probability to pay dividends and seems to increase the target payout ratio. Leverage and the strength of bank presence has a small negative effect on the probability of paying dividends and a strong negative effect on the size of dividends. Investment opportunities on the industry level have a negative effect both on the probability to pay dividends and on the target payout ratio. The large positive effect of dividend history (on average 0.59, significant at the 1% level in all specifications) supports the use of Lintner's model. The decrease of dividend income tax positively contributes to the probability to pay dividends. Finally, earnings-per-total assets and sales to staff costs measures of efficiency have a positive and weakly significant effect on the probability to pay dividends.

7. Robustness Checks

7.1. Variables Definition

The use of different earnings measures in Equations (3) and (5): operating profit before income tax, profit including/excluding extraordinary items, or after tax profit has no impact on results reported in Tables V and VI.

We use total sales instead of total assets as a measure of a firm's size, bank loans as a fraction of total assets instead of total liabilities as a fraction of total assets as an alternative measure of leverage, and cash holdings including or excluding marketable securities[37]. These changes in control variables have again no impact on our results in Tables V and VI.

7.2. Investment Opportunities

As alternative measures of investment opportunities we use the growth rate of total assets, earnings, or value added in the industry the firm is part of (excluding the firm itself). We tried growth rates both between the current year and the following year, and between the previous year and the current year. In all these specifications the results are unchanged.

Finally, we use the firm-level growth rate of total assets (or total sales) in combination with industry dummy variables instead of various industry-level growth rates. Tables VII and VIII have the same structure as Tables V and VI, respectively, and report results from these regressions. The coefficients in front of ownership variables remain to a large extent unchanged and confirm corporate dividend behaviour found in the main specification: Firms with a dominant majority owner pay dividends less often and their target payout ratio is small. In contrast, firms with a majority owner and at least one strong minority owner pay dividends more often and the target payout ratio is large.

---

[37] We add the item "Cash and investments" to the cash variable used in the main specification. In Czech accounting statements this item includes short-term investments in very liquid financial assets.

7.3. Decision to Pay Dividends

We estimate the stage 1 decision to pay dividends using a probit regression:

$$\mathbf{I}[D_{i,t} > 0] = g\Big(const, OWN\,(j)_{i,t}\,, CONTROLS_{i,t}, \\ EFFICIENCY_{i,t}, TAX_{96-98}, DIV_{i,t-1}) + \xi_{i,t}, \qquad (6)$$

where variables on the right hand side are the same as in (3). To account for the endogeneity of ownership we estimate predictions of ownership variables $OWN\,(j)_{i,t}$ from a reduced form equation and plug them into the decision to pay dividends equation (6). Since $\Pr(\widehat{ownership} = j)$ converge to $\Pr(ownership = j)$, by inserting the predicted values of the ownership variables into (6) we obtain consistent estimates of average partial effects. Formally, we run the following probit regression:

$$OWN(j)_{i,t} = h\,(INSTRUMETS_{i,t}) + \varsigma_{i,t}, \qquad (7)$$

with the same instrumental variables as in the main specification (3). Predicting ownership dummy variables is difficult since in some ownership groups we have a small number of observations and thus to receive feasible estimates we have to broaden the ownership categories. We employ this estimation approach as it is used in the literature and we are aware of all problems described in Blundel and Smith (1994). Also, correcting for sample selection bias—important in our sample—is not possible in this case.

Table IX has the same structure as Table V and reports results from regression (6). The results are broadly consistent with the one from the linear probability model: in the probit case Czech majority owners decrease the probability of paying dividends (marginal effect -0.03, significant at the 1% level) whereas foreign majority owners increase the probability of paying dividends (marginal effect 0.02, significant at the 10% level). The results for concentration are not significant. This is because the lack of observations prevents us from combining concentration with domicile and therefore the effects of Czech and foreign ownership are mixed. Financial institutions have a positive effect on the probability of dividend payments (consistent with the results in Table V) while industrial firms have a negative effect. Results for individuals and the state are not significant in this specification.

8. Conclusion

The key agency costs in firms with concentrated ownership shift from the traditional owner-manager conflict to the dominant shareholder's incentive to consume private benefits at the expense of other minority shareholders. The question whether this rent extraction takes place, how significant is it, and whether minority shareholders are able to monitor large shareholders in order to preclude such consumption is answered in this paper.

We find that corporate dividend policy in an emerging market economy depends on concentration and domicile of ownership. Firms with a dominant majority owner pay dividends less often and their target payout ratio is small. In contrast, firms with a majority owner and at least one strong minority owner pay dividends more often and the target payout ratio is large. We interpret these results as evidence that dominant owners extract rents from firms and that strong minority shareholders can prevent this behaviour. This dividend pattern holds both for domestic and foreign largest owners though domestic owners do enjoy significantly higher rents. The results are robust to alternative definition of key ownership variables, the way we measure firms' investment opportunities and efficiency, and use of an alternative estimation technique.

Our analysis of expropriation from the perspective of dividends does provide quantitative evidence on the expropriation that takes place within Czech companies. Expropriation by corporate insiders is not simply a matter of redistribution amongst shareholders only. It is damaging more generally as corporate insiders might choose to invest in projects with low or negative returns just because they create opportunities for expropriation. Investment decisions are hence distorted and corporate growth is slower than it could be. Such inefficient investment behaviour, if undertaken by a large number of firms, has adverse effects on the whole economy. This is of an exceptional interest in countries like Czech Republic which struggle to catch up with developed economies of Western

Europe. Each dollar available for investing should be allocated to growth opportunities with the highest returns and the investment decision should not be based on what projects make expropriation easy. To address these problems regulators should, first, strengthen the rights of minority shareholders to enable them to limit expropriation. Second, and more importantly, regulators should support the development of sound and transparent financial markets prevalent in Western Europe as they seem, based on extensive both anecdotal and research evidence, to police dominant owners most effectively. We expect similar results to hold in countries with a comparable institutional framework, i.e., where fundamental ownership rights are honoured but capital markets and corporate governance mechanisms are underdeveloped.

9. References

Allen, F., Bernardo, A., and Welch, I. (2000) A Theory of Dividends Based on Tax Clienteles, *Journal of Finance* **55**, 2499-2536.

Amemiya, T. (1985), *Advanced Econometrics, Cambridge*, Harvard University Press.

Amihud, Y. and Murgia, M. (1997) Dividend, Taxes, and Signaling: Evidence from Germany, *Journal of Finance* **52**, 397-408.

Angrist, J.D. and Krueger, A.B. (2001) Instrumental variables and the search for identification: From supply and demand to natural experiments, *Journal of Economic Perspectives* **15**, 69-85.

Barclay, M. and Holderness, C. (1989) Private Benefits from Control of Public Corporations, *Journal of Financial Economics* **25**, 371-395.

Barclay, M. and Holderness, C. (1992) The Law and Large-block Trades, *Journal of Law and Economics* **35**, 265-294.

Benartzi, S., Michaely, R., and Thaler, R. (1997) Do Changes in Dividends Signal the Future or the Past? *Journal of Finance* **52**, 1007-1034.

Bergström, C. and Rydqvist, K. (1990) Ownership of Equity in Dual-class Firms, *Journal of Banking and Finance* **14**, 255-269.

Bernheim, D. (1991) Tax Policy and the Dividend Puzzle, *Rand Journal of Economics* **22**, 455-476.

Bernheim, D. and Wantz, A. (1995) A Tax-Based Test of Dividend Signaling Hypothesis, *American Economic Review* **85**, 532-551.

Bhattacharya, S. (1979) Imperfect Information, Dividend Policy, and 'The Bird in the Hand' Fallacy, *Bell Journal of Economics* **10**, 259-270.

Blundel, R. and Smith, R. (1994) Coherency and estimation in simultaneous models with censored or qualitative dependent variables, *Journal of Econometrics* **64**, 355-373.

Bolton, P. and von Thadden, E. L. (1998) Blocks, Liquidity, and Corporate Control, *Journal of Finance* **53**, 1-25.

Brook, Y., Charlton, W., and Hendershott, R. (1998) Do Firms Use Dividends to Signal Large Future Cash Flow Increases? *Financial Management* **27**, 46-57.

Burkart, M., Gromb, D., and Panunzi, F. (2000) Agency Conflicts in Public and Negotiated Transfers of Corporate Control, *Journal of Finance* **55**, 647-677.

Claessens, S. (1997) Corporate Governance and Equity Prices: Evidence from the Czech Republic, *Journal of Finance* **52**, 1641-1658.

Claessens, S. and Djankov, S. (1999) Ownership Concentration and Corporate Performance in the Czech Republic, *Journal of Comparative Economics* **27**, 498-513.

Claessens, S., Djankov, S., and Pohl, G. (1997) Ownership and Corporate Governance, Policy Research Working Paper No. 1737, World Bank.

Coffee, J. (1996) Institutional Investors in Transitional Economies: Lessons from the Czech Experience, in: R. Frydman, C. Gray, and A. Rapaczynski (eds.), *Corporate governance in Central Europe and Russia*, Vol 1. Banks, funds, and foreign investors, Central European University Press, Budapest, 111-186.

Cull, R., Matesová, J., and Shirley, M. (2001) Ownership Structure and the Temptation to Loot, Policy Research Working Paper No. 2568, World Bank.

DeAngelo, H., DeAngelo, L., and Skinner, D. (2004) Are dividends disappearing? Dividend concentration and the consolidation or earnings, *Journal of Financial Economics* **72**, 425-456.

DeAngelo, H., DeAngelo, L., and Stulz, R. (2004) Dividend Policy, Agency Costs, and Earned Equity, NBER Working Paper No. 10599.

Demsetz, H. and Lehn, K. (1985) The Structure of Corporate Ownership: Causes and Consequences, *Journal of Political Economy* **93**, 1155-1177.

Desai, M., Foley, F., and Hines, R. (2002) Dividend Policy Inside the Firm, NBER Working Paper No. 8698.

Dewenter, K. and Warther, V. (1998) Dividends, Asymmetric Information, and Agency Conflicts: Evidence from a Comparison of the Dividend Policies of Japanese and U.S. Firms, *Journal of Finance* **53**, 879-904.

Dhaliwal, D., Erickson, M., and Trezevant, R. (1998) A Test of the Theory of Tax Clienteles for dividend Policies, working paper, University of Arizona.

Easterbrook, F. (1984) Two Agency-Cost Explanations of Dividends, *American Economic Review* **74**, 650-659.

Faccio, M., Lang, L., and Young, L. (2001) Dividends and Expropriation *American Economic Review* **91**, 54-78.

Fama, E. and French, K. (2001) Disappearing dividends: Changing characteristics or lower propensity to pay? *Journal of Financial Economics* **60**, 3-43.

Feldstein, M. and Green, J. (1983) Why Do Companies Pay Dividends? *American Economic Review* **73**, 17-30.

Gugler, K. (2003) Corporate Governance, Dividend Payout Policy, and the Interrelation Between Dividends, R&D, and Capital Investment, *Journal of Banking and Finance* **27**, 1297-1321.

Gugler, K. and Yurtoglu, B. (2003) Corporate Governance and Dividend Pay-out Policy in Germany, *European Economic Review* **47**, 731-758.

Gupta, N., Ham, J., and Švejnar, J. (2001) Priorities and Sequencing in Privatization: Theory and Evidence from the Czech Republic, The William Davidson Institute Working Paper No. 323.

Hanousek, J., Kočenda, E., and Švejnar, J. (2004) Ownership, Control and Corporate Performance after Large-Scale Privatization, The William Davidson Institute Working Paper No. 652.

Hausman, J. (1978) Specification Tests in Econometrics, *Econometrica* **46**, 1251-1271.

Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica* **47**, 153-161.

Hines, J. (1996) Dividends and Profits: Some Unsubtle Foreign Influences, *Journal of Finance* **51**, 661-689.

Jensen, M. (1986) Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers, in The Market for Corporate Control Papers and Proceedings of the Ninety-Eighth Annual Meeting of the American Economic Association, *American Economic Review* **76**, 323-329.

John, K. and Williams, J. (1985) Dividends, dilution, and taxes: A Signaling Equilibrium, *Journal of Finance* **40**, 1053-1070.

Johnson, S., McMillan, J., and Woodruff, C. (2002) Property Rights and Finance, *American Economic Review* **92**, 1335-1356.

Kočenda, E. (1999) Residual State Property in the Czech Republic, *Eastern European Economics* **37**, 6-35.

Kotrba, J. (1995) Privatization Process in the Czech Republic: Players and Winners, in: J. Švejnar, *The Czech Republic and economic transition in Easter Europe*, Academic Press, San Diego, London, and Toronto, 159-198.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R. (2000) Agency Problems and Dividend Policies around the World, *Journal of Finance* **55**, 1-33.

La Porta, R., Lopez-de-Silanes, F., and Shleifer, A. (1999) Corporate Ownership around the World, *Journal of Finance* **54**, 471-517.

Lang, L. and Litzenberger, R. (1989) Dividend Announcement: Cash Flow Signaling vs. Free Cash Flow Hypothesis? *Journal of Financial Economics* **24**, 181-191.

Lintner, J. (1956) Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes, *American Economic Review, Papers and Proceedings of the Sixty-eight Annual Meeting of the American Economic Association* **46**, 97-113.

Miller, M. and Rock, K. (1985) Dividends Policy under Asymmetric Information, *Journal of Finance* **40**, 1031-1051.

Miller, M. and Scholes, M. (1978) Dividend and Taxes, *Journal of Financial Economics* **6**, 333-364.

Offer, A. and Siegel, D. (1987) Corporate Financial Policy, Information, and Market Expectations: An Empirical Investigation of Dividends, *Journal of Finance* **42**, 889-911.

Shleifer, A. and Vishny, R. (1986) Large Shareholders and Corporate Control, *Journal of Political Economy* **94**, 461-488.

Shleifer, A. and Vishny, R. (1989) Management Entrenchment: The Case of Manager-Specific Investments, *Journal of Financial Economics* **25**, 123-139.

Shleifer, A. and Vishny, R. (1997) A survey of corporate governance, *Journal of Finance* **52**, 737-783.

Švejnar, J. and Singer, M. (1994) Using Vouchers to privatize an Economy: The Czech and Slovak Case, *Economics of Transition* **2**, 43-64.

Wooldridge, J. M. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, Massachusetts.

Zingales, L. (1994) The Value of the Voting Right: A Study of the Milan Stock Exchange Experience, *Review of Financial Studies* **7**, 125-148.

Zwiebel, J. (1996) Dynamic Capital Structure under Managerial Entrenchment, *American Economic Review* **86**, 1197-1215.

10. Appendix

Table I. Ownership concentration: descriptive statistics

The sample consists of 1,664 firms over the period 1996-2003. These firms are all medium and large companies privatized in the Czech Republic by 1994. Ownership concentration structures are: Majority: The company is controlled by a single majority owner (more than 50 percent of equity) and the next largest owner holds less than 10 percent of equity. Monitored majority: The majority owner (more than 50 percent of equity) is checked by the presence of at least one significant minority owner (either blocking minority, more than 33.3 of equity, or legal minority owner, more than 10 percent of equity). Dispersed: All shareholders have less than 10 percent of equity. Column "Obs" shows the number of firm-years observations in a given category. Variables Profit/Total assets, Liabilities/Total assets, Sales/Total assets, and Sales/Staff costs are weighted by Total assets. Only firms with liabilities less than twice the size of total assets are included.

	Ownership concentration	Mean	Std	Obs
Total assets (mil. CZK)	Majority	1.009	7.935	1,775
	Monitored majority	1.431	8.167	2,235
	Dispersed	1.920	9.037	1,866
Dividend / Profit	Majority	0.040	0.681	1,775
	Monitored majority	0.026	0.291	2,235
	Dispersed	0.032	0.158	1,866
Profit / Total assets	Majority	0.019	0.156	1,719
	Monitored majority	0.042	0.242	2,204
	Dispersed	-0.005	0.120	1,853
Liabilities / Total assets	Majority	0.398	0.283	1,719
	Monitored majority	0.626	0.358	2,204
	Dispersed	0.347	0.238	1,853
Sales / Total assets	Majority	0.935	0.781	1,719
	Monitored majority	1.441	0.874	2,204
	Dispersed	0.799	0.580	1,853
Sales / Staff costs	Majority	8.003	37.294	1,719
	Monitored majority	15.915	38.511	2,204
	Dispersed	6.310	7.718	1,853

Table II. Domicile of ownership: descriptive statistics

The sample consists of 1,664 firms over the period 1996-2003. These firms are all medium and large companies privatized in the Czech Republic by 1994. Domicile of ownership is classified according to the single largest owner. Column "Obs" shows the number of firm-years observations in a given category. Variables Profit/Total assets, Liabilities/Total assets, Sales/Total assets, and Sales/Staff costs are weighted by Total assets. Only firms with liabilities less than twice the size of total assets are included.

	Domicile of ownership	Mean	Std	Obs
Total assets (mil. CZK)	Czech	1.044	7.399	5,786
	Foreign	1.803	7.571	844
Dividend / Profit	Czech	0.012	1.252	5,786
	Foreign	0.068	0.273	844
Profit / Total assets	Czech	-0.009	0.286	5,688
	Foreign	0.051	0.153	827
Liabilities / Total assets	Czech	0.479	0.350	5,688
	Foreign	0.434	0.338	827
Sales / Total assets	Czech	1.103	0.903	5,688
	Foreign	0.954	0.654	827
Sales / Staff costs	Czech	11.420	46.776	5,688
	Foreign	9.624	19.125	827

Table III. Type of ownership: descriptive statistics

The sample consists of 1,664 firms over the period 1996-2003. These firms are all medium and large companies privatized in the Czech Republic by 1994. Type of ownership is classified according to the single largest owner. State ownership includes ownership by municipalities and various state agencies. Owners from the financial sector are predominantly banks, bank-sponsored funds, and insurance companies. Column "Obs" shows the number of firm-years observations in a given category. Variables Profit/Total assets, Liabilities/Total assets, Sales/Total assets, and Sales/Staff costs are weighted by Total assets. Only firms with liabilities less than twice the size of total assets are included.

	Type of ownership	Mean	Std	Obs
Total assets (mil. CZK)	State	6.998	25.537	435
	Individual	0.222	0.441	1,035
	Industrial	0.838	3.692	4,656
	Financial	0.764	2.241	498
Dividend / Profit	State	0.129	1.387	435
	Individual	0.052	1.318	1,035
	Industrial	0.002	1.181	4,656
	Financial	0.011	0.106	498
Profit / Total assets	State	0.029	0.073	429
	Individual	0.009	0.103	1,021
	Industrial	0.038	0.107	4,563
	Financial	0.002	0.099	496
Liabilities / Total assets	State	0.389	0.195	429
	Individual	0.459	0.227	1,021
	Industrial	0.467	0.256	4,563
	Financial	0.490	0.247	496
Sales / Total assets	State	0.427	0.327	429
	Individual	0.875	0.634	1,021
	Industrial	0.791	0.588	4,563
	Financial	0.724	0.489	496
Sales / Staff costs	State	9.048	5.876	429
	Individual	8.238	18.766	1,021
	Industrial	10.857	28.553	4,563
	Financial	6.616	5.541	496

Table IV. Number of dividend-paying companies

The sample consists of 1,664 firms over the period 1996-2003. These firms are all medium and large companies privatized in the Czech Republic by 1994. Column "Obs" shows the number of positive dividend payments observed in a given year. Only firms with liabilities less than twice the size of total assets are included.

Year	Obs
1996	71
1997	86
1998	75
1999	61
2000	63
2001	58
2002	54
Total	468

Table V. STAGE 1: Decision to pay dividends
Dependent variable: 0/1 indicating whether dividends are paid or not.

The sample consists of 1,664 firms over the period 1996-2003 for a total of 5,437 firm-years observations. These firms are all medium and large companies privatized in the Czech Republic by 1994. The dependent variable in all regressions is zero-one variable; one if a firm pays a dividend in a given year and zero otherwise. All estimates are 2SLS/IV estimates with White heteroskedasticity-consistent standard errors reported in parentheses under the coefficient estimates. We use data from 1991-1994 (privatization period) as instrumental variables that allow us to control for the endogeneity of ownership. The last but one row reports the results of the Sargan test of the overidentifying restrictions. All regression equations contain the full set of ownership structure dummies and the residual group of firms not assigned to any category is denoted as "Other". Detailed description of ownership variables and control together with instrumental variables is in section 4.1, 4.3.1, and 4.3.2, respectively. *, **, *** denotes a significant at the 10%, 5%, and 1% level, respectively.

Ownership	Domicile		Concentration		Type	
	Czech	Foreign	Czech	Foreign	Czech	Foreign
All sample	0.110***	0.352***				
	(0.033)	(0.051)				
Majority			0.095**	0.261***		
			(0.050)	(0.075)		
Monitored majority			0.161***	0.578***		
			(0.048)	(0.211)		
Minority			0.064	0.427***		
			(0.044)	(0.144)		
Financial					0.236***	1.223***
					(0.070)	(0.415)
Industrial					0.145***	
					(0.038)	
Individual					0.063	
					(0.065)	
State					0.257***	
					(0.048)	
Dispersed or unknown	0.172***		0.175***		0.185***	
	(0.036)		(0.04)		(0.04)	
Other			0.119***		0.061	
			(0.037)		(1.091)	
Total assets	0.002***		0.002***		0.001*	
	(0.001)		(0.001)		(0.001)	
Total liabilities / Total assets	-0.009		-0.006		0.007	
	(0.011)		(0.012)		(0.013)	
Bank loans / Total liabilities	-0.006		-0.015		-0.029	
	(0.025)		(0.026)		(0.031)	
Cash / Total assets	-0.024		-0.083		-0.041	
	(0.105)		(0.124)		(0.124)	
Investment opportunities (industry-level)	-0.040*		-0.051**		-0.044*	
	(0.022)		(0.024)		(0.025)	
Dividend 1 year before dummy	0.592***		0.590***		0.585***	
	(0.022)		(0.023)		(0.023)	
Tax dummy (1996-1998)	-0.014		-0.019*		-0.027***	
	(0.010)		(0.011)		(0.011)	
Earnings / Total assets	0.042*		0.045*		0.069**	
	(0.022)		(0.026)		(0.033)	
Sales / Total assets	-0.003		-0.001		-0.001	
	(0.004)		(0.005)		(0.005)	
Sales / Staff costs	0.102*		0.107*		0.111	
	(0.062)		(0.066)		(0.071)	
Mills (Sample selection)	-0.083***		-0.077***		-0.091***	
	(0.024)		(0.023)		(0.028)	
Number of observations	5,437		5,437		5,437	
Test overidentif. (p-value)	1.16 (.160)		1.05 (.366)		1.10 (.268)	
Adjusted R^2	0.42		0.39		0.37	

Table VI. STAGE 2: Conditional dividends payments
Dependent variable: Dividend paid in year t by company i.

The sample consists of 1,664 firms over the period 1996-2003 for a total of 468 firm-years observations with a positive dividend payment. These firms are all medium and large companies privatized in the Czech Republic by 1994. The dependent variable in all regressions is the dividend paid in year t by company i. Coefficient α represents dividend smoothing and τ is a target dividend payout ratio in the Lintner-type model. All estimates are OLS estimates with standard errors reported in parentheses under the coefficient estimates. For each specifications we perform Hausman endogeneity test and according to results we treat ownership as exogenous. All regression equations contain the full set of ownership structure dummies and the residual group of firms not assigned to any category is denoted as "Other". Detailed description of ownership variables and control variables is in section 4.1 and 4.3.1, respectively. *, **, *** denotes a significant at the 10%, 5%, and 1% level, respectively.

Ownership	Domicile		Concentration		Type	
	α	τ	α	τ	α	τ
Czech	0.490***	0.125**				
	(0.027)	(0.062)				
Foreign	0.600***	0.464***				
	(0.105)	(0.093)				
Czech majority			0.473***	0.134		
			(0.031)	(0.319)		
Czech monitored majority			0.451***	0.823***		
			(0.139)	(0.236)		
Czech minority			0.801 ***	0.138*		
			(0.075)	(0.085)		
Foreign majority			0.715***	0.607***		
			(0.106)	(0.065)		
Foreign monitored majority			0.853**	0.858***		
			(0.380)	(0.258)		
Financial					0.998***	0.540***
					(0.101)	(0.083)
Industrial					0.471***	0.563***
					(0.029)	(0.089)
Individual					0.112	-0.081
					(1.570)	(1.713)
State					0.128	0.498
					(0.592)	(2.279)
Dispersed or Unknown	0.748***	0.966***	0.711***	0.925***	0.704***	0.925***
	(0.127)	(0.121)	(0.12)	(0.11)	(0.124)	(0.117)
Other			-0.303	-0.201		
			(0.408)	(0.319)		
Total assets	0.392		0.456		0.438	
	(0.455)		(0.042)		(0.454)	
Total liabilities / Total assets	-89.1**		-76.8**		-95.8**	
	(44.9)		(40.6)		(44.8)	
Bank loans / Total liabilities	9.30		8.97		7.97	
	(15.90)		(14.88)		(15.74)	
Cash / Total assets	-1,651		-1,603		-2,113	
	(1,388)		(1,497)		(1,352)	
Investment opportunities (industry-level)	-29.4		-70.9		-36.1	
	(61.8)		(57.4)		(61.0)	
Year dummies	YES		YES		YES	
Mills (Sample selection)	35.8***		48.7***		48.5***	
	(12.3)		(11.9)		(12.7)	
Number of observations	468		468		468	
Hausman test (p-value)	0.66		0.26		0.71	
Adjusted R^2	0.60		0.66		0.61	

Table VII. STAGE 1: Decision to pay dividends, firm-level growth rates and industry dummies
Dependent variable: 0/1 indicating whether dividends are paid or not.

The sample consists of 1,664 firms over the period 1996-2003 for a total of 6,188 firm-years observations. These firms are all medium and large companies privatized in the Czech Republic by 1994. The dependent variable in all regressions is zero-one variable; one if a firm pays a dividend in a given year and zero otherwise. All estimates are 2SLS/IV estimates with White heteroskedasticity-consistent standard errors reported in parentheses under the coefficient estimates. We use data from 1991-1994 (privatization period) as instrumental variables that allow us to control for the endogeneity of ownership. The last but one row reports the results of the Sargan test of the overidentifying restrictions. All regression equations contain the full set of ownership structure dummies and the residual group of firms not assigned to any category is denoted as "Other". Detailed description of ownership variables and control together with instrumental variables is in section 4.1, 4.3.1, and 4.3.2, respectively. Alternative measures of growth opportunities: firm-level growth rates and industry dummies are described in section 7.2. *, **, *** denotes a significant at the 10%, 5%, and 1% level, respectively.

Ownership	Domicile		Concentration		Type	
	Czech	Foreign	Czech	Foreign	Czech	Foreign
All sample	0.120***	0.386***				
	(0.028)	(0.057)				
Majority			0.100**	0.357***		
			(0.047)	(0.086)		
Monitored majority			0.224***	0.454***		
			(0.048)	(0.176)		
Minority			0.062	0.401***		
			(0.044)	(0.153)		
Financial					0.290***	0.845***
					(0.069)	(0.339)
Industrial					0.157***	
					(0.028)	
Individual					0.013	
					(0.045)	
State					0.265***	
					(0.044)	
Dispersed or unknown	0.181***		0.188***		0.187***	
	(0.041)		(0.04)		(0.04)	
Other			0.120***		0.480	
			(0.035)		(0.863)	
Total assets	0.002***		0.002**		0.001*	
	(0.001)		(0.001)		(0.001)	
Total liabilities / Total assets	0.003		0.003		0.008*	
	(0.003)		(0.004)		(0.004)	
Bank loans / Total liabilities	0.015		0.003		-0.013	
	(0.210)		(0.022)		(0.025)	
Cash / Total assets	-0.002		-0.066		-0.017	
	(0.080)		(0.113)		(0.072)	
Investment opportunities (firm-level)	0.004		0.474		0.315	
	(0.200)		(0.313)		(0.403)	
Dividend 1 year before dummy	0.515***		0.510***		0.517***	
	(0.022)		(0.023)		(0.023)	
Tax dummy (1996-1998)	-0.012		-0.015		-0.030***	
	(0.010)		(0.012)		(0.011)	
Earnings / Total assets	0.003		0.004		0.008*	
	(0.003)		(0.004)		(0.005)	
Sales / Total assets	-0.005		-0.005		0.003	
	(0.004)		(0.005)		(0.005)	
Sales / Staff costs	0.054*		0.060*		0.054	
	(0.031)		(0.033)		(0.034)	
Industry dummies	YES		YES		YES	
Mills (Sample selection)	0.031***		0.035***		0.024***	
	(0.005)		(0.006)		(0.005)	
Number of observations	6,188		6,188		6,188	
Test overidentif. (p-value)	1.20 (.139)		0.94 (.598)		1.24 (.104)	
Adjusted R^2	0.38		0.32		0.37	

Table VIII. STAGE 2: Conditional dividends payments, firm-level growth rates and industry dummies
Dependent variable: Dividend paid in year t by company i.

The sample consists of 1,664 firms over the period 1996-2003 for a total of 467 firm-years observations with a positive dividend payment. These firms are all medium and large companies privatized in the Czech Republic by 1994. The dependent variable in all regressions is the dividend paid in year t by company i. Coefficient α represents dividend smoothing and τ is a target dividend payout ratio in the Lintner-type model. All estimates are OLS estimates with standard errors reported in parentheses under the coefficient estimates. For each specifications we perform Hausman endogeneity test and according to results we treat ownership as exogenous. All regression equations contain the full set of ownership structure dummies and the residual group of firms not assigned to any category is denoted as "Other". Detailed description of ownership variables and control variables is in section 4.1 and 4.3.1, respectively. Alternative measures of growth opportunities: firm-level growth rates and industry dummies are described in section 7.2. *, **, *** denotes a significant at the 10%, 5%, and 1% level, respectively.

Ownership	Domicile		Concentration		Type	
	α	τ	α	τ	α	τ
Czech	0.486***	0.044				
	(0.028)	(0.072)				
Foreign	0.588***	0.366***				
	(0.105)	(0.112)				
Czech majority			0.442***	-0.232		
			(0.032)	(0.369)		
Czech monitored majority			0.481***	0.776***		
			(0.153)	(0.226)		
Czech minority			0.810***	0.127		
			(0.076)	(0.088)		
Foreign majority			0.682***	0.532***		
			(0.105)	(0.079)		
Foreign monitored majority			0.890**	0.858***		
			(0.383)	(0.248)		
Financial					1.016***	0.538***
					(0.100)	(0.084)
Industrial					0.453***	0.470***
					(0.030)	(0.109)
Individual					-0.227	0.144
					(1.627)	(0.930)
State					-0.251	-0.001
					(0.610)	(0.168)
Dispersed or unknown	0.780***	0.908***	0.723***	0.879***	0.705***	0.902***
	(0.134)	(0.118)	(0.130)	(0.110)	(0.131)	(0.124)
Other			-0.443	0.002		
			(0.438)	(0.085)		
Total assets	0.426		0.553		0.517	
	(0.475)		(0.442)		(0.468)	
Total liabilities / Total assets	-110.3**		-109.5***		-121.8***	
	(47.7)		(42.5)		(47.4)	
Bank loans / Total liabilities	8.05		10.13		8.08	
	(16.54)		(15.62)		(16.31)	
Cash / Total assets	-1,751		-1,863		-2,152*	
	(1,368)		(1,501)		(1,325)	
Investment opportunities (firm-level)	111.9**		165.9***		156.7***	
	(46.8)		(45.1)		(46.9)	
Year dummies	YES		YES		YES	
Industry dummies	YES		YES		YES	
Mills (Sample selection)	37.1***		49.4***		50.9***	
	(12.4)		(12.2)		(12.8)	
Number of observations	467		467		467	
Hausman test (p-value)	0.97		0.98		0.32	
Adjusted R^2	0.62		0.66		0.63	

Table IX. STAGE 1: Decision to pay dividends, PROBIT
Dependent variable: 0/1 indicating whether dividends are paid or not.

The sample consists of 1,664 firms over the period 1996-2003 for a total of 5,437 firm-years observations. These firms are all medium and large companies privatized in the Czech Republic by 1994. The dependent variable in all regressions is zero-one variable; one if a firm pays a dividend in a given year and zero otherwise. All estimates are probit estimates. Ownership variables are predicted by estimating probit regressions using data from 1991-1994 (privatization period) to control for the endogeneity of ownership as described in section 7.3. Standard errors are reported in parentheses under the coefficient estimates. Ceteris paribus marginal effects are reported in the column "Marginal". Detailed description of ownership variables and control together with instrumental variables is in section 4.1, 4.3.1, and 4.3.2, respectively. *, **, *** denotes a significant at the 10%, 5%, and 1% level, respectively.

Ownership	Domicile		Concentration		Type	
	Coefficient	Marginal	Coefficient	Marginal	Coefficient	Marginal
Czech	-0.32***	-0.03				
	(0.10)					
Foreign	0.28*	0.02				
	(0.17)					
Majority			-0.11	-0.01		
			(0.33)			
Monitored majority			-0.65	-0.05		
			(0.59)			
Minority			-0.23**	-0.02		
			(0.12)			
Financial					0.37*	0.03
					(0.21)	
Industrial					-0.23***	-0.02
					(0.07)	
Individual					-0.65	-0.05
					(0.46)	
State					-0.06	-0.01
					(0.18)	
Dispersed or unknown			0.33***	0.03		
			(0.10)			
Constant	-1.57***	-0.13	-1.84***	-0.15	-1.73***	-0.14
	(0.13)		(0.11)		(0.11)	
Total assets	0.01***	0.00	0.01***	0.00	0.01***	0.00
	(0.00)		(0.00)		(0.00)	
Total liabilities / Total assets	-0.05	0.00	-0.08	-0.01	-0.05	0.00
	(0.16)		(0.16)		(0.16)	
Bank loans / Total liabilities	-0.06	0.00	-0.06	0.00	-0.05	0.00
	(0.08)		(0.08)		(0.08)	
Cash / Total assets	-9.68*	-0.79	-9.09	-0.74	-10.35*	-0.85
	(5.97)		(5.85)		(5.89)	
Investment opportunities	-0.64**	-0.05	-0.65**	-0.05	-0.65**	-0.05
	(0.29)		(0.29)		(0.29)	
Dividend 1 year before dummy	2.15***	0.17	2.14***	0.17	2.16***	0.18
	(0.07)		(0.07)		(0.07)	
Tax dummy (1996-1998)	-0.22***	-0.02	-0.23***	-0.02	-0.22***	-0.02
	(0.08)		(0.08)		(0.08)	
Earnings / Total assets	3.85***	0.31	3.97***	0.32	4.07***	0.33
	(0.41)		(0.41)		(0.41)	
Sales / Total assets	-0.11*	-0.01	-0.11*	-0.01	-0.12**	-0.01
	(0.06)		(0.06)		(0.06)	
Sales / Staff costs	0.68	0.06	0.67	0.05	0.64	0.05
	(0.51)		(0.51)		(0.50)	
Number of observations	5,437		5,437		5,437	
Log likelihood	-838.48		-839.36		-840.27	
Standardized R^2	0.31		0.31		0.31	

Testing for logistic and exponential smooth transition cointegration with an application to monetary exchange rate models

Joanna Bruzda
Nicolaus Copernicus University
Toruń, Poland
Joanna.Bruzda@umk.pl

Abstract

In the paper several tests of the hypothesis of no cointegration against logistic smooth transition (LSTR) cointegration are suggested. The tests are similar in their underpinning to the Kapetanios, Shin and Snell (2006) tests against exponential smooth transition (ESTR) cointegration. Furthermore, four test statistics to jointly test for LSTR and ESTR cointegration are considered. In the empirical part of the paper the sticky-price monetary exchange rate models are examined. The exchange rates considered are the official Czech Koruna and Polish Zloty nominal exchange rates against Euro and U.S. Dollar, announced by the appropriate national banks. Our nonlinear cointegration tests are able to find stable long-term relationships, whereas the linear-based tests fail. Due to this our empirical investigation makes it possible to explain the observed long-lasting misalignment of exchange rates with economic fundamentals called an "exchange rate disconnect puzzle".

Keywords: exchange rates modelling, smooth transition cointegration
JEL: C12, C22, F31

1. Introduction

Over the last twenty years the interest in nonlinear time series analysis has been steadily increasing. In macroeconomic and financial applications the most popular models are piecewise linear models which allow for regime-switching behavior. Among these specifications a very important class of models constitute smooth transition autoregressive (STAR) processes. A general STAR model is given by:

$$z_t = k + \sum_{j=1}^{p} \pi_j z_{t-j} + \left(k^* + \sum_{j=1}^{p} \pi_j^* z_{t-j} \right) F(s_t; \gamma, c) + \varepsilon_t , \tag{1}$$

where the ε_t's are assumed to be a martingale difference sequence with respect to the history of the process and the transition function $F(\cdot)$ is a continuous function, usually bounded between 0 and 1. In the STAR models discussed in [22] the transition variable s_t is assumed to be a lagged endogenous variable, i.e. $s_t = y_{t-d}$. Following [23] we relax this assumption here allowing s_t to be a function of lagged endogenous variables. The most popular specifications of the transition function $F(\cdot)$ are the U-shaped exponential function:

$$F(s_t) = 1 - \exp\left[-\gamma(s_t - c)^2\right], \quad \gamma > 0, \tag{2}$$

the first-order logistic function:

$$F(s_t) = \{1 + \exp[-\gamma(s_t - c)]\}^{-1}, \quad \gamma > 0, \tag{3}$$

and the second-order logistic function, which is also U-shaped:

$$F(s_t) = \{1 + \exp[-\gamma(s_t - c_1)(s_t - c_2)]\}^{-1}, \quad c_1 \le c_2, \gamma > 0. \tag{4}$$

The transition function of interest here is the first-order logistic function (3) and the resulting STAR model is the first-order logistic STAR (LSTAR) model. The function (3) changes monotonically form 0 to 1 as s_t increases, whereas the parameter γ controls the smoothness of the transition from one regime to the second. As γ becomes large, the transition becomes almost instantaneous at $s_t = c$ and, as a result, the function (3) approaches the Heaviside indicator function $I_{\{s_t > c\}}$. Due to this the first-order LSTAR models nest two-regime threshold and momentum-threshold autoregressive models that might be seen as a limiting case of LSTAR processes. This two-regime type of behaviour is convenient especially for modelling business cycles asymmetry to distinguish expansions and recessions, while the U-shaped transition functions (2) and (4) allow for the existence of a band around an equilibrium value accounting for the presence of transaction costs. It is worth noticing that as $\gamma \to \infty$, a second-order logistic STAR model approaches a restricted three-regime SETAR model, with the restriction that the outer regimes are identical.

It is now well recognised that standard unit root and cointegration tests like the ADF test and the Engle-Granger and Johansen procedures are not appropriate for investigating nonlinear processes (see, for example, [8]). In the context of the interplay between nonstationarity, cointegration and nonlinearity two main problems arise. Firstly, the standard linear tests lack their power in the case of stationary nonlinear processes. Secondly, even more dangerous problem seems to be serious size distortions in the case of nonstationary and nonlinear processes, which can be wrongly recognised as stationary ones. The paper contributes to a strand of research attempting to shed more light on the two questions, with special emphasise on the first one. It is quite obvious that misspecifying a stable nonlinear process as a nonstationary one can give misleading impulse responses and a worse post-sample performance. Additionally, allowing for nonlinear adjustment processes enables to find long-term relationships where linear cointegration tests fail. Several authors have addressed this questions and suggested unit root tests against specific nonlinear alternatives (see, for example, [3], [9], [11], [12]) as well as tests of the hypothesis of no cointegration against an alternative assuming a particular nonlinear stationary dynamics (see [10] and [13]). In the paper four tests of the hypothesis of no cointegration against logistic smooth transition (LSTR) cointegration are developed. The tests fill the gap in the existing econometric literature being similar in spirit to the Kapetanios, Shin and Snell [13] tests against exponential smooth transition (ESTR) cointegration. Further four test statistics are suggested to jointly test for LSTR and ESTR cointegration[1]. The theoretical considerations in the paper lead to a specification procedure, which can be helpful to distinguish between linear, LSTR and ESTR cointegration. The modeling procedure resembles the well-known Teräsvirta procedure (see [22]) to test for LSTR and ESTR linearity of stationary time series.

The rest of the paper is organized as follows. Section 2 develops in detail the tests for LSTR cointegration. Section 3 outlines the specification technique and indicates possible modifications and extensions of the tests. In Section 4 empirical examples concerning exchange rate models are provided. The empirical investigation makes it possible to explain the observed long-lasting misalignment of exchange rates with economic fundamentals called an "exchange rate disconnect puzzle". Finally, Section 5 shortly concludes.

2. Tests for LSTR cointegration

To begin with we consider a zero mean LSTAR process in the form

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-1}\left(\frac{1}{1+e^{-\gamma(s_t - c)}} - \frac{1}{2}\right) + \varepsilon_t, \qquad (5)$$

where $\gamma > 0$, $s_t = z_{t-1}$ or $s_t = \Delta z_{t-1}$, and $\varepsilon_t \sim iid(0,\sigma^2)$. Such a process may be seen as a generalization of a two-regime TAR(1) or M-TAR(1) (momentum-TAR) process (see [9]). After rearranging (5) we obtain

$$\Delta z_t = \rho_1 z_{t-1} + \rho_2 z_{t-1}\left(\frac{1}{1+e^{-\gamma(s_t - c)}} - \frac{1}{2}\right) + \varepsilon_t, \qquad (5)$$

with $\rho_1 = \pi_1 - 1$ and $\pi_2 = \rho_2$. As $\gamma \to \infty$, the process approaches a TAR process in the form

$$\Delta z_t = \begin{cases} (\rho_1 - \frac{\rho_2}{2})z_{t-1} + \varepsilon_t & \text{for } s_t \le c \\ (\rho_1 + \frac{\rho_2}{2})z_{t-1} + \varepsilon_t & \text{for } s_t > c \end{cases}. \qquad (6)$$

Conditions for the process (5) to be stationary are the following (see [9], pp. 59-60):

[1] As will be explained later in the text, by smooth transition cointegration we understand a linear long-term relationship with a nonlinear adjustment process of a STAR type. However, smooth transition cointegration may be also thought as a nonlinear long-term relationship in the form of a smooth transition regression. Such a concept was developed by Choi and Saikkonen (see [6]) and the authors refer to it as cointegrating smooth transition regression.

$$\rho_1 - \frac{\rho_2}{2} < 0,$$
$$\rho_1 + \frac{\rho_2}{2} < 0, \tag{7}$$
$$\left(\rho_1 - \frac{\rho_2}{2} + 1\right)\left(\rho_1 + \frac{\rho_2}{2} + 1\right) < 1.$$

Testing for linearity of the process (5) consists in testing for H₀: $\gamma = 0$. However, the test cannot be performed directly as under the null hypothesis the parameter ρ_2 is not identified. To overcome this problem we follow Luukkonen et.al. (see [17]) and use the Taylor series approximation to the logistic transition function around $\gamma = 0$. The first-order Taylor approximation to the function $F(x) = (1+\exp(-x))^{-1} - 1/2$, where $x = \gamma(s_t - c)$, is given by $T_1(x) = 1/4\, x$, while the third-order approximation is $T_3(x) = 1/4\, x - 1/48\, x^3$. The two approximations give the following auxiliary equations:

$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1} s_t + \varepsilon_t, \tag{8}$$

where $\alpha_1 = \rho_1 - c\gamma\rho_2/4$, $\alpha_2 = \gamma\rho_2/4$, and

$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1} s_t + \alpha_3 z_{t-1} s_t^2 + \alpha_4 z_{t-1} s_t^3 + \varepsilon_t, \tag{9}$$

where $\alpha_1 = \rho_1 - c\gamma\rho_2/4 + c^3\gamma^3\rho_2/48$, $\alpha_2 = \gamma\rho_2/4 - c^2\gamma^3\rho_2/12$, $\alpha_3 = c\gamma^3\rho_2/12$ and $\alpha_4 = -\gamma^3\rho_2/48$. Now we can derive joint tests for a unit root and linearity (a linear unit root). As our transition variable is either z_{t-1} or Δz_{t-1}, the test equations are the following:

$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}^2 + \varepsilon_t, \tag{10}$$
$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1} \Delta z_{t-1} + \varepsilon_t, \tag{11}$$
$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}^2 + \alpha_3 z_{t-1}^3 + \alpha_4 z_{t-1}^4 + \varepsilon_t, \tag{12}$$
$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1} \Delta z_{t-1} + \alpha_3 z_{t-1} \Delta z_{t-1}^2 + \alpha_4 z_{t-1} \Delta z_{t-1}^3 + \varepsilon_t \tag{13}$$

with the null hypothesis H₀: $\alpha_1 = \alpha_2 = 0$ in the case of the regressions (10) and (11), and H₀: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ in the more general settings (12) and (13). To test the null hypothesis of a linear unit root we suggest using an *F* statistic in the form

$$F = \frac{(SSR_0 - SSR_1)/k}{SSR_0/(T-k)}, \tag{14}$$

where $SSR_0 = \sum_{t=1}^{T} z_t^2$, SSR_1 is the sum of squared residuals from the appropriate test equation and *k* stands for the number of estimated parameters. Further in the text we use the notations F_{02}, F_{02}^{Δ}, F_{04} and F_{04}^{Δ} for test statistics based on regressions from (10) to (13), respectively.

It is worth noticing that the first-order Taylor approximation to the exponential transition function leads to the test equation in the form[2]

$$\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1} s_t + \alpha_3 z_{t-1} s_t^2 + \varepsilon_t , \qquad (15)$$

where $\alpha_1 = \rho_1 + c^2 \gamma \rho_2$, $\alpha_2 = -2c\gamma\rho_2$ and $\alpha_3 = \gamma\rho_2$, which under $\rho_1 = c = 0$ reduces to

$$\Delta z_{t-1} = \alpha_3 z_{t-1} s_t^2 + \varepsilon_t .$$

In the case of the second-order logistic function we obtain the same test equation (15) with $\alpha_1 = \rho_1 + {}^{c_1 c_2 \gamma \rho_2}\!/_4$, $\alpha_2 = -{}^{(c_1+c_2)\gamma\rho_2}\!/_4$ and $\alpha_3 = {}^{\gamma\rho_2}\!/_4$.

Following [12], when the process under study has a nonzero mean or a nonzero linear trend, we suggest using the de-meaned and de-trended data, respectively.

Now we can turn to tests for cointegration[3]. To this end we take the pragmatic residual-based two-step approach, where in the first step the OLS residuals $u_t = y_t - \boldsymbol{\beta}\mathbf{x}_t$ are computed (y_t is an endogenous I(1) process, $\boldsymbol{\beta}$ is a $1 \times k$ vector of parameters and $\mathbf{x}_t$ is a $k \times 1$ vector of weakly exogenous with respect to the parameters $\boldsymbol{\beta}$ I(1) processes. To accommodate for deterministic components we consider also regressions based on the de-meaned and de-trended data[4], i.e.

$$y_t^* = \boldsymbol{\beta}\mathbf{x}_t^* + u_t^* , \qquad (16)$$

$$y_t^+ = \boldsymbol{\beta}\mathbf{x}_t^+ + u_t^+ , \qquad (17)$$

where the superscripts $*$ and $+$ denote the de-meaned and de-trended data, respectively. To test for logistic smooth transition (LSTR) cointegration we suggest using the equations (10)-(13), in which the variable z_t is replaced with the residuals u_t or their de-meaned or de-trended counterparts defined in (16) and (17). Taking into consideration that the Engle-Granger testing procedure lacks its power when the so called common factor (COMFAC) restrictions are not fulfilled[5], we introduce also ECM-based tests for LSTR cointegration, in which the following test equations are utilized:

$$\Delta y_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1}^2 + \boldsymbol{\psi}_0 \Delta\mathbf{x}_t + \sum_{i=1}^{p} \boldsymbol{\psi}_i \Delta\mathbf{w}_{t-i} + \varepsilon_t , \qquad (18)$$

$$\Delta y_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1} \Delta u_{t-1} + \boldsymbol{\psi}_0 \Delta\mathbf{x}_t + \sum_{i=1}^{p} \boldsymbol{\psi}_i \Delta\mathbf{w}_{t-i} + \varepsilon_t , \qquad (19)$$

$$\Delta y_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1}^2 + \alpha_3 u_{t-1}^3 + \alpha_4 u_{t-1}^4 + \boldsymbol{\psi}_0 \Delta\mathbf{x}_t + \sum_{i=1}^{p} \boldsymbol{\psi}_i \Delta\mathbf{w}_{t-i} + \varepsilon_t , \qquad (20)$$

---

[2] Compare [12] and [13].
[3] In our testing framework we analyze one cointegrating vector only.
[4] We assume a linear deterministic trend - compare [13].
[5] See, for example, [15].

$$\Delta y_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1}\Delta u_{t-1} + \alpha_3 u_{t-1}\Delta u_{t-1}^2 + \alpha_4 u_{t-1}\Delta u_{t-1}^3 + \boldsymbol{\psi}_0 \Delta \mathbf{x}_t + \sum_{i=1}^{p} \boldsymbol{\psi}_i \Delta \mathbf{w}_{t-i} + \varepsilon_t , \quad (21)$$

where $\mathbf{w}_t' = [y_t, x_{1t}, \ldots, x_{kt}]$, $\boldsymbol{\psi}_0$, $\boldsymbol{\psi}_i, i = 1, \ldots, p$ are appropriate vectors of parameters and ε_t is an *iid* error term. As previously, the residuals form the first step of the testing procedure may also come from the regressions (16) and (17). To test for the joint hypothesis of no cointegration and linearity we suggest using F tests of the hypotheses $\alpha_1 = \alpha_2 = 0$ and $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ in (18) and (19), respectively, where the test statistics are computed as

$$F = \frac{(SSR_0 - SSR_1)/k}{SSR_0 /(T - k - p - 1)} , \quad (22)$$

where SSR_0 and SSR_1 stand for the sums of squared residuals in the restricted and unrestricted models, respectively, and k is the number of restrictions. Further in the text the LSTR cointegration test statistics based on equations (10)-(13) will be denoted by F_{EGk} and F_{EGk}^{Δ} to underline similarity to the well-known Engle-Granger approach, while the ECM-based tests - by F_{NECk} and F_{NECk}^{Δ} - as they utilize specific nonlinear error correction models.

It is worth noticing that the equation (20) nests the test equation for ESTR cointegration suggested in [13] that has the form

$$\Delta y_t = \alpha_1 u_{t-1} + \alpha_2 u_{t-1}^2 + \alpha_3 u_{t-1}^3 + \boldsymbol{\psi}_0 \Delta \mathbf{x}_t + \sum_{i=1}^{p} \boldsymbol{\psi}_i \Delta \mathbf{w}_{t-i} + \varepsilon_t . \quad (23)$$

Due to this our F_{NEC4} test might be thought as a joint test for LSTR and ESTR cointegration. We attribute a similar interpretation to the F_{NEC4}^{Δ} test as well as to the corresponding Engle-Granger-type tests, i.e. F_{EG4} and F_{EG4}^{Δ}.

3. General remarks, modifications and extensions

Having derived the testing framework several remarks are at place. Firstly, we notice that our tests of the joint hypothesis of a unit root and linearity will generally have power against some nonstationary and nonlinear processes. Although it is not very likely that an economic process is generated by an equation in the form

$$\Delta z_t = \alpha z_{t-1}^2 + \varepsilon_t , \quad (24)$$

if it is the case, our tests based on (10) and (13) should reject the null hypothesis. It is, however, straightforward to see that the process (24) is nonstationary even in the first moment. A more realistic case refers to a process in the form

$$\Delta z_t = \alpha z_{t-1} \Delta z_{t-1} + \varepsilon_t , \quad (25)$$

which approximates the unit root bilinear process suggested in [5] to describe bubbles in stock prices under small bilinearity. Another example concerns the so-called partial unit root process (see [3]), which can be thought as a limiting case of the process (5) under $\rho_1 + \rho_2/2 = 0$ or $\rho_1 - \rho_2/2 = 0$. The partial unit root processes and the resulting partial cointegration analysis were suggested in [2] as a way to describe deviations of stock prices from their long-term values determined, *e.g.*, by the present value relationship, as well as relationships between short- and long-term interest rates.

The above remark relates to a general problem of size distortions in unit root and cointegration tests under nonlinear processes (see, for example, [5] and [8], pp. 58-68, for a discussion about size distortions under nonlinear processes in the case of the standard Dickey-Fuller test). As it is easily seen from the derivation of the parameter α_1 in the equations (8) and (9), the ADF test may show size distortions in the case of nonstationary LSTAR processes and its particular performance will depend on the value of the parameter *c*. We notice also that the same problem concerns the F_{NEC} test of Kapetanios et. al. (see [13]) utilizing the test equation (23). To circumvent the problem one may be interested in supplementing the testing procedure with certain co-mixing tests as suggested in [4] and [8], pp. 217-219. However, as our main motivation was to derive tests for LSTR cointegration, we are much more concerned with a power evaluation of our cointegration tests. The power of the tests was investigated in a companion paper [1]. Similarly to [13], the companion paper documents significant power gains over the standard testing framework.

The tests we developed so far let us suggest a specification procedure to distinguish between linear, ESTR and LSTR cointegration in a way similar to the well-known Teräsvirta (1994) procedure to test for STAR nonlinearity. The specification technique takes advantage of the general to specific modeling and starts with the most general F_4 tests of the hypothesis H_{04}: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. If the null is rejected, the general conclusion is that the process under scrutiny is not a linear unit root process[6]. The conclusion combined with results of some co-mixing tests can be thought as an evidence on (linear or nonlinear) cointegration. The next step consists in testing for the significance of the last parameter α_4 in the regressions (12), (13), (20) or (21) with the ordinary *t* statistic, which under stationarity of the underlying process and normally and identically distributed errors has the usual Student's *t* distribution. The test of the hypothesis H_0: $\alpha_4 = 0$ against the two-sided alternative will provide two main indications. If the null is rejected, first of all we conclude that LSTR cointegration takes place. Secondly, the sign of the parameter α_4 will give additional information about the LSTAR adjustment process, i.e. if it is negative, the process under scrutiny is more mean-reverting in the first regime than in the second, while if it is positive, the opposite is true.

[6] Taking into consideration that the F_4 tests may have low power as they are testing four coefficients, in the case when the null is not rejected, we suggest executing also the subsequent F_3 and F_2 tests as well as the standard Dickey-Fuller test.

If the null is accepted, we turn to testing the conditional hypothesis H_{03}: $\alpha_3 = 0 | \alpha_4 = 0$. Rejection of H_{03} after accepting H_{04} may be treated as an indication for the presence of ESTR cointegration, but only if the parameter α_3 is negative. If the null H_{03} is accepted, the subsequent conditional hypothesis should be tested, i.e. H_{02}: $\alpha_2 = 0 | \alpha_3 = \alpha_4 = 0$. Rejection of H_{02} after accepting H_{03} may be treated as an evidence on the existence of LSTR cointegration. This time, if the sign of the parameter α_2 is negative, one can conclude that the LSTAR adjustment process is more mean-reverting in the second regime, while if it is positive, the parameter ρ_2 is greater than 0 and the process behaves 'more stationary' in the first regime. If the null H_{02} is accepted, the most likely dynamics is the linear one and, as a result, we conclude that linear cointegration takes place.

Having discussed the specification procedure for STAR adjustment processes one additional remark is at place. As rejections of our null hypotheses might generally be due to the presence of certain nonstationary nonlinear dynamics, it seems useful to separately simulate appropriate critical values for tests of significance of the last parameters in our test equations under the null of a linear unit root. Such tests might be useful to state what kind of nonlinear dynamics is directly responsible for the rejection of the hypothesis of a linear nonstationary dynamics, with special emphasis on the bilinear and partial unit root cases. We do not develop this idea here leaving it as a subject for further studies.

Now we shortly comment on further modeling issues and possible modifications of our tests. Firstly, we notice that in the case of serially correlated errors the test equations (10)-(13) and (15) can be augmented with lagged differences of the process under scrutiny with the number of lags chosen according to standard model selection criteria. We treat the linear augmentation as a first-order approximation to a possibly nonlinear dynamics of the error terms. Alternatively, a semi-parametric correction advocated by Phillips and Perron (see [20]) might be used.

Furthermore, a direct generalization to the multivariate framework is also possible. To this end we suggest applying the idea of Seo (2004) and Krishnakumar and Neto (2005), who developed the appropriate methodology to test for threshold cointegration in threshold vector error correction models. Let us consider a smooth transition vector error correction model in the form

$$\Delta\mathbf{x}_t' = \mathbf{x}_{t-1}'\boldsymbol{\beta}\boldsymbol{\Theta}_1 + \mathbf{x}_{t-1}'\boldsymbol{\beta}\boldsymbol{\Theta}_2 F(s_t;\lambda,c) + \Delta\mathbf{x}_{t-1}'\boldsymbol{\Phi}_1 + \ldots + \Delta\mathbf{x}_{t-p}'\boldsymbol{\Phi}_p + \boldsymbol{\varepsilon}_t, \tag{26}$$

where $\mathbf{x}_t$ is a $k \times 1$ vector of zero-mean I(1) variables, $\boldsymbol{\beta}$ is a $k \times 1$ cointegrating vector that may be known or estimated by the OLS method, Θ_1 and Θ_2 are $k \times 1$ vectors of parameters, $\Phi_i, i = 1,\ldots,p$, are $k \times k$ matrices of dynamic coefficients and $\boldsymbol{\varepsilon}_t$ is a multivariate white noise. The variable s_t is assumed to be either $\mathbf{x}_{t-1}'\boldsymbol{\beta}$ or $\Delta\mathbf{x}_{t-1}'\boldsymbol{\beta}$. Denoting $z_t = \mathbf{x}_t'\boldsymbol{\beta}$ and using the third-order Taylor approximation to the logistic transition function we can replace (26) with an auxiliary model in the form

$$\Delta \mathbf{x}'_t = z_{t-1}\mathbf{A}_1 + z_{t-1}s_t\mathbf{A}_2 + z_{t-1}s_t^2\mathbf{A}_3 + z_{t-1}s_t^3\mathbf{A}_4 + \Delta\mathbf{x}'_{t-1}\mathbf{\Phi}_1 + \ldots + \Delta\mathbf{x}'_{t-p}\mathbf{\Phi}_p + \boldsymbol{\varepsilon}_t . \quad (27)$$

To jointly test for linear, ESTR and LSTR cointegration we put the following hypothesis H_0: $\mathbf{A}_1 = \mathbf{A}_2 = \mathbf{A}_3 = \mathbf{A}_4 = \mathbf{0}$, which can be tested with the standard Wald test. Further considerations lead to a specification procedure similar to the one described before and to allowing for deterministic components in the processes under scrutiny.

Our last remark refers to the problem of constructing a test of a linear unit root against a stationary alternative of an LSTAR type, i.e. a test which will directly test stationarity in the LSTAR framework. To this end we consider an LSTAR process in the form

$$z_t = \pi_1 z_{t-1} + \pi_2 z_{t-1}\left(1 - \frac{b}{1+e^{-\gamma(s_t - c)}}\right) + \varepsilon_t , \quad (28)$$

where $\gamma > 0$, $b < 1$, $s_t = z_{t-1}$ or $s_t = \Delta z_{t-1}$, and $\varepsilon_t \sim iid(0,\sigma^2)$. The transition function in (28) takes its values in the intervals $(1-b,1)$ if $b > 0$, and $(1,1-b)$ if $b < 0$, while under $b = 0$ the process (28) becomes a linear autoregression. To simplify our consideration we assume $c = 0$ [7] and after rearranging (28) we obtain

$$\Delta z_t = \rho_1 z_{t-1} + \rho_2 z_{t-1}\left(1 - \frac{b}{1+e^{-\gamma s_t}}\right) + \varepsilon_t , \quad (29)$$

with $\rho_1 = \pi_1 - 1$ and $\pi_2 = \rho_2$. The LSTAR model (29) nest as a limiting case an TAR (M-TAR) model in the form

$$\Delta z_t = \begin{cases} (\rho_1 + \rho_2) z_{t-1} + \varepsilon_t & \text{for } s_t \le 0 \\ [\rho_1 + \rho_2(1-b)] z_{t-1} + \varepsilon_t & \text{for } s_t > 0 \end{cases} . \quad (30)$$

Conditions for stationarity of (30) are the following

$$\begin{aligned} &\rho_1 + \rho_2 < 0, \\ &\rho_1 + \rho_2(1-b) < 0, \\ &(\rho_1 + \rho_2 + 1)[\rho_1 + \rho_2(1-b) + 1] < 1. \end{aligned} \quad (31)$$

As the process (29) is somewhat overparametrised, we further assume that $\rho_1 = 0$ and consider a process in the form

$$\Delta z_t = \rho\, z_{t-1}\left(1 - \frac{b}{1+e^{-\gamma s_t}}\right) + \varepsilon_t , \quad (32)$$

for which the condition for stationarity is $0 > \rho > \frac{b-2}{1-b}$. Now we can construct a test of a unit root against a stationary alternative of an LSTAR type that will consist in testing for H_0: $\rho = 0$ against H_1: $\rho < 0$ by utilizing the standard *t* statistic for the parameter ρ. As under H_0 the parameters *b* and γ are

[7] The assumption can be easily relaxed leading to a bit more complicated and by far more computationally intensive testing procedure. However, we notice that the assumption seems to be quite natural in cointegration analysis, where there can be expected that an adjustment process behaves differently for negative and positive deviations from a long-term value or, alternatively, for negative and positive increments. Examples of such

unidentified, to overcome the so-called nuisance parameters problem we suggest adopting the idea of Kiliç and de Jong R. (see [14]), and define a test statistic in the form

$$\inf t = \inf_{(b,\gamma)\in B\times\Gamma} \hat{t}_{\rho=0}(b,\gamma), \qquad (33)$$

which takes the lowest possible value over a space of relevant values for b and γ, i.e. $B=(\underline{b},\overline{b})$ and $\Gamma=(\underline{\gamma},\overline{\gamma})$. We notice that a natural choice for the interval B is (-1, 1), which includes also the linear autoregression case. Due to this we might expect significant power gains over the standard Dickey-Fuller test[8]. As previously, the equation (32) can be augmented with lagged differences of z_t.

4. Empirical example

In our empirical example we model the U.S. Dollar and Euro exchange rates with the help of linear long-term relationships with a possibly nonlinear adjustment. A nonlinear adjustment process enables taking into account different lengths of overvaluation and undervaluation periods or different forms of the adjustment with regard to the size of deviations from parity. Moreover, it makes it possible to explain the observed long-lasting misalignment of exchange rates with the economic fundamentals (comp., for example, [4] and [7]). This misalignment is called in the literature an “exchange rate disconnect puzzle” and was coined by Obstfeld and Rogoff (see [19]). The exchange rates considered come from national banks of two Central European countries willing to enter the Euro zone – the Czech Republic and Poland. We concentrate on the sticky-price monetary model in the form

$$s_t = \alpha_0 + \alpha_1(m_t - m_t^*) + \alpha_2(y_t - y_t^*) + \alpha_3(r_t^r - r_t^{r*}) + \alpha_4(\pi_t^e - \pi_t^{e*}) + \eta_t, \qquad (34)$$

where m and y are the natural logarithms of the domestic stock of money and real income, respectively, r_t^r is the short-term real interest rate, π_t^e stands for the expected inflation rate and an asterisk denotes corresponding variables in the foreign country[9].

The data used in this study were obtained from three sources: the Czech National Bank, the National Bank of Poland and the IMF’s International Financial Statistics database. The data are monthly and span the period 1998.01-2005.07, which gives the total of 91 observations. Inflationary expectations were proxied using inflation rates over the preceding three months, calculated on the yearly basis. Changes in the appropriate Consumer Price Indices were used as measures of inflation. The short-term interest rate is the three-month money market rate. Real interest rates were computed according to the well known formula $r^r = (1+r)/(1+\pi) - 1$. The monetary aggregate used is M1, while the income measure is the monthly index of industrial production. The indices of industrial

---

cointegrating regressions can be the present value relationship or the relationship between short- and long-term interest rates.

[8] The power evaluation of the test is currently under investigation.

[9] Comp., for example, [18].

production, money supply, inflationary expectations as well as nominal exchange rates were seasonally adjusted with the help of the additive version of the moving average method. The exchange rates, output and money were transformed into logarithms. Further in the text we use the following abbreviations: *czk_eur*, *czk_usd*, *pln_eur*, *pln_usd* stand for the appropriate exchange rates, *p_cz_ea*, *p_cz_usa*, *p_p_ea*, *p_p_usa* denote the relative industrial productions (logarithms of the ratio of industrial productions), where *cz* stands for the Czech Republic, *p* for Poland, *ea* for Euro area and *usa* for the United States. Further, in the same manner we denote the relative money supplies: *m_cz_ea*, *m_cz_usa*, *m_p_ea*, *m_p_usa*, real interest rate differentials: *rr_cz_ea*, *rr_cz_usa*, *rr_p_ea*, *rr_p_usa* and inflation differentials: *i_cz_ea*, *i_cz_usa*, *i_p_ea*, *i_p_usa*.

In the first step we tested our series for stationarity with the standard ADF test. The results (see Table 1) indicate that all series are I(1) variables, except maybe for the inflation rate differentials for the Czech Republic that can also be treated as stationary.

Table 1. Results of the ADF unit root tests

Variable	ADF statistic for levels (augmentation)	ADF statistic for first differences (augmentation)
czk_eur	-1,108 (1)	-12,239 (0) **
czk_usd	-0,784 (0)	-10,453 (0) **
pln_eur	-2,263 (1)	-7,502 (0) **
pln_usd	-1,453 (0)	-9,003 (0) **
p_cz_ea	1,472 (5)	-6,072 (4) **
p_cz_usa	0,686 (5)	-5,719 (4) **
p_p_ea	0,259 (2)	-11,921 (1) **
p_p_usa	-0,435 (3)	-5,853 (2) **
m_cz_ea	-0,772 (0)	-10,380 (0) **
m_cz_usa	-0,499 (0)	-10,547 (0) **
m_p_ea	-1,588 (1)	-11,954 (0) **
m_p_usa	-1,069 (1)	-12,050 (0) **
rr_cz_ea	-2,264 (5)	-5,393 (4) **
rr_cz_usa	-1,961 (3)	-11,521 (2) **
rr_p_ea	-2,454 (3)	-7,413 (2) **
rr_p_usa	-2,764 (3)	-7,475 (2) **
i_cz_ea	-2,871 (7)	-4,050 (6) **
i_cz_usa	-3,015 (7) *	-3,981 (6) **
i_p_ea	-2,415 (3)	-7,552 (2) **
i_p_usa	-2,805 (3)	-7,810 (2) **

The tests allow for a constant, but not a trend in the testing equations. The augmentation was based on the Ljung-Box statistics. Critical values for the ADF tests for $T = 90$ are: -2,584 (10%), -2,894 (5%), -3,504 (1%).
'*' and '**' denote rejection of the null hypothesis at the 5% and 1% significance levels, respectively.

In the next step we estimated the models (34) by the DSUR (dynamic seemingly unrelated regressions) method. We performed this estimation for 2 two-equation systems, each of which consisted of one equation for Euro and one equation for Dollar for a particular country. The estimation outputs are given in Table 2. To check the long-run validity of the estimated relationships, firstly we executed the ADF test on dynamic SUR residuals of the estimated models. In what follows we use the additional abbreviations: *res_czk_eur*, *res_czk_usd* – residuals from Model 1 and *res_pln_eur*, *res_pln_usd* – residuals from Model 2. The results of the linear cointegration test are presented in Table 3.

Table 2. Estimation outputs for the potential long-term relationships

Dependent variables		Regressors			
		Output differential	Money supply differential	Real interest rate differential	Inflation differential
Model 1	*czk_eur*	-0,14801 (0,39705)	-0,16638 (0,26788)	0,00392 (0,00793)	0,03996 (0,03593)
	czk_usd	-1,83240 (0,50737)	0,24145 (0,25056)	0,01290 (0,00779)	0,03242 (0,04654)
Model 2	*pln_eur*	-0,01239 (0,18859)	0,46977 (0,32861)	-0,01111 (0,00490)	0,00807 (0,01216)
	pln_usd	-1,12291 (0,73398)	0,95392 (0,66910)	0,02906 (0,01920)	0,12078 (0,08920)

Figures in parenthesis are standard errors. To save space we do not report constant terms.

Table 3. Results of the linear cointegration test

Variable	ADF statistic (augmentation)
res_czk_eur	-3,270 (0)
res_czk_usd	-4,758 (0) **
res_pln_eur	-2,808 (0)
res_pln_usd	-2,902 (0)

Critical values for the ADF test for T = 90 and 4 regressors are: -4,250 (10%), -4,571 (5%), -5,207 (1%). '*', '**' denote rejection of H_0 at the 5% and 1% significance levels, respectively.

As we pointed out in the methodological part of the paper, the standard testing methodology is not valid any longer if adjustment processes are of nonlinear nature. Due to this we executed our F tests of the hypothesis of a linear unit root. The results are summarized in Tables 4 and 5.

Table 4. Results of the nonlinear cointegration test with z_{t-1} as the transition variable

Variable	F statistic	t statistic for the last parameter (p-value)
Panel 1: test equation $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}^2 + \alpha_3 z_{t-1}^3 + \alpha_4 z_{t-1}^4 + \varepsilon_t$, F_{EG4} statistic		
res_czk_eur	5,070***	1,265 (0,209)
res_czk_usd	5,572***	0,196 (0,845)
res_pln_eur	3,951**	2,237 (0,028)
res_pln_usd	3,556	0,700 (0,286)
Panel 2: test equation $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}^2 + \alpha_3 z_{t-1}^3 + \varepsilon_t$, F_{EG3} statistic		
res_czk_eur	6,182**	-2,367 (0,020)
res_czk_usd	7,504***	-0,535 (0,594)
res_pln_eur	3,103	-1,204 (0,218)
res_pln_usd	4,660*	-1,148 (0,234)
Panel 3: test equation: $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}^2 + \varepsilon_t$, F_{EG2} statistic		
res_czk_eur	6,131	1,227 (0,223)
res_czk_usd	11,209***	1,212 (0,204)
res_pln_eur	3,909	-0,154 (0,878)
res_pln_usd	5,995	1,668 (0,116)

Critical values for tests in Panel 1: 4,706 (1%), 3,948 (5%), 3,574 (10%); critical values for tests in Panel 2: 6,187 (1%0, 5,161 (5%), 4,648 (10%); critical values for tests in Panel 3: 9,031 (1%), 7,477 (5%), 6,715 (10%). The critical values were obtain by simulating 100000 regressions on de-meaned independent random walks. '*', '**', '***' denote rejection of H_0 at the 10%, 5% and 1% significance levels, respectively.

Table 5. Results of the nonlinear cointegration test with Δz_{t-1} as the transition variable

Variable	F statistic	t statistic for the last parameter (p-value)
Panel 1: test equation $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}\Delta z_{t-1} + \alpha_3 z_{t-1}\Delta z_{t-1}^2 + \alpha_4 z_{t-1}\Delta z_{t-1}^3 + \varepsilon_t$, F_{EG4}^{Δ} statistic		
res_czk_eur	7,578***	3,048 (0,003)
res_czk_usd	5,8781***	0,307 (0,760)
res_pln_eur	2,059	0,686 (0,495)
res_pln_usd	2,539	0,928 (0,356)
Panel 2: test equation $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}\Delta z_{t-1} + \alpha_3 z_{t-1}\Delta z_{t-1}^2 + \varepsilon_t$, F_{EG3}^{Δ} statistic		
res_czk_eur	6,356***	-2,833 (0,006)
res_czk_usd	7,764***	-0,006 (0,995)
res_pln_eur	2,605	-0,178 (0,860)
res_pln_usd	3,103	0,428 (0,670)
Panel 3: test equation: $\Delta z_{t-1} = \alpha_1 z_{t-1} + \alpha_2 z_{t-1}\Delta z_{t-1} + \varepsilon_t$, F_{EG2}^{Δ} statistic		
res_czk_eur	5,086	-0,828 (0,410)
res_czk_usd	11,790***	1,427 (0,157)
res_pln_eur	3,939	0,210 (0,834)
res_pln_usd	4,608	1,488 (0,141)

Critical values for tests in Panel 1: 4,833 (1%), 4,102 (5%), 3,716 (10%); critical values for tests in Panel 2: 6,289 (1%0, 5,300 (5%), 4,793 (10%); critical values for tests in Panel 3: 9,132 (1%), 7,622 (5%), 6,859 (10%). The critical values were obtain by simulating 100000 regressions on de-meaned independent random walks. '*', '**', '***' denote rejection of H_0 at the 10%, 5% and 1% significance levels, respectively.

Interestingly, according to our testing procedure, 3 out of 4 relationships can be treated as cointegrating regressions. In the case of the Euro equation for the Czech Republic we found an evidence on the existence of ESTR cointegration, while in the case of the USD equation the testing procedure outlined in Section 3 leads to linear cointegrating, being in accordance with the result of the Engle and Granger test. In the case of the Euro equation for Poland we conclude that LSTR cointegration might be present (with a more mean-reverting behavior in the first regime). We executed also the ECM-based tests. Their results were similar but somewhat worse as the number of estimated parameters was greater. For this reason we do not present them here. We comment only that in the case of the F_{NEC} tests there exists a kind of a trade-of between estimating more parameters and fulfilling the COMFAC restrictions. Due to this we recommend to perform them jointly with the F_{NEG} tests.

The empirical example presented above shows the potential usefulness of the suggested testing procedure which can be applied to study many other phenomena like, for example, the present value relationship, the term structure of interest rates, money demand or business cycles asymmetry. In the case of exchange rates it makes it possible to explain the exchange rate disconnect puzzle, which can be solved by considering nonlinear adjustment process of a STAR type. Similarly, many other macroeconomic or financial hypotheses can be revisited. This will be analyzed in our future work. It is worth noticing that our tests are simple to perform in standard econometric packages and do not require any new sophisticated software.

5. Conclusions

The investigation of nonstationarity in conjunction with nonlinearity plays a prominent role in recent econometric studies. As standard unit root and cointegration tests lack their power and show serious size distortions when nonlinearities are involved, there is a need to develop new statistical tests, which might be helpful to find long-term relationships where linear-based tests fail. Our LSTR and joint LSTR and ESTR cointegration tests inscribe to this strand of research.

Our empirical investigation concerning the monetary approach to exchange rate determination enables to explain the observed long-lasting misalignment of exchange rates with economic fundamentals. Our explanation of this misalignment relates to the presence of nonlinear adjustments which can be effectively described by logistic and exponential smooth transition autoregressive processes.

References

[1] Bruzda J., The Power Evaluation of the Logistic Smooth Transition Cointegration Tests, mimeo, paper accepted for presentation at the *European Meeting of Statisticians*, Toruń, July 24-28, 2006.

[2] Bruzda J., Górecka D., Koźliński T. (2006), Examination of the Term Structure of Interest Rates in Poland - Linear and Non-linear Cointegration Analysis, paper presented at the 5th Annual International Conference *Forecasting Financial Markets and Economic Decision-Making*, May 11-13, 2006, Łódź.

[3] Caner M. Hansen B. E. (2001), Threshold Autoregression with a Unit Root, *Econometrica*, 69, 1555-1596.

[4] Chaouachi S., Dufrénot G., Mignon V. (2004), Modelling the Misalignments of the Dollar-Sterling Real Exchange Rate: A Nonlinear Cointegration Perspective*, Economics Bulletin*, 3, 1-11.

[5] Charemza W. W., M. Lifshits, S. Makarova (2005), Conditional Testing for Unit-Root Bilinearity in Financial Time Series: Some Theoretical and Empirical Results, *Journal of Economic Dynamics and Control*, 29, 63-96.

[6] Choi I., Saikkonen P. (2004), Testing Linearity in Cointegrating Smooth Transition Regressions, *Econometrics Journal*, 7, 341-365.

[7] Dufrénot G., Mathieu L., Mignon V., Peguin-Feissolle A. (2002), Persistent Misalignments of the European Exchange Rates: Some Evidence from Nonlinear Cointegration, *Economics Working Paper Archive EconWPA, International Finance*.

[8] Dufrénot G., Mignon V. (2002), *Recent Developments in Nonlinear Cointegration with Applications to Macroeconomics and Finance*, Kluwer Academic Publishers, Boston.

[9] Enders W., Granger C. W. J. (1998), Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest rates, *Journal of Business and Economic Statistics*, 16, 304-311.

[10] Enders W., Siklos P. L. (2001), Cointegration and Threshold Adjustment, *Journal of Business and Economic Statistics*, 19, 166-176.

[11] Kapetanios G., Shin Y. (2006), Unit Root Tests in Three-Regime SETAR Models, Econometrics, in press.

[12] Kapetanios G., Shin Y., Snell A. (2003), Testing for a Unit Root in the Nonlinear STAR Framework, *Journal of Econometrics*, 112, 359-379.

[13] Kapetanios G., Shin Y., Snell A. (2006), Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models, *Econometric Theory*, 22, 279-303.

[14] Kiliç R., de Jong R. (2004), Testing a Linear Unit Root Against a Stationary ESTAR Process, Working paper, Georgia Institute of Technology.

[15] Kremers J. J. M., Ericsson K. R., Dolado J. J. (1992), The Power of Cointegration Tests, *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

[16] Krishnakumar J., Neto D. (2005), Testing Unit Root in Threshold Cointegration, Working Paper no 2005.04, Univerité de Genève.

[17] Luukkonen R., Saikkonen P., Teräsvirta T. (1988), Testing Linearity Against Smooth Transition Autoregressive Models, *Biometrika*, 75, 491-499.

[18] Meese R. A., Rose A. K. (1991), An Empirical Assessment of Non-Linearities in Models of Exchage Rate Determination, *Review of Economic Studies*, 58, 603-619.

[19] Obstfeld M., Rogoff K. S. (2001), The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?, *Economics Working Paper Archive EconWPA, International Trade*.

[20] Phillips P. C. B., Perron P. (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, 75, 335-346.

[21] Seo M. (2004), Testing for the Presence of Threshold Cointegration in a Threshold Vector Error Correction Model, Working Paper, University of Wisconsin.

[22] Teräsvirta T. (1994), Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, *Journal of the American Statistical Association*, 89, 208-218.

[23] van Dijk D., Teräsvirta T., Franses P. H. (2002), Smooth Tramsition Autoregressive Models –A Survey of Recent Developments, *Econometric Reviews*, 21, 1-47.

Multi-Agent Approaches in Economics

Tomáš Cahlík, Jiří Hlaváček, Julie Chytilová, Jana Marková, Natálie Reichlová, Petr Švarc
Charles University in Prague – Faculty of Social Sciences, Institute of Economic Studies, Opletalova 26, 110 00 Prague 1, cahlik@fsv.cuni.cz

Abstract

In social sciences and economics out of the mainstream, reality is often understood as a complex system characterized – among others - by distributed interactions among heterogeneous agents. A somewhat different approach is in the mainstream economics, where representative agents – firms or households – are usually used. The advantage of the latter is that quite a lot of important results can be obtained analytically. If we work with heterogeneous agents, obtaining of analytical results is usually too complicated and computer simulation must be used. The aim of this paper is:

- to show that multi-agent approaches have been part of economic thinking for a long time,
- to illustrate the multi-agent approach in economics with some applications.

From the well-known economists, the complex – multi-agent approach is clearly visible in the work of Friedrich A. Hayek and Herbert A. Simon. Their multi-agent thinking is illustrated in the first part of this paper.

In the second part of this paper, the possibilities of the multi-agent approach in economics are shortly demonstrated on four applications. Applications concern the analyses of migration and migration networks and the analyses of the system of universities both with optimizing agents and agents with procedural rationality. The specific multi-agent model and simulation experiments for each application are described. The theoretical background is shortly summarized wherever necessary: motivation theory of A.H. Maslow and a simple neural network – perceptron – for the analyses of migration and migration networks and the difference between optimization and procedural rationality in the analyses of the system of universities.

Keywords: Multi-agent approach, simulation, migration, education
JEL: C61, F22, I21

1. Introduction

Economic systems can be understood as networks of agents on different aggregation levels, for example:

- The Economy of the EU as a network of national economies
- A national economy as a network of industrial branches or firms
- An economy as a network of firms and consumers
- An economy as a network of firms and workers
- The educational sector (in which the human capital as one of the main factors of production is created) as a network of schools
- The R&D sector (in which another main factor of production is created - knowledge and technology) as a network of research institutions
- Knowledge as a network of keywords or citations in scientific articles.

In these networks, spontaneous order is quite often created. This order results from the behavior of single agents that follow some set of rules. The regularity in the behavior of agents produces the regularity in the behavior of the whole system. Some prominent economists have made research on the spontaneous order, for example Friedrich von Hayek (Hayek, 1949) and Herbert Simon (Simon, 1976).

The spontaneous order concept has a lot of common features with the concept of a complex adaptive system. This concept has been elaborated for example in the Santa Fe Institute (Stuart Kaufmann, 2004). In this framework, a lot of mathematical models for the research of biological, social and economic systems have been constructed. A lot of methods of artificial intelligence were used in these models, for example cellular automata, neural networks and genetic programming.

In the mainstream economics, representative agents – firms or households – are usually used. These representative agents are homogenous. The advantage of the use of homogenous agents is that quite a lot of important results can be obtained analytically. In complex systems – in networks - reality is usually characterized by distributed interactions among heterogeneous agents. If we work with heterogeneous agents, obtaining of analytical results is usually too complicated and computer simulation must be used.

It is interesting that in the last years the number of economic articles in which networks of heterogeneous agents are explicitly used has been increasing. In the Web of Science, there are about 30 articles in all the periods covered, about one half is from the last five years. These articles describe usually applications, but some of them are about the dynamics of networks of generalized economic agents (Fagiolo, 2005; Zajac, 2004; Weizel and Konig, 2003; Tesfatsion, 2002 – among others).

The aim of this paper is to illustrate the multi-agent approach in economics with four applications. Applications concern the analyses of migration and migration networks and the analyses of the system of universities both with optimizing agents and agents with procedural rationality. The specific multi-agent model and simulation experiments for each application are described. The theoretical background is shortly summarized wherever necessary: motivation theory of A.H. Maslow and a simple neural network – perceptron – for the analyses of migration and migration networks and the difference between optimization and procedural rationality in the analyses of the system of universities.

2. Applications of the Multi-agent Approach

With new countries joining European Union the question of both economic and social effects of migration becomes more important. While some see migration as a solution for the problem of ageing population in former EU members, opponents of migration stress the possible negative impacts on a host country labor market, public finances and social conditions. In both cases, understanding migration flows can be the key determinant in formulating adequate political tools to its regulation. Two multi – agent models concerning migration are in 2.1 and 2.2.

With the Bologna process for education and Lisbon process for R&D, the questions about the impact of different ways of controlling universities in different EU members have become very important. Two multi – agent models concerning the evolution of a university system are in 2.3 and 2.4.

2.1 Model of Migration

A traditional approach to explain migration flows from one state or region to another is to employ *homo oeconomicus* and let him compare economic conditions in different regions.

Our model regards social networks as having an influence on migration. We consider not only social networks created in the destination region, but we also consider individuals as part of social networks in the home region. We base our theoretical model on Maslow's motivational theory, which suggests that wages, social networks, and feelings of stability provided by the home region may be key factors in migration decision making.

According to Maslow, there are five stages of needs fulfillment that an individual can experience. The first situation is wherein physiological needs are not gratified. Then an individual will move location provided such action serves to decrease hunger or thirst. Second, the individual has enough food but lives in an unsafe, threatening surrounding, then he or she will seek another location with a better level of safety, predictability, and order. On the other hand, safety needs are an important factor binding people to their native land. Third, both safety and physiological needs are fulfilled, but the individual suffers from the absence of family, friends, or colleagues. Social needs may encourage migration, especially in cases when family members have moved to a new destination already. The reunification of families is a fundamental stimulus of migration. On the other hand, the same strong force that motivates people to follow family members to a foreign country may otherwise persuade them to remain in their native land, surrounded by their families, friends, neighbors, and colleagues. The fourth factor that might motivate people to migrate is their longing for esteem, reputation, or glory and the last motive for movement may be, according to Maslow's theory, the desire for self-actualization.

In our analysis, we simplify Maslow's approach and employ only three motives of behavior. Furthermore, we assume that there are only two levels of decision making. At the first level, only physiological needs are taken into account. When an individual reaches some threshold level of saturation of physiological needs, safety and social needs follow. If wage levels are such that they meet physiological needs, the individual seeks to secure all needs simultaneously.

Model

In our model, agents have individual characteristics as described above. The environment in which the agents are situated comprises three regions. Each region has twenty times twenty cells and is convoluted into a torus shape. The toruses represent three regions with various wage levels. Wage level in region i at time t is given by the equation:

$$w_{it} = \frac{W_i}{n_{it}}, \ \ i = A, B, C ,$$

where W_i is the predefined wage parameter (the part of income of the region that goes to wages) and n_{it} is number of agents present in region i at time t.

There are 399 agents. At the beginning of the simulation, 133 agents are placed in each home region and their exact positions are determined randomly. It is possible to start different runs with the identical initial distribution of agents.

In our model, agents are maximizing their utility through wage maximization. They also appreciate living in the home region as well as the direct contact with other agents that are socially valuable to them.

The social value of agent k for agent j can be expressed as:

$$
\begin{aligned}
s_t^{jk} &= s_{t-1}^{jk} + \sigma \quad \text{if } k \text{ is present in } j\text{'s Moore neighborhood in period } t \\
s_t^{jk} &= s_{t-1}^{jk} - \sigma \quad \text{if } k \text{ is not present in } j\text{'s Moore neighborhood in period } t \\
s_t^{jk} &= 1 \qquad \text{for } s_{t-1}^{jk} + \sigma > 1 \\
s_t^{jk} &= 0 \qquad \text{for } s_{t-1}^{jk} - \sigma < 0 \\
s_0^{jk} &= 0 \qquad \text{for all agents}
\end{aligned}
$$

, where $t = 1,2,3...$ is the time variable and $\sigma \in \langle 0,1 \rangle$ is the coefficient that determines the speed of the establishment and the abandonment of social ties between agents.

Utility of agent j at time t is expressed separately for different wage levels as:

$$
u_t^j(w,b,S) = \begin{cases} (1+w_{it})^{1-\alpha-\beta}(1+b_t^j)^{\alpha}(1+S_t^j)^{\beta} \\ \quad \text{if agent is in home region} \\ (1+w_{it})^{1-\alpha-\beta}(1+S_t^j)^{\beta} \\ \quad \text{otherwise} \end{cases} \qquad \text{for } w_{it} > T
$$

$$
u_t^j(w,b,S) = w_{it} \qquad \text{for } w_{it} \le T
$$

where w_{it} is the wage an agent receives in the region in which he/she is present at time *t*. $\alpha \in \langle 0,1 \rangle$ is the parameter of the utility function indicating sensitivity to safety needs. Parameter $\beta \in \langle 0,1 \rangle$ expresses sensitivity to the social variable. The benefit from living in the home region is expressed by the variable b_j.

$$
b_t^j(\tau^j) = \frac{1}{\tau^j + 1},
$$

where τ^j is the number of periods agent j spent abroad. We assume that the additional utility gained from living in the home region decreases with the time spent abroad.

$S_t^j = s_t^{j1} + s_t^{j2} + ... + s_t^{j8}$ and $j1, j2 \ldots, j8$ are cells in the Moore neighborhood of agent j. The variable *T* is the physiological threshold. Below this level of wage agents are interested only in wage. The first period in which agents make their decision about migration may be postponed (to allow them to create social ties with other agents).

Simulation Results

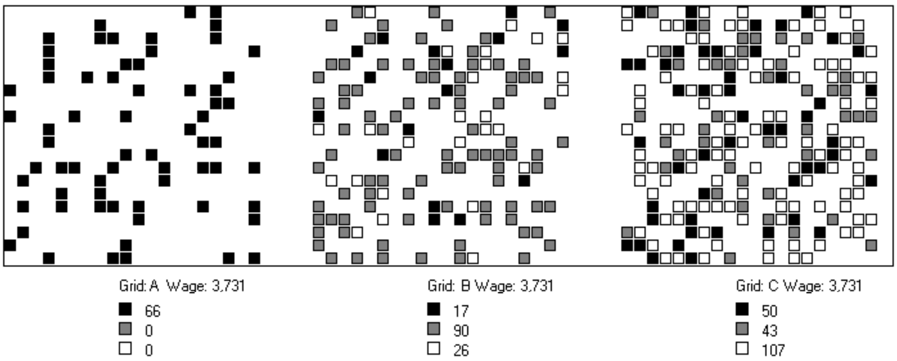
There are three regions—A, B, and C. Wages in regions A, B, and C are in Figures 1,2,3 indicated below each region, together with the number of agents of each color present in the region.

In the case of simple wage maximization and initial wages equal to:

$$W_A = 250 \quad W_B = 500 \quad W_C = 750,$$

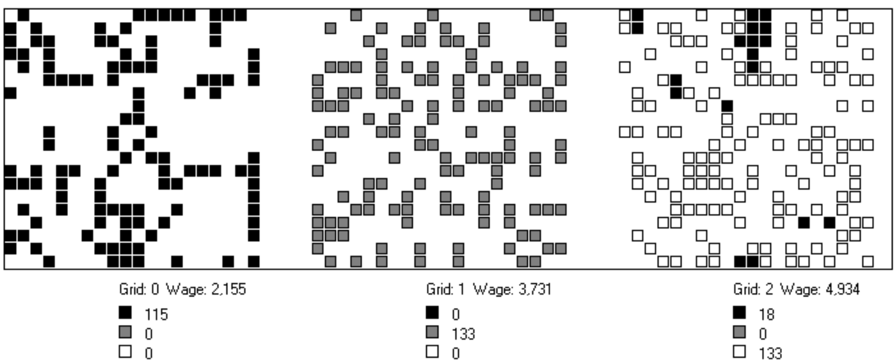
136 agents moved within the first period from their home region to another, and wages equalized in all three regions (see Figure 1). Hence the result is exactly what neoclassical theory predicts.

Figure 1: Simple wage maximization



Let us now discuss the model in which both safety and social needs are active for the same initial wage parameters and coefficients α = 0.3 and β = 0.3. The speed of establishment of social ties σ is equal to 0.1, threshold *T = 0* and postponement of the first migration decision *P = 10*. The result of simulation is in Figure 2.

Figure 2: Model with safety and social needs – P = 10, T = 0



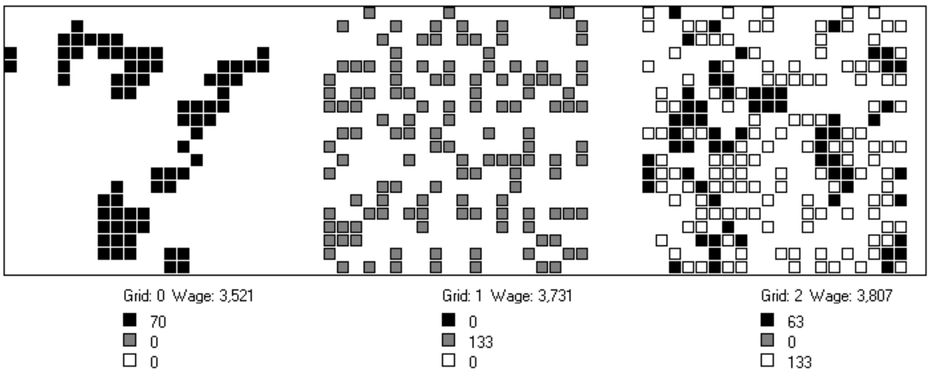
Only eight agents migrated in the 11th period, followed by one, one, three and five agents in subsequent periods. Then the system reached a stable state with zero migration. The wage level in Region B remained unchanged. We can see that the combination of social ties created within the first ten periods and safety needs leads to some kind of "conservatism"; that is, agents are less mobile and less willing to leave their home region. Agents from the region with the lowest wage moved to the region with the highest.

For threshold $T = 3.5$ and $P = 10$, an intriguing situation occurred (Figure 3). Stable state was not reached even in 1000 periods. Migration flows settled on two agents making their moves in each period.

Wages stabilized at the following level:

$$w_A = 3.521 \quad w_B = 3.731 \quad w_C = 3.807 .$$

Figure 3: Model with safety and social needs – P = 10, T = 3.5



Summary

In comparison with other migration models, we are able to explicitly work with a preference for the known, familiar environment, and an appreciation of the proximity of friends, family, and other socially tied individuals. These factors are, in the majority of models, hidden under the all-inclusive term "barriers." In fact, to understand the real factors influencing migration would be no doubt crucial in formulating policy measures aimed at migration.

Our model leads to the following conclusions:

1) People should move from countries where wages are below the physiological threshold.

2) If people appreciate living in their home country compared with a foreign country, and if their income in their home country is higher than the physiological minimum, then migration flows might stop even if wage differentials between states (regions) exist.

3) If people appreciate the proximity of those they know well and if their income is higher than the physiological minimum, then migration flows might stop even if wage

differentials between states (regions) exist.

The implications for real-world economies are quite apparent. First, if countries provide social-security benefits above the physiological threshold, people also take into account other than economic factors in their decision-making. Hence, people are less mobile and are less willing to move from a current location due to economic reasons.

Second, real migration flows depend on individual valuations of social ties and safety. These features might be largely determined by cultural habits and customs. Therefore, identical wage differences might induce different migration flows in various regions.

Third, wage differences may persist even though no barriers to migration exist. The way to induce labor mobility then lies in the reduction of the native-country preference through, for example, language education or closer international social ties.

2.2 Model of Migration Networks

One of the main characteristics of international migration and maybe the most striking one is forming of ethnic clusters of migrants: for example Turks in Germany, Tamils in Switzerland, Moroccans in the Netherlands and Belgium, Italians in Argentina, Greeks in Australia and Ukrainians in Canada. The prevailing explanation of this phenomenon is the so called network effect, the existence of externalities created by the former migrants. These externalities can strongly impact on the situation of new ones.

This network effect can be both positive and negative. Existing network positively influences future migration in many ways: for example by providing useful information about situation on the host country labor market and increasing the amount of ethnic goods available in the location. On the other hand, increasing number of migrants with similar characteristics can lower wages due to increasing competition on the specific labor market. This can have negative effect on future migration. We can expect that as the number of migrants increases the positive effect first dominates but later is dominated by the negative effect, resulting in the inverted U-shaped relationship between the number of migrants in a certain region and the probability of migrating to that region.

Aim of the model is to study formation and evolution of migration networks and how the existence of these networks influences migration patterns created. Resulting dynamics of the model is studied and compared with empirical findings.

Model

In our model, world is populated with artificial agents. They are homogenous in the sense they are equipped with the same cognitive capabilities but heterogeneous with respect to different initial settings of parameters representing individual knowledge.

Agents in our model are divided into several ethnic groups. The number of ethnic groups and regions does not need to be the same. The ethnics define groups of agents that could form a migration network. No agent from some ethnic group can form a network with any agent from other ethnics. Agents can also use other members of the same ethnic to gather information about the specific labor market.

Agents have to learn their environment. They react to the stimuli given by the environment. If agent's reaction to this stimuli is correct he/she will earn money in the form of wage available in the location he/she is living. In the case of failure he/she gets nothing.

To increase the probability of the right respond to the stimuli agents learn from their previous experiences and/or form networks with other agents living in the same region and belonging to the same ethnic group. If agent is a member of a network he/she can coordinate his/her actions with other members of the net. The network decides what will be the response to the stimuli of the environment and all members of the network will have the same response and reward.

After some time, agents compare their wealth level with other members of the ethnic group they belong to and decide whether or not to move to some different location. The only criterion in the decision process is the average wage experienced by other agents from the ethnics.

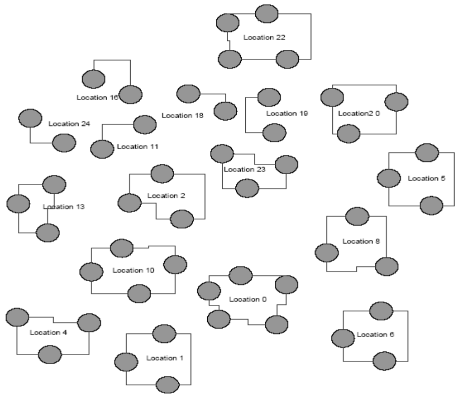
The cognitive capabilities of individual agents are modeled here by a single-layer neural network (called perceptron) with two input neurons and one output neuron. If the classification task is complex enough, forming networks could be the only way how agents can successfully learn it.

We use the Boolean exclusive-OR function to generate the complexity in the classification task. This specific function was chosen as a simplest linearly non-separable function that cannot be, taken singularly, simulated (or classified) by the perceptron neural network, hence successively learned by a single agent.

Simulation Results

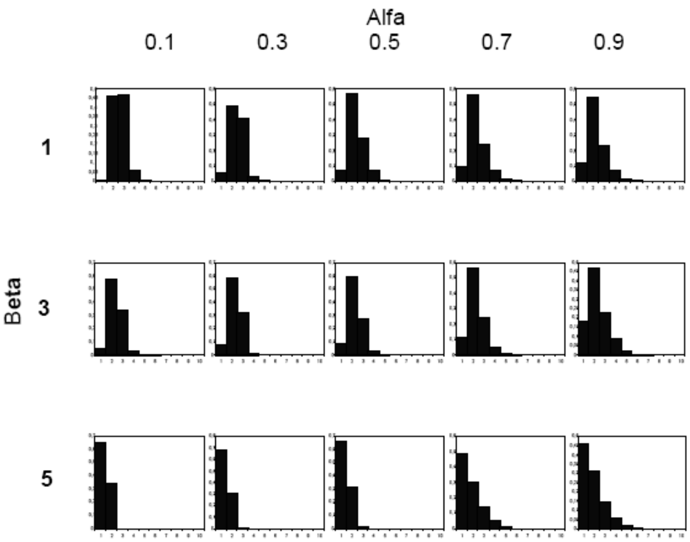
Formation of networks and its effect on the migration of agents is the key aspect of our model. Typical situation generated during simulation is displayed in Figure 4.

Figure 4: Typical networks generated during simulation



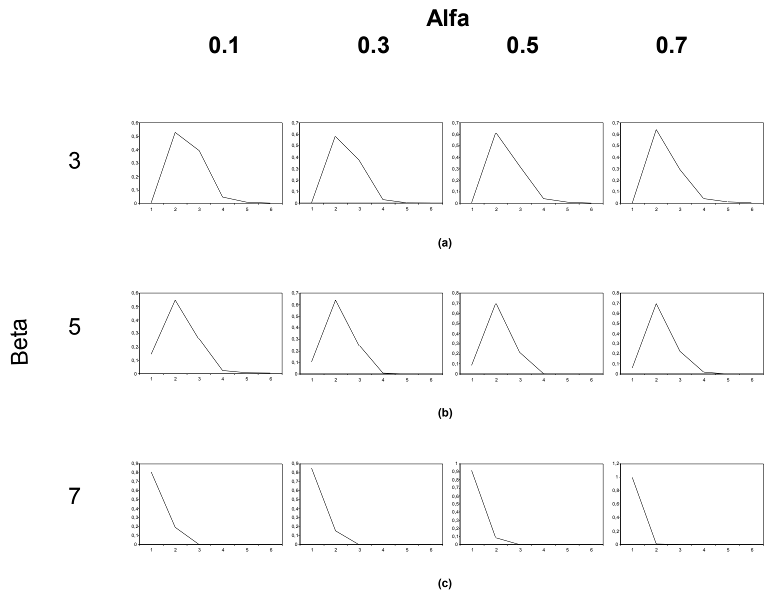
Interesting question concerning migration networks is what will be the typical or average size of networks and how it will change with the changes of main parameters of the model. Figure 4 shows histograms of the size of networks recorded after 1000 periods of simulation for three different levels of acceptable number of mistakes β (rows) and five different levels of the sensitivity of wages on the number of migrants α (columns). The results tell us that leaving other parameters unchanged the size of network decreases substantially with the increasing acceptable number of mistakes and increases when wages are less sensitive to the number of workers from the same ethnic group.

Figure 5: Histograms of the size of networks for different parameters



Finally we can turn to the main task of the model and explore how the formation of networks influences migration of agents. Figure 6 shows how often agents in our simulation chose to migrate to the locations with different number of members from the same ethnic group. Rows and columns represent different values of parameters β and α. The emergence of the inverse U-shaped relation between the number of migrants in certain location and the probability of migration to that location is typical in the situations where the value of parameter β is relatively low. On the other side, with high value of β agents preferred to migrate to the locations with smaller number of previous migrants.

Figure 6: Frequency of migration as function of the number of members of the same ethnic group in the region for different parameters



Summary

In our model, we combine the agent-based modeling with the neural network technology to simulate the formation and evolution of migration networks and we analyse how these networks impact on the migration of agents.

Despite the simplicity of the model, simulation results successfully replicate some real world characteristics of migration patterns reported in empirical studies concerning the effects of migration networks. For example, the U-shaped relation between the size of the network and the frequency of choosing particular location is empirically well known for low-skill and illegal migrants.

2.3 Model of Optimizing Universities

The aim of this agent-based model is to compare different ways of financing the university education and to construct the supply function of a university, i.e. the dependence of the supply (the number of places for students) on its price, i.e. on the revenue of the university (either from scholarship or from state subsidy) per student.

We derive our results from an optimization model and we have encountered the problem of criterion. *Ex definitione* the profit criterion cannot be used. We have been thinking about some other production criterion, for example the number of students (school-leavers). But there is another problem here: optimum is situated on a frontier of the set of available solutions. That leads to maximal number of students, maximal tuition-fee (not less then the fee that fills in the school capacity), minimal teachers´ salary (not more then the salary which is enough for the minimal necessary number of teachers). Such a university could lose its accreditation, because the short-run optimality threatens the long-run survival here.

The best solution could be understood as the solution maximizing the probability of survival. This criterion does not contradict the profit criterion, but it is more general: when an agent is threatened by missing money only, the mentioned two criteria are identical.

Model

On the one hand the universities are threatened by insufficient income, on the other hand with the loss of accreditation through the escape of teachers. Control variables are the size of the tuition-fee and the teachers' salary. In competition with other universities: too high fee could lead to half-empty class-rooms; too low teachers' salaries could cause their escape. However, too low fee or too high teachers' salaries could lead to the lack of money. Optimum is a compromise that takes into account the behavior of competitors (average fee and average salary in the system). That is why our model is an iterative system, in which the results of optimization in the step n influence the formulation of the optimization problem of all competitors in the step n+1.

The probability of survival is dependent on the level of the critical variable. We suppose that:

- it is equal to zero for income equal to some threshold, respectively with income below this threshold,
- it approaches one as the critical variable (income) grows to infinity,
- it is bigger if the relative reserve of the critical variable grows in relation to the threshold of the definite downfall.

These assumptions are fulfilled by the Pareto distribution of the first degree (Figure 7). Let us denote y as the value of the critical variable whose decrease under the threshold $b>0$ would lead to certain downfall. Distribution function (probability of the downfall for the critical variable equal to y) is thus for this distribution:

$F(y)=(y-b)/y$ for $y \geq b$,

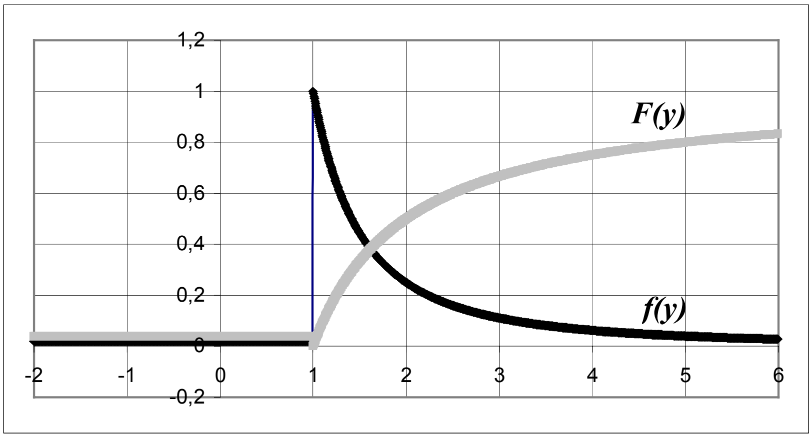
$F(y)=0$ for $y<b$

Corresponding probability density function is

$f(y)=b/y^2$ for $y \geq b$

$f(y)=0$ for $y<b$

Figure 7: Pareto distribution of the first degree for the threshold of the definite downfall of b = 1: distribution function F(y), probability density function f(y)



Pareto distribution of the first degree has median equal to the double of the border of the unavoidable downfall (on graph y equal to 2). Probability of survival is here proportional to relative reserve, thus the probability of avoiding downfall due to low level of critical variable y is for example for $y= 1.6$ triple relative to $y= 1.2$: relative reserve is in the first case 60%, in second 20%.

Simulation Results

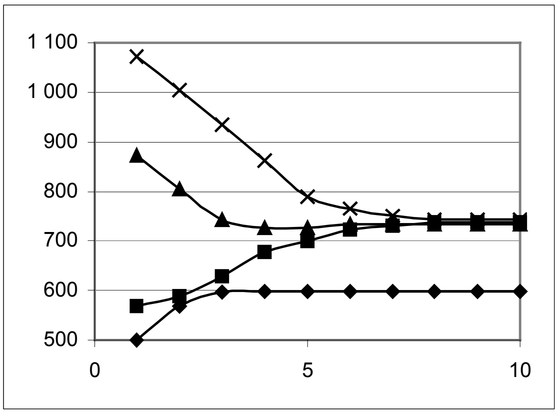
In our simulation experiments, we analyze impact of three alternatives of financing of universities: only from the tuition-fee, only from the state subsidy and combined financing as the third alternative. Main results are presented in Table 1 and Figures 8, 9, 10.

Table 1: Results of simulation experiments - comparing alternatives of he financing of universities

	alternative of financing	*university No.* 1	2	3	4	*in the system*	*difference to the combined alternative*
number of students	*combined financing*	597	737	733	743	2 810	-
	only from tuition-fee	599	780	778	781	2 938	+ 128
	only from state subsidy	556	754	1 100	1 200	3 610	+ 800
number of teachers	*combined financing*	68	85	99	108	360	-
	only from tuition-fee	66	80	90	99	335	- 25
	only from state subsidy	56	76	110	125	367	+ 7
size of tuition-fee	*combined financing*	2,8	2,8	2,8	2,8	2,8	-
	only from tuition-fee	7,3	7,3	7,3	7,3	7,3	+ 4,5
	only from state subsidy	0	0	0	0	0	-
teachers´ salary	*combined financing*	10,8	10,8	10,8	10,8	10,8	-
	only from tuition-fee	10,4	10,4	10,4	10,4	10,4	- 0,4
	only from state subsidy	10,5	10,5	10,5	10,5	10,5	- 0,3

Figure 8: Results of simulation experiments – financing only from the tuition-fee

number of students



demand for teachers

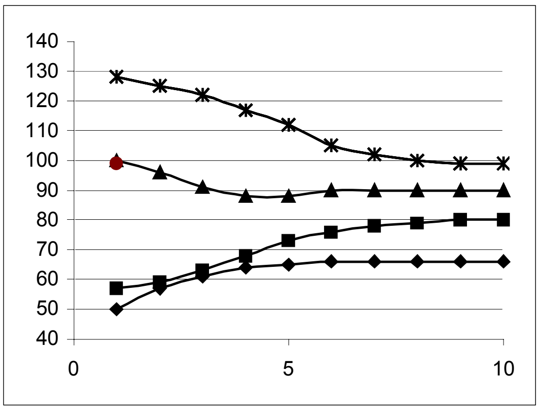


Figure 9: Results of simulation experiments – financing only from the state subsidy

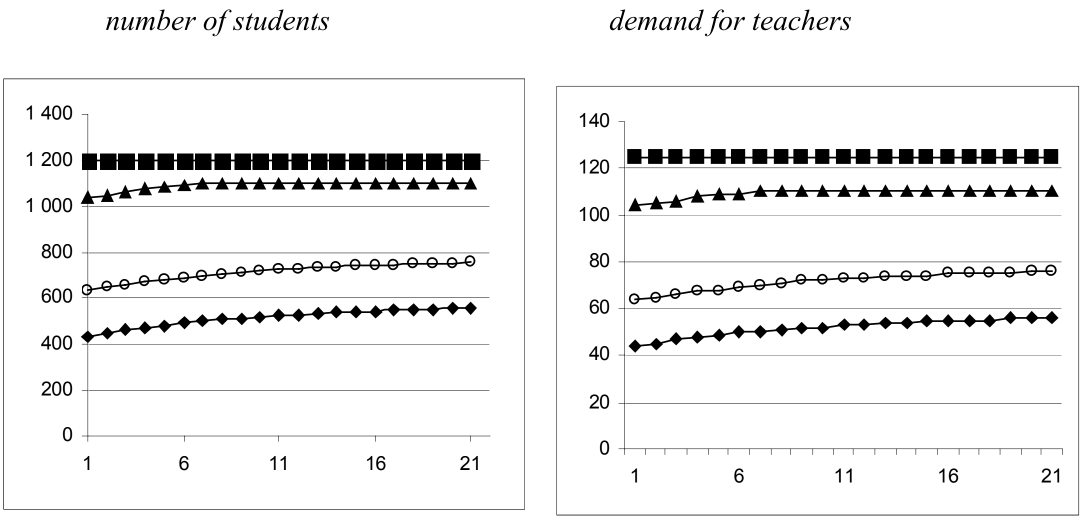
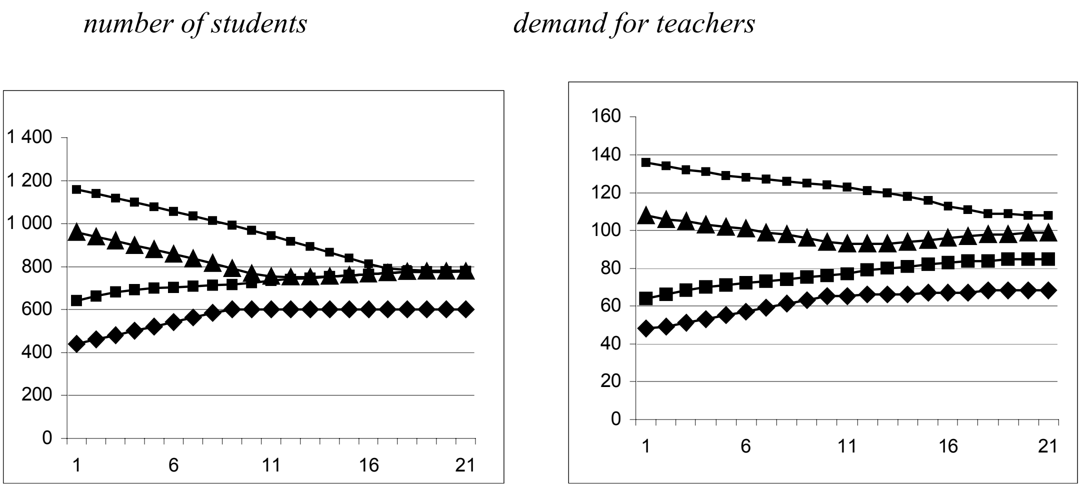


Figure 10: Results of simulation experiments – combined financing



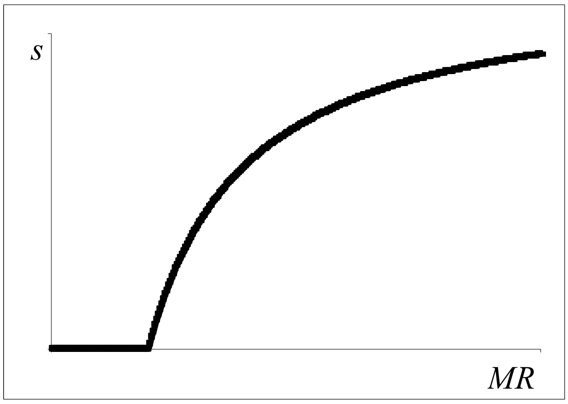
Summary

Surprisingly *ceteris paribus* (including the same incomes):

- the alternative „only tuition fees“ is the worst as far as the teachers' salary is concerned,
- there are significantly less teachers in the alternative „only tuition fees“,
- there are significantly more students in the alternative „only state subsidy“.

This model enables us to derive the supply function of the university (Figure 11). We can understand it as the capacity of university multiplied by the probability of its survival depending on marginal revenues (i.e. tuition-fee + state subsidy per student):

Figure 11: Supply function of a university



The supply function is growing and strictly concave when its value is positive. Its graph is similar to the graph of the standard supply function when only profit is maximized. The difference to the standard profit maximization model lies in the fact that the optimum does not appear on the productivity frontier.

2.4 Model of Universities with Procedural Rationality

In this part, impacts of different ways of university financing are analyzed as well as in the preceding part, but universities are not optimizing agents, they obey the procedural rationality. They implement some reaction rule in each situation developed during the simulation in the model. The reaction rule in our model is quite simple, it defines how each university reacts to the difference between its capacity and the number of applicants. Capacity depends on the number of teachers. We assume that other constraints (space capacity of lecture halls etc.) are not important.

The three alternatives of financing of universities can be again: only from the tuition-fee, only from the state subsidy and combined financing.

We assume that the state subsidy depends on the number of students, in other words the state defines the subsidy per student. Variable costs are dependent on the number of students as well, fixed costs depend on the capacity of the specific university. In some experiments, we implement a research subsidy that depends on the number of teachers.

Model is realized in Excel and is available on samba.fsv.cuni.cz/~cahlik in the directory „Articles´ Support“.

Model

The reaction rule for all universities is "the more students the best". They accept all applicants and increase capacity immediately – they hire new teachers. They behave in this way even in the case in which another strategy could bring them higher profits. We can imagine they believe that universities with higher number of students will be more protected by the state in the future, will be just "too big to fail". So our reaction rule describes the behavior of universities maximizing the subjective probability of survival in the long time horizon.

Algorithm is iterative, in each step (academic year) each university:

- can bankrupt,
- can increase its teachers' salary if teachers are missing,
- can increase its capacity.

After each iteration, we summarize which universities have survived and how the possible bankruptcies impact on the demand for studies at survived universities.

Simulation Results

We describe here the simulation of a system of five universities, in which two universities S1 and S2 can be financed only from tuition fees and three universities S3, S4 and S5 only from state subsidies.

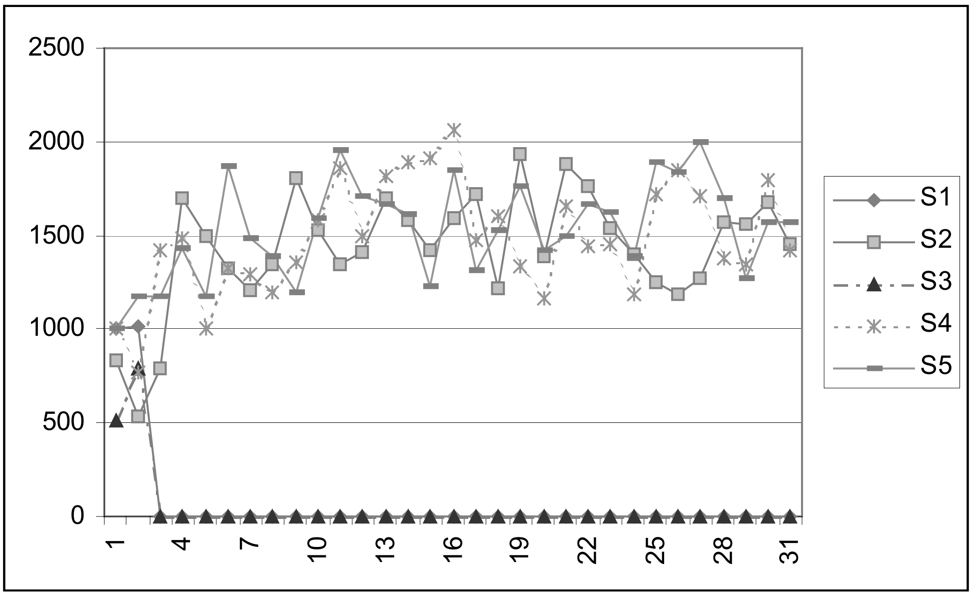
Our basic question is the possibility of survival of universities in different member countries of the European Union, where in some member countries tuition fees are not allowed.

In simulation experiments, we have changed different parameters for single universities – fixed costs, capacity and research subsidy.

Different fixed costs

We set fixed costs high for universities S1 (tuition fee) and S3 (state subsidy) and low for S2 (tuition fee), S4 and S5 (state subsidy). The bankruptcy of some of high fixed costs universities in the most of experiments occurs, independently on the way of financing. A typical development is in Figure 12. Number of students is on the vertical axis, no students are the case of bankruptcy.

Figure 12: Bankruptcy of university S3 in the third period



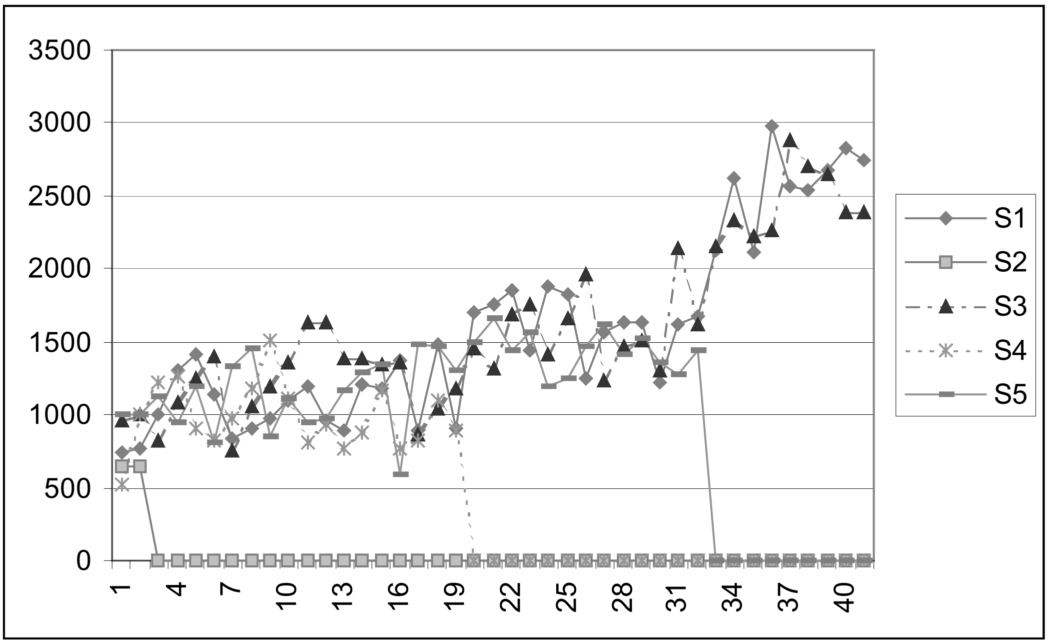
Different capacity of schools

Capacity depends on the number of teachers. We set the number of teachers low for S1 (tuition fee) and S3 (state subsidy), high for S2 (tuition fee) and S5 (state subsidy) and medium for S4 (state subsidy). In experiments, we have not found any relationship between bankruptcy and differences in capacities or ways of financing.

Different research subsidies

We implement specific state subsidies for research (subsidies depend on the number of teachers) and set them high in S1 (tuition fee) and S3 (state subsidy). Universities with research subsidies usually survive, independently on the way of financing. Other universities usually bankrupt. A typical development is in Figure 13.

Figure 13: Survival of S1 and S3



Summary

In our simulation experiments, we have not found any relationship between the way of financing and bankruptcy. Other parameters – fixed costs or availability of research subsidies – seem to be more important for the survival.

3. Conclusions

Multi-agent approaches are relatively new to economics, it is usually accepted that they started in the mid-nineties by a group of economists interested in studying evolutionary processes in economy in the bottom-up fashion. It is quite interesting that in Czechoslovakia, before the change of the social and economic system there had been research groups in this specific area – Zeman's group in Prague, Šujan's group in Bratislava and Kindler's group in Pilsen – among others.

So-called agent-based computational economics (ACE) was defined by Leigh Tesfatsion (2002) as: "the computational study of economies modeled as evolving systems of autonomous interacting agents. Starting from initial conditions, specified by the modeler, the computational economy evolves over time as its constituent agents repeatedly interact with each other and learn from these interaction."

Not only in economics, but in other social sciences as well, reality is often understood as a complex system. Arthur, Durlaf and Lane (1997) characterized complexity with a set of features that complex systems ought to fulfill. These features are as follows:

a) Dispersed local interactions among heterogeneous agents.

b) Non-existence of a global regulator, which would use up all interactions in an economy, even if there may exist weak global interactions.

c) Hierarchical structure with many complicated feedbacks and mutual transversality.

d) Adaptation through learning and adapting agents.

e) Permanent innovations motivated by new markets, technologies, rules of behavior and institutions.

f) Non – equilibrium dynamics ether with no equilibrium or many equilibriums and the system probably not being close to a global optimum.

If the Simon's approach to complex systems and characteristics of Arthur, Durlauf and Lane were compared, parallels in majority of points would be found. Although Simon does not use directly the term „agent“ or „heterogeneous“, his perception of elements of complex system is very similar. Moreover, he perceives interactions between agents as a feature contributing to the complexity of the system. According to Simon, majority of existing systems has a hierarchical structure, evolution by natural selection is hardly possible for other than hierarchically structured system. Furthermore, all the learning takes place inside an agent. An organization learns in two ways: through learning of its members or through acceptance of a new member, who brings new knowledge. Simon also perceives innovations as a part of complex system, which can not be predicted in advance. If the system is complex and its environment changes all the time (i.e. conditions under which biological and social evolution takes place), there is no guarantee that the system will be located near some stable equilibrium, local or global. It is obvious that Simon perceives the real world as a complex system, on the other hand he warns against attributing overestimated importance to complexity.

In this paper, four different applications of the multi-agent approach in economics are summarized. In (Cahlík et al, 2006), these applications are described in more detail and two other applications are added, the analysis of the dynamics of market structure and the analysis of the dynamics of consumption structure. Farther on, the link between the multi – agent approach and artificial intelligence is broadly discussed in this book as well.

References

[1] Arthur, W. B., S. N. Durlaf, a D. Lane (1997): The Economy As an Evolving Complex System II.; Addison-Wesley, Reading, MA.

[2] Cahlík, T. et al. (2006): Multiagentní přístupy v ekonomii. Připraveno k publikaci pro nakladatelství Karolinum, (Multi-Agent Approaches in Economics, is to be published in Karolinum Publishing House).

[3] Fagiolo, G. 2005): Endogenous neighborhoods formation in a local coordination model with negative network externalities. Journal of Economic Dynamics and Control.

[4] Hayek, Friedrich A. (1949): Individualism and Economic Order. Routledge and Kegan Paul Ltd., London.

[5] Kauffman, S (2004): Investigations. Oxford University Press, 2000. In Czech: Čtvrtý zákon. Paseka.

[6] Simon, H. A. (1976): How Complex are Complex Systems? Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 1976; Volume Two: Symposia and Invited Papers.

[7] Tesfatsion, L. (2002): Agent-based computational economics: Growing economies from the bottom up. Artificial life.

[8] Weitzel, T., Konig,W. (2003): Computational economics and the formulation of an interdisciplinary network theory. Wirtschaftsinformation.

[9] Zajac, J. (2004): Agent-based simulation of economics. Ekonomický časopis, 2004.

Stability of Bayes Actions

Jana Čerbáková[1]
Charles University, Prague, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, 186 75 Prague, Czech Republic
janacerb@karlin.mff.cuni.cz

Abstract

The paper deals with the Bayesian decision problem to find an action minimizing the expected loss on the set of all plausible actions. The loss function expresses the consequences of choosing a particular decision for a concrete realization of random parameter. We assume that the set of all actions does not depend on the choice of probability distribution. The main goal of the paper is to study the behavior of minimal losses and corresponding optimal actions with respect to small changes in the probability distribution. This helps us to evaluate an error which can be caused by using an approximated or perturbed probability distribution. We also show how these results can be related to stability with respect to weak convergence.

Keywords

Bayes action, stability, weak convergence, probability metric
JEL: C11, C44

1 Introduction

Incomplete or unprecise knowledge of input parameters of economic models influences the quality of the obtained optimal decisions which may be then quite different from the truly optimal actions. In Bayesian models, the uncertainties are incorporated into the model and there is a chance to analyze stability of decisions with respect to the perturbed input, new information, etc. Simple economic applications of Bayes methods have been frequently used in practice, see e.g. [13]. This paper is devoted to stability analysis for Bayes decision models.

In Bayes decision model (see [1]) the only unknown quantity is the parameter $\theta \in \Theta$, where the set of admissible values Θ is a non-empty closed subset of $\mathbb{R}^n$. We assume that θ is random with probability distribution P belonging to a class of all probability distributions $\mathcal{P}$ defined on $(\Theta, \mathcal{B}(\Theta))$, where $\mathcal{B}(\Theta)$ denotes Borel σ-algebra of Θ. The decision maker chooses his action (decision) a from the set of all admissible actions $\mathcal{A}$, where $\mathcal{A}$ is supposed to be a non-empty closed subset of $\mathbb{R}^m$. He makes his decision on the basis of random lower semicontinuous loss function $L : \mathcal{A} \times \Theta \to \mathbb{R}$ which represents the loss caused by action a when the true value of random parameter is θ.

Definition 1.1 (Bayes action)**.** An action $a^* \in \mathcal{A}$ is called Bayes if and only if it minimizes the expected loss

$$a^* \in \operatorname*{argmin}_{a \in \mathcal{A}} \int_{\Theta} L(a, \theta) \, \mathrm{d}P(\theta), \tag{1}$$

where P is the assumed probability distribution at the time of decision making. The set of all Bayes actions, i.e. the set of optimal solutions of (1) with respect to P, is denoted by $\mathcal{A}^*(P)$.

Distribution P can represent the prior probability distribution or, in statistical decision problems, the posterior probability distribution after observing the data. The posterior distribution combines the prior information with the sample information represented by the likelihood function according to Bayes theorem, see [1]. In view of stability discussed in the paper, it is not important to distinguish between P representing a prior or a posterior distribution.

In real problems we usually do not know the exact probability distribution of random parameters. We have to estimate them. Therefore, it is very important to be able to calculate an error, which can be caused by using

[1]This work was supported by the Grant Agency of the Czech Republic (grant 201/05/H007). The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s.

estimated distribution or be confident the error will be sufficiently small. And this is problem of stability, i.e. small modifications of underlying probability distribution or problem formulation are supposed to cause only small changes of solutions.

In this paper we shall give not only the usual continuity results (see definition 1.2 below) but we shall also quantify the errors in minimal expected loss and in the Bayes actions due to perturbations. Such results are of importance in real-life problems, e.g. in robustness analysis of the obtained results.

Definition 1.2 (Stable action)**.** We say that a Bayes action $a^* \in \mathcal{A}^*(P)$ is stable if for every sequence of probability distributions $\{P_n, n \in \mathbb{N}\}$ weakly converging to P, $P_n \xrightarrow[n\to\infty]{w} P$, where $P, P_n \in \mathcal{P}, \forall n$, and for every sequence of loss functions $\{L_n, n \in \mathbb{N}\}$ converging (in some topology) to L

$$\left[\int_\Theta L_n(a^*, \theta)\, \mathrm{d}P_n(\theta) - \inf_{a\in\mathcal{A}} \int_\Theta L_n(a, \theta)\, \mathrm{d}P_n(\theta)\right] \xrightarrow[n\to\infty]{} 0 \tag{2}$$

holds true.

It was shown in [5] that for $L_n(a, \theta)$ converging to $L(a, \theta)$ uniformly in a and θ, the condition (2) is equivalent to

$$\left[\int_\Theta L(a^*, \theta)\, \mathrm{d}P_n(\theta) - \inf_{a\in\mathcal{A}} \int_\Theta L(a, \theta)\, \mathrm{d}P_n(\theta)\right] \xrightarrow[n\to\infty]{} 0, \tag{3}$$

which can be rewritten as

$$\inf_{a\in\mathcal{A}} \int_\Theta \left[L(a, \theta) - L(a^*, \theta)\right] \mathrm{d}P_n(\theta) \xrightarrow[n\to\infty]{} \inf_{a\in\mathcal{A}} \int_\Theta \left[L(a, \theta) - L(a^*, \theta)\right] \mathrm{d}P(\theta) = 0.$$

The stability of a^* becomes then the question of the continuity, at P, of the infimum integral functional

$$\inf \left\{ \int_\Theta F(a, \theta)\, \mathrm{d}P(\theta) : a \in \mathcal{A} \right\}, \tag{4}$$

where $F(a, \theta) := L(a, \theta) - L(a^*, \theta)$. We will assume that the loss function L does not depend on n, thus (3) can be used in definition of stability instead of (2). All presented results can be extended to uniformly convergent $\{L_n, n \in \mathbb{N}\}$.

Other formulations of stability of a Bayes decision problem can be found in [6]. The authors introduce two definitions of Strong Stability for ε-minimal solutions of Bayes problem and derive sufficient conditions for their equivalence. They also prove stability results with respect to weak convergence of probability distributions based on the work of [2], [3], [5] and [12]. The most important findings will be mentioned in section 4.

Using general stability results of [11] in the context of Bayesian decision analysis we shall be able to obtain error bounds for optimal values (minimal expected losses) and for the solution sets (Bayes actions) caused by perturbations of the underlying distribution. This concept of stability is formulated in section 2, the main results on improved distances of solutions sets are presented in theorems 3.1 and 3.2.

2 Problem formulation

According to definition 1.1, Bayesian decision analysis deals with the problem

$$\min \left\{ \int_\Theta L(a, \theta)\, \mathrm{d}P(\theta) : a \in \mathcal{A} \right\}, \tag{5}$$

namely with the behavior of the set of optimal solutions $\mathcal{A}^*(P)$ and optimal values $\vartheta(P)$ in dependence on small changes of probability distribution P. Together with the original problem (5) we consider a pertubated model with another distribution $Q \in \mathcal{P}$ instead of P. We apply the following notation:

$$\vartheta_{\mathcal{U}}(Q) := \inf \left\{ \int_\Theta L(a, \theta)\, \mathrm{d}Q(\theta) : a \in \mathcal{A} \cap \mathrm{cl}\,\mathcal{U} \right\} \quad \text{the optimal value of perturbed model,}$$

$$\mathcal{A}^*_{\mathcal{U}}(Q) := \left\{ a \in \mathcal{A} \cap \mathrm{cl}\,\mathcal{U} : \int_\Theta L(a, \theta)\, \mathrm{d}Q(\theta) = \vartheta_{\mathcal{U}}(Q) \right\} \text{the set of optimal solutions of perturbed model.}$$

To measure the distance of probability distributions we define for any nonempty and open subset $\mathcal{U}$ of $\mathbb{R}^m$ the set

$$\mathcal{P}_{L_\mathcal{U}} := \left\{ Q \in \mathcal{P} : -\infty < \int_\Theta \inf_{a \in \mathcal{A} \cap r\mathbb{B}} L(a,\theta)\,\mathrm{d}Q(\theta) \quad \forall r > 0, \sup_{a \in \mathcal{A} \cap \mathrm{cl}\,\mathcal{U}} \int_\Theta L(a,\theta)\,\mathrm{d}Q(\theta) < \infty \right\} \tag{6}$$

to ensure all mentioned optimization problems are well defined and on $\mathcal{P}_{L_\mathcal{U}}$ the following probability pseudo-metric

$$d_{L_\mathcal{U}}(P,Q) := \sup_{a \in \mathcal{A} \cap \mathrm{cl}\,\mathcal{U}} \left| \int_\Theta L(a,\theta)\,\mathrm{d}P(\theta) - \int_\Theta L(a,\theta)\,\mathrm{d}Q(\theta) \right|. \tag{7}$$

A uniform distance of the form (7) is called a distance having ζ-structure.

Example 2.1. An important class of probability metrics with ζ-structure are the *Fortet-Mourier metrics* defined for $p \geq 1$ by

$$\zeta_p(P,Q) := \sup_{L \in \mathcal{L}_p} \left| \int_\Theta L(\theta)\,\mathrm{d}P(\theta) - \int_\Theta L(\theta)\,\mathrm{d}Q(\theta) \right|,$$

where $P, Q \in \mathcal{P}_p := \{Q \in \mathcal{P} : \int_\Theta \|\theta\|^p\,\mathrm{d}Q(\theta) < \infty\}$ and $\mathcal{L}_p$ denotes the classes of locally Lipschitz continuous functions that increase with p, i.e.

$$\mathcal{L}_p := \left\{ L : \Theta \to \mathbb{R} : |L(\theta) - L(\bar\theta)| \leq \max\{1, \|\theta\|, \|\bar\theta\|\}^{p-1}\|\theta - \bar\theta\|, \forall \theta, \bar\theta \in \Theta \right\}.$$

In the one-dimensional case we can use the following explicit formula

$$\zeta_p(P,Q) = \int_{-\infty}^{\infty} \max\{1, |t|^{p-1}\}|G(t) - H(t)|\,\mathrm{d}t,$$

where G, H are distribution functions associated with P, Q, see [4], [9]. For example for two 0-1 random variables

$$X_i = \begin{cases} 0 & \text{with probability } p_i, \\ 1 & \text{with probability } 1 - p_i, \end{cases}$$

$p_i \in [0,1]$, with probability distributions P_i, $i \in \{1,2\}$ we obtain $\zeta_p(P_1, P_2) = |p_1 - p_2|$ for $p \geq 1$.

3 Stability theorems

To state the main stability results for optimal decisions we need to introduce the *growth function*

$$\psi_P(\tau) := \min \left\{ \int_\Theta L(a,\theta)\,\mathrm{d}P(\theta) - \vartheta(P) : d(a, \mathcal{A}^*(P)) \geq \tau, a \in (\mathcal{A} \cap \mathrm{cl}\,\mathcal{U}) \right\}$$

and its inversion $\psi_P^{-1}(t) := \sup\{t \in \mathbb{R}_+ : \psi_P(\tau) \leq t\}$.

Theorem 3.1. *Let $L : \mathbb{R}^m \times \Theta \to \bar{\mathbb{R}}$ be a random lower semicontinuous function, $\mathcal{A}^*(P) \neq \emptyset$ and $\mathcal{U} \subset \mathbb{R}^m$ be an open bounded neighbourhood of $\mathcal{A}^*(P)$, where $P \in \mathcal{P}_{L_\mathcal{U}}$.*

Then the multifunction $\mathcal{A}^*_\mathcal{U} : (\mathcal{P}_F, d_F) \to \mathbb{R}^m$ is upper semicontinuous at P and for any $Q \in \mathcal{P}_{F_\mathcal{U}}$, the following properties hold

$$|\vartheta(P) - \vartheta_\mathcal{U}(Q)| \leq d_{L_\mathcal{U}}(P,Q), \tag{8}$$

$$\emptyset \neq \mathcal{A}^*_\mathcal{U}(Q) \subset \mathcal{A}^*(P) + \psi_P^{-1}(2d_{L_\mathcal{U}}(P,Q))\mathbb{B}. \tag{9}$$

Proof. For more general problem a similar result is proved in [11], theorem 5 and theorem 9. We present here a version of proof for our special problem where the set $\mathcal{A}$ does not depend on probability distribution an we obtain a tighter bound (9) for optimal decisions.

For $a \in \mathcal{A} \cap \mathrm{cl}\,\mathcal{U}$ and $Q \in \mathcal{P}_{L_\mathcal{U}}$ define the function f from $(\mathcal{A} \cap \mathrm{cl}\,\mathcal{U}) \times \mathcal{P}_{L_\mathcal{U}}$ to $\mathbb{R}$ by $f(a,Q) := \int_\Theta L(a,\theta)\,\mathrm{d}Q(\theta)$. The function is lower semicontinuous and finite with respect to (6), see theorem 3 in [11]. Hence, $\mathcal{A}^*_\mathcal{U}(Q)$ is nonempty for each $Q \in \mathcal{P}_{L_\mathcal{U}}$. For $a^* \in \mathcal{A}^*(P)$, $Q \in \mathcal{P}_{L_\mathcal{U}}$ and $\bar a \in \mathcal{A}^*_\mathcal{U}(Q)$ the inequalities (8) follows:

$$\begin{aligned} |\vartheta(P) - \vartheta_\mathcal{U}(Q)| &\leq \max\left\{ \left|\int_\theta L(a^*,\theta)(Q-P)\,(\mathrm{d}\theta)\right|, \left|\int_\theta L(\bar a,\theta)(P-Q)\,(\mathrm{d}\theta)\right| \right\} \\ &\leq d_{L_\mathcal{U}}(P,Q). \end{aligned}$$

The mapping $\mathcal{A}^*_{\mathcal{U}}$ is closed at $P \in \mathcal{P}_{L_{\mathcal{U}}}$ and, hence, upper semicontinuous at P.

By definition of ψ, $d_{L_{\mathcal{U}}}(P,Q)$ and with $\bar{a} \in \mathcal{A}^*_{\mathcal{U}}(Q) \subset (\mathcal{A} \cap \mathrm{cl}\,\mathcal{U}) =: \mathcal{A}_{\mathcal{U}}(P)$ we derive

$$\begin{aligned}
\psi(d(\bar{a}, \mathcal{A}^*(P))) &\le \Big|\int_\Theta L(\bar{a},\theta)\,\mathrm{d}P(\theta) - \vartheta(P)\Big| \\
&\le \Big|\int_\Theta L(\bar{a},\theta)(P-Q)\,(\mathrm{d}\theta) + \vartheta_{\mathcal{U}}(Q) - \vartheta(P)\Big| \\
&\le \Big|\int_\Theta L(\bar{a},\theta)(P-Q)\,(\mathrm{d}\theta)\Big| + |\vartheta_{\mathcal{U}}(Q) - \vartheta(P)| \\
&\le 2d_{L_{\mathcal{U}}}(P,Q).
\end{aligned}$$

From here we obtain $d(\bar{a}, \mathcal{A}^*(P)) \le \psi_P^{-1}(2d_{L_{\mathcal{U}}}(P,Q))$, which implies (9). □

Theorem 3.1 stands as a basic tool for measuring errors caused by employing inaccurate probability distribution. We illustrated under which conditions it can be declared that small changes of underlying distribution do not evoke significant distance of Bayes actions (9) and difference in suffered losses (8).

If, in particular, the problem (5) has *k-order growth* at the solution set $\mathcal{A}^*(P)$ for some $k \ge 1$, i.e. $\psi_P(\tau) \ge \gamma\,\tau^k$ for each small $\tau \in \mathbb{R}^+$ and some $\gamma > 0$, then for $\bar{a} \in \mathcal{A}^*_{\mathcal{U}}(Q)$ and $P, Q \in \mathcal{P}_{L_{\mathcal{U}}}$,

$$\gamma\, d(\bar{a}, \mathcal{A}^*(\mathcal{P}))^k \le \psi(d(\bar{a}, \mathcal{A}^*(P))) \le 2d_{L_{\mathcal{U}}}(P,Q).$$

Hence,

$$\emptyset \ne \mathcal{A}^*_{\mathcal{U}}(Q) \subset \mathcal{A}^*(P) + \left(\frac{2}{\gamma} d_{L_{\mathcal{U}}}(P,Q)\right)^{\frac{1}{k}} \mathbb{B}.$$

Localized optimal values $\vartheta_{\mathcal{U}}(Q)$ and solution sets $\mathcal{A}^*_{\mathcal{U}}(Q)$ can be replaced by their global versions $\vartheta(Q)$ and $\mathcal{A}^*(Q)$, e.g. if the problem (5) is convex, $\mathcal{A}^*_{\mathcal{U}}(Q) \subset \mathcal{U}$ and $\exists \delta > 0$ such that $\forall Q \in \mathcal{P}_{L_{\mathcal{U}}} : d_{L_{\mathcal{U}}}(P,Q) < \delta$ (cf. [11]).

In the next theorem we combine convexity with properties of locally Lipschitz functions. A similar theorem can be also found in [11].

Theorem 3.2. *Let the assumptions of theorem 3.1 be satisfied. Furthermore, let $\mathcal{A}$ be convex and $F(\cdot,\theta)$ be convex on $\mathcal{A}$ for each $\theta \in \Theta$. If there exist constants $K > 0$, $p \ge 1$ such that $\frac{1}{K}L(a,\cdot) \in \mathcal{L}_p$ for each $a \in \mathcal{A} \cap cl\mathcal{U}$ then $\exists \delta > 0$ such that*

$$|\vartheta(P) - \vartheta(Q)| \le K\zeta_p(P,Q),$$

$$\emptyset \ne \mathcal{A}^*(Q) \subset \mathcal{A}^*(P) + \psi_P^{-1}(2K\zeta_p(P,Q))\mathbb{B},$$

whenever $P, Q \in \mathcal{P}_p$ and $\zeta_p(P,Q) < \delta$.

Proof. The statement follows by application of theorem 3.1 and the fact that $\frac{1}{K}L(a,\cdot) \in \mathcal{L}_p$ implies $d_{L_{\mathcal{U}}}(P,Q) \le K\zeta_p(P,Q)$. □

For convex model (5) it can be proved (see [10], theorem 7.69 and [11], theorem 13) that the ε-minimal solution sets behave Lipschitz continuously in terms of the Pompeiu-Hausdorff distance $\mathbb{D}_\infty(C,D) := \inf\{\eta \ge 0 : C \subset D + \eta\mathbb{B}, D \subset C + \eta\mathbb{B}\}$ defined for nonempty closed sets $C, D \subset \mathbb{R}^m$.

Theorem 3.3. *Let F be a random lower semicontinuous convex function, $\mathcal{A}$ closed convex, $P \in \mathcal{P}_{L_{\mathcal{U}}}$ and $\mathcal{A}^*(P)$ be nonempty and closed. Then there exist constants $\rho > 0$ and $\bar{\varepsilon} > 0$ such that the estimate*

$$\mathbb{D}_\infty(\mathcal{A}^*_\varepsilon(P), \mathcal{A}^*_\varepsilon(Q)) \le \frac{2\rho}{\varepsilon} d_{L_{\mathcal{U}}}(P,Q)$$

holds for $\mathcal{U} := (\rho + \bar{\varepsilon})\mathbb{B}$ and any $\varepsilon \in (0,\bar{\varepsilon})$, $Q \in \mathcal{P}_{L_{\mathcal{U}}}$ such that $d_{L_{\mathcal{U}}}(P,Q) < \varepsilon$.

4 Stability with respect to weak convergence

Let us return to the definition 1.2 of stability of Bayes actions with respect to weak convergence of $\{P_n \in \mathcal{P}, n \in \mathbb{N}\}$ to $P \in \mathcal{P}$. In section 1 we derived that the stability of Bayes action is under uniform convergence of loss functions equivalent to the convergence

$$\inf_{a\in\mathcal{A}} \int_\Theta F(a,\theta)\,\mathrm{d}P_n(\theta) \xrightarrow[n\to\infty]{} \inf_{a\in\mathcal{A}} \int_\Theta F(a,\theta)\,\mathrm{d}P(\theta). \tag{10}$$

In previous section we introduced how we can measure the distance of the two optimal values from (10), cf. (8). Now we show under which assumptions this distance converge to 0, i.e. when $a^* \in \mathcal{A}^*(P)$ is stable.

The most cited conditions of stability come from [12], we present them in the next theorem.

Theorem 4.1. *Assume that*

(i) $L : \mathcal{A} \times \Theta \to \mathbb{R}$ *is lower semicontinuous on* $\mathcal{A} \times \Theta$*,*

(ii) $L(a^*, \cdot)$ *is continuous on* Θ*,*

(iii) L *has locally equi-lower-bounded growth, i.e.* $\forall a \in \mathcal{A}$ *there exist a neighbourhood* $\mathcal{U}(a)$ *of* a *and* $b(a) \in \mathbb{R}$ *such that for all* $\bar{a} \in \mathcal{U}(a)$*,*

$$L(\bar{a}, \theta) - L(a^*, \theta) \geq b(a), \quad \forall \theta \in \Theta.$$

Then $a^* \in \mathcal{A}^*(P)$ *is stable if and only if for any sequence* $P_n \xrightarrow[n\to\infty]{w} P$ *and every* $\varepsilon > 0$ *the sequence*

$$\left\{ \inf_{a\in\mathcal{A}} \int_\Theta F(a,\theta)\, dP_n(\theta),\, n \in \mathbb{N} \right\}$$

has a bounded sequence of ε*-minimal solutions. It means that for any* $\varepsilon > 0$ *there exist a compact subset* $K_\varepsilon \subset \mathcal{A}$ *and a sequence* $\{a_n \in K_\varepsilon, n \in \mathbb{N}\}$ *such that for all* n

$$\int_\Theta F(a_n,\theta)\, dP_n(\theta) < \inf_{a\in\mathcal{A}} \int_\Theta F(a,\theta)\, dP_n(\theta) + \varepsilon.$$

Assumptions (i) and (ii) imply that F (defined in (4)) is lower semicontinuous on $\mathcal{A} \times \Theta$. Condition (iii) is trivially satisfied for $L(a,\theta)$ bounded on $\mathcal{A} \times \Theta$. The existence of bounded sequence of ε-minimal decisions is guaranteed e.g. when $\mathcal{A}$ is compact.

For discrete set $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ the following theorem was proved in [6].

Theorem 4.2. *Let* $L(a,\theta)$ *be lower semicontinuous in* θ *for all* $a \in \mathcal{A}$*. If* $P(\mathcal{A}_L) = 0$*, where* $\mathcal{A}_L := \{\theta \in \Theta :$ θ *is discontinuity point of* $L(a^*, \cdot)$*,* $a^* \in \mathcal{A}^*(P)\}$*, then* a^* *is stable.*

To derive other sufficient conditions for stability in sense of definition 1.2 we employ the following representation

$$\begin{aligned}
\int_\Theta L(a^*,\theta)\,\mathrm{d}P_n(\theta) - \inf_{a\in\mathcal{A}} \int_\Theta L(a,\theta)\,\mathrm{d}P_n(\theta) &= \sup_{a\in\mathcal{A}} \left[\int_\Theta L(a^*,\theta)\,\mathrm{d}P_n(\theta) - \int_\Theta L(a,\theta)\,\mathrm{d}P_n(\theta) \right] \\
&\leq \left| \int_\Theta L(a^*,\theta)\,\mathrm{d}P_n(\theta) - \int_\Theta L(a^*,\theta)\,\mathrm{d}P(\theta) \right| \\
&+ \left| \int_\Theta L(a^*,\theta)\,\mathrm{d}P(\theta) - \inf_{a\in\mathcal{A}} \int_\Theta L(a,\theta)\,\mathrm{d}P(\theta) \right| \\
&+ \sup_{a\in\mathcal{A}} \left| \int_\Theta L(a,\theta)\,\mathrm{d}P_n(\theta) - \int_\Theta L(a,\theta)\,\mathrm{d}P(\theta) \right| \\
&\leq 2d_L(P_n, P),
\end{aligned}$$

where

$$d_L(P_n, P) := \sup_{a\in\mathcal{A}} \left| \int_\Theta L(a,\theta)\,\mathrm{d}P_n(\theta) - \int_\Theta L(a,\theta)\,\mathrm{d}P(\theta) \right|.$$

The problem of stability of Bayes action $a^* \in \mathcal{A}$ becomes now the task of P-uniformity of a class $\mathcal{L}_a := \{L_a(\cdot) := L(a,\cdot)\,,\ a \in \mathcal{A}\}$, i.e. under which conditions

$$\lim_{n\to\infty} d_L(P_n, P) = 0 \tag{11}$$

holds true for every P_n weakly convergent to P.

In [3] necessary (A1) and sufficient (A2) conditions ensuring (11) were introduced:

(A1) $\sup_{a\in\mathcal{A}} \sup_{\theta_1,\theta_2\in\Theta} |L(a,\theta_1) - L(a,\theta_2)| < \infty$,

(A2) $\lim_{\delta\downarrow 0} \sup_{a\in\mathcal{A}} \left[\int_\Theta \sup_{\theta_1,\theta_2\in\mathcal{U}(\theta,\delta)} |L(a,\theta_1) - L(a,\theta_2)|\,\mathrm{d}P(\theta) \right] = 0.$

Condition (A1) is satisfied if $\mathcal{L}_a$ is the class of equi-bounded functions.

The assumption of equi-continuity of $\mathcal{L}_a$ then implies (A2). Equi-continuity is fulfilled for $L(a,\theta)$ continuous in θ uniformly in a. Other sufficient condition for equi-continuity is local Lipschitz continuity, i.e. existence of function $g(a,\theta)$, $\alpha > 0$, $\varepsilon > 0$:

(a) $|L(a,\theta) - L(a,\theta')| \leq g(a,\theta) \parallel \theta - \theta' \parallel^{\alpha}$,

(b) $\sup\limits_{a \in \mathcal{A}} \int\limits_{\Theta} g(a,\theta)\,\mathrm{d}P(\theta) < \infty$.

hold true $\forall \theta' \in \mathcal{U}(\theta, \varepsilon)$.

For detailed discussion of above mentioned requirements on stability see [3], [5], [6], [8] and [12].

Moreover, results presented in section 3 can be also applied to sample based Bayes actions (solutions of (5) with respect to empirical probability distributions).

5 Conclusion

General stability results applied to stability of Bayes decision problems provide error bounds for the minimal expected loss and for the optimal Bayes actions related to perturbations of the input. In addition, they may be used to quantify classical stability properties with respect to weak convergence.

References

[1] Berger, J. O.: Statistical Decision Theory: Foundations, Concepts and Methods. Springer Series in Statistics, Springer Verlag, New York 1980.

[2] Berger, J. O., Salinetti, G.: Approximations of Bayes Decision Problems: The Epigraphical Approach. *Annals of Operations Research* **56** (1995), 1–13.

[3] Billingsley, P.,Topsøe, F.: Uniformity in weak convergence. *Zeitschrift für Wahrschenkeitlichkeitstheorie und verwandte Gebiete* **39** (1967), 1–30.

[4] Dupačová, J., Römisch, W.: Quantitative Stability for Scenario-based Stochastic Programs. Prague Stochastics '98 (M. Hušková et al. eds.), JČMF 1998, 119-124.

[5] Kadane, J., Chuang, D. T.: Stable Decision Problems. *Annals of Statistics* **10** (1978), 1095–1110.

[6] Kadane, J., Srinivasan, C.: Bayes Decision Problems and Stability. *Indian Journal of Statistics* **60** (1998), 383–404.

[7] Kadane, J., Salinetti, G., Srinivasan, C.: Stability of Bayes Decisions and Applications. Robust Bayesian Analysis, Lecture Notes in Statistics **152**, Springer Verlag, New York 2000, 187–196.

[8] Lucchetti, R., Salinetti, G., Wets, R, J-B: Uniform Convergence of Proability Measures: Topological Criteria. *Journal of Multivariate Analysis* **51** (1994), 252–264.

[9] Rachev, S. T.: Probability Metrics and the Stability of Stochastic Models, Wiley, Chichester 1991.

[10] Rockafellar, R. T., Wets, R. J-B: Variotional Analysis. A Series of Comprehensive Studies in Mathematics **317**, Springer Verlag, New York 1998.

[11] Römisch, W.: Stability of Stochastic Programming Problems. Handbooks in Operation Research and Management Science: Stochastic Programming, A. Ruszczyński, A. Shapiro eds., Elsevier Science B.V. **10** 2003, 483–554.

[12] Salinetti, G.: Stability of Bayesian Decisions. *Journal of Statistical Planning and Inference* **40** (1994), 313–329.

[13] Wonnacott, T. H., Wonnacott, R.J.: Decision Trees. Inductory Statistics for Business and Economics, 4th Edition, John Wiley & Sons **21** 1990, 639–663.

SOMA application to the Traveling Salesman Problem

Zuzana Čičková, Ivan Brezina
Dolnozemská cesta 1/b, 852 35 Bratislava
cickova@euba.sk, brezina@euba.sk

Abstract

This article describes the application of Self Organizing Migrating Algorithm (SOMA) to the well-known optimization problem - Traveling Salesman Problem (TSP). SOMA is a new optimization method based on Evolutionary Algorithms that are originally focused on solving non-linear programming tasks containing continuous variables. The TSP has model character in many branches of Operation Research because of its computational complexity; therefore the use of Evolutionary Algorithm requires some special approaches to guarantee feasibility of solutions. In this article two concrete examples of TSP as 8 cities set and 25 cities set are given to demonstrate the practical use of SOMA method. Firstly, the penalty approach is applied as a simple way to guarantee feasibility of solution. Then, new approach that works only on feasible solutions is presented.

Keywords

Traveling Salesman Problem, Evolutionary Algorithms, Self Organizing Migrating Algorithm
JEL: C61

1 INTRODUCTION

The Traveling Salesman Problem (TSP) is well known in optimization. A traveling salesman has number of n cities to visit. A tour (the sequence in which the salesman visits different cities) should be such that every city on the list is visited once and only once, except that salesman returns to the city from which he starts. The goal is to find a tour that minimizes the cost (usually the total distance), among all the tours that satisfy this criterion. The problem can be visualized on graph. Each city becomes a node. Edge lengths correspond to the distance between the attached cities (we assume complete weighted graph). TSP can be formulated as finding a Hamiltonian cycle in a weighted graph with the minimal length.

TSP is one of the most discussed problems in literature. Many algorithms were applied with more or less success. There are various ways to classify algorithms, each with its own merits. One way to classify algorithms is by implementation principle:

Explicit enumeration. It leads to reconnaissance all possible solutions of problems, therefore is applicable for small problem size.

Deterministic methods. These algorithms base only on rigorous methods of „classical" mathematics. Some additional information, such as gradient, convexity etc. is usually needed (Branch and Bound Algorithm, Cutting Plane Method, Dynamic Programming etc.).

Stochastic methods. These algorithms work on probabilistic methods to solve problems. Stochastic algorithms work slowly and are applicable only for „guessing"(Monte Carlo, Random search Walk, Evolutionary Computation etc.).

Combined methods. Combined methods are comprised by stochastic and deterministic composition. Various metaheuristics algorithm has been devised (Ant Colony Optimization, Memetic Algorithms, Genetic algorithms etc.). Metaheuristics consist of general search procedures whose principles allow them to escape local optimality using heuristics design. Evolutionary algorithms are significant part of metaheuristics.

2 PRINCIPLES OF SOMA

Evolutionary Algorithms (EA) are relatively new optimization techniques that use mechanisms inspired by biological evolution, such as reproduction, mutation, recombination and natural selection in focus to Genetic Algorithms, which are the most popular types of EA. Self Organizing Migrating Algorithm (SOMA) was created in 1999 [6]. It can be classified as an evolutionary algorithm, despite the fact that no new individuals are created during the computation, and only the position of individuals in the search place is changed (the behavior of a social group of individuals). Even through SOMA is not based on the philosophy of evolution, the final result, after one migration loop, is equivalent to the result from one generation derived by EA algorithms.

SOMA, as well other EA algorithms, is working on a population of individuals (*np*–number of individuals in the population). A population can be viewed as a $np\times(d+1)$ matrix, where the columns represent individuals. Each individual represents one candidate solution for the given problem, i.e. a set of arguments of objective function. Associated with each individual is also the fitness $f_c(\mathbf{x}_i)$, $i=1,2,...,n$ which represents the relevant value of objective function. The fitness does not take part in the evolutionary process itself, but only guides the search.

	$f_c(\mathbf{x}_i)$	1	2	....	d
$\mathbf{x}_1$	$f_c(\mathbf{x}_1)$	x_{11}	x_{12}		x_{1d}
$\mathbf{x}_2$	$f_c(\mathbf{x}_2)$	x_{21}	x_{22}		x_{2d}
. . .					. . .
$\mathbf{x}_{np}$	$f_c(\mathbf{x}_{np})$	x_{np1}	x_{np2}	....	x_{npd}

Table 1 Population

A population $P^{(0)}$ is usually randomly initialized at the beginning of evolutionary process used so-called Specimen, which is defined for each parameter.

$$\text{Specimen} = \left\{\left\{\text{type},\left\{\text{Lo,Hi}\right\}\right\}_1, \left\{\text{type},\left\{\text{Lo,Hi}\right\}\right\}_2, ..., \left\{\text{type},\left\{\text{Lo,Hi}\right\}\right\}_d\right\} \tag{1.1}$$

Type (integer, real, discrete etc.)

Lo – lower border

Hi – upper border

The borders define the allowed range of values for each parameter of individuals at the beginning and also during migration loops. The initial population $P^{(0)}$ is generated as follows:

$$\mathrm{P}^{(0)} = x_{i,j}^{(0)} = \mathrm{rnd}\left(x_{i,j}^{(Hi)} - x_{i,j}^{(Lo)}\right) + x_{i,j}^{(Lo)} \quad i=1,2,...,np \quad j=1,2,...,d \tag{1.2}$$

If during the migration loop some parameters of individual exceed specimen's borders, that parameters are changed according to the rule:

$$x_{i,j}^{t+1} = \begin{cases} rnd\left(x_{i,j}^{(Hi)} - x_{i,j}^{(Lo)}\right) + x_{i,j}^{(Lo)}, \text{ ak } x_{i,j}^{t+1} < x_{i,j}^{(Lo)} \text{ alebo } x_{i,j}^{t+1} > x_{i,j}^{(Hi)} \\ \\ x_{i,j}^{t+1}, \text{ otherwise} \end{cases} \tag{1.3}$$

SOMA, as other EA algorithms, is controlled by a special set of parameters:

d - dimensionality. Number of arguments of objective function.

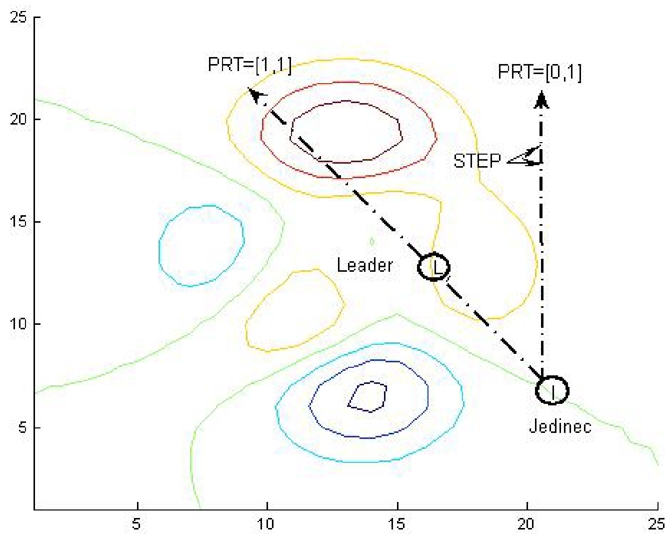
np – population size. It depends of user and his hardware.
m – migrations. Represent the maximum number of iteration.
mass – path length, $mass \in \langle 1.1, 3 \rangle$. Represents how far an individual stops behind the leader.
step– $step \in \langle 0.11, mass \rangle$. Defines the granularity with what the search space is sampled.
prt – perturbation, $prt \in \langle 0,1 \rangle$. Determines whether an individual travel directly towards the leader or not.

SOMA was inspired by the competitive–cooperative behavior of intelligent creatures solving a common problem. SOMA works in migration loops. Basic principle is shown in Figure 1.1. Each individual is evaluated by cost function and the individual with the highest fitness – *Leader* is chosen for the current migration loop. According to the *step*, other individuals begin to jump towards the Leader according the rule:

$$x_{i,j}^{mk+1} = x_{i,j,start}^{mk} + (x_{L,j}^{mk} - x_{i,j,start}^{mk})tprt_j \qquad t \in \langle 0, \text{by } step \text{ to } mass \rangle \tag{1.4}$$

Each individual is evaluated after each jump using the objective function. The jumping continues, until new position defined by the Mass is reached. Then the individual returns to that position, where the best fitness was found:

$$x_{i,j}^{mk+1} = \min\left\{ f_c(x_{i,j}^{mk}), f_c(x_{i,j,start}^{mk} \right\} \tag{1.5}$$



Before individual begins to jump, a random number for each individual component is generated and is compared with *prt*. If the random number is larger then *prt*, then the associated component of the individual is set to 0. Hence, the individual moves in *n-k* dimensional subspace. This fact establishes a higher robustness of the algorithm.

$$\mathrm{PRT}_j = \begin{cases} 1, \text{ ak } rand_j \langle 0,1 \rangle > prt \\ \\ 0, \text{ otherwise} \end{cases} \tag{1.6}$$

The general convention used is known as *AllToOne* strategy. In literature [5], [6] can be found different working strategies of SOMA. All versions are fully comparable with each other in the sense of finding optimum of objective function.

3 SOMA for TSP

TSP is a discrete optimization problem. By solving, a natural representation of individual, known from genetic algorithm, is used. Using this representation, the cities are listed in the order in which they are visited. Fitness represents the cost of corresponding tour. Firstly, example of 8 cities set, as Slovak region cities, is used. The penalty approach is used as a simple way to guarantee feasibility of solutions (In case duplicity appears, the penalty constant, which is greater as longest distance between couple of cities, is adding as many times as duplicity appears). The values of d is fixed according to problem size to 8, parameters np = 80 and mig = 300, $mass$ = 3 are used during the simulation. Thus, one simulation takes about 20 minutes on PC AMD64 3200+ with 512Mb RAM. For computation, system Matlab 8.1 with some changed function in Toolbox Evolutionary Algorithms was used. To choose the other parameters efficiently, 168 simulations were carried out and some statistical methods (ANOVA, Kruskal –Wallis etc.) were applied. The best parameter values obtained: prt = 0.8; $step$ = 0.9 Results for 8 simulations gives Table 2.

	1	2	3	4	5	6	7	8	Mean	Variance
Tour length	848	848	857	848	848	857	857	848	851,375	21,67

Table 2

The tour length 848 was obtaining also by using system GAMS (model tsp42 from model library). In order to use penalty approach for solving 25 cities set (capitals of EU), the parameters of np = 150 and mig = 3000, was not found single one feasible solution (cca 17 hours of computing). Because of the penalty approach works on infeasible solution too, complexity is increased. For that reasons, new approach that works only on feasible solutions was developed. A population of individuals is randomly initialized at the beginning as follows:

$$\mathrm{P}^{(0)} = x_{i,j}^{(0)} = \mathrm{randperm}\left(d\right) \qquad i = 1,2,\ldots,np \;\; j = 1,2,\ldots,d \tag{1.7}$$

This function is assigning each individual with a random permutation vector of integers size d. (random permutation of cities index on the salesman route). During the SOMA's migration loop some parameters of individual exceeds specimen's borders, in that case, only a valid part of individual is preserved, and that part is completed to a permutation size d, following idea of validity of each individual.

1) let **m** is a vector of parameters of individual size d with k different elements. If $d - k = 0$, go to step 7)
2) **p** is a d-k size vector of rand permutation of d-k elements that don't contain in vector **m**
3) if number of nonzero elements of vector p=0, go to step 7)
4) m_c is the first repetitive element of vector **m**
5) p_k is the firth nonzero element of vector **p**, then $m_c = p_k$
6) let p_k =0 and back to 3)
7) return **m**

This approach was tested on 8 cities set task with number of migration loops mig = 50 (one simulation takes cca 1 min.). Results included number of migration loop, in which was firth time the tour length 848 appeared, give table 3:

	1	2	3	4	5	6	7	8
Tour length	848	848	848	848	848	848	848	848
mig	6	2	7	10	11	7	7	7

Table 3

Increasing number of migration loops, more alternative solutions are a part of final population. In this case, all 16 cyclic permutations are components of final migration loop. For solving problem of 25 cities set, parameters *np* = 150, *mig* = 2000, *prt* = 0.8, *step* = 0.9, *mass* = 2 are used. The results are compared with optimal solution obtained by GAMS (14118 km tour length).A result gives table 4:

SOMA	Tour length	%dev.opt.
1	15077	6,79%
2	14760	4,54%
3	15077	6,79%
Mean	14971	6,04%

Table 4

The results of this article show, that presented approach is completely used for small size problems, for problems of larger size SOMA gives relatively good solution. Problems in solving large instance problems are similar in appearance to other algorithms. In spite of it, SOMA is usable for larger size problems. Perhaps, another special variants of SOMA for solving TSP could lead to increasing efficiency of computation.

References

[1] Brezina, I.: Kvantitatívne metódy v logistike, EKONÓM, Bratislava, 2003

[2] Čičková, Z.: Aplikácia evolučných algoritmov na riešenie úlohy obchodného cestujúceho, Dizertačná práca, KOVE FHI, EU, Bratislava, 2005

[3] Haupt, L. Randz - Haupt Sue Ellen: Practical Genetic algorithm, A John Wiley&Sons,Inc. Publication, New Jersey, 2004

[4] Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs, Springer – Verlag Berlin Heidelberg, USA, 1996

[5] Onwubolu, G.C., Babu, B.V: New Optimization Techniques in Engineering, Studies in Fuzziness and Soft Computing, Berlin, Heidelberg, New York, Hong Kong, London, Milano, Paris, Tokyo, Sringer, Volume 14, 2004

[6] Zelinka, I.: Umělá intelligence v problémech globální optimalizace, BEN-technická literatúra, 2002

Real option application for modular project valuation

Miroslav Čulík[1]
VŠB – Technical University Ostrava, Faculty of economics, Finance department,
miroslav.culik@vsb.cz

Abstract

This paper concentrates on a problem faced by a firm operating in the electricity generating sector. Specifically, the problem is to evaluate and decide, if to invest in building of large capacity power plant, which is irreversible investment, or in a flexible sequence of smaller modular power plants, whose total capacity equals to the capacity of the large project. Modular project is valued by employing real option methodology as a compound growth option to invest in the following module; calculation is made under price uncertainty.

Key words

Option, real option, modularity, compound growth option, Black – Scholes model, investment, flexibility. mean-reversion process, energy sector.

JEL: G31,

1. Introduction

Electricity generation and supply has been regarded for many years as a sector, which was best run as a monopoly and in most cases as a state-owned monopoly. If private utilities have been allowed, they were tightly regulated. Over the last ten years, this view of electricity markets has changed and in most countries the electricity supply industry has undergone some reform. Restructuring the power sector is a very complex problem influenced by national energy strategies and policies, macroeconomic developments and national conditions. In Europe, all EU member countries have been gradually liberalising their electricity markets in accordance with the Directive for Unification of Electricity Markets.

Electricity market liberalisation process usually leads to the increase in uncertainty. Originally, under the monopoly situation, the only uncertainties were considered in fuel prices and electricity demand. After the liberalisation process, companies producing electricity face other market risks, particularly electricity prices and companies' competition. Moreover, both under monopoly situation and on liberalised markets, electricity suppliers face legal risks (legal environmental controlling, etc.).

This tendency is apparent not only in the Czech Republic, but other European countries as well. Moreover, transition economy process and economy restructuring are additional specific features of energy sector development. The changes have influenced the conditions of decision-making and several previous decisions had to be re-evaluated.

Originally, electricity generators preferred to invest in the large capacity electricity generating units to make benefit of the size effect resulting in more attractive production costs (e.g. economy of scale). Due to the liberalisation of electricity markets and above mentioned results accompanying this process, attention is slowly focused on the modular low capacity projects which are valued as a sequence of dependent projects. Such flexible project sequences can better react on market uncertainties and eliminate partly the market risk and limiting potential producers´ loses.

The main objective of this study is to solve the traditional investment problem, i.e. if to invest in a large capacity power plant or in a modular project and quantify the modularity value of such a flexible project under price uncertainty.

This paper is organised as follows: second chapter concentrates on problem description solved in this study. Next, electricity prices features and modelling possibilities are described. Valuation process

[1] This paper is supported by the Grant Agency of the Czech Republic (GACR): 402/04/1357 and within a project MMS 6198910007.

of large capacity project and modular flexible project is the contents of the chapter 4 and 5, Results, comments and general conclusion statements are in the last chapter.

2. Modularity valuation -problem description

Traditional problem faced by the electricity generators is solved here by applying real option methodology.

In this study, we want to compare large capacity project with maximal output of 1200 MWh with modular sequential project consisting of two low capacity units (capacity of 400 and 800 MWh) and show that modular project in a competitive energy environment with strong uncertainty about the electricity prices is more suitable than large capacity power plant.

Valuation process is in both cases based on the real option methodology application. Here, financial option valuation models are, under some circumstances, applied on real assets valuation; general description of this methodology including applications is deeply described in Dluhošová (2004), Trigeorgis (2004) etc.

Large capacity power with the possibility to defer the investment is considered. Thus, the project is valued as a European call option on the project. Here, the underlying asset is gross project value (calculated as the present value of cash flow) with exercise price equals to investment outlay.

Modular project is valued as a compound growth option, i.e. producer makes sequential decisions in the future. In such a sequential investment, starting construction of the first unit creates option to invest in the subsequent unit resulting in the increase of the total capacity installed. The producer, after completing of the first unit, decides, if to start the construction of the following unit immediately or to delay it. Thus, the producer has the opportunity to stop it at a certain level, delay the investment or invest immediately in the subsequent unit. Moreover, in the case of modular project it is supposed that the construction of subsequent unit can not be started before the first unit is completed. The completion of the first units lasts 3 years.

Basic parameters for both investment variants are in the following Table 1.

		Output (kWh/year)	Investment outlay	Years of operation	Variable costs per kWh
Large capacity project		1200	7000	50	2,7
Modular project	1 unit	400	2000	50	2,8
	2 unit	800	5000	50	2,75

Table 1: Modular vs. large capacity project (basic parameters)

3. Electricity price behavior modelling

There are several models in finance applied frequently for various variables forecasting and their application depends on many factors.

Electricity prices have several characteristics, which distinct them from other financial variables and can be summarised as follows:

- Mean-reversion – power prices tend to fluctuate around values determined by the long run and the level of demand.
- Seasonality – power spot prices change by time of day, week, month and year in response to cyclical fluctuation in demand.
- Non-storability – electricity can not be stored and once it is generates, it must be almost immediately consumed.
- Price spikes – power prices series exhibit occasionally price spikes due to the several reasons (shocks in demand, transmission constraints, unexpected outages etc.).
- Regional differences – due to the fact, that the electricity is non storable and transmission constrains, there may be regional differences between spot and forward electricity prices.

In our study, we consider competitive electricity markets where the price of electricity varies randomly in time. For the purpose of the prices forecasting, we have applied the geometric mean-reversion process, which suits the best for electricity behaviour.

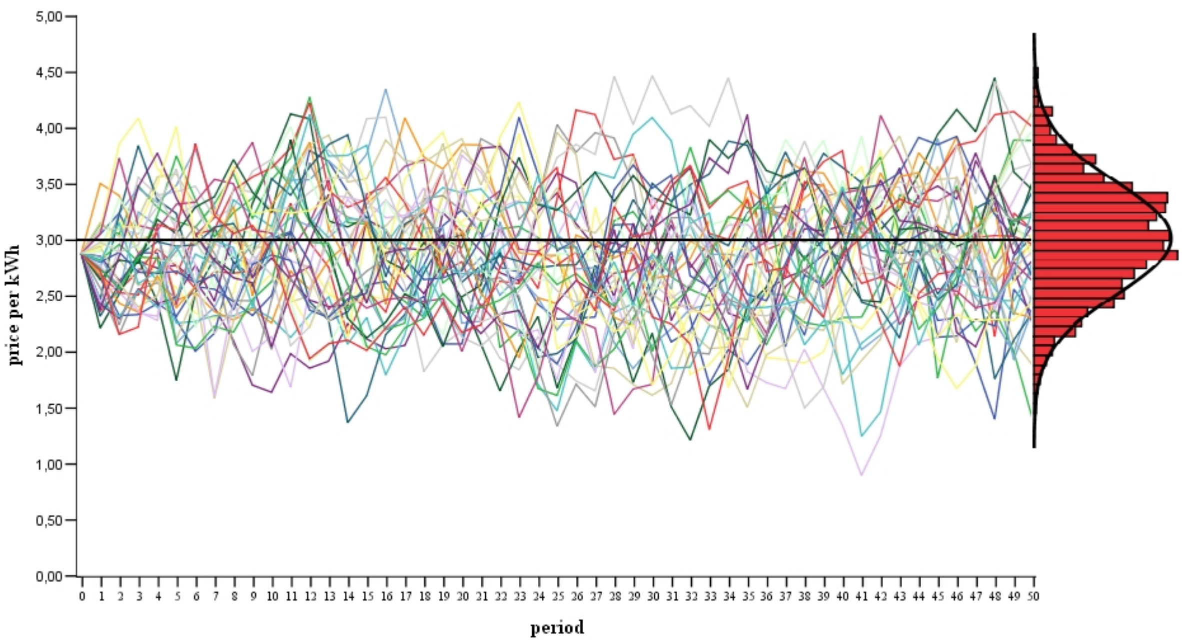
Mathematically, the general form of the mean-reversion model applied has this form,

$$S_{t+1} - S_t = \eta \cdot (\overline{S} - S_t) + \sigma \cdot dz, \tag{1}$$

where $\overline{S}$ is the long-run mean electricity price which the price tend to revert, η is the speed of reversion, σ is the electricity standard deviation and dz is specific Wiener process.

We made thousand price simulations in discrete sub-intervals over the horizon analysed to estimate the expected electricity price at discrete time *t.* A sample of 40 price paths including density function is depicted in the Figure 1.

Figure 1: A sample of 40 electricity price paths and density function for the terminal values

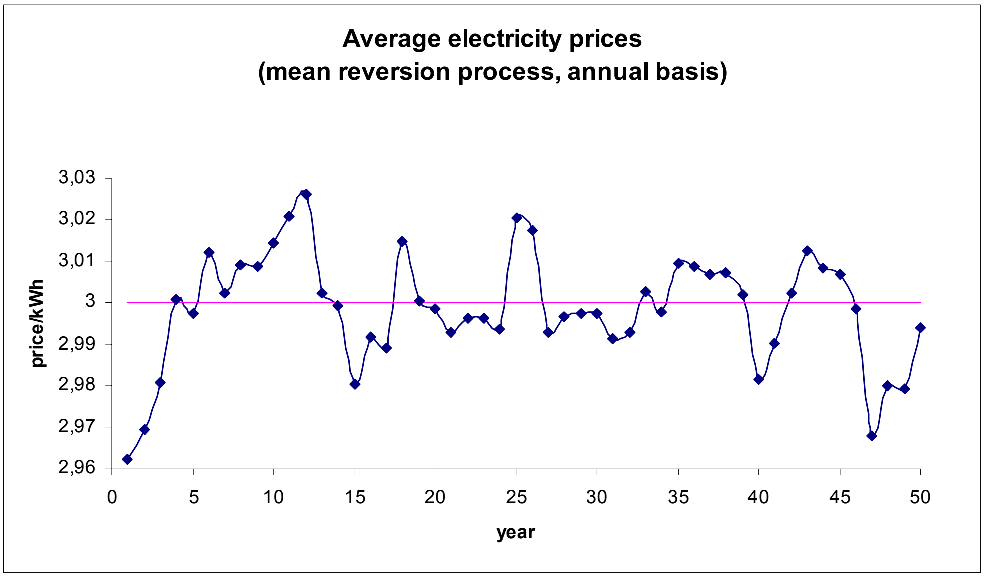


4. NPV calculation of the large capacity project as a European call option

This chapter is focused on the valuation of the large capacity project with maximal output of 1200 kWh per year. Recall, it is assumed electricity price uncertainty.

By applying above mentioned model for electricity price modelling, averaging the prices achieved by simulation at the end of discrete subintervals, *t*, expected average prices in a given year over the horizon analysed are obtained, see Figure 2. (model parameters: S_0 = 2,9; $\overline{S}$ = 3, σ = 40 %, η = 0,7).

Figure 2: Average electricity prices at the end of the sub interval (annual basis).



Generally, the cash flow generated by the project in a given year *t*, CF_t, is calculated according to this formula,

$$CF_t = [E(P_t) \cdot Q - (VC_t \cdot Q + D_t)] \cdot (1 - T) + D_t, \tag{2}$$

where $E(P_t)$ is expected electricity price in a given year *t*, *Q* is total capacity installed, VC_t is variable cost of production, D_t is depreciation in a given year and *T* is income tax rate.

Gross project value (calculated as the present value of subsequent stream of cash flow over the expected time of project operation) is calculated as a difference of two perpetuities, i.e.

$$V_t = \frac{CF_t}{\mu} - \frac{CF_t}{\mu} \cdot (1 + \mu)^{-(t+L)}, \tag{3}$$

where μ is appropriate discount rate (expressed as % per year) and *L* is total years of project in operation (here it is assumed 50 years). By subtracting the investment outlays, we would get the NPV calculated by traditional way, i.e.

$$NPV_t = V_t - I_t.$$

Nevertheless, we have calculated the project NPV as a call option of European type by employing Black-Scholes model with the possibility to defer the project initiation. In this case, the project NPV can be calculated by this way,

$$F_0(P_t^L) = V_0 \cdot N(d_1) - I \cdot e^{-r \cdot t} \cdot N(d_2), \tag{4}$$

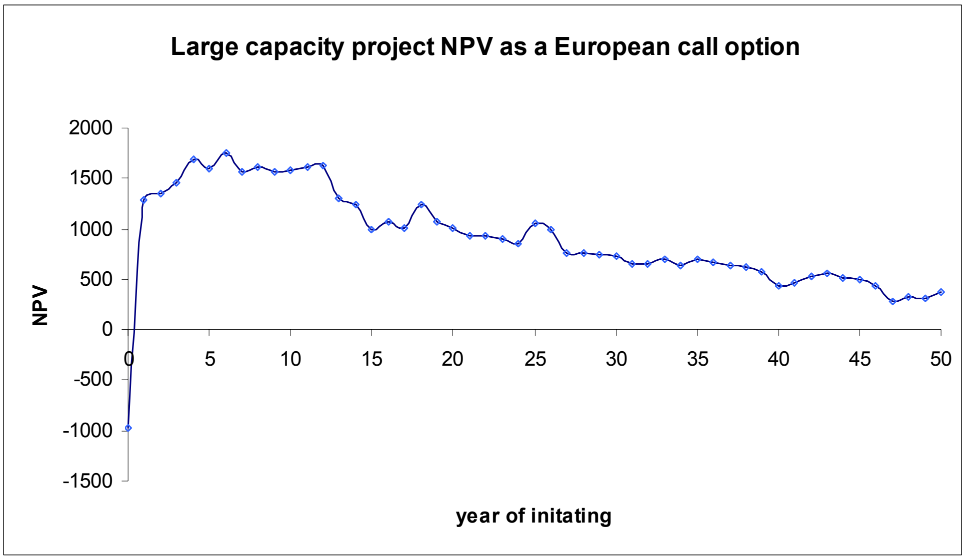
where $F_0(P_t^L)$ is project NPV started in year *t*, V_0 is present value of cash flow generated by exercising the option to invest in a year *t* (underlying asset) *I* is investment outlay (exercise price), *r* is risk free rate, and $N(\cdot)$ is cumulative probability of the normal distribution N (0,1).

Optimal project initiating time can be written as follows,

$$H_0^L = max_t [V_0 - I_0; F_0(P_t^L)], \tag{5}$$

calculation results are graphically depicted in the Figure 1.

Figure 3: NPV of the large capacity project as a European call option



It is obvious, that immediate project starting would invoke net loose of about -980 mil. currency units. Viewing the project as a European real option, it is optimal to delay it and start it 6 years from now.

5. NPV calculation of the modular project as a growth option

As mentioned above, modular project consists of the two low capacity units with maximum output of 400 and 800 kWh, respectively. For the comparison purposes, total investment outlays are equal to the investment outlay of the large capacity project. Both low capacity units have the same expected life of operation as the large project (see Table 1)

To be more realistic, it is assumed that both units can not be started at the same time. Precisely, second unit of the modular project can be initiated only after completing of the first unit. Construction time of the first unit takes 3 years.

It is important to realize, that from the real option methodology point of view, modular project consists of starting investment and growth compound option. Here it is also supposed that starting both units can be deferred to maximise expected NPV of the global total value of the modular project. The only constraint condition is that the time difference between initial investment and subsequent unit is at least 3 years.

Cash flow generated by the modular units in a given year and gross unite values are calculated the same way as in the case of the large capacity project, e.g. by employing (2) and (3), respectively.

NPV of the first unit, n_1, of the modular project, $F_0\left(P^M_{t,n_1}\right)$, was calculated as a option to defer, i.e. with possibility to postpone the initiating; mathematically,

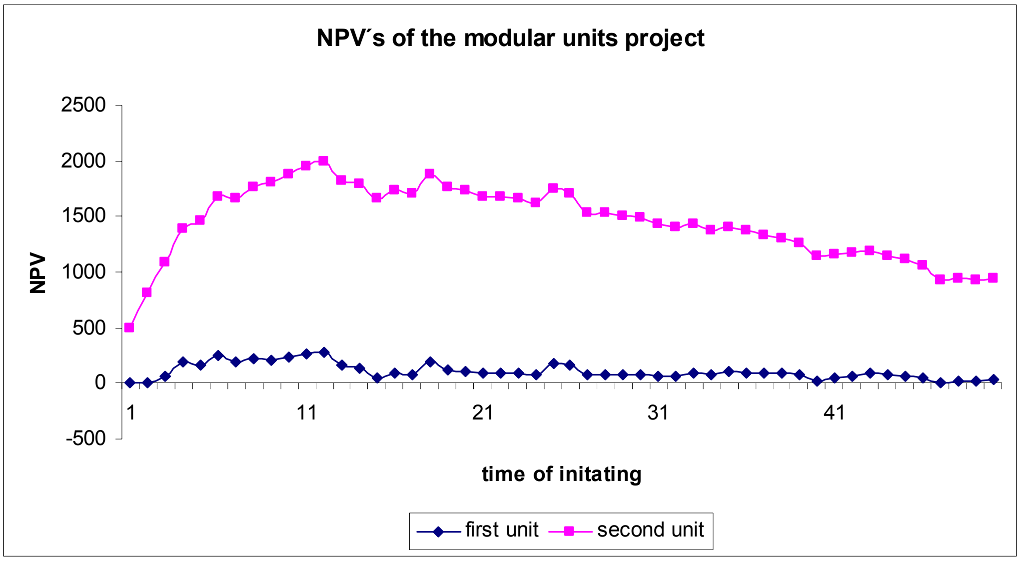
$$F_0\left(P^M_{t,n_1}\right) = max_t\left[V^M_{0,n_1}, -I^M_{t,n_1}; PV\left(V^M_{t,n_1} - I^M_{t,n_1}\right); 0\right], \tag{6}$$

NPV of the second unit, $F_0\left(P^M_{t,n_2}\right)$, which is contingent on the completing of the first unit and creates growth option is calculated by applying (4),i.e.

$$F_0\left(P^M_{t,n_2}\right) = V^M_{0,n_2} \cdot N(d_1) - I^M_{t,n_2} \cdot e^{-r \cdot t} \cdot N(d_2). \tag{7}$$

Graphical presentation of the results of the NPV´s is in the following Figure 4.

Figure4: NPV´s of the units of the modular project



Total NPV of the flexible modular project with two subsequent units is formally written as follows,

$$F_0\left(P_t^M\right) = max_t\left[F\left(P_{t,n_1}^M\right) + F\left(P_{t,n_2}^M\right)\right], \tag{8}$$

or by substituting (6) and (7) into (8) and rearranging,

$$F_0\left(P_t^M\right) = max_t\left[E\left(V_{t,n_1}^M - I_{t,n_1}^M\right) \cdot e^{-r \cdot t} + V_{0,n_2}^M \cdot N(d_1) - I_{t,n_2}^M \cdot e^{-r \cdot t} \cdot N(d_2)\right].$$

Due to the fact, that the second unit can not be started before or at the same time as the first one, but at least after completing of the first unit, the firm when maximizing (8) must find the optimal combination for initiating the first and the subsequent second unit, t_i and t_j maximizing (8).

Problem is solved by function SOLVER in MS EXCEL, objective function and constraints are defined as follows,

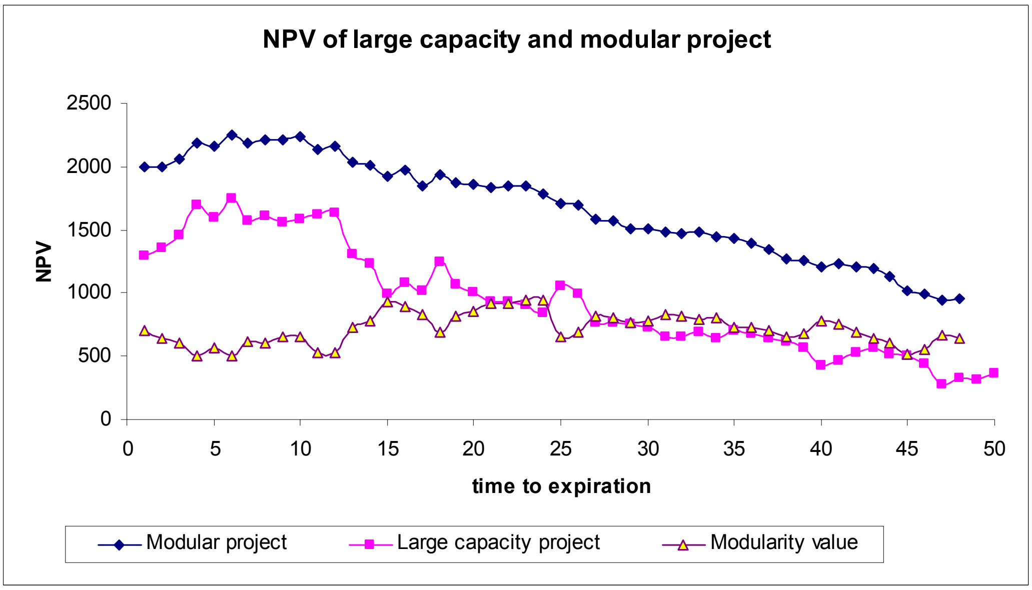
Objective function: $F\left(P_{t,n_1}^M\right) + F\left(P_{t,n_2}^M\right) \rightarrow max\,imize$

Constraints:
$$t_i, t_j \geq 0$$
$$t_i - t_j \geq 3 \;\; for\; i < j$$

The objective function maximizes the sum of the both generating units of the modular project, first constraint reflects the requirements on non-negative solution results, the second one the minimum time difference between the initiating generating units.

Solution has been found for the following combination: $t_1 = 6$, $t_2 = 12$, total NPV value of modular project is 2249 mil. c.u. Figure 4 shows all maximal combinations for a given t_i and t_j.

Figure 5: NPV of the large capacity project, modular project and modularity value.



6. Conclusion

The objective of this paper was to demonstrate the possibility valuation projects in the energy sector.

Primary attention was focused on the traditional problem solved by the electricity producers: if to invest in a large capacity power plant or in a modular power plant with several subsequent generating units.

For comparison purposes, we supposed, that the total maximum generating capacity of the large power plant is equal to the sum of capacities of modular plant units; moreover, the same is true about the investment outlays and expected years of operations. The main distinction is the production (variable) costs: we take into account the economy of scale, i.e. the larger project usually generates electricity with lower production costs than smaller generating units.

Both types of investments were valued as real options of European type by applying Black-Scholes model. It means that besides the project NPV, we also solved the problem of investment timing. From real option perspective, option to delay an investment was included.

In the case of the large capacity power plant, under the price uncertainty, it is the best to start the project 6 years from now, see Figure 3. Except for the immediate project starting, any other deferring would lead to the positive NPV, but not maximal.

Modular project consists of two generating units with the capacity 400 and 800 kWh, respectively. Here we supposed that unit two is contingent on the unit one, i.e. it can not be started before previous unit is completed. In valuation of both units the possibility to defer the investment was supposed.

Calculation of the NPV of modular unit is described in Chapter 5, graphical results are in Figure 4. This Figure represents NPV of both generating units calculated by Black-Scholes model with the possibility to defer. To meet the constraint referring to the time gap between the initiating both units, we had to find optimal combination of units starting maximising the total NPV of modular project. Using SOLVER tool in MS EXCEL spreadsheet we found the solution for $t_1 = 6$, $t_2 = 12$ years from now.

There are in Figure 5 all the maximal NPV sum combinations of both generating units, modularity value is expressed a difference between NPV of the large capacity project and modular project, i.e.

$$MV = max_t\left[\left(F_0\left(P_t^M\right) - F_0\left(P_t^L\right)\right); 0\right].$$

We can state following general conclusions on the basis of the results achieved:

- general knowledge referring to the real options were confirmed, i.e. it is not always optimal to invest in a project immediately even if its NPV is positive. Project deferring can bring valuable additional value resulting from resolving uncertainty in the future. On the other hand, it does not have to be always optimal to wait. On competitive markets, electricity producer may risk losing his market if he waits too long.
- modular projects enables flexibility making it possible to adapt easier to changing uncertain environment. Such projects, valued as growth compound options, provide to the electricity producers sequences of opportunities, which can be exercised under only pre specified conditions.
- if investment outlays, capacity and other factor are the same, the lower production costs does not necessarily means advantage leading to preferring a project.

Acknowledgement

This paper is supported by the Grant Agency of the Czech Republic (GAČR): 402/04/1357 and within a project MMS 6198910007.

References

[1] AMRAM, M., KULATILAKA, N.: *Real Options: managing strategic investments in an uncertain world*. Harvard Business School Press, Boston, 1999.

[2] BRENNAN, M.J., TRIGEORGIS, L.: *Project Flexibility, Agency and Competition: new developments in the theory and application of Real Options*. Oxford University Press, London, 2000.

[3] COPELAND, T.E., ANTIKAROV, V.: *Real Options: practitioner's guide*. Texere, New York, 2000.

[4] DLUHOŠOVÁ, D. et al.: *New approaches and financial instruments in financial decision-making.* EkF VŠB-TU Ostrava, 2004.

[5] DIXIT, A.K., PINDYCK, R.S.: *Investment under Uncertainty*. University Press, 1994.

[6] HOWEL, S. et al.: *Real options. Evaluating Corporate Investment Opportunities in a Dynamic World*. Prentice Hall, London, 2001.

[7] HULL, J.C. *Option, futures and other derivatives.* 5th edition. Prentice Hall, USA, 2002.

[8] MUN, J.: *Real Option Analysis. Tool and Techniques for Valuing Strategic Investments and Decisions.* John Wiley and Sons, Inc., New Jersey, USA, 2002.

[9] TRIGEORGIS, L.: *Real Options in Capital Investments: models, strategies and applications*. Praeger, Westport, 2000.

[10] TRIGEORGIS, L.: *Real Options: managerial flexibility and strategy in resource allocation*. The MIT Press, Cambridge, 1998.

[11] TRIGEORGIS, L., SCHWARTZ, E.S.: *Real Options and Investments under Uncertainty*. The MIT Press, Cambridge, 2001.

[12] VOLLERT, A.: *A Stochastic Control Framework for Real Options in Strategic valuation*. McKinsey Comp., Boston, 2003.

[13] ZMEŠKAL, Z. et al.: *Financial models*. EkF VŠB-TU Ostrava. 2004.

Optimization under Exogenous and Endogenous Uncertainty

Jitka Dupačová
Charles University, Prague, Faculty of Mathematics and Physics, Dept. of Probability and Math. Statistics, Sokolovská 83, CZ-18675 Prague

Abstract

Customary stochastic programs aim at the best feasible decision made before the realization of the random element is observed. The common assumption is that the probability distribution does not depend on decisions — the case of the exogenous uncertainty. This paper focuses on stochastic programming models for which through decisions, a decision-dependent, endogenous randomness is put into effect. Problem structure then becomes important. Examples point out at tractable cases and solution techniques.

Keywords

Stochastic programs, distributions dependent on decisions, exogenous and endogenous uncertainty, contamination
JEL: C61, C44

1 Stochastic programming problems

Customary stochastic programs, see e.g. [3], [5], [9], [14], aim at the best feasible decision x made before the relevant random element $\omega \in \Omega$ is unveiled. The cost of choosing x depends on ω as quantified by a real-valued function $f(x, \omega)$. The problem is

$$\text{minimize } F(x; P) := E_P f(x, \omega) \tag{1}$$

over a closed nonempty subset $\mathcal{X}$ of a finite-dimensional (Euclidean) space. P denotes a known probability distribution of ω on Ω which is x-independent — the case of the exogenous uncertainty. In practice, however, there are important decision problems in which through decisions, a decision-dependent, *endogenous* randomness is put into effect; see e.g. [6], [7], [8], [10].

One may try to *remove the dependence of P on x by formulating a simpler model*: The settlement of revenues of a pension fund is influenced by the attained fund investments profitability whose probability distribution depends on the investment decisions. To an extent, this problem may be circumventioned by fixing the valorization of the accumulated wealth of individual participants to a predetermined guaranteed minimal level and penalizing the deviations; cf. [13]. This, however, is not a general approach.

In this paper we shall deal with stochastic programs of the form

$$\text{minimize } F(x) = \int_{\Omega} f(x, \omega) P_x(d\omega) \quad \text{on } \mathcal{X} \tag{2}$$

which differ from the standard version (1) in making explicit a possible dependence of probability distributions on decisions. We shall assume for simplicity that the expectations $\int_{\Omega} f(x, \omega) P_x(d\omega)$ are finite for all $x \in \mathcal{X}$ and an optimal solution exists.

Under specific assumptions, dependence of P on x in (2) can be removed by a suitable *transformation of the decision-dependent probability distribution P_x*, cf. [16] and the Push-in technique explained in [11], [15]:

[1]This work is partly supported by the project “Methods of modern mathematics and their applications” – MSM 00216120839 and by Grant Agency of the Czech Republic (grants 201/05/2340 and 402/05/0115).

Assume that there *exist densities* $p(x, \omega)$ *of probability distributions* P_x *with respect to a common probability measure* Q. Then the objective function in (2) can be rewritten as

$$F(x) = \int f(x, \omega) p(x, \omega) Q(d\omega)$$

Thus we recover the common form $F(x) := \int_\Omega \tilde{f}(x, \omega) Q(d\omega)$ with $\tilde{f}(x, \omega) := f(x, \omega) p(x, \omega)$ and with a decision-independent probability distribution Q. However, it is obtained at the cost of losing convenient properties of the original random objective function $f(x, \omega)$. The properties of the resulting objective function depend on the structure of the problem, namely, on type of dependence of P on x. For example, assume that

$$P_x(B) = Q(B \oplus Hx) \tag{3}$$

for every Borel set $B \subset \Omega$, with Q a probability distribution, $\oplus$ the direct sum and H a given matrix of the matching dimension. Changing variables in $\int f(x, \omega) P_x(d\omega)$ transforms the objective function to $\int f(x, \zeta - Hx) Q(d\zeta)$, whose properties depend on properties of $f(x, u)$ viewed as a function of (x, u) *jointly*.

The acceptance of the decision-dependent model (2) may cause various technical difficulties: For instance, if $f(x, \omega)$ is a convex function of x for each ω, then so is $F(x; P)$, whereas the convexity property of $F(x) := F(x; P_x)$ may be lost. This in turn puts limitations on the choice of numerically tractable optimization techniques even if evaluation of the objective function at any point x is no more complicated then for the classical model (1). Depending on the structure, recursive optimization methods and search techniques, cf. [11], numerical enumeration techniques including branch-and-bound method and disjunctive programming can be used.

The decisions may partly aim at *enhancement of the knowledge of the probability distribution*: For instance, in simple inventory-type stochastic programs the demand observed in the first stage may serve to collect more precise information about the probability distribution of the future demand. However, a demand higher than a certain cut-off point, such as the supply available for the first stage, will not be observed. The wish to obtain as precise information as possible may lead to increasing the order for the first stage. Such decision process can be then formulated by means of *sensors*, cf. [1] [2].

Special attention is needed for multistage problems with a decision-dependent probability distribution. Here not only the first-stage decision, but also the later decisions affect the information about the probability distribution available to the decision maker as they may influence the marginal and conditional probability distributions in subsequent stages; [2] displays examples of this kind. Moreover, decisions may influence the time at which uncertainty gets resolved, i.e., *nonanticipativity conditions may be decision dependent*; see [6], [7].

We shall see that tractability of problems (2) depends essentially on their structure and that there are several favorable problem classes, e.g.,

- The probability distribution is of a known type and the decisions influence only its parameters, see Section 2;
- There is a fixed finite set of probability distributions, see Section 3. The dependence on decisions may often be modeled by Boolean variables and the decisions may be partly related to the choice of a probability distribution from the given set.

Stability of the optimal solution of problem (2) with respect to changes of the probability distribution will be discussed in Section 4.

2 Decision-dependent parameters

2.1 Stochastic PERT problem

Several modeling issues on the subject of the stochastic PERT problem are discussed in [9], [14]. The primary concern is to minimize the expected duration of a project defined as a set of activities which consume time and resources and have to reflect certain temporal precedence relationships. The project can be described by an acyclic directed network with nonnegative arc lengths and with two specific nodes "Start" and "End" of the project.

If the durations of individual activities (lengths of arcs) $g := (g_1, \ldots, g_n)$ are known, the shortest time in which the project can be completed while observing the prescribed preference relations is equal to the length $l^*(g)$ of the critical path, the longest path connecting the Start and End nodes, which is a nonnegative convex function of g. Durations g_j of activities $j = 1, \dots, n$, may be reduced for an additional cost. More specifically, assume that g_j are convex functions of parameters $x_j := (x_{j1}, \dots, x_{jn_j}),\ j = 1, \dots, n$; then the project duration — the composite function $f(x) := l^*(g_j(x_j), j = 1, \dots, n)$ — is convex in x. As to the additional cost $k(x)$, assume that it is convex and separable in individual components $x_{ji},\ i = 1, \dots, n_j,\ j = 1, \dots, n$. The problem is to choose the best parameter values with respect to both the project duration and the additional costs and considering constraints. Thanks to convexity, it is possible to rewrite it in the form with one aggregated objective function, such as $\lambda_1 f(x) + \lambda_2 k(x)$, with parameters $\lambda_1, \lambda_2 > 0$.

Assume now that durations of individual activities are random, $\omega := (\omega_1, \dots, \omega_n)$, and that *parameters*, say $x_j,\ j = 1, \dots, n$, of their probability distributions may be changed for a cost. In [12], this problem is discussed in detail for the class of independent uniform distributions with a fixed spread around changing expected values x_j and for independent triangular distributions determined by the lower/upper bounds $a_j x_j$ / $b_j x_j$ and moduses $m_j x_j,\ j = 1, \dots, n$. These parameters are subject to linear constraints which reflect limits on resources and on activity durations. The applied cost function is $k(x) := \sum_{j=1}^{n} k_j x_j^{-1}$ and the objective function reflects then two convex criteria: Minimize the expected project completion time when using probability distribution identified by parameters $x_j,\ j = 1, \dots, n$, and minimize the costs $k(x)$ for chosen parameter values $x_j\ \forall j$. In [12], sample-path optimization is presented as an efficient solution method.

2.2 Queuing networks

Let $\omega = (\omega(t), 0 \le t \le T)$ be a stochastic process in continuous time controlled by the parameter $x = (x_1, x_2, \dots, x_r)^\top$, the probability distribution of ω being thus dependent on x, denoted P_x. We should minimize (or maximize) the expectation of a functional $f(x, \omega)$ of that process. For instance, ω may describe performance of a queuing network where customers pass through r service stations, according to definite rules. The inter-arrival times as well as service times are exponentially distributed, the formers with some fixed intensities, the latters with intensities $x_1, x_2, \dots, x_r$. The set $\mathcal{X}$ may be given by budget limitations as $\{x \ge 0 : c(x) \le K\}$, $c(x)$ the cost of running the system under control parameter x. The functional f may be the number of customers whose service was completed during the time interval $[0, T]$. For a fixed x which may be chosen or controlled by the manager we can get the value of $f(x, \omega)$ by simulating histories of all customers who entered the system.

To get the optimal decision, one needs to use a recursive optimization method, such as stochastic approximation procedures; see [11]. Sometimes, a random process can be simulated which leads to P_x in the limit. For example, [10] applies a stochastic quasigradient algorithm in the context of optimal control of a system with a decision dependent transition operator and with an unknown steady-state probability law P_x, which may be found by simulation.

3 Decision-dependent scenario trees

Assume now that there is a finite number of possible probability distributions and that each of them has been approximated by a discrete distribution carried by a finite number of atoms — scenarios. For multistage stochastic programs, each of these discrete probability distributions is used to create a scenario tree which takes into account the related topology of stages and path probabilities are attached to the scenarios; cf. [5]. The resulting scenario trees and the path probabilities are indexed by a finite number of indices d. In principle, one may apply a full enumeration with respect to d, or a version of the branch-and-bound method, cf. [8], or disjunctive programming techniques, cf. [6], [7].

3.1 Probing for information

This simple example illustrating the case of decision-dependent probabilities has been motivated by Chapter 21 in [17]:

A large oil exploration company holds a lease that must be either sold out immediately for a know market price, or after one year for a price which depends on an exogenous factor — uncertainty in future oil prices, or sold after some exploration, e.g. after an experimental drilling. There are three possible outcomes of drilling — discovery of a dry or wet well or a gusher. The crucial issue is to determine their probabilities. In principle, the decision maker may use probabilities based on past experience. However, the cost of drilling is high and it would be useful to eliminate drilling if the well is dry. The suggestion is to precede drilling by another, substantially cheaper exploration method — seismic analysis. Based on its outcome probabilities of the three possible scenarios can be revised.

With the finite and small action space this problem can be modeled using decision trees and no special optimization technique is needed.

3.2 Sizes problem [8]

A production line must meet the demand for a certain number of products ordered according to their attributes (e.g., sizes) $a_1 \prec \ldots \prec a_r$. If the demand for a given product cannot be satisfied, after an additional treatment (e.g., cutting) it is possible to exploit a higher category product for additional substitution costs. Production costs consist of fixed set-up costs for every category that will be actually produced, of random per unit costs of the initial production and of random substitution costs whenever it applies. The additional substitution costs depend on the first-stage decision which determines the initial production levels $x_1, \ldots, x_r$.

The decision about producing category j or not is modeled by a Boolean variable d_j. Hence, there is a collection indexed by vectors $d \in \mathcal{D} = \{0, 1\}^r$ of a finite number of stochastic programs

$$\min_x E_{P_d} f(x, \omega) \text{ subject to } x \in \mathcal{K}(d), \tag{4}$$

where $\mathcal{K}(d)$ denotes coupling constraints on x given d (for example, $d_j = 0$ implies $x_j = 0$, or $\sum_j x_j d_j$ equals the total demand). The random vector ω consists of components related with the production costs on the both initial production and additional treatment levels. Its joint probability distribution P_d depends on the decision d which determines the structure of production.

The producer tries to select the best production scheme, i.e., to decide according to the best stochastic program

$$\min_{d \in \mathcal{D}} \{ \min_{x \in \mathcal{K}(d)} E_{P_d} f(x, \omega) \}. \tag{5}$$

The choice of d (and of x as well) is the first-stage decision, the second-stage decisions concerning the additional treatment enter the random production costs $f(x, \omega)$. Random demand, an exogenous uncertainty, can also be incorporated.

3.3 Project selection

Consider the possibility of investment into I projects of uncertain potential which may be initiated at time instants $t = 1, \ldots, T$. When a project starts, it cannot be interrupted or closed before the considered investment horizon T. There exist several versions of each of these projects, at most one of them can be used. The probability distribution of the uncertain characteristics θ_i of project i is discrete, carried by scenarios θ_i^s, $s \in \mathcal{S}_i$ with probabilities p_i^s; the scenario and its time evolution gets revealed only after the investment is started. The problem is to decide when and to which project (or project version) to invest and at what level of investment subject to various cost, capacity or technological constraints so that the expected net present value of the total investment over the whole horizon is maximal.

Exploration of new oil fields, cf. [6], belongs into this class of problems. It involves decisions about building work platforms at certain places, pipelines networks, production platforms, etc. Oil fields are characterized by their initial capacities γ_i and deliverabilities δ_i. They are the endogenous source of uncertainty because their realizations $\theta_i^s = (\gamma_i^s, \delta_i^s)$ can be observed only after the project i was accepted and started. The time evolution of capacity and deliverability over time is described by a linear reservoir model. This means that all uncertain parameters and the related coefficients are known after their initial values were observed at the starting time, say t_i, of field i exploration. Let the versions $v \in \mathcal{V}_i$ of project i differ only by different time points $t_i = 0, 1, \ldots T - 1$, at which the exploration of the field i starts.

Decision variables d_{iv}, $i = 1, \ldots, I$, $v = 0, \ldots, T - 1$ equal 1 if version v of project i is chosen (i.e., if project i starts at $t_i = v$) and 0 otherwise. These $0 - 1$ variables have to fulfil conditions

$$\sum_v d_{iv} \leq 1 \, \forall i.$$

Operational variables for version v of project i at time t and for scenario s are continuous, denoted as y_{iv}^{ts}, $s \in \mathcal{S}_i$, $v \in \mathcal{V}_i$, $i = 1, \ldots, I$. Assume for simplicity that they are subject to a system of linear inequalities

$$q^t \leq \sum_i \sum_{v \in \mathcal{V}_i} f_{iv}^{ts} y_{iv}^{ts} d_{iv} \leq Q^t, \; t = 0, \ldots, T - 1, \; s \in \mathcal{S}_i, \; i = 1, \ldots, I$$

with coefficients f_{iv}^{ts} known for $t \geq v$ and capacity constraints

$$L d_{iv} \leq y_{iv}^{ts} \leq c_i^s d_{iv}, \; t = 0, \ldots, T - 1, \; s \in \mathcal{S}_i, \; i = 1, \ldots, I$$

$$\sum_t y_{iv}^{ts} d_{iv} \leq \gamma_i^s,\; s \in \mathcal{S}_i,\; i = 1, \ldots, I.$$

Nonanticipativity conditions for each field separately are influenced by the selected version of the i-th project and reflect the simple fact that for project i the decisions d_{iv} and for $d_{iv} = 1$ also y_{iv}^{ts} for $t \leq v$ are equal for all scenarios $s \in \mathcal{S}_i$. The reason is that individual scenarios θ_i^s cannot be distinguished before the project i starts and the form of the nonanticipativity conditions is thus influenced by decisions d_{iv}.

With discounted unit costs r_{iv}^{ts} corresponding to version v of project i at period t and for scenario $s \in \mathcal{S}_i$, the objective function is

$$\sum_t \sum_i \{\sum_{v \in \mathcal{V}_i} \sum_{s \in \mathcal{S}_i} p_i^s r_{iv}^{ts} y_{iv}^{ts} d_{iv}\}.$$

4 Stability of optimal solutions

Consider first the decision-independent case (1). The influence of changes in the probability distribution P can be modeled using the *contamination approach*, see e.g. [4], i.e., using contaminated distributions

$$P_\lambda = (1 - \lambda)P + \lambda \tilde{P}, \tag{6}$$

with $\lambda \in [0, 1]$ and with $\tilde{P}$ another probability distribution under consideration. We suppose that for all considered distributions, stochastic program (1) has an optimal solution.

The objective function $F(x; \lambda) := E_{P_\lambda} f(x, \omega)$ is linear in λ and its derivative with respect to λ equals $E_{\tilde{P}} f(x, \omega) - E_P f(x, \omega)$.

Define the optimal value function

$$\varphi(\lambda) := \min_{x \in \mathcal{X}} F(x; \lambda).$$

If $\hat{x}$ is the *unique* minimizer of (1), then under mild conditions (e.g. [4]) the one-sided derivative exists and

$$\varphi'(0^+) = E_{\tilde{P}} f(\hat{x}, \omega) - E_P f(\hat{x}, \omega), \tag{7}$$

i.e., the local change of the optimal value function caused by a small change of P in the direction $\tilde{P} - P$ is asymptotically the same as that of the objective function at $\hat{x}$. Moreover, $\varphi(\lambda)$ is a concave function of λ on [0, 1], hence the bounds

$$(1 - \lambda)\varphi(0) + \lambda\varphi(1) \leq \varphi(\lambda) \leq \varphi(0) + \lambda\varphi'(0^+)$$

are valid for all $\lambda \in [0, 1]$.

Consider now the decision-dependent case (2). Let $\hat{x}$ be the true or approximated minimizer of $F(x) = F(x, P_x)$ on $\mathcal{X}$. If the probability distributions $(P_x, x \in \mathcal{X})$ are contaminated by $(\tilde{P}_x,\; x \in \mathcal{X})$ as in (6), then the derivative of $F(\hat{x}, P_{\hat{x}})$ in the direction $\tilde{P}_{\hat{x}} - P_{\hat{x}}$ measures again sensitivity of the objective function at $\hat{x}$ against small changes of $P_{\hat{x}}$ in that direction. However, the assertion (7) about the optimal value function $\varphi(\lambda)$ is no longer true in general. It is true in some special cases, e.g., in the case (3), where

$$P_x(\bullet) = Q(\bullet \oplus Hx),\; \tilde{P}_x(\bullet) = \tilde{Q}(\bullet \oplus Hx),\; x \in \mathcal{X}.$$

Hence, sensitivity analysis for the decision-dependent case (2) would require development of new quantitative stability results.

The directional derivative can be exploited algorithmically: Assume for example that a search technique for solving problem (2) was stopped at a point $\hat{x}$ which is the true or approximate minimizer of $F(x) = F(x, P_x)$ on the set $\mathcal{X}$. In the context of example 3.2 it means that $\hat{x}$ solves a problem akin to (4) with the probability distribution P_d. To analyze the effect of perturbations of the applied vector d and, possibly, to change d to get an improvement, one can exploit directional derivatives $F(\hat{x}, P_{\tilde{d}}) - F(\hat{x}, P_d)$ of the objective function $F(x, P_d)$ at $\hat{x}$ in the direction of $P_{\tilde{d}} - P_d$ for $\tilde{d} \in \mathcal{D}$; cf. [8].

References

[1] Artstein Z (1994) Probing for information in two-stage stochastic programming and the associated consistency. In: Mandl, P, Hušková M (eds.) *Asymptotic Statistics*, Proc. of the 5th Prague Symposium 1993, Physica-Verlag, Heidelberg, pp. 21–33.

[2] Artstein Z (1999) Gains and costs of information in stochastic programming. *Annals of Oper. Res.* 85: 129–152.

[3] Birge JR, Louveaux F (1997) *Introduction to Stochastic Programming*, Springer-Verlag (Series in Operations Research), New York.

[4] Dupačová J (1990) Stability and sensitivity – Analysis for stochastic programming. *Annals of Oper. Res.* 27: 115–142.

[5] Dupačová J, Hurt J, Štěpán J (2002) *Stochastic Modeling in Economics and Finance*, Kluwer Acad. Publ., Dordrecht Boston London.

[6] Goel V, Grossmann IE (2004) A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. *Computers & Chemical Engineering* 28: 1409–1429.

[7] Goel V, Grossmann IE (2005) A class of stochastic programs with decision dependent uncertainty, SPEPS 2004–23.

[8] Jønsbraten TW, Wets RJ-B, Woodruff DL (1998) A class of stochastic programs with decision dependent random elements. *Annals of Oper. Res.* 82: 83–106.

[9] Kall P, Wallace SW (1994) *Stochastic Programming*, J Wiley & Sons, Chichester.

[10] Pflug GCh (1990) On-line optimization of simulated Markovian processes. *Math. Oper. Res.* 15: 381–395.

[11] Pflug GCh (1999) *Optimization of Stochastic Models. The Interface between Simulation and Optimization*, 2nd Printing, Kluwer Acad. Publ., Boston Dordrecht London.

[12] Plambeck EL, Fu B-R, Robinson SM, Suri R (1996) Sample-path optimization of convex stochastic performance functions. *Math. Progr.* 75: 137–176.

[13] Polívka J (2002) Liability side of asset–liability management model. In: Šafránková J (ed.) *WDS'02 Part I*, Matfyzpress, Prague, pp. 61–67.

[14] Prékopa A (1995) *Stochastic Programming*, Kluwer Acad. Publ., Dordrecht Boston London and Akadémiai Kiadó, Budapest.

[15] Rubinstein RY, Shapiro A (1993) *Discrete Event Systems. Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, J. Wiley&Sons, Chichester.

[16] Varayia P, Wets RJ-B (1990) Stochastic dynamic optimization approaches and computation. In: Iri M, Tanabe K (eds.) *Mathematical Programming. Recent Developments and Applications*, KTKSci Publ., Tokyo, pp. 309–332.

[17] Wonnacott TH, Wonnacot RJ (1990) *Introductory Statistics for Business and Economics*, 4th edition, J. Wiley&Sons, New York.

Dynamic Traveling Salesman Problem

Jan Fábry
University of Economics Prague, W. Churchill sq. 4, 130 67 Prague 3
fabry@vse.cz

Abstract

Existing competitive business forces distribution firms to offer immediate services to their customers. While a static version of traveling salesman problem does not accept additional requirements of customers, a dynamic version enables dispatcher to change the planned route of the vehicle after an occurrence of on-line requests. In this paper, the minimization of the total length of travel is examined as the objective of the distribution firm with use of all available information about advance-request and immediate-request customers. Both the optimization algorithm and heuristic method are proposed to find the solution of this problem. Time windows constraints are included in extended model together with the cost function considering the travel costs and lateness costs as the penalties in case the time windows are violated. The solution to dynamic traveling salesman problem is demonstrated on generated data using LINGO as a linear programming solver and VBA in MS Excel as an interface and the output environment. Because of diversity of business activities and policies, specific model has to be formulated for each distribution firm. Possible extensions of distribution problems are indicated.

Keywords

Distribution problem, vehicle routing problem, dynamic traveling salesman problem, time windows, LINGO, VBA in MS Excel

1 Introduction

Recent boom of communication and information technology enables distribution firms to provide a flexible reaction to customers' requirements. Real-life distribution problems involve pick-up and delivery of material, goods, people, mail, etc. Many vehicle routing applications deal with the firms offering messenger, courier, residential utility repair or emergency services. Because a quality of the service and its price are comparable for firms operating in the same area of business, success of a firm and its attractiveness for potential customers often depend on its ability to accept on-line requests. At the beginning of a planning horizon (e.g. in the morning), roughly half portion of customers' requirements, to be covered in a day, is known. The rest will arise randomly over time. In this paper, the locations and time points of those orders are not known beforehand and even cannot be forecasted with any probability.

Size of the order is not considered and therefore a vehicle capacity has no limitations. In addition, the only vehicle in a depot is prepared for service. From this point of view, Traveling Salesman Problem is discussed and extended. While in the static version of the problem a set of customers is known with certainty and cannot be changed, in a dynamic problem it is modified in real time. Using all available information, the optimal route can be updated over the planning horizon.

As to the solvability, Traveling Salesman Problem, being approached with methods of integer programming, belongs to the class of NP-hard problems. Because in case of Dynamic Traveling Salesman Problem a computation time follows exponentially a number of new requests, it is important to develop effective techniques used in vehicle routing optimization. Although time necessary for finding an optimal solution is essential for the firm's flexible reaction to be successful on the market, the easy solvability of the problems is not the key issue of this paper. The emphasis is placed on mathematical models and their classification on a practical point of view. The possibility of using heuristic methods is indicated.

2 Traveling Salesman Problem

A static version of Traveling Salesman Problem (TSP) is based on the knowledge of all customers before the vehicle drives out from a central depot. A dispatcher is supposed to know distances between all pairs of customers and between the depot and customers. At the beginning of each day, the dispatcher can determine the optimal route, e.g. with use of the Miler-Tucker-Zemlin's mathematical model (see [5]):

$$\text{minimize} \quad z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \ , \tag{1}$$

subject to

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1,2,...,n, \tag{2}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1,2,...,n, \tag{3}$$

$$u_i - u_j + n x_{ij} \le n - 1, \quad i = 1,2,...,n, \quad j = 2,3,...,n, \quad i \ne j, \tag{4}$$

$$x_{ij} \in \{0,1\}, \quad i, j = 1,2,...,n, \tag{5}$$

where n is a number of locations (including the depot denoted by index 1) that the vehicle must visit over day, c_{ij} is a shortest distance between locations i and j, and x_{ij} is a binary decision variable, the value of which is 1 if the vehicle visits location j immediately after location i, and 0, otherwise. The objective (1) is to minimize a total distance that the vehicle travels on the tour. Constraints (2) and (3) assure that the vehicle visits each location exactly once, and inequalities (4) with variables u_i and u_j do not allow to generate partial cycles.

3 Dynamic Traveling Salesman Problem

After the optimal solution of the mathematical model (1)-(5) is found, the dispatcher will set the route and will prepare a report for a driver of the vehicle. A new requirement can occur after the vehicle starts its travel. If the dispatcher accepts the requirement, he/she will make a decision about the integration of the new location into the planned route. In Dynamic Traveling Salesman Problem (DTSP), two approaches can be used for the integration of new customers: re-optimization of the route and heuristic methods such as the insertion algorithm described in this paper.

3.1 Re-optimization of the Route

The new customer is added to a set of customers that still have not been visited, and the route that has been planned before is re-optimized. This route has to start in the location, being approached when the new customer is calling, i.e. the vehicle cannot change its direction. The re-optimization method minimizes the total distance as an objective function. However, it is quite impossible to use it in case of a high number of customers to be visited. A re-optimization model can be defined as follows:

$$\text{minimize} \quad z = \sum_{i \in U_N} \sum_{j \in U_N} c_{ij} x_{ij} \ , \tag{6}$$

subject to

$$\sum_{\substack{j \in U_N \\ j \ne j_{next}}} x_{ij} = 1, \quad i \in U_N - \{1\}, \tag{7}$$

$$\sum_{\substack{i \in U_N \\ i \ne 1}} x_{ij} = 1, \quad j \in U_N - \{j_{next}\}, \tag{8}$$

$$u_i - u_j + |U_N| . x_{ij} \le |U_N| - 1, \quad i \in U_N, \quad j \in U_N - \{j_{next}\}, \quad i \ne j, \tag{9}$$

$$x_{1j} = 0, \quad j \in U_N - \{j_{next}\}, \tag{10}$$

$$x_{ij_{next}} = 0, \quad i \in U_N - \{1\}, \tag{11}$$
$$x_{ij} \in \{0,1\}, \quad i, j \in U_N. \tag{12}$$

Set U_N contains all non-visited locations[1], $|U_N|$ is a total number of them. The number $j_{next} \in U_N$ is an index of the location that the vehicle is approaching when the new requirement arrives. This location is an initial point for the route that is going to be re-optimized[2].

As the alternative, a modified model of TSP can be also used for the re-optimization:

$$\text{minimize} \quad z = \sum_{i \in U_N} \sum_{j \in U_N} c_{ij} x_{ij} \,, \tag{13}$$

subject to

$$\sum_{j \in U_N} x_{ij} = 1, \quad i \in U_N, \tag{14}$$
$$\sum_{i \in U_N} x_{ij} = 1, \quad j \in U_N, \tag{15}$$
$$u_i - u_j + |U_N| . x_{ij} \le |U_N| - 1, \quad i \in U_N, \quad j \in U_N - \{1\}, \quad i \ne j, \tag{16}$$
$$x_{1 j_{next}} = 1, \tag{17}$$
$$x_{ij} \in \{0,1\}, \quad i, j \in U_N. \tag{18}$$

Before using this model, the distance $c_{1 j_{next}}$ between the depot and location j_{next} has to be changed to the value of total distance that the vehicle has traveled[3]. As the vehicle is approaching location j_{next}, constraint (17) has to be respected.

Example 1
Let us consider depot 1 and five customers 2-6 with their generated Cartezian coordinates (see Table 1). While these customers are known before, the requirement of last customer 7 arrives after the vehicle is on the route. In Table 2 all distances are calculated.

Location	X	Y
1	46	39
2	24	70
3	70	4
4	13	66
5	96	23
6	54	5
7	84	40

Table 1 – Generated locations

c_{ij}	1	2	3	4	5	6	7
1	0	38.01	42.44	42.64	52.50	34.93	38.01
2	38.01	0	80.45	11.70	85.98	71.59	67.08
3	42.44	80.45	0	84.22	32.20	16.03	38.63
4	42.64	11.70	84.22	0	93.48	73.50	75.61
5	52.50	85.98	32.20	93.48	0	45.69	20.81
6	34.93	71.59	16.03	73.50	45.69	0	46.10
7	38.01	67.08	38.63	75.61	20.81	46.10	0

Table 2 – Symmetric matrix of distances (in km)

The length of the optimal route including customers 2-6 is approximately 223.49 km (see Figure 1). It is quite important, for the integration of a new customer, when its requirement arrives, or what position the vehicle has at that time. First, assume the vehicle is going from customer 2 to customer 5, i.e. a new customer can be visited after the vehicle's visit of customer 5. Figure 2 shows the result of the re-optimization. The length of the optimal route 5-7-3-6-1 is approximately 110.39 km, the length of the route 1-4-2-5 is 140.33 km. Thus, the total distance is 250.72 km.

[1] The new customer is included in the sequence together with the depot as the last visited location.
[2] Re-optimization is supposed to be finished before visiting this location. Otherwise, as an initial point for a new route the successive location will be taken.
[3] Location j_{next} is included in such completed route.

Example 2
If a requirement of customer 7 was known before the vehicle's start from the depot, the length of the optimal route 1-4-2-7-5-3-6-1 would be approximately 225.40 km, i.e. more than 25 km shorter than in the previous example. The same route will be generated if the last requirement arrives before customer 2 is visited.

Example 3
In case that last requirement arrives after customer 5 has been visited, the total distance will be obviously lengthened. If the vehicle's position is anywhere between customers 5 and 3, the total distance will be 272.67 km.

The examples demonstrate that it is advantageous to know all requirements as soon as possible.

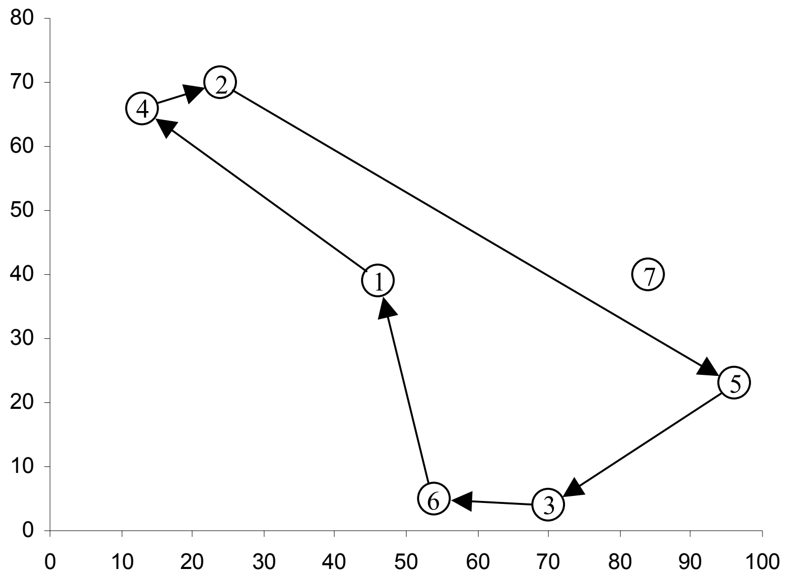


Fig. 1 – Optimal route for customers being known before a start of vehicle

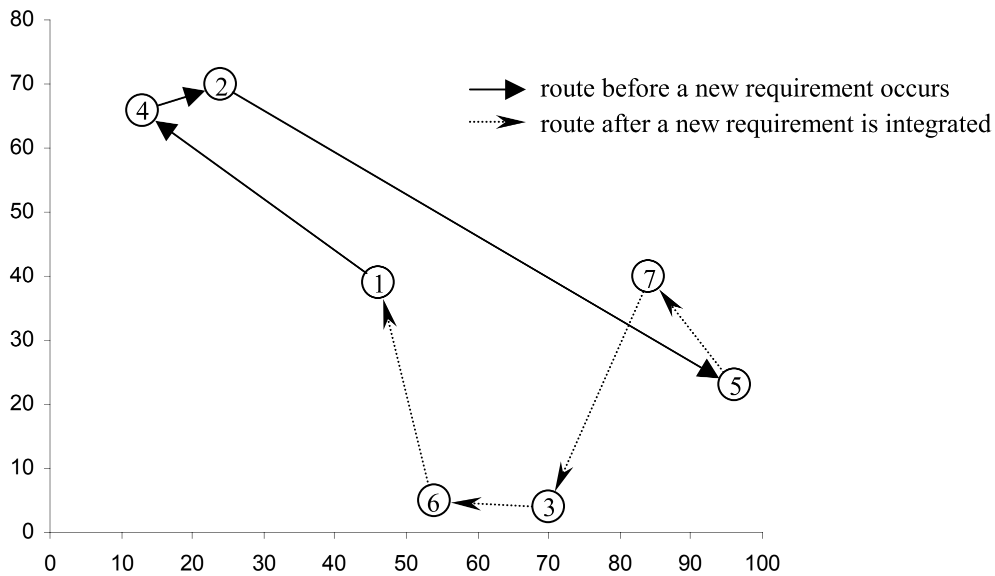


Fig. 2 – Re-optimization of the route when a new requirement is accepted

3.2 Insertion Algorithm

In the case of a high number of customers or too frequent arrivals of new requirements it is not possible to repeat the re-optimization procedure so often. Therefore, heuristic algorithms are used in real applications. The principle of an insertion method is to find the most advantageous pair of customers between which a new customer is inserted.

Let $U_N = \{i_1, i_2, ..., i_m\}$ be a sequence of m locations[4] ($i_m = 1$) remaining to be visited in compliance with the planned route, and $r \notin U_N$ be the index of a new customer. Then, the extension of the current route after inserting this customer between locations i_k and i_{k+1} can be calculated as follows:

$$\Delta z_k = c_{i_k, r} + c_{r, i_{k+1}} - c_{i_k, i_{k+1}}, \quad k = 1,2,..., m-1. \tag{19}$$

The objective is to find such *t*, which minimizes the value (19):

$$\Delta z_t = \min_{k=1,2,...,m-1} \Delta z_k . \tag{20}$$

Hence, customer r will be visited immediately after customer i_t before customer i_{t+1}.

Example 4

Let us consider Example 1. A new requirement of customer 7 arrives when the vehicle is approaching customer 5. The following table is generated for the sequence $U_N = \{5,3,6,1\}$ with use of (19):

i_k	i_{k+1}	$c_{i_k,7}$	$c_{7,i_{k+1}}$	$c_{i_k,i_{k+1}}$	Δz_k
5	3	20.81	38.63	32.20	27.23
3	6	38.63	46.10	16.03	68.69
6	1	46.10	38.01	34.93	49.18

Table 3 – Possible extensions of the current route (in km)

The minimal extension 27.23 km occurs when the new customer is inserted between customers 5 and 3. This is the identical result we have obtained using the re-optimization algorithm.

4 Traveling Salesman Problem with Time Windows

In definition of standard Traveling Salesman Problem with Time Windows (TSPTW), similarly to TSP, all the customers are known before the vehicle starts its travel. In addition, each customer has its requirement for being visited in a specified time interval. A time window of customer *i* is open at the earliest arrival time e_i and closed at the latest leaving time l_i. The mathematical model given below includes variable τ_i as real arrival time of the vehicle to the location *i*. An inequality $\tau_i \geq e_i$ does not allow to start a service of a customer before the earliest arrival time. If the vehicle arrives before, it will have to wait for opening the window. Besides distance value c_{ij}, travel time t_{ij} is given for each pair of locations. In the model the zero service duration is assumed for each customer.

$$\text{minimize} \quad z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} , \tag{21}$$

subject to

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1,2,..., n, \tag{22}$$

[4] The sequence corresponds to the initial optimal route or to the changed route after inserting one or several new customers.

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1,2,...,n, \tag{23}$$

$$e_i \le \tau_i \le l_i, \quad i = 2,3,...,n, \tag{24}$$

$$\tau_i + t_{ij} - M(1 - x_{ij}) \le \tau_j, \quad i = 1,2,...,n, \quad j = 2,3,...,n, \quad i \ne j, \tag{25}$$

$$\tau_1 = 0, \tag{26}$$

$$\tau_i \ge 0, \quad i = 2,3,...,n, \tag{27}$$

$$x_{ij} \in \{0,1\}, \quad i, j = 1,2,...,n. \tag{28}$$

When an average speed of the vehicle is 60 km.h^{-1}, it is possible to use parameter t_{ij} instead of c_{ij} in the objective function (21). Thus, the total travel time will be minimized. Constraints (24) and (25) ensure feasibility of the time schedule (M is a large constant value). In [1] the following nonlinear inequalities are also used instead of (25):

$$x_{ij}(\tau_i + t_{ij} - \tau_j) \le 0, \quad i = 1,2,...,n, \quad j = 2,3,...,n, \quad i \ne j. \tag{25a}$$

Denote by $w_j \ge 0$ ($j = 2,3,...,n$) a variable representing the waiting time before the vehicle starts a service of customer j. Then, the objective function is defined as follows:

$$\text{minimize} \quad z = \sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij} x_{ij} + \sum_{j=2}^{n} w_j \tag{21a}$$

and inequalities (25) are replaced by equations

$$\tau_i + t_{ij} - M(1 - x_{ij}) + w_j + v_{ij} = \tau_j, \quad i = 1,2,...,n, \quad j = 2,3,...,n, \quad i \ne j. \tag{25b}$$

Variables v_{ij} in (25b) must satisfy the following conditions:

$$0 \le v_{ij} \le M(1 - x_{ij}), \quad i = 1,2,...,n, \quad j = 2,3,...,n, \quad i \ne j. \tag{29}$$

5 Dynamic Traveling Salesman Problem with Time Windows

Similarly to DTSP, Dynamic Traveling Salesman Problem with Time Windows (DTSPTW) is processed in the following steps:

- a static optimization of TSPTW,
- a start of the travel following the planned route,
- an integration of new customers in the route.

Denote by n_K a number of the customers (the depot is included) being known before the vehicle's travel. Then, the optimal value of a total travel time (21a) is

$$T = \min\left(\sum_{i=1}^{n_K} \sum_{j=1}^{n_K} t_{ij} x_{ij} + \sum_{j=2}^{n_K} w_j \right). \tag{30}$$

Setting the optimal route, the vehicle is leaving the depot, while a dispatcher is going to accept new requirements.

5.1 Re-optimization of the Route

A re-optimization model of DTSPTW, being based on the model (6)-(12), can be formulated as follows:

$$\text{minimize} \quad z = \sum_{i \in U_N} \sum_{j \in U_N} t_{ij} x_{ij} + \sum_{j \in U_N} w_j\,, \tag{31}$$

subject to

$$\sum_{\substack{j\in U_N\\ j\neq j_{next}}} x_{ij} = 1, \quad i\in U_N - \{1\}, \tag{32}$$

$$\sum_{\substack{i\in U_N\\ i\neq 1}} x_{ij} = 1, \quad j\in U_N - \{j_{next}\}, \tag{33}$$

$$e_i \le \tau_i \le l_i \quad i\in U_N, \tag{34}$$

$$\tau_i + t_{ij} - M(1-x_{ij}) + w_j + v_{ij} = \tau_j, \quad i\in U_N - \{1\}, \quad j\in U_N - \{j_{next}\}, \quad i\neq j, \tag{35}$$

$$0 \le v_{ij} \le M(1-x_{ij}), \quad i\in U_N - \{1\}, \quad j\in U_N - \{j_{next}\}, \quad i\neq j, \tag{36}$$

$$x_{1j} = 0, \quad j\in U_N - \{j_{next}\}, \tag{37}$$

$$x_{ij_{next}} = 0, \quad i\in U_N - \{1\}, \tag{38}$$

$$x_{ij} \in \{0,1\}, \quad i,j\in U_N, \tag{39}$$

$$\tau_{j_{next}} = const., \tag{40}$$

$$\tau_j \ge 0, \quad j\in U_N - \{j_{next}\}, \tag{41}$$

$$w_j \ge 0, \quad j\in U_N - \{j_{next}\}, \tag{42}$$

$$w_1 = 0, \tag{43}$$

where U_N is a set of locations remaining to be visited[5] (including a new customer), $|U_N|$ is a number of them. Index j_{next} is associated with the location the vehicle is approaching when the new requirement arrives. The value $\tau_{j_{next}}$, defined as the constant (40) in the model, is determined in the previous optimization of the route as time at which location j_{next} will be reached. Respecting the constraints (34), (35) and (41), return time τ_1 of the vehicle to the depot has to satisfy the depot's time window $< e_1, l_1 >$. This interval depends on a working day of a distribution firm.

5.2 Insertion Algorithm

Assume $U_N = \{i_1, i_2, ..., i_m\}$ is a sequence of locations that have to be visited. If new customer $r \notin U_N$ is inserted between locations i_k and i_{k+1}, service time of this customer can be determined as follows:

$$\tau_r = \max\,(\tau_{i_k} + t_{i_k,r}, e_r). \tag{44}$$

The following inequalities have to be, of course, satisfied:

$$e_r \le \tau_r \le l_r. \tag{45}$$

The waiting time for the service of customer r is

$$w_r = \max\,(0, e_r - \tau_r). \tag{46}$$

Because the remaining part of the route is changed after inserting the new customer, service time and waiting time values have to be recalculated:

$$\tau_{i_{k+1}} = \max\,(\tau_r + t_{r,i_{k+1}}, e_{i_{k+1}}), \tag{47}$$

$$\tau_{i_{s+1}} = \max\,(\tau_{i_s} + t_{i_s,i_{s+1}}, e_{i_{s+1}}), \quad s = k+1, k+2, ..., m-1, \tag{48}$$

$$w_{i_s} = \max\,(0, e_{i_s} - \tau_{i_s}), \quad s = k+1, k+2, ..., m. \tag{49}$$

However, it is necessary to check the feasibility of the modified route:

$$e_{i_s} \le \tau_{i_s} \le l_{i_s}, \quad s = k+1, k+2, ..., m. \tag{50}$$

[5] The depot will be the last visited location on the route.

Travel time of the route after inserting customer r between customers i_k and i_{k+1} can be expressed as

$$z_k = \sum_{s=1}^{k-1} (t_{i_s,i_{s+1}} + w_{i_{s+1}}) + (t_{i_k,r} + w_r) + (t_{r,i_{k+1}} + w_{i_{k+1}}) + \sum_{s=k+1}^{m-1} (t_{i_s,i_{s+1}} + w_{i_{s+1}}) \,. \tag{51}$$

The objective is to find index v with the minimal value (51):

$$z_v = \min_{k=1,2,\dots,m-1} z_k \,. \tag{52}$$

6 Computational Experiments

The system involving models of DTSP was developed in MS Excel with use of VBA in MS Excel. It is linked to the system LINGO as a linear programming solver. Experiments have been executed on generated data. Figure 3 shows the generated example with a depot and 4 customers being known before a start of the vehicle (5 known locations) and 5 new customers with their requirements arriving over a working day.

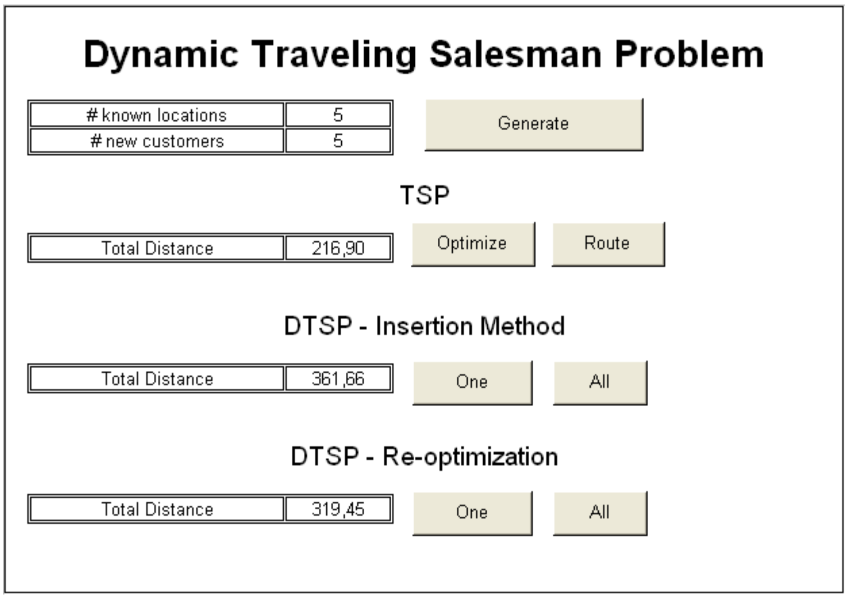


Fig. 3 – MS Excel interface for DTSP

First, static TSP has been optimized. Total distance of the optimal route (see Figure 4) has been calculated as approximately 216.90 km. The square represents the depot, while circles represent the customers.

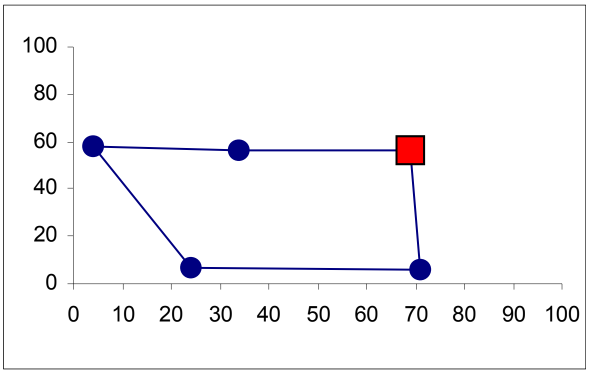


Fig. 4 – Optimal route of TSP

Introducing new customers, the modification of the vehicle's route can be monitored for each customer (button One), or the final route after integrating all customers is displayed (button All). The insertion method forms the final route of total length 361.66 km (see Figure 5), while the re-optimization algorithm brings the improvement to 319.45 km (see Figure 6).

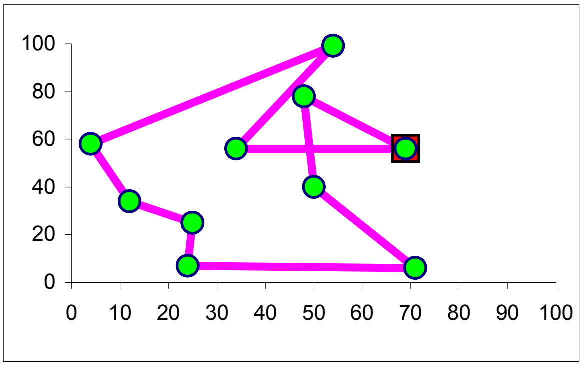


Fig. 5 – DTSP, insertion method

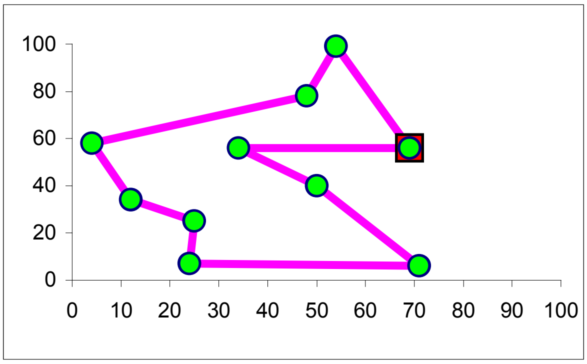


Fig. 6 – DTSP, re-optimization

7 Conclusions and Future Work

In this paper, a dynamic version of Traveling Salesman Problem is discussed. Diversity of distribution firms' business and strategies requires many various models to be formulated. As the basic application, Traveling Salesman Problem with Time Windows is extended into its dynamic version. Instead of hard constraints $e_i \le \tau_i \le l_i$ soft conditions $e_i \le \tau_i$ can be used. However, in case of the vehicle's delay ($\tau_i > l_i$), penalty has to be paid for the violation of the customer's requirement. Then, the objective is to minimize a total cost function, consisting of the total distance and the total penalty with appropriate weights.

Multiple vehicles and/or multiple depots can be included in other dynamic extensions of distribution problems. In many real applications, rather than total distance, time necessary for all customers being served is more important. In Dynamic Vehicle Routing Problems, each customer specifies a size of its requirement. Thus, a vehicle capacity must be respected and the possibility to split pick-up volumes should be considered. Different strategies for the movement of an idle vehicle can be analyzed when it is possible to forecast an arrival time and a location of a new requirement.

Because of computation time complexity of distribution models, improvement of current algorithms and development of new efficient approaches can be the key issues for most of distribution firms. Although firms desire the optimal route, a heuristic method is often the only reasonable approach to manage the problem.

This publication was supported by the project 402/06/0123 founded by the Grant Agency of the Czech Republic.

References

[1] Desrosiers, J., Dumas, Y., M. M. Solomon and F. Soumis (1995): *Time Constrained Routing and Scheduling*. In Ball, M. O., Magnanti, T. L., C. L. Monma and G. L. Nemhauser (eds.): *Network Routing, Handbooks in Operations Research and Management Science*. Vol. 8, pp. 35-139. North-Holland, Amsterdam.

[2] Gendreau, M., Guertin, F., J.-Y. Potvin and E. Taillard (1999): *Parallel Tabu Search for Real-Time Vehicle Routing and Dispatching*. Transportation Science, Vol. 33, No. 4, pp. 381-390.

[3] Ichoua, S., M. Gendreau and J.-Y. Potvin (2000): *Diversion Issues in Real-Time Vehicle Dispatching*. Transportation Science, Vol. 34, No. 4, pp. 426-438.

[4] Larsen, A., O. B. G. Madsen and M. M. Solomon (2004): *The A Priori Dynamic Traveling Salesman Problem with Time Windows*. Transportation Science, Vol. 38, No. 4, pp. 459-472.

[5] Pelikán, J. (2001): *Diskrétní modely v operačním výzkumu*. Professional Publishing, Praha.

Models of Regulation in Network Industries
(In the Field of Slovak National Electricity Markets)

Eleonora Fendeková
Michal Fendek
University of Economics, Bratislava, Slovak Republic
nfendek@dec.euba.sk,, fendek@dec.euba.sk

Abstract

Creation of new regulatory framework was the important part of restructuring market with electricity, gas and other goods of network industries. In August 2001 the Regulatory Office for Network Industries (hereinafter referred as to "RONI") was established which task was an issue of licenses and regulation of prices and quality standards for goods of network industries. RONI started issuing licenses and creating quality standards in year 2001, but it started executing the price regulation from 1.1.2003. The primary aim of RONI was to prepare new regime of price regulation for goods of network industries. RONI already defined regime of new price regulation for electricity distribution and worked out the system of new regulation rules for price making of goods of network industries. Models for determination of maximum prices and tariffs of goods of network industries were created over the years 2001-2002 and on their basis RONI proceeded in creation of regulatory and legal framework within the area of price regulation, which determined the method of calculation of maximum prices and tariffs for an item of goods or of a services, which delivery and provision is considered as a performing of activities subject to regulation.

The goal of the submitted paper is to present the results of application of models for determination of maximum prices and tariffs of goods of network industries in Slovak national electricity markets.

Keywords

Network industries, Regulatory Office for Network Industries, multicriteria evaluation regulated prices, reasonable profit in regulated industries, rate of return regulation, Performance Based Regulation, PROMETHEE outranking methods.

1 Introduction

The Slovak Republic has opened its national electricity market gradually. The basic framework has been defined by an Act No. 70/1998 Z.z on Energy, which became effective as of 1.7.1998, the Act No. 276/2001 Z.z on Regulation of Network Industries, which became effective as of 1.8.2001 and on its basis Regulatory Office for Network Industry was established and further amendments to Act No. 70/1998 Z.z on Energy, when competences of Ministry of Economy of the Slovak Republic (hereinafter referred to as „Ministry“) were transferred to RONI in the area of regulation and issuing licenses. .

RONI took the responsibility for regulation of energetics and the responsibility for economic policy in network industries and execution of property rights in regulated enterprises remained to government. The regulation is executed by Act on Regulation, which determines the method of calculation of maximum prices, or determination of maximum prices or tariff, whereby such a determined price must reflect economically justified costs and adequate profit from performing regulating activity.

The Regulatory Office for Network Industries, in compliance with its primary aim that is a technical and price regulation of enterprising in activities subject to regulation of network industries, had to provide for effective solution of two tasks:

(a) Creation of effective, and at the same time it is necessary to stress also competitive market environment with goods and services of network industries by putting into effect standard regulatory mechanisms especially in context of entry of Slovak Republic into EU and gradual adoption to conditions of electricity markets within integrated Europe;

(b) To prepare such a analytical apparatus for price regulation of network industries, which would granted effective development of regulative subjects, whereby in first phase the identically important task was to eliminate deformation in prices of goods of network industries.

2 Methodological aspects of electricity market regulation

In Slovak Republic, the company National Economic Research Associates is dealing with preparation of methodological apparatus of tariffs regulation for electricity distribution. Project's results are documented in extensive and analytically detailed reports and also related models for calculation of price regulation are part of project's solutions and they are prepared in environment of software Microsoft Excel.

Methodologies, which were presented among realized solutions to these projects, served as significant support by creating series of RONI's decisions over the years 2001-2002 and price regulation of goods of network industries in Slovak Republic were managed according to them. Finally, 30.7.2003 RONI issued 4 regulations which defined concrete methods for price regulation in network industries (e.g. in energetic, gas industry, thermal economy and water management) and determined the scope of economically justified costs and adequate profit of regulated subjects. It is possible to consider this phase of regulatory methodology creation to be completed and in the next period it will be necessary to assess its efficiency and in case of need to specify, modify or supplement it.

The company National Economic Research Associates proposed for price regulation of electricity in Slovakia the method, which combines method of Rate of Return Regulation with a method of Performance Based Regulation.

The basic relations of model for price regulation of electricity in Slovak Republic were suggested by company National Economic Research Associates in compliance with theoretically reasoned methods of natural monopoly regulation and after certain elaboration were published in „ Regulation of Regulatory Office for Network Industries from 30.7.2003 No. 1/2003, which lays down further information about methods for electricity price regulation.

It is appropriate to remind that the standard method of Rate of Return Regulation (RoR) is aimed to ensure that regulated subject set the price of goods or services for its customers in such a way to cover from its revenues all adequate and cautiously arisen costs as well as regulated return of its investment. At the same time, it is necessary to realize that based on H. Averch and L. Johnson model, the company is regulated according to principal of Rate of Return on capital Regulation. Therefore in effort to increase its permitted "adequate" profit, companies tend to inadequately and uselessly raise their investments.

In the conditions of electricity price regulation is this methodology applicable by quantification of adequate revenues of energy company according to relation which is derived on the basis of modification of formula (content as well as methodical) for Rate of Return Regulation (ROR).

Within this general methodology, burden of argumentation about justifiability of costs to satisfaction of regulatory authority is the task and responsibility on the side of regulated subjects. Regulated subject must univocally demonstrates that each element of cost item arose cautiously and it is also an adequate cost needed for fulfilling energetically needs of economy.

Investments into fixed assets must be expended circumspectly, „used and useful" by provision of energy in such a way to become a component of a charge basis of company.

Analogically, elements of operating costs must be classified as a cautious and needed for provision of service, to be implemented in to subcategory EO&M.

The main advantage of the price regulation method on the basis of the method RoR is that prices, based on tested year, are steady and so they are fixed till the next tariff's procedure. On the other hand, there are also disadvantages, from which the most substantial is that fact that this methodology supports by regulated energy company tendency to build up new power plants and equipment, which perhaps are not necessary but they pretend investments in to fixed assets. Of course, it is an extreme situation.

The certain modification of Rate on Return Regulation (RoR) is method of Performance Based Regulation (PBR). Method of PBR eliminates some „not motivating" features of method RoR and creates the system of stimulus for increasing performance of regulated subject. The method of Performance Based Regulation in initial phase on the basis of RoR defines initial, perhaps starting, requirements for revenues as well as basic tariffs for goods and services.

The initial year for setting starting parameters of system for energy market in Slovak Republic was year 2002. These requirements for revenues and basic tariffs are regularly updated by using RoR methodology. Interval of tariffs revision is usually 3-10 years. Prices are regulated between revisions according to special formulas, which take into account character of industry and economic priorities of the whole managed system.

Resulted from the made analysis aimed on regulation of used methodology we can state that consulting company National Economic Research Associates chose regulation methods theoretically reasoned, by professional literature documented and in practise tried methods of price regulation, which motivate regulated subjects to increase of their production efficiency.

3 The effects of changes of regulated prices in energetics on industries profitability

Ordinances, which have been constituted by Regulatory Office for Network Industries in the area of network industries regulation over the last years, have directly related to economic activities and their results in single industries in Slovak Republic. Therefore it is necessary scientifically to evaluate impacts of changes in energy prices realized over the years 2000 till 2003. The effective way is method of multi-criterion evaluation of variants (VVV) for judgment economic potential development of crucial industries in Slovak Republic.

The analysis was made for following industries over the years 2000 -2004.

C	Mining of mineral resources
D	Industrial production
E	Generation and distribution of electricity, steam and water
STRP	Engineering industry
ELP	Electro technical industry
CHFP	Chemical and pharmaceutical industry
DSP	Wood-processing industry
LP	Light industry
HUTP	Metallurgical industry
SPRP	Processing industry

We compared and evaluated economic potential of industries in respective years with a help of multi-criterion evaluation of variants and statistic analysis, which provided us with information about their development tendency after realized price changes. The aim was to evaluate the price change influence on monitored industries.

We used method of evaluation of variants PROMETHEE for solution of the set problem. The aim of this method was the ordering of variants evaluated according to several criteria. From the entries we took the state of industries in the year 2000, 2001, 2002, 2003 and 2004.

From the individual economic parameters we took criteria, on basis of which and with a given method, we evaluated the states of industries in respective years.

We chose two groups of criteria for the analysis- main and supplementary. The main criteria include export, profit/loss before tax, costs, return on revenues and return on costs. The supplementary criteria include total revenues, revenues from own products and services, total revenues, return on fixed assets, return of current assets, return on equity and return on not-own capital. We made an analysis not only for main but also for all criteria in the industries C, D, E and in the total industry of SR. We used main criteria for selected industries of processing industry in the sphere of producing and network industries, because they have the most significant influence on economic potential. When we compare the results for industries C, D, E and total industry of SR we could see that the ordering of variants is nearly identical in both kinds of criteria.

On the basis of importance of influence of individual criteria we set the weights that are shown in the table below.

	Weight
Export	0.15
Costs	0.10
Profit/loss before taxation	0.15
Return on revenues	0.10
Return on costs	0.10
Total revenues	0.05
Revenues from own products and services	0.05
Added values	0.05
Revenues	0.05
Return on fixed assets	0.05
Return on current assets	0.05
Return on equity	0.05
Return on not-own capital	0.05

Resulted from the made analysis we can state that overall economic potential of Slovak industry is growing, even if in certain industries there is a noticeable influence of price change of energies or other economic attributes.

Development of economic potential in Industries of Slovakia during the years 2000 – 2003 we presented in the following Table.

Development of economic potential in Industries					
	2000	2001	2002	2003	2004
C	2	1	4	5	3
D	4	5	3	2	1
E	5	4	3	2	1
Industry SR	**5**	**4**	**3**	**2**	**1**
STRP	5	4	1	3	2
ELP	5	3	4	2	1
CHFP	5	1	4	3	2
DSP	4	1	5	3	2
LP	5	3	1	4	2
HUTP	5	3	4	2	1
SPRP	5	3	4	2	1

Taking the results from analysis we can state, that regulation methods, that were used by Regulatory Office for Network Industries in Slovak electricity and power industry market have their motivational aspect for effective development of regulated participants apart from their stabilizing function in the market environment of products of network industries.

4 Conclusion

There are of course also other forms of price regulation, which influence the reasonable profit of the firm directly on the basis of the volume of its production, the level of product sale of the regulated firm respectively on the basis of the amount of its total costs. The aim is to support the effective development of the regulated subject by help of regulation mechanisms.

According to our opinion a certain lack that can cause a delay of full development of electricity market is non-existent operator of electricity market. Certain role of this according to our opinion a very important participant of electricity market is preformed by Slovenská prenosová sústava, a.s. (further only" SEPS"). SEPS is a joint-stock company owned by the state established for transmission system operation. SEPS performs as a system operator and ministry authorized SEPS to execute a function of electricity market operator, until the market powers create an independent market operator. Slovak Republic has chosen a different approach and for instance in contrast to the Czech Republic, the government of Slovak Republic did not create (established) the market operator as an independent subject.

Further problem for full liberalization of the market seems to be cross -border connections of Slovak Republic with neighbouring countries and non existence of interstate connection of Slovak Republic and Austria.

On the basis of a made analysis we can state that realized price changes in the product market of network industries have not a significant negative influence on individual industries in generally. In the industry E "Production and distribution of electricity, steam and water" have even significantly positive effect. According to our opinion the most important impact of price change of energy we can see on the light industry, which after its positive trend in the years 2001, 2002 recorded in the year 2003 a rapid decline (worsening) in its situation below its level in 2001 after the most significant price change.

References

[1] Bien, F.: Systemwechsel im Europäischen Kartelrecht. In: Der Betrieb N' 46/2000. Düsseldorf Verlagsgruppe Handelsblatt, 2000.

[2] Fendeková, E.: Oligopoly a regulované monpoly. IURA Edition, Bratislava, 2006

[3] Fendeková, E. - Fendek, M.: Multiple Criteria Evaluation of Economic Potential of Selected Conversion Firms in Slovakia. Central European Journal for Operations Research and Economics, 4/1996, No.4. Bratislava: Karprint.

[4] Scherer, F.M.-Ross, D.: Industrial Market Structure and Economic Performance. Boston: Houghton Miffin Company, 1990.

[5] Tirole, J.: The Theory of Industrial Organisation. Massachusetts, Cambridge: The MIT Press, 1997.Act No. 291/2002 Coll. dated May 21, 2002 about the State Treasury and modification and completion of some other acts

[6] Zákon č. 137/1990 Zb. z 27. februára 2001o ochrane hospodárskej súťaže a o zmene a doplnení zákona Slovenskej národnej rady č. 347/1990 Zb. o organizácii ministerstiev a ostatných ústredných orgánov štátnej správy Slovenskej republiky v znení neskorších predpisov.

Doc. Ing. Eleonora Fendeková, PhD., Department of Business Economics, University of Economics Bratislava, Dolnozemská 1, 852 35 Bratislava, Slovakia, e-mail:nfendek@dec.euba.sk
Prof. Ing. Michal Fendek, PhD., Department of Operations Research and Econometrics, University of Economics Bratislava, Dolnozemská 1, 852 35 Bratislava, Slovakia, e-mail:fendek@dec.euba.sk

Modeling of Network Competition

Petr Fiala, Lenka Flusserová, Václav Kořenář

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Prague 3

pfiala@vse.cz, lflusser@vse.cz, korenar@vse.cz

Abstract

The purpose of this paper is to develop a formal framework and models for analyses of competition between interconnected networks providing services. The formal framework is based on game and negotiation model approaches. The models are devoted to analysis of specific issues. Critical elements of competition are terms and conditions of the access arrangements between networks. Access fees are determined either by a regulator or by competition. The models analyze when network competition is sustainable without regulation, and how regulation can promote sustainable network competition when it is not. The sustainability of competition between networks is influenced by initial market share allocations, propensity of subscribers to switch networks, fixed costs of operations, and a market's price. The impact of competition between networks on negotiation and integration within a network is analyzed also.

Keywords

Network, Competition, Game, Oligopoly, Regulation, Vertical integration, Negotiation

JEL: L14, L22, C70, C78

1 Introduction

The network economy is a term for today's global relationship among economic agents characterized by massive connectivity (Shapiro, Varian, 1999). Many industries involve networks competing with others. In some cases, these networks are interconnected with others and engaged in competition for subscribers. The network industry typically requires access to rival networks to provide services or to satisfy its customers. Such examples include networks for communication services, electricity transmission, gas transportation, banks' ATM networks, etc. This feature is what distinguishes the network industries from others in that interconnected firms try to take dominant position not only by competing in prices but also by deteriorating competing network by charging excessive access fees. There are still many open research questions in the network competition analysis. Further development of network competition theory can help resolve these issues and improve policy. The paper presents a theoretical framework for analyzing of network competition. Specific models analyze interconnections of network competition, regulation and integration.

2 Basic theoretical framework

The general framework is related to the literature on competition in networks, e.g. (Armstrong 1998), (Laffont-Rey-Tirole 1998a, 1998b), and (Dessein 2003). Laffont, Rey and Tirole (LRT) have analyzed a model of two local network companies that possess different attributes for consumers. In their model, the two companies, given access charges, set the local prices competitively. The customers of a network are charged the same price independent of the network which completes their call. The networks compete only in prices since the other attributes are assumed to be fixed. They use the standard Hotelling (1929) location model. LRT model have provided a basic theoretical framework to analyze the network competition and interconnection issues.

The simplicity of the framework suggests that it should be possible to extend it in a number of different directions. The framework can be extended in terms of the number of networks, economic instruments, cost structures, price discrimination, asymmetric structures, etc. There is a vast literature about the possible extensions, e.g. (Chemla, 2003), (De Fontenay, Gans, 2004). The general framework contains the extension possibilities and modeling instruments.

The modeling instruments for analyses of network competition are:

- game theory models,
- oligopoly models,
- negotiation models.

Game theory is the basis for development of network competition models, non-cooperative and cooperative models as well. The Cournot, Stackelberg and Bertrand models are representations of oligopolistic behavior. Nash equilibrium concept is used for solution. Cartel models are representations of cooperative behavior. Negotiations take place in cooperative solutions of competition problems. There are approaches based on game theory and other approaches including ones based on multicriteria evaluations.

The theoretical framework serves as a common basis for developing special models for analyzing specific features in network competition. In the next two sessions there are presented two simple models for analyzing relations of network competition and regulation and network competition and integration.

3 Model of competition and regulation

The model is based on simplified assumptions. There are two interconnected symmetric networks in the market. The firms provide network services which are close substitutes. The Bertrand model is a base for the situation. Networks compete in prices (p_1, p_2) and are assumed to have the following linear demands

$$q_i(p_i, p_j) = b_0 - b_1 p_i - b_2 p_j, \quad i, j = 1, 2,$$

where b_0, b_1, b_2 are parameters of the demand function. The number of calls originating on a network and completed on the network is equal to the demand for the service.

Networks have the same cost structure. There is a fixed cost f of providing services. A network incurs a marginal cost c_0 per payment at the originating and terminating ends of the payment and marginal cost c_1 in between. The total marginal cost of a payment is thus $c = 2c_0 + c_1$.To provide interconnection services, it is necessary to provide essential input services to its competing network. Networks charge fees for this service called access charges (a_1, a_2). Networks may set their access fees non-cooperatively and therefore possibly asymmetrically in nonreciprocal access pricing.

Objectives of networks are to maximize their profits. The network profit is given by

$$z_i(p_i, p_j, a_i, a_j) = [(p_i - c)q_i - f] + [(a_i - c_0)q_j - (a_j - c_0)q_i], \quad i, j = 1, 2.$$

Two different market environments are analyzed:

1. access fees are determined by a regulator and networks compete non-cooperatively in prices,
2. in the first stage networks compete in access fees and in the second stage they compete in prices.

Regulation of access fees

A regulator sets a reciprocal access charge a_r for both networks. The networks compete in prices by given access charge. The network profit is given by

$$z_i(p_i, p_j) = [(p_i - c)q_i - f] + [(a_r - c_0)q_j - (a_r - c_0)q_i], \quad i, j = 1, 2.$$

Nash equilibrium is computed by solving following equations

$$\frac{\partial z_i}{\partial p_i} = 0 \ , \quad i = 1, 2.$$

There is a symmetric Nash equilibrium

$$p_r = p_1 = p_2 = \frac{b_0 + b_1 c}{2b_1 - b_2} + \frac{(b_1 + b_2)(a_r - c_0)}{2b_1 - b_2}.$$

The final prices consist from two parts. The first part is determined by price competition and the second part is determined by access fees. The access charge depends on an objective of the regulator. If the access charge is set $a_r = c_0$, then the final price is the same as the one under the typical Bertrand competition

$$p^* = \frac{b_0 + b_1 c}{2b_1 - b_2}.$$

If the objective is to maximize consumer welfare under constraints that network profit is zero, then we get the access charge a_0 and the price p_0.

It can be shown that it holds

$$p_0 < p^* \text{ and } a_0 < c_0.$$

The access fee lower than marginal costs ($a_0 < c_0$) does not imply that networks face deficits of providing access. At equilibrium, flows in and out of each network are balanced, and thus there is no deficit in access revenue. The regulator can effectively control the price by regulating the access fee.

Competition in access fees

The networks set access fees non-cooperatively. The model is based on two stage game where the networks first set access fees and second compete in prices. The analysis starts with the second price competition stage, taking nonreciprocal access charges as given, and then continues with the first access charge competition stage backward.

For given access fees, networks' profit maximizing first order conditions are given by

$$\frac{\partial z_i}{\partial p_i} = 0 \ , \quad i = 1, 2.$$

From these conditions the second stage Nash equilibrium given access charges can be derived. Using backward induction, it can be expressed each network's profit in terms of access fees. Then it can be easily find the Nash equilibrium for this two stage game

$$\frac{\partial z_i}{\partial a_i} = 0 \ , \quad i = 1, 2.$$

There is a symmetric equilibrium where the access fee $a_c = a_1 = a_2$ and the final price is determined by

$$p_c = p_1 = p_2 = \frac{b_0 + b_1 c}{2b_1 - b_2} + \frac{(b_1 + b_2)(a_c - c_0)}{2b_1 - b_2} .$$

Difference between p_r and p_c differs only from the way how access charges are determined, if they are determined by a regulator or by the market. The equilibrium price increases as the equilibrium access charge increases.

It can be shown that it holds

$$p_0 < p^* < p_c \text{ and } a_0 < c_0 < a_c.$$

The price is higher when the networks compete in both access charges and prices than when access fees are regulated, since the regulator sets access charges below the marginal costs. The access charges determined by market forces are above the marginal costs. The networks has an incentive to lower the price in order to attract more consumers when markets are more competitive, but at the same time has an incentive to increase rival's prices by increasing access charges. The networks are not only competitors in the final product market but also input suppliers as providers of facilities. The regulator's intervention in the access pricing can facilitate a tacit collusion.

4 Model of competition and integration

The impact of competition between networks on negotiation and integration within a network is analyzed. The model is based on simplified assumptions which make possible to analyze investigated effects. There is a case of m oligopolists O_i , $i = 1, 2, \ldots, m$, each of whom sells a input to n_i downstream units, r_i of them are integrated with the oligopolist. The hole system consists of m networks N_i , $i = 1, 2, \ldots, m$. The network $N_i = (O_i, n_i, r_i)$ composed from the oligopolist O_i , n_i downstream units, r_i of them are integrated with the oligopolist.

For simplicity, we assume that the oligopolist provides an input a unit of which can be converted by into a unit of the final product. Downstream units have limited capacity; they can produce at most one unit of the final product. The products are perfect substitutes. Downstream units produce a total quantity $n = \sum_{i=1}^{m} n_i$, this results in a market price of $p(n)$. The final product price linearly depends on the total number of products

$$p(n) = b - n ,$$

where b is a parameter of the demand function. The system has following cost structure. Each downstream unit has fixed costs f. There are no additional marginal costs associated with producing the unit other than those arising from payments upstream. A vertically integrated downstream unit has an additional fixed cost g.

The gross profit of non-integrated unit and for of integrated unit equals $z = p - f$, $z = p - f - g$, respectively. The profit accruing to the network N_i, $i = 1, 2, \ldots, m$, is

$$Z_i = p(n)\, n_i - fn_i - gr_i.$$

For comparison, we take the classical Cournot oligopoly model. Nash equilibrium is computed by solving following equations

$$\frac{\partial Z_i}{\partial n_i} = 0\ , \quad i = 1, 2, \ldots, m.$$

There is a symmetric Nash equilibrium

$$n_i^* = \frac{b - f}{m + 1}, \quad i = 1, 2, \ldots, m.$$

The modified negotiation model is based on a negotiation game in each network. The negotiation game has the following stages:

1. system network structure,
2. negotiations,
3. competition.

System structure

Each oligopolist O_i chooses the number of integrated units r_i in its network and the number $(n_i - r_i)$ of independent downstream units enter to be supplied by O_i (see Fig. 1). The key assumption here is that these decisions happen simultaneously. The oligopolist integrates for strategic reasons. The initial set of units cannot be replaced.

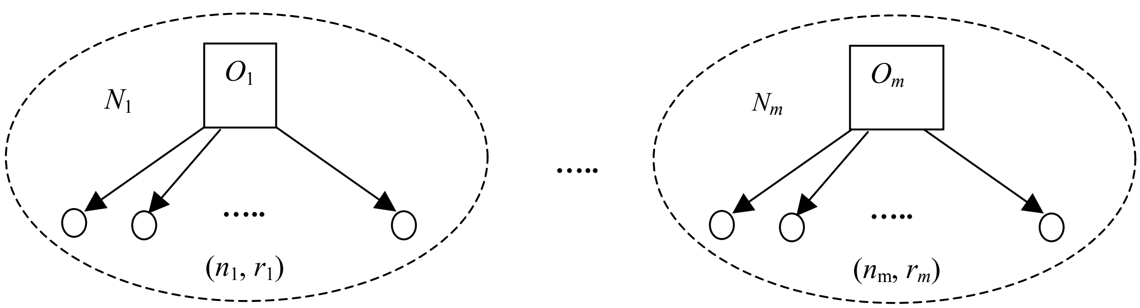


Fig. 1 System network structure

Negotiations

Oligopolists negotiate with their independent downstream units over a supply contract. The negotiations are about quantities n_i and transfer supply prices $a_i(n_i, r_i)$. The supply prices can be interpreted also as access fees for downstream units to be members of the network. For negotiation process can be used various negotiation models. When negotiations with an individual firm break down the oligopolist must also renegotiate pricing arrangements with other firms and may face a competitive response from rival networks.

There is a symmetric Nash equilibrium for negotiation model without integration

$$n_i^0 = \frac{3(b-f)}{3m+1} > n_i^* = \frac{b-f}{m+1}, \quad i = 1, 2, \ldots, m.$$

The resulting supply prices for independent units in case without integration $a_i^0 = a_i\ (n_i, 0)$. The incentive to integrate is given by this schema:

- for $g = 0$, there is complete integration, $r_i = n_i = n_i^*$,
- for $g < a_i^0$, there is partial integration, $r_i < n_i$,
- for $g \geq a_i^0$, there is no integration, $r_i = 0$.

Integration will occur if it is relatively cheap, and it will result in a reduction in network and total output. Integration allows the oligopolist to negotiate higher supply prices to independent downstream units.

There is a symmetric Nash equilibrium for negotiation model with integration

$$n_i^I = \min\left(\frac{b-f+2g}{m+1}, \frac{3(b-f)}{3m+1}\right),$$
$$r_i^I = \max\left(0, \frac{b-f-(3m+1)g}{m+1}\right).$$

Competition
Competition of downstream units occurs and payoffs are realized. Important managerial implications result from the model. Issues of competition and its impact on integration are considered.The increased inter-network competition can mitigate incentives for inefficient integration. As the level of upstream competition increases, integration is less likely to occur. Increased network competition both improves competitive outcomes and reduces inefficiencies that might arise from inefficient strategic vertical integration.

5 Conclusions

Network competition is the important subject of an intensive economic research. The paper presents a basic theoretical framework for analyses of network competition. The framework makes possible to develop simple models for analyzing specific features in network competition. Two simple models are presented. The simple models have very important managerial implications indeed. The combination of such models can give more complex views on the problem of network competition.

The research project was supported by Grant No. 402/05/0148 from the Grant Agency of Czech Republic „Network economy – modeling and analysis“.

References

[1] Armstrong, M. (1998): Network Interconnection in Telecommunications. The Economic Journal 108, 545–564.

[2] Chemla, G. (2003): Downstream competition, foreclosure and vertical integration. Journal of Economics and Management Strategy 12, 261-289.

[3] De Fontenay, C. C., Gans, J. S. (2004): Can vertical integration by a monopsonist harm consumer welfare? International Journal of Industrial Organization 22, 821-834.

[4] Dessein, W. (2003): Network Competition in Nonlinear Pricing. RAND Journal of Economics 34, No. 4.

[5] Hotelling, H. (1929): Stability in Competition. Economic Journal 39, 41-57.

[6] Laffont, J.-J., Rey, P., Tirole J. (1998a): Network Competition: I. Overview and Nondiscriminatory Pricing. Rand Journal of Economics 29, 1-37.

[7] Laffont, J.-J., P. Rey, P., J. Tirole J. (1998b): Network Competition: II. Price Discrimination. RAND Journal of Economics 29, 38-56.

[8] Shapiro, C., Varian, H. (1999): Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press, Boston.

Prof.RNDr.Ing. Petr Fiala, CSc., MBA, Ing. Lenka Flusserová, Ing. Václav Kořenář, CSc.
Katedra ekonometrie, Vysoká škola ekonomická, nám. W. Churchilla 4, 130 67 Praha 3

Analysis of panel data with binary dependent variable

Zuzana Fíglová[1]

Department of Econometrics
University of Economics Prague
W. Churchill Sq. 4, 130 67 Prague 3

Abstract

This paper presents an econometric analysis of panel data when dependent variable in the model is supposed to be binary. Binary choice model is based on random utility theory of consumer behavior where each consumer maximizes his utility through the consumption of goods and services. That is the reason why discrete choice models have become popular and widely used in many diverse areas such as marketing research, consumer behavior, psychology, transportation and others. We consider a panel data set that represents data where multiple cases were observed at two or more time periods. In our study we analyzed nonlinear logit model with fixed effects. In the case of continuous dependent variable and no autoregressions, the fixed effects model gives consistent estimates of the regression coefficients. That is not the case when dependent variable is qualitative and there are only a few time-series observations per individual where maximum likelihood estimation provides inconsistent estimates of the parameters. In the logit model one can obtain consistent estimates when applying conditional likelihood approach. This method does not produce computational simplifications that all fixed effects must be estimated as a part of the estimation procedure. The fixed effects models have several disadvantages. They may contain too many cross-sectional units of observations requiring too many dummy variables for their specification. Too many dummy variables may sap the model of degrees of freedom for adequately powerful statistical test, for example White test for heteroskedasticity. There can be also a problem of multicollinearity. Finally, the classical assumptions for the error term may have to be modified. The logit model was applied to data from the Household Budget Surveys 2000-2004 carried out by the Czech Statistical Office in order to analyze choice behavior of households. The data on households included demographic characteristics of individuals, housing, household amenities, net income, and opinions of households about their own socioeconomic situation. We analyzed the role of income as determinant of ownership of PC and its change through the observed period of time. Another approaches could be used for analysis, for example the random effects model or error components model.

Keywords

binary choice model, panel data, household budget survey, durable goods
JEL: C23, C25

1 Binary Choice Model with Panel Data

Observations on many individual economic units (firms, households, geographical areas, and the like) over a period of time are said to be a "panel data set". Panel data may be contrasted with pure cross section data, observations on individual units at a point in time, and with pure time-series data, observations, usually of an aggregate nature, over time without any "longitudinal" dimension. Panel data offer several important advantages over data sets with only a temporal or only a longitudinal dimension.

We combine a panel data approach with discrete choice models. Discrete choice models have become very useful and popular in a various areas as consumer behavior, marketing research, energy, psychology, telecommunications, biomedical research, to name a few.

[1] Financial support of GACR 402/04/0756 is gratefully acknowledged by the author.

We define discrete choice (or qualitative response) models as models, where the dependent variable takes a discrete and finite set of values. That means consumer or decision maker faces up to a choice among a set of alternatives. In this set, the number of alternatives is finite, the alternatives are mutually exclusive, and finally the set of alternatives is exhaustive. We will concentrate on models, where dependent variable takes a dichotomous value and this model is called binary choice model.

Let consider the data consist of observations of n subjects, and the i^{th} subject has T_i observations. Denote the t^{th} observation of the i^{th} subject by $(y_{it}, \boldsymbol{X}_{it})$, where y_{it} is a binary value and $\boldsymbol{X}_{it} = (x_{it0}, x_{it1})^T$ is the covariate matrix with column vectors x_{it0} and x_{it1} associated with values 0 or 1 of the dependent variable y_{it}. In the context of consumer behavior mentioned above, y_{it} denotes ownership of the durable goods chosen by the i^{th} subject (in our case household by decile) at the t^{th} time of asking. If the individual-specific effect α_i, is assumed to be fixed we can define the following fixed effect regression model based on a latent-variable formulation (Wooldridge, 2002)

$$y_{it}^* = \boldsymbol{X}_{it}^T \boldsymbol{\beta} + \alpha_i + \eta_{it} \qquad i = 1,2,\ldots,n, \quad t = 1,2,\ldots,T, \tag{1}$$

where y_{it}^* is a latent variable that is not observed by the researcher. A consumer chooses $y_{it} = 1$ if the latent variable is positive and 0 otherwise, hence

$$y_{it} = \begin{cases} 1 & \text{if } \ y_{it}^* > 0, \\ 0 & \text{if } \ y_{it}^* \leq 0. \end{cases} \tag{2}$$

The latent variable can be y_{it}^* interpreted as the utility difference between choosing $y_{it} = 1$ and 0. For example, in a binary logit model y_{it} can be modeled as independent across all subjects and across all repeated observations and the probability that the y_{it} takes on the value 1 is

$$P_i = E_i(y_{it} = 1 \,|\, \boldsymbol{X}_{it}, \alpha_i) = \frac{\exp(\boldsymbol{X}_{it}^T \boldsymbol{\beta} + \alpha_i)}{1 + \exp(\boldsymbol{X}_{it}^T \boldsymbol{\beta} + \alpha_i)}. \tag{3}$$

We can define binary probit model very similar when using a cumulative distribution function for standard normal.

Methods of estimation

There are several possibilities of estimation of panel data with binary dependent variable.

Least square dummy variable estimator - this is computed using by least squares regression of $y_{it}^* = (y_{it} - \bar{y}_i)$ on the same transformation of $\boldsymbol{X}_{it}$ where the averages are group specific means. The individual specific dummy variable coefficients can be estimated using group specific averages of residuals. We used this method of estimation for linear probability model. In this case we have decided to use a one-way model that means time effects play an important role in the model according to the chi-squared statistics. Then we have to make a decision between fixed effects and random effects model. We have used a Hausman test (Greene, 2003) and decided to use a fixed effects model. The fixed effects model has several disadvantages. They may contain too many cross-sectional units of observations requiring too many dummy variables for their specification. Too many dummy variables may sap the model of degrees of freedom for adequately powerful statistical test, for example White test for heteroskedasticity. There can be also a problem of multicollinearity. Finally, the classical assumptions for the error term may have to be modified.

- **Conditional maximum likelihood estimator** – this method of estimation we have used for estimatig of binary logit and probit models of panel data with fixed effects (Chamberlain, 1980). Maximum likelihood function has the following form

$$L = \prod_{i=1}^{N} P\left(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, ..., Y_{iT} = y_{iT} \middle| \sum_{t=1}^{T} Y_{it} \right). \tag{4}$$

  This estimation technique provides consistent estimates of parameters but it is computationally very difficult and it is hard to extend it to more complex econometric models.

- **Maximum score estimator** – it is defined as the maximizer of the objective function

$$n^{-1} \sum_{i=1}^{n} (2y_{it} - 1)\{\boldsymbol{X}_{it}^T \boldsymbol{\beta} \geq 0\}. \tag{5}$$

  Given that there is no log-likelihood function underlying the fitting criteria, there is no information matrix to provide a method of obtaining standard errors for the estimates. A method that is used to provide at least some idea of the sampling variability of the estimator is bootstrapping.

- **Kernel estimator** – in the previous method, there is little information about any relationship between the response and the exogenous variables based on estimation results. For a kernel estimator for a nonparametric regression function the function value is estimated with

$$F(z^*) \approx \frac{\sum_{i=1}^{n} w_i(z^*) y_i}{\sum_{i=1}^{n} w_i(z^*) y_i}, \tag{6}$$

  where $w_i(z^*) = K[(z^* - z_i)/(\lambda s)]$,

  $K(r_i) = P(r_i)[1 - P(r_i)]$,

  $P(r_i) = [1 + \exp(-c r_i)]^{-1}$.

  The constant $c = (\pi / \sqrt{3})^{-1} \approx 0.5513$ is used to standardize the logistic distribution that is used for the kernel function. The parameter λ is the smoothing or bandwidth parameter.

2 Econometric analysis of panel data

The data for analysis of panel data with a binary dependent variable were obtained from the *Household Budget Survey* in the Czech Republic that was carried out by the Czech Statistical Office. We analyzed the observed period 2000-2004. It contains the data from households distributed by net income per person in deciles, ownership of major household durable goods such as refrigerator, microwave, personal computer etc. Our pooled data set contains a total of 10*5=50 observations. In other words ten household groups by deciles are followed for five years and are sampled annually.

We estimated econometric models to investigate whether or not net income, household group by decile or time raise the probability that a household owns a personal computer. We analyzed five econometric models: linear probability model with fixed effects (LPMFE), logit model with fixed effects (LOGFE), probit model with fixed effects (PROBFE), model of binary choice estimated by semiparametric method the maximum score method (MSCORE) and model of binary choice estimated by semiparametric kernel estimator (KERNEL). Independent dummy (0-1) variables in the models are years (*Year2000,…,Year2004*) and because we wanted to avoid falling into the dummy-variable trap (i.e. the situation of perfect collinearity) there is no dummy for *Year2000.* We are treating HD1 as the base, whose intercept value is given by β_0. The results contain Table 1.

From Table 1 is obvious that all estimated parameters, except *Year2001,* in model LPMFE are statistically significant, there is a high value of R^2 and F-test leads to the rejection of the hypothesis that dummy variables *Year2000,…,Year2004* have no effect on ownership of PC in households at the significance level α = 0.05.

The positive signs of the parameters mean that dummy variables *Year2000,…,Year2004* and *Income* increase the probability of ownership of PC. We can not reject the hypothesis of heteroskedasticity in model LPMFE when applying the White test with ignoring the cross–product terms, $\chi^2_{(6)}$ = 17.017 with p-value equals 0.009 and $F_{(6,37)}$ = 3.182 with p-value equals 0.013. We can not also reject the hypothesis of no autocorrelation using the Lagrange multiplier test where $\chi^2_{(1)}$ = 23.580 with p-value equals 0.000 and $F_{(1,43)}$ = 38.376 with p-value equals 0.000.

Parameter estimates in the other models are not directly comparable; they could be made comparable by transforming them to the marginal effects. But MSCORE model does not produce a marginal effect.

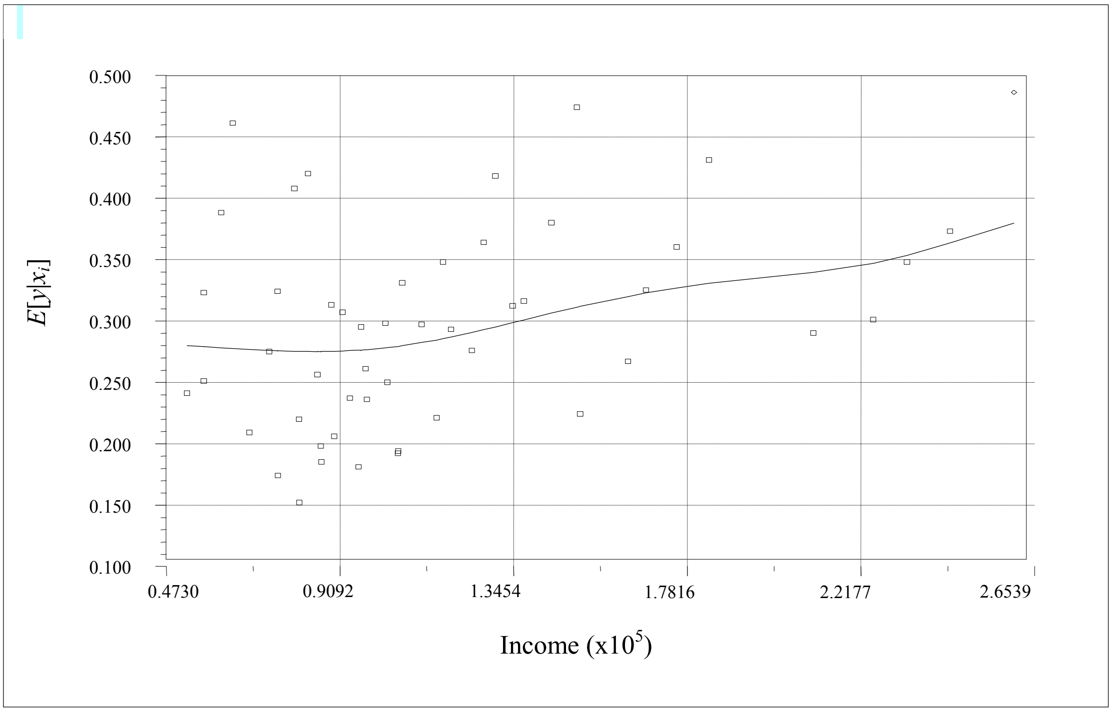
Table 1 Estimated parameters for the models of panel data

Variable	LPMFE	LOGFE	PROBFE	MSCORE
Intercept	0,1660			
	(7,889)			
Income	0,4567E-06	0,3197E-05	0,1762E-05	-0,9285
	(3,308)	*(0,056)*	*(0,051)*	*(-107,738)*
Year2001	0,0343	0,2081	0,1244	-0,1857
	(1,586)	*(0,133)*	*(0,135)*	*(-3,368)*
Year2002	0,0651	0,2792	0,1594	-0,1857
	(3,009)	*(0,111)*	*(0,106)*	*(-3,347)*
Year2003	0,1195	0,6170	0,3685	-0,1857
	(5,497)	*(0,427)*	*(0,432)*	*(-3,422)*
Year2004	0,1725	0,8389	0,5055	-0,1857
	(7,905)	*(0,608)*	*(0,616)*	*(-3,969)*
	R^2 = 0.6977			
	$F_{(5,44)}$ = 20.31			

[a] In parenthesis are *t*–values of estimated coefficients for significance level α=0.05.

Figure 1 shows a plot of estimate the regression function for $E[y|x_i]$ and the coefficients are the MSCORE estimates given in the Table 1. The plot is produced by computing fitted values for 100 equally spaced points in the range of $\boldsymbol{X}_{it}^T\boldsymbol{\beta}$.

Figure 1 Nonparametric regression



3 Conclusion

We concerned about estimation binary choice models with panel data in this study. We analyzed the data about the household's ownership of a PC. As the independent variables we chose net income and time in years 2000 to 2004. We estimated linear probability model with fixed effects, logit model with fixed effects, probit model with fixed effects, model of binary choice estimated by semiparametric method the maximum score method and model of binary choice estimated by semiparametric kernel estimator. There could be also included more variables as education, number of children or age in the model. Another approaches could be used for analysis, for example the random effect model that differs from the fixed effects model in the specification of the disturbance term.

References

[1] Chamberlain, G.: Analysis of Covariance with Qualitative Data, *The Review of Economic Studies 47/1*, Econometrics Issue, pp. 225-238
[2] Greene, W.A.: *Econometric Analysis*, Prentice Hall, New Jersey, 2003
[3] Hsiao, C.: *Analysis of Panel Data*, Cambridge University Press, 2003.
[4] Hušek, R., Moravová, J.: Analýza a prognóza na základě modelů binární diskrétní volby. *Politická ekonomie*, 2002, L/4, pp. 579–593
[5] Hušek, R., Pelikán, J.: *Aplikovaná ekonometrie. Teorie a praxe.* Professional Publishing, 2003.
[6] Maddala, G.S.: Limited Dependent Variable Models Using Panel Data, *The Journal of Human Resources 22/3,* str.307-338
[7] Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.
[8] www.czso.cz

Bayesian Estimation of Closed and Open Czech Economy Model*

Hana Fitzová

Department of Applied Mathematics and Computer Science
Masaryk University
Faculty of Economics and Administration
Lipová 41a, 602 00 Brno
Czech Republic
e-mail: pytelova@mail.muni.cz

Abstract

This paper illustrates behaviour of two models estimated on the Czech economy data. Presented models are DSGE models with monetary policy rule, inflation targeting and rational expectations. The first model is a forward-looking closed economy model with Taylor rule. The second model is a small scale new open economy forward-looking model also with Taylor rule. Both models are estimated by Bayesian estimation technique, concretely by the Metropolis-Hastings algorithm in Matlab. The estimation of the closed economy model does not represent any larger problem because the model is very robust. The open economy model better represents the Czech economy, but the estimation is more disputable. Important results of the estimation are discussed and illustrated in graphs.

Keywords: macroeconomic model, DSGE model, closed economy, open economy, Taylor rule, Bayesian estimation, Metropolis-Hastings algorithm

JEL classfications: C15, C51, E12, E17

1 Introduction

This paper presents an application of two conceptual frameworks of economy. It illustrates the behaviour of the two presented models on the Czech economy data. The first framework is a forward-looking closed economy model with forward-looking Taylor rule. The second model is a small scale new open economy model. Both presented models are models with monetary policy rule, inflation targeting and rational expectations. Both models are estimated by Bayesian estimation technique, concretely by the Metropolis-Hastings Markov Chain Monte Carlo method for simulation of posterior densities of parameters. In this case all inferences about parameters are included in the estimated posterior density of parameter vector. Estimation was realized by Dynare in Matlab. Advantages and weak points of both models are discussed in next parts.

2 Closed Economy Model

Monetary policy significantly influences the short-term course of the real economy due to temporary nominal price rigidities. The first framework is a forward-looking dynamic stochastic general equilibrium (DSGE) model of economy with money and temporary nominal price rigidities based on [2]. The aggregate behaviorial equations evolve from optimization by households and firms. It is very important that current economic behaviour depends on expectations of the future course of monetary policy, as well as on current and past policy actions.

Let x_t be the output gap, which is the deviation of actual output from its potential level (in natural logarithms). Let π_t^n be the period t deviation of the net inflation rate from its target, let i_t be the deviation of the nominal interest rate from its equilibrium level and let E_t be the expected mean based on information available

*This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172

at time t. ϕ, δ, λ, β, γ_π, γ_x, μ and ρ are parameters. The structure of the first model is following:

$$x_t = -\phi(i_t - E_t\pi^n_{t+1}) + (1-\delta)E_t x_{t+1} + \delta x_{t-1} + g_t \tag{1}$$
$$\pi^n_t = \lambda x_t + \beta E_t \pi^n_{t+1} + v_t \tag{2}$$
$$i_t = \gamma_\pi E_t \pi^n_{t+1} + \gamma_x x_t \tag{3}$$
$$g_t = \mu g_{t-1} + \hat{g}_t \tag{4}$$
$$v_t = \rho v_{t-1} + \hat{v}_t \tag{5}$$

Equation (1) represents a forward-looking IS curve that relates the output gap negatively to the real interest rate and positively to the expected output gap. It describes demand side of economy. Parameter $\delta \in \langle 0; 1\rangle$ describes the measure of looking backward and forward in actual production. Delta close to unity means predominant influence of the past development of production; on the other hand, delta close to zero represents main influence of the expected development of production.

Equation (2) is a forward-looking Phillips curve. It describes positive relation between the inflation rate and both the output gap and the expected future inflation.

Equation (3) represents forward-looking Taylor rule. It is a reaction of the monetary authority to the state of the economy. It demonstrates real behaviour of many central banks. It is simple and much more transparent than rules based on optimization. When expected inflation or production is above the target level, the central bank rises the interest rate to move the economy back to the steady state and vice versa.

The last two equations (4) and (5) describe random (demand and supply) shocks. Shocks g_t and v_t are independent and identically distributed random variables with zero mean and variances σ^2_g and σ^2_v, respectively, and $\mu, \rho \in \langle 0, 1)$.

3 Open Economy Model

The second framework is a small scale DSGE new open economy model (NOEM) based on [3] also with forward-looking Taylor rule. It is very similar to the first one, but it is "open", so it contains an exchange rate equation. Furthermore the equation of inflation is augmented with import inflation.

Notation: x_t is the output gap; π_t denotes the period t net domestic inflation rate (deviation from its target) in domestic country (only domestic goods), π^c_t is total CPI inflation rate in domestic country, π^*_t is the inflation rate in foreign country; i_t is the nominal interest rate (deviation from its equilibrial level), i^*_t is the same in foreign country; E_t is the expected mean based on information available at time t; q_t denotes the real exchange rate (an increase in q_t represents a real depreciation); g_t, v_t and ϵ_t are white noise shocks. ϕ, δ, λ, β, γ_π, γ_x and α are parameters. The second model is following:

$$x_t = -\phi(i_t - E_t\pi_{t+1}) + (1-\delta)E_t x_{t+1} + \delta x_{t-1} + g_t \tag{6}$$
$$\pi_t = \lambda x_t + \beta E_t \pi_{t+1} + v_t \tag{7}$$
$$\pi^c_t = \pi_t + \frac{\alpha}{1-\alpha}(q_t - q_{t-1}) \tag{8}$$
$$q_t = \kappa E_t q_{t+1} - (1-\alpha)(i_t - E_t\pi_{t+1} - i^*_t + E_t\pi^*_{t+1}) + (1-\alpha)\epsilon_t \tag{9}$$
$$i_t = \gamma_\pi E_t \pi_{t+1} + \gamma_x x_t \tag{10}$$

The first two equations are the same as in the preceding model. Equation (8) describes relation between domestic inflation π_t and CPI inflation π^c_t; parameter α describes openness of the domestic economy. Equation (9) represents uncovered real interest rate parity (UIP) condition. It shows that expected real depreciation is dependent on the real interest rate differential.

4 Results of the Estimation

The models are identified on the Czech economy data. Data of the output gap, the net inflation rate, the net inflation target, the short-term nominal interest rate and the real exchange rate are quarterly seasonally adjusted values of the data from the Czech National Bank (first quarter of 1996 — first quarter of 2006). Data from foreign country are German data because German economy is considered as the main trade partner of the Czech economy.

The estimated systems are nonlinear, because states and parameters are estimated simultaneously. The systems are interdependent and they contain rational expectations of production and inflation. Both models

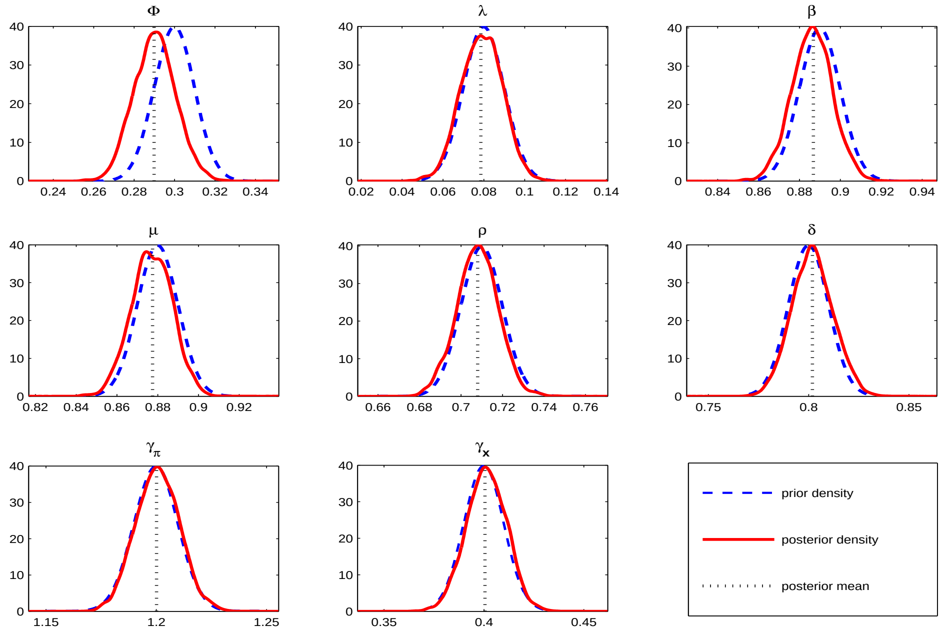


Figure 1: Prior and posterior density of parameters – closed economy model

contain two forward-looking variables and have two eigenvalues greater than one, which means that a unique steady state exists (i.e. Blanchard-Kahn condition holds).

The first model was estimated without considerable problems because it is very robust. But there were some complications during the estimation process of the second model. It was possible to realize the estimation of the second model with rational expectations after some modifications. The coefficient κ was added in front of the term $E_t q_{t+1}$ and it should be near to unity according to economic theory. But if it is one, the Blanchard-Kahn condition is violated and the second system cannot be estimated as a system with two rational expectations. So the coefficient κ was fixed to 0.9999. Exchange rate expectations were modeled as adaptive expectations.

The simultaneous estimation of both systems was realized by Bayesian estimation technique, concretely by the Metropolis-Hastings algorithm. It was done by Dynare in Matlab. 50000 iterations and 3 blocks were used for each model time. All the estimated parameters of both models are statistically significant at the level of 95%.

4.1 Closed Economy Model

Figure 1 represents prior and posterior densities of the parameters of the closed economy model, and posterior mean of the estimated parameters. Most of the posterior distributions seems to be normal. Only posterior distributions of the parameters λ and μ does not seem normal, they are not even unimodal. If we change some prior characteristics of the system nearly always we get some non unimodal distribution of some parameters. This fact magnifies using of Metropolis-Hastings approach instead of classical methods like Kalman filter. If we focus on relative size of the estimated parameters, we can say that they are in accordance with economic intuition.

The interest rate elasticity (ϕ) should be greater than zero and its estimation is around 0.29. It represents a negative relation between the output gap and the real interest rate in the Czech economy. The Phillips curve coefficient (λ) should be greater than zero, and the result is around 0.079. It means nearly zero dependence between the output gap and the rate of inflation. The discount factor (β) should be between zero and unity, near to unity, and its estimation is around 0.89, which means that agents discount future (today is more than tomorrow). The random shock parameters (μ and ρ) should be between zero and unity, and the results are around 0.88 and 0.71 respectively. It implies a great persistence of demand and supply shocks. Parameter δ is estimated around 0.8. It means that agents' predicitions of actual output gap are mainly based on the past development. Monetary policy coefficients γ_x and γ_π are estimated around 0.4 and 1.2 respectively. It shows, that the central bank is much more concerned with the inflation rate deviations from its target than with the output gap deviations.

4.2 Open Economy Model

Figure 2 depicts prior and posterior densities of the parameters of the open economy model. Most of the posterior densities seems to look like normal densities. In addition to the first model there is parameter α. Parameter α describes openness of the domestic economy (ratio of imports to GDP) and is about 0.39. The higher α, the higher influence of changes in the exchange rate and conditions in the foreign country on the domestic economy. The other estimations are similar to those from the closed economy model.

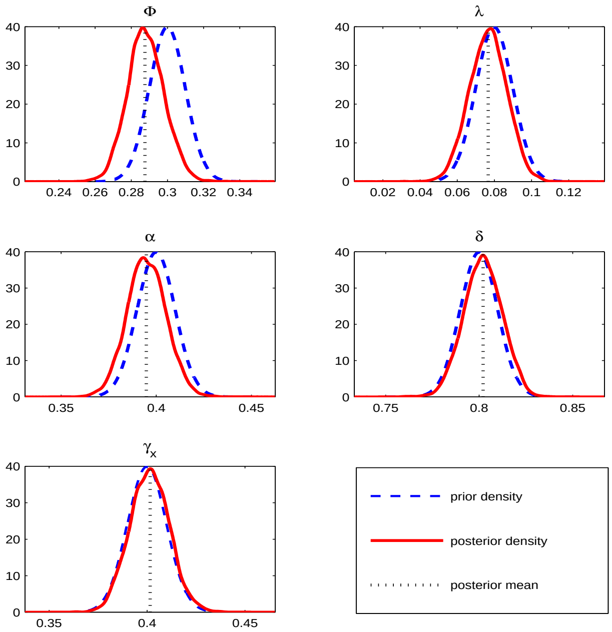


Figure 2: Prior and posterior density of parameters – open economy model

5 Other Characteristics

5.1 Univariate diagnostic

Univariate diagnostic was realized according to [1]. The course of the univariate diagnostic of the model parameters and the univariate diagnostic of the model as a whole imply convergence of the estimations. All parameter estimations in both models converge in all three paralel sequences of samples of length 50000.

5.2 Smoothed shocks development

The next model characteristic is development of the smoothed shocks. In the closed economy model, the demand shock is moving around zero, concretely in the interval $\langle -1\%; 1\%\rangle$ and the supply shock is moving about zero, concretely in the interval $\langle -2\%; 2\%\rangle$. On the other hand, the monetary shock is nearly all the time above zero and the maximum values $(10 - 15\%)$ correspond to the second half of the year 1997 and the first half of year 1998 which was the time of large monetary and exchange rate regime changes.

In the open economy model, the inflation shock is moving around zero, particularly it is in the interval $\langle -5\%; 5\%\rangle$, but there is a large jump up to 10% and then down in the period of crisis 1997-98. Interest rate shock is all the time above its equilibrial zero level, the largest fluctuation is again evident in 1997-98. Then it slowly converges to the zero level. It corresponds to the inflation shock development. Behaviour of the output gap shock is very similar to the interest rate shock behaviour, so it can signify lots of sources of demand shocks (because of the large openness of the economy) or very persistent demand shocks in the Czech economy. The exchange rate shock is all the time above the zero level and it converges to zero at the end of the period. It means that the exchange rate was undervaluated whole the time and it gradually appreciated.

5.3 Impulse responses

From the course of the impulse response functions (IRF) of both models can be seen that they are all in accordance with economic intuition and after a shock (demand, supply or moneatary shock) all of the model variables return back to their steady state values in few periods. All deviations are in the range of $\pm 0 - 3\%$ for the closed economy model, for the open economy model they are larger (maximally 8%). Periods of time are measured on the horizontal axis, deviation from equilibrium is measured on the vertical axis.

The three graphs on the left hand side of the Figure 3 belong to the closed economy model and represent the adjustment process of the economy after a demand shock. We can see that a unity demand shock induces rising production and inflation. The central bank reacts by rising the interest rate. All the variables rise next few periods and then return back to their equilibrium levels quite slowly because the coefficient μ is near to unity and it causes large inertia of demand shocks.

The three graphs on the right hand side of the Figure 3 belong to the closed economy model and represent the adjustment process of the economy after a supply shock. We can see that a unity supply shock induces rising inflation and falling production. The central bank reacts by rising the interest rate and the production falls again. All variables return back to they equilibrium levels quite quickly because the coefficient ρ is smaller than μ.

A unity monetary shock (not shown here) causes decline of production and inflation and all variables return back to their equilibrial levels in 16 periods.

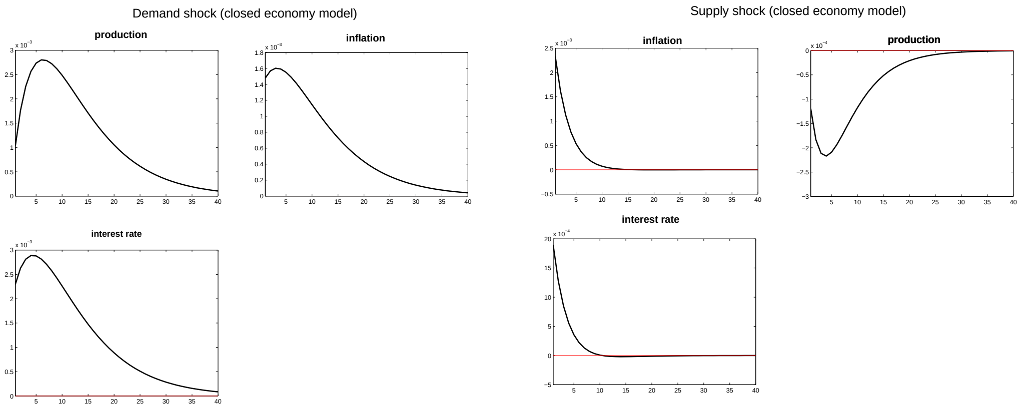


Figure 3: Impulse response function to a unit demand and supply shock – closed economy model

IRF of the open economy model are depicted in the Figure 4. The left upper four frames refer to a demand shock. Demand shock induces rising production and inflation. The central bank reacts by rising the interest rate. The exchange rate appreciate because the UIP condition claims expected depreciation in the next period. In the second period the central bank lowers the interest rate because inflation calms and it encourages production to return back to the equilibrium. All variables return back to they equilibrial levels quite slowly.

The right upper four frames refer to a supply shock. We can see that a unit supply shock induces rising inflation. The central bank reacts by rising the interest rate so the production falls again. The exchange rate appreciate because the UIP condition claims expected depreciation in the next period.

The lower four frames refer to a monetary shock. Monetary shock (i.e. increase of the interest rate) causes decline of production and inflation, the exchange rate appreciate because the UIP condition claims expected depreciation in the next period. Then all variables return back to their equilibrial levels.

The exchange rate shock (not shown here) only causes increase of the CPI inflation rate in the first period, in the next periods inflation falls and slowly returns to its equilibrium.

5.4 Forecast

Both model give quite reasonable predictions of the future development of the model variables. The principal results of both models are very similar. Figure 5 depicts forecast of production, inflation (domestic and total CPI) and interest rate of the open economy model. We can see that all the depicted variables should be around their steady state values in the future but their 95% confidence intervals are quite wide ($\pm 10\%$) so the predictions are quite uncertain and dependent on external influences. The advantage of the close economy model is that confidence intervals for predictions are narrower ($\pm 5\%$).

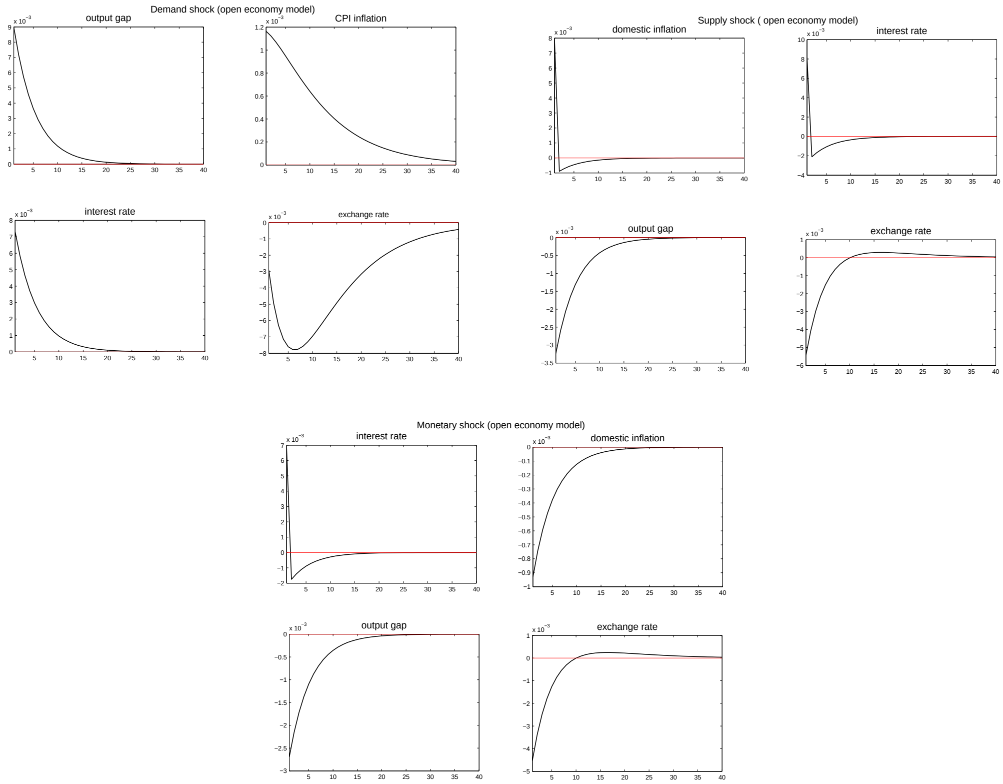


Figure 4: Impulse response function to a unit demand, supply and monetary shock – open economy model

6 Conclusion

The paper presents closed and open economy model estimation. The models were applied on the Czech economy. Bayesian estimation technique, concretely the Metropolis-Hastings algorithm, was used for estimation of the models.

The estimated models can be considered as a rough approximation of the behaviour of the real economic system. According to the estimated parameters, all of the causal relations are identified correctly. The closed economy model is very robust and predict impulse responses to shocks in accordance with economic intuition. The open economy model is more adequate for the description of the Czech economy but it is more complicated to estimate it and evaluate gained results. Both models incorporate rational expectations and a stable unique solution was found according to the Blanchard-Kahn condition. Properties of the models were analyzed by impulse responses to the demand, supply, monetary and exchange rate shocks and they seem to be in accordance with reality as well as forecasts of both models. Estimated central bank reaction function, i.e. Taylor rule also corresponds to reality.

The gained estimation can be accepted from the point of economic theory. The course of the distribution of the parameters was predominantly unimodal with considerable evidence of a mean value. On the other hand, some of the estimated parameters do not seem unimodal. This fact magnifies the significance of using the Bayesian estimation technique instead of classical methods like Kalman filter. The further reason for its utilization is unsufficient length of the available time series and dramatical changes in the economic development in transitive countries especially in years 1997-98.

References

[1] Brooks, S. P.; Gelman, A.: General Methods for Monitoring Converence of Iterative Simulations. Journal of Computational and Graphical Statistics, Volume 7, Number 4, pages 434-455; 1998.

[2] Clarida, R.; Galí, J.; Gertler, M.: The Science of Monetary Policy: A New Keynesian Perspective. NBER Working Paper 7147, National Bureau of Economic Research, May 1999.

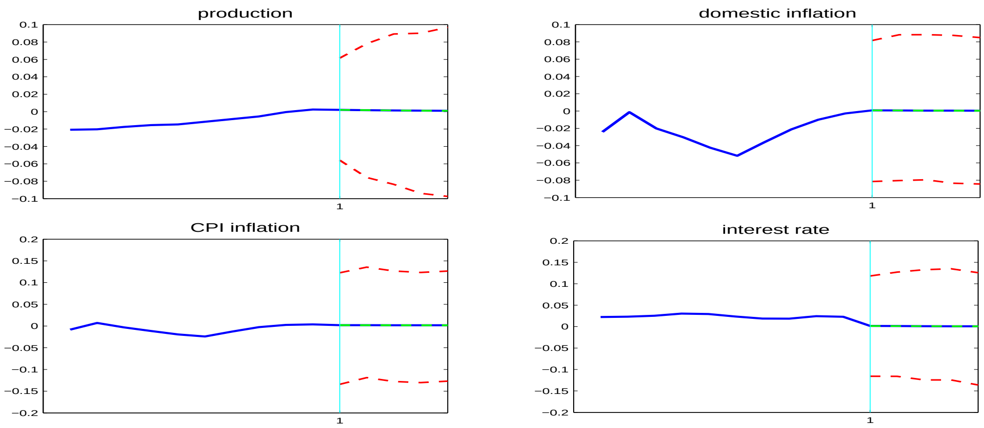


Figure 5: Model forecasts – open economy model

[3] Dennis, R.: Optimal Policy in Rational Expectations Models: New Solution Algorithms. NBER Working Paper 2001-09, National Bureau of Economic Research, September 2005.

[4] Galí, J.; Monacelli, T.: Monetary Policy and Exchange Rate Volatility in a Small Open Economy. Review of Economic Studies 72, pages 707-734, 2005.

[5] Ireland, P. N.: A Method for Taking Models do the Data. Boston College and NBER, February 2003.

[6] Lubik, T.; Shorfheide, F.: A Bayesian Look at New Open Economy Macroeconomics. NBER Macroeconomics Annual 2005, vol. 20, pages 313-366, MIT Press, 2005.

[7] Malley, J.: Lecture Notes on the Theory, Calibration and Estimation of Dynamic Stochastic General Equilibrium Models. University of Glasgow, November 30, 2004.

[8] Pytelová, H.; Vašíček, O.: Estimation of Alternative Monetary Policy Rules and their Comparison. Bulletin of the Czech Econometric Society, July 2004, Volume 11, Issue 20, pages 15-28. ISSN 1212 - 074x.

[9] Söderlind, P.: Lecture Notes for Monetary Policy (PhD course at UNISG). University of St. Gallen, Switzerland, 2003.

Volatility of prices in a multiple relation of composite commodities

Jiří Frank*, Martin Gavalec†
University of Hradec Králové, Czech Republic
jiri.frank@uhk.cz, martin.gavalec@uhk.cz

Abstract

Volatility of prices and traded volumes of composite commodities in a multiple relation is studied. Two tradable inputs are related in the complementarity form and, at the same time, each of these inputs is related in the input-output form to a common final product. Conditions for larger or smaller volatility of prices and traded volumes, in comparison to dually related goods, or to independent goods, have been formulated in the previous work [4]. The presented results are a continuation and an extension of the previous research. New results contain a more detailed analysis of the relations between volatilities of prices and traded volumes in case of multiple related, dually related or independent goods. The results are supported by numerical examples.

Keywords

Volatility of price, volatility of traded volume, related goods, composite commodities, substitution, complementarity, input/output relation

JEL: C30, C51, F19

1 Introduction

As indicated in our previous work, applications of composite-commodity theorem for explanations of international trade flows have been rare so far and have been typically focused (Chen and others [2, 3]) on examining individual possible different relations of pairs of commodities (specified there as substitutive or complementary, input-output connected or joint products, and some other relations). It was also stressed there that all these examinations had been performed separately, according to these individual mentioned types of related commodities. And, what was stressed especially, was that what had not been taken into consideration in these distinctions of relevant relations was a possible interpretation of related commodities as tradable factors of production, among which the dominance of complementarity exists ([5], p. 98).

*This work was supported by Czech Science Foundation # 402/04/0642.
†This work was supported by Czech Science Foundation # 402/06/1071.

So, our earlier work was a natural extension of possible applications of composite commodity theorem. In the mentioned earlier work, we have tested two above defined relations of traded commodities making together a first multiple relation and have created its relevant model in the framework of composite commodities. We have found there that the volatility of prices and traded volumes in the considered multirelational model is larger than in the model with one single relation. However, this observation was found holding true only when the coefficients of the model satisfied one simple condition, and without this condition the relation between models was ambiguous.

2 Relations between tradable commodities

In the previous work, it was demonstrated that there is a broader variety of possible relations among internationally traded commodities. It was also shown that individual traded commodities are not only very often related to each other but they can be observed and classified in this respect also with regard to their position or role in some relevant production function.

Two or more traded commodities can be mutually related as standard inputs (e.g. raw materials, components, intermediate products) for a specific final product in the form of substitution or in the form of complementarity. Some traded commodities, on the other hand, can be characterized in their supply as discrete goods due to their, let us say, specific role as specific inputs in the relevant production function (supplies of capital goods, relevant process technology).

In recent literature for example, effort has been developed to examine whether trade in capital goods is responsible for the observed volatility of both net exports and terms of trade of trading countries [1]. These examinations, of course, are performed using aggregate figures and commodities and are of quite different nature than those performed here. Our ambition is different - to show the above relations including those of capital goods and technologies on the trade flows on the level of relevant industries and business transactions. Capital goods, therefore, remain in this present work as a part of input-output relation in the framework of examined multiple relation model.

Finally, there is the relevant final product itself presented as a central traded commodity having a relationship to other commodities which are all possible inputs (standard or specific) in its relevant production function.

3 Input-output and complementarity multiple relation model

We consider a specific multiple relation between commodities, which has been studied in [4]. It is a combined relation consisting of two parallel input-output relations and one complementarity relation. A tradable commodity C_1 is a common output of two input commodities C_2, C_3, which are related to each other in a complementarity relation. The situation is described by the linear model

$$\mathbf{Q} = \mathbf{a}_0 + \mathbf{AP} + \mathbf{u}$$
$$\mathbf{S} = \mathbf{b}_0 + \mathbf{BP} + \mathbf{v}$$

with vector variables: $\mathbf{Q}$ denoting the demands, $\mathbf{S}$ denoting the supplies and $\mathbf{P}$ denoting the prices of commodities C_1, C_2, C_3.

$$\mathbf{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

Vector variables $\mathbf{u}$ ($\mathbf{v}$) are random disturbances in the demand (supply) called non-systematic risks. All risk variables are stochastically independent of each other, and their expected values are equal to zero, i.e. $E(u_i) = E(v_i) = 0$ for $i = 1, 2, 3$. The coefficient matrices $\mathbf{A}, \mathbf{B}$ in the model are of the form

$$\mathbf{A} = \begin{bmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & -a_{23} \\ a_{31} & -a_{32} & -a_{33} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & -b_{12} & -b_{13} \\ b_{21} & b_{22} & 0 \\ b_{31} & 0 & b_{33} \end{bmatrix}$$

The values of all coefficients in the model are assumed to be positive. The negative influence of coefficients a_{11}, a_{22}, a_{33} in the model is shown by the minus sign. Similar convention is used for coefficients a_{23}, a_{32}, expressing the complementarity relation between C_2 and C_3 and coefficients b_{12}, b_{13} expressing the parallel input-output relations.

The equilibrium state of the model is characterized by the equality of demand and supply, i.e. by the equation $\mathbf{Q} = \mathbf{S}$, which implies

$$\mathbf{a}_0 + \mathbf{AP} + \mathbf{u} = \mathbf{b}_0 + \mathbf{BP} + \mathbf{v}$$

Hence, the equilibrium prices $\mathbf{P}^\star$ fulfill the matrix equation

$$\mathbf{MP}^\star = \mathbf{c}_0 + \mathbf{w}$$

with $\mathbf{M} := \mathbf{A} - \mathbf{B}$, $\mathbf{c}_0 := \mathbf{b}_0 - \mathbf{a}_0$ and $\mathbf{w} := \mathbf{v} - \mathbf{u}$. We may assume that the equilibriuim prices are uniquely determined. Then the matrix $\mathbf{M}$ is regular and the prices $\mathbf{P}^\star$ are described by formula

$$\mathbf{P}^\star = \mathbf{M}^{-1}(\mathbf{c}_0 + \mathbf{w}) \tag{1}$$

The formula (1) was computed in more detail in [4], by the well-known formula for computation of inverse matrices using algebraic minors of $\mathbf{M}$. In view of this computation, the equilibrium prices can be expressed by putting the coefficients from $\mathbf{M}$ and $\mathbf{c}_0$ into formulas for $P_1^\star$, $P_2^\star$, $P_3^\star$ below.

$$\mathbf{M} = \begin{bmatrix} -a_{11} - b_{11} & b_{12} & b_{13} \\ a_{21} - b_{21} & -a_{22} - b_{22} & -a_{23} \\ a_{31} - b_{31} & -a_{32} & -a_{33} - b_{33} \end{bmatrix} \qquad \mathbf{c}_0 = \begin{bmatrix} b_{10} - a_{10} \\ b_{20} - a_{20} \\ b_{30} - a_{30} \end{bmatrix}$$

$$P_1^\star = |\mathbf{M}|^{-1}\Big((m_{22}m_{33} - m_{23}m_{32})(c_{10} + w_1) + (m_{13}m_{32} - m_{12}m_{33})(c_{20} + w_2) + \\ +(m_{12}m_{23} - m_{13}m_{22})(c_{30} + w_3)\Big)$$

$$P_2^\star = |\mathbf{M}|^{-1}\Big((m_{23}m_{31} - m_{21}m_{33})(c_{10} + w_1) + (m_{11}m_{33} - m_{13}m_{31})(c_{20} + w_2) + \\ +(m_{13}m_{21} - m_{11}m_{23})(c_{30} + w_3)\Big)$$

$$P_3^\star = |\mathbf{M}|^{-1}\Big((m_{21}m_{32} - m_{22}m_{31})(c_{10} + w_1) + (m_{12}m_{31} - m_{11}m_{32})(c_{20} + w_2)+ \\ +(m_{11}m_{22} - m_{12}m_{21})(c_{30} + w_3)\Big)$$

The vector of equilibrium trade volumes $\mathbf{Q}^\star$ can be computed by putting the above values of $\mathbf{P}^\star$ into the formula $\mathbf{Q}^\star = \mathbf{a}_0 + \mathbf{A}\mathbf{P}^\star + \mathbf{u}$. Hence, we get

$$\mathbf{Q}^\star = \mathbf{a}_0 + \mathbf{A}\mathbf{M}^{-1}(\mathbf{c}_0 + \mathbf{w}) + \mathbf{u} \tag{2}$$

and the detailed formulas for $Q_1^\star$, $Q_2^\star$, $Q_3^\star$ are computed analogously as formulas for $P_1^\star$, $P_2^\star$, $P_3^\star$.

4 Comparison of variances in a multiple relation model

J.R. Chen in [2] considered binary relation between tradable commodities, and compared the variance of theri prices and trade volumes, with the variances in the case when the commodities were independent. His conclusion was that the variance in the case of related commodities is larger that the variance in the independent case. We shall reffer to this inequality as to Chen's inequality.

In [4] a generalization of Chen's inequality for multiple related commodities was studied. A sufficient condition was foumd under which the generalized Chen's inequality is fulfilled in the input-output and complementarity multiple relation model described in section 3. The main goal of this paper is to discuss the question whether the sufficient condition described in [4] is also necessary.

We shall compare the variances of equilibrium prices $\mathbf{P}^\star$ and equlibrium trade volumes $\mathbf{Q}^\star$ in the multiple relation model from the previous section, with the variances of equilibrium prices $\mathbf{P}^{\star\star}$ and equlibrium trade volumes $\mathbf{Q}^{\star\star}$ in the binary relation model with independent commodity C_1 and with the complementarity relation between commodities C_2, C_3. The binary relation model is described by the vector equations

$$\begin{aligned} \mathbf{Q} &= \mathbf{a}_0 + \widetilde{\mathbf{A}}\mathbf{P} + \mathbf{u} \\ \mathbf{S} &= \mathbf{b}_0 + \widetilde{\mathbf{B}}\mathbf{P} + \mathbf{v} \end{aligned}$$

with coefficient matrices

$$\widetilde{\mathbf{A}} = \begin{bmatrix} -a_{11} & 0 & 0 \\ 0 & -a_{22} & -a_{23} \\ 0 & -a_{32} & -a_{33} \end{bmatrix} \qquad \widetilde{\mathbf{B}} = \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix}$$

The equilibrium values $\mathbf{P}^{\star\star}$ and $\mathbf{Q}^{\star\star}$ are computed analogously as in section 3. Denoting

$$\widetilde{\mathbf{M}} = \begin{bmatrix} -a_{11} - b_{11} & 0 & 0 \\ 0 & -a_{22} - b_{22} & -a_{23} \\ 0 & -a_{32} & -a_{33} - b_{33} \end{bmatrix}$$

we get

$$\begin{aligned} \mathbf{P}^{\star\star} &= \widetilde{\mathbf{M}}^{-1}(\mathbf{c}_0 + \mathbf{w}) \\ \mathbf{Q}^{\star\star} &= \mathbf{a}_0 + \widetilde{\mathbf{A}}\widetilde{\mathbf{M}}^{-1}(\mathbf{c}_0 + \mathbf{w}) + \mathbf{u} \end{aligned}$$

Hence, we have for $P_1^\star$ and $P_1^{\star\star}$

$$\begin{aligned}\operatorname{Var}(P_1^\star) &= |\mathbf{M}|^{-2}\big(|M_{11}|^2\operatorname{Var}(w_1)+|M_{21}|^2\operatorname{Var}(w_2)+|M_{31}|^2\operatorname{Var}(w_3)\big)\\ \operatorname{Var}(P_1^{\star\star}) &= |\widetilde{\mathbf{M}}|^{-2}\big(|M_{11}|^2\operatorname{Var}(w_1)\big)\end{aligned}$$

and for $Q_1^\star$ and $Q_1^{\star\star}$

$$\begin{aligned}\operatorname{Var}(Q_1^\star) &= |\mathbf{M}|^{-2}\Big(a_{11}^2\big(|M_{11}|^2\operatorname{Var}(w_1)+|M_{21}|^2\operatorname{Var}(w_2)+|M_{31}|^2\operatorname{Var}(w_3)\big)+\\ &\qquad +\operatorname{Var}(u_1)\Big)\\ \operatorname{Var}(Q_1^{\star\star}) &= |\widetilde{\mathbf{M}}|^{-2}\Big(a_{11}^2\big(|M_{11}|^2\operatorname{Var}(w_1)\big)+\operatorname{Var}(u_1)\Big)\end{aligned}$$

Our aim is to compare the above variances. As the matrices $\mathbf{M}$, $\widetilde{\mathbf{M}}$ are regular, we have $|\mathbf{M}| \neq 0$ and $|\widetilde{\mathbf{M}}| \neq 0$. First we describe a necessary and sufficient condition under which the inequality

$$|\widetilde{\mathbf{M}}|^{-2} \leq |\mathbf{M}|^{-2} \tag{3}$$

holds true. The inequality (3) is fulfilled, if one of the inequalities

$$|\widetilde{\mathbf{M}}| \geq |\mathbf{M}| > 0 \tag{4}$$

$$|\widetilde{\mathbf{M}}| \leq |\mathbf{M}| < 0 \tag{5}$$

holds true. Using the notation

$$\begin{aligned}|\widetilde{\mathbf{M}}| &= -(a_{11}+b_{11})\,|M_{11}| = X\\ |\mathbf{M}| &= -(a_{11}+b_{11})\,|M_{11}| - b_{12}\,|M_{12}| + b_{13}\,|M_{13}| = X - Y\end{aligned}$$

it is easy to see that inequality (4) is fulfilled if

$$X > 0 \text{ and } Y \geq 0 \tag{6}$$

and (5) is fulfilled if

$$X < 0 \text{ and } Y \leq 0 \tag{7}$$

In other words, the inequality (3) holds true if and only if one of the conditions (6) and (7) is fulfilled.

If the condition (3) is satisfied, then the generalized Chen's inequalities

$$\operatorname{Var}(P_1^\star) > \operatorname{Var}(P_1^{\star\star}), \quad \operatorname{Var}(Q_1^\star) > \operatorname{Var}(Q_1^{\star\star}) \tag{8}$$

hold true. This means that if one of the conditions (6) and (7) is satisfied, then the volatility of the price $P_1^\star$ in the multiple-related model is larger than the volatility of the price $P_1^{\star\star}$ in the model without the input-output relation. This is an extension of the result presented in [4], where only condition (6) was considered.

However, if none of the conditions (6) and (7) holds true, then the relations between prices $\operatorname{Var}(P^\star)$ and $\operatorname{Var}(P^{\star\star})$, as well as the relations between the traded volumes $Q_1^\star$ and $Q_1^{\star\star}$, are ambiguous. We discuss this question below in more detail. We show in the next section that the condition (3) need not be fulfilled, but the generalized Chen's inequalities (8) still can hold or not hold, in dependence on the coefficients of the model.

Similar conclusions can be stated about the prices and the traded volumes of the commodities C_2, C_3.

5 Numerical examples

It is a question of interest, whether the condition (3) must be always true. If the condition (3) should fail, then one of the conditions (6), (7) would not hold. Hence, we would have either

$$X < 0 \text{ and } Y \geq 0 \tag{9}$$

or,

$$X > 0 \text{ and } Y \leq 0 \tag{10}$$

We show on numerical examples that such a situation can occur, in dependence on the values of vectors and matrices

$$\mathbf{a}_0 = \begin{bmatrix} a_{01} \\ a_{02} \\ a_{03} \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & -a_{23} \\ a_{31} & -a_{32} & -a_{33} \end{bmatrix}$$

$$\mathbf{b}_0 = \begin{bmatrix} b_{01} \\ b_{02} \\ b_{03} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} b_{11} & -b_{12} & -b_{13} \\ b_{21} & b_{22} & 0 \\ b_{31} & 0 & b_{33} \end{bmatrix}$$

The inequalities (9) are valid e.g. for values

$$\mathbf{a}_0 = \begin{bmatrix} 13000 \\ 14000 \\ 15000 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 5 & 0 & 0 \\ 6 & 17 & 40 \\ 8 & 67 & 10 \end{bmatrix} \tag{11}$$

$$\mathbf{b}_0 = \begin{bmatrix} 11500 \\ 7500 \\ 8500 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 25 & 83 & 2 \\ 15 & 6 & 0 \\ 11 & 0 & 27 \end{bmatrix} \tag{12}$$

The inequalities (10) are valid e.g. for values

$$\mathbf{a}_0 = \begin{bmatrix} 13000 \\ 14000 \\ 15000 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 5 & 0 & 0 \\ 14 & 15 & 40 \\ 8 & 30 & 2 \end{bmatrix} \tag{13}$$

$$\mathbf{b}_0 = \begin{bmatrix} 11500 \\ 7500 \\ 8500 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 25 & 18 & 21 \\ 15 & 17 & 0 \\ 11 & 0 & 40 \end{bmatrix} \tag{14}$$

This does not mean that the generalized Chen's inequalities (8) do not hold true for the above values. The detailed formulas for the variances of equilibrium prices $P^\star, P^{\star\star}$ and trade volumes $Q^\star, Q^{\star\star}$ show that the validity of (8) depends also on values of $\mathrm{Var}(w_1)$, $\mathrm{Var}(w_2)$ and $\mathrm{Var}(w_3)$. This is demonstrated on the numerical examples below. For the sake of brevity, we assume $\mathrm{Var}(w_2) = \mathrm{Var}(w_3)$.

If the coefficients of the model take values (11),(12), then the inequalities (8) are not satisfied for sufficiently large values of $\mathrm{Var}(w_2) = \mathrm{Var}(w_3)$, in comparison with $\mathrm{Var}(w_1)$. The minimal such values are shown in Table 1 below.

Var(w_1)	Var(w_2) = Var(w_3)
1	10
2	21
3	33
4	43
5	54
6	65
7	75
8	86
9	97
10	107

Table 1

On the other hand, if the coefficients of the model take values (13),(14), then the inequalities (8) are not satisfied for sufficiently large values of Var(w_1), in comparison with Var(w_2) = Var(w_3). The minimal such values are given in Table 2.

Var(w_1)	Var(w_2) = Var(w_3)
61	1
121	2
181	3
241	4
301	5
361	6
421	7
481	8
541	9
601	10

Table 2

6 Conclusions

The results presented in this contribution are a continuation of the work started in [4]. The volatilities of prices and traded volumes in the considered multirelational model are shown to be larger than the corresponding volatilities in the model with one single relation, if one specific assumption concerning the coefficients of the model is satisfied (inequality (3)). A necessary and sufficient condition for this assumption is presented. It is also shown on numerical examples that if the above specific assumption does not hold true, then the inequality between volatilities can be oriented in both directions, in dependence on the values of non-systematic risks.

References

[1] M. Boileau, Trade in capital goods and the volatility of net exports and the terms of trade, J. International Economics **48** (1999), 347-365.

[2] J.R. Chen, Modelling of Related Commodities and Its Implications for Stabilization Policy of the World Market, J. Agricultural Economics **55** (1994), 139-166.

[3] J.R. Chen, H. Stocker, Composite Commodities and the Prebisch-Singer Thesis, World Development **25** (1997), 1863-1871.

[4] J. Frank, M. Gavalec, Multiple relations of composite commodities, Proc. of the 23rd Conf. Mathem. Methods in Economics, University of Hradec Králové, Hradec Králové (2005), 104-110.

[5] J.R. Hicks, Value and Capital, Oxford University Press, Oxford, 1st ed. 1939, reprinted 1941.

Using the Weibull Distribution for Simulation of Machine Lifetime

Jana Friebelová, Ludvík Friebel

ZF JU České Budějovice
friebelova@zf.jcu.cz

Abstract

This article deals with the use of Weibull distribution when modelling failure-free time of machine run. All stages of machine lifetime can be described by Weibull distribution with different continuous scale parameters. This distribution can be used also in case of economic evaluation of machine replacement or maintenance. Simulations based on this distribution can help companies to find an optimal interval of machine maintenance in order to minimize costs. One of the simulation techniques was applied on a concrete machine and implemented in SW @RISK.

Keywords

Maintenance interval, Exponential distribution, Weibull distribution, failure-free run, replacement, economic evaluation, simulation, regression analysis, correlation analysis.

JEL: C13, C15

Introduction

Weibull distribution is widely used for describing lifetime of machines components and other devices, whose lifetime do not fit an exponential distribution. This distribution is typically used in case of material wastage and fatigue.

Weibull distribution can describe various stages of machine lifetime by changing its continuous scale parameters. This distribution can be successfully used in case of economic evaluation of machine replacement or maintenance. Simulations based on this distribution can help companies to find an optimal interval of machine maintenance in order to minimize costs.

The Weibull distribution

Density function of Weibull distribution W(δ, c) can be written as

$$f(x) = \frac{cx^{c-1}}{\delta^c} \exp\left[-\left(\frac{x}{\delta}\right)^c\right], x > 0, c > 0, \delta > 0, \tag{1}$$

$$f(x) = 0, x \leq 0.$$

Where c is scale parameter and β is shape parameter, respectively.

Cumulative density function is defined:

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\delta}\right)^c\right], x > 0, \tag{2}$$

$$F(x) = 0, x \leq 0.$$

Special case of this distribution is Exponential distribution E($0,\delta$) for c=1 and Rayleigh distribution Ra(δ) for c=2, for c=2.5, Weibull distribution approximates lognormal distribution, for c=3.6, Weibull distribution approximates normal distribution. Thanks to its flexibility, many empirically obtained characteristics can be modeled by Weibull distribution with success.

Quantiles of Weibull distribution are defined as

$$x_p = \delta[-\ln(1-P)]^{\frac{1}{c}} . \tag{3}$$

Moments:

$$E(X) = \Gamma\left(\frac{1}{c}+1\right)\delta , \tag{4}$$

$$D(X) = \left[\Gamma\left(\frac{2}{c}+1\right) - \Gamma^2\left(\frac{1}{c}+1\right)\right]\delta^2 . \tag{5}$$

Estimation of parameters of Weibull distribution

One method of parameter estimation is based on moments. Estimations of parameters can be obtained by solution of given system (4) and (5) [3]. We can compute $\hat{c}$ from equation (6) given from (4) and (5).

$$\frac{s^2}{\bar{x}^2} = \frac{\Gamma\left(\frac{2}{\hat{c}}+1\right)}{\Gamma^2\left(\frac{1}{\hat{c}}+1\right)} - 1 \tag{6}$$

After finding $\hat{c}$ it is possible to compute parameter $\hat{\delta}$:

$$\hat{\delta} = \frac{\bar{x}}{\Gamma\left(\frac{1}{\hat{c}}+1\right)} . \tag{7}$$

Another way to obtain the estimations of parameters is the Maximum Likelihood Estimate (MLE). For complete random sample (x_1, x_2, …, x_n) from the Weibull distribution described by (1), the likelihood equations can be written in the following form [4]:

$$\frac{\sum_{i=1}^{n} x_i^{\hat{c}} \ln x_i}{\sum_{i=1}^{n} x_i^{\hat{c}}} - \frac{1}{\hat{c}} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 0 . \tag{8}$$

After solving equation (8) for parameter c estimation, estimation for parameter δ is automatically obtained:

$$\hat{\delta} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\hat{c}}\right)^{\frac{1}{\hat{c}}} . \tag{9}$$

Estimation of parameters can be obtained with the least square method using Weibull probability plot. This plot is based on transformation of cumulative density function to a straight line.
The last and the least mathematically intensive method for parameter estimation is the graphical method based on Weibull probability plot too, see [2].

Example

Many companies have to decide whether to replace a bearing in the machine before failure and break of this machine. The planned replacement of the ball bearing requires less downtime than an unplanned replacement. The costs will be lower regarding the planned replacement than an unplanned replacement, but the planned replacement interval must not be too frequent. The aim of our paper is to find the optimal time period of replacement by the simulation.
We need to know an appropriate distribution to model lifetime of a ball bearing, in order to use @RISK to ensure optimal time period of replacement. The lifetime of ball bearing follows Weibull distribution [5].
We have sample values of lifetime of concrete bearing. The empirical distribution of sampled values is depicted in figure 1. We can compute the mean value and variance:

$$\bar{x} = 629.795, s = 477.489 .$$

Let's suppose that every hour of downtime costs the company 500 CZK. In case of the planed replacement, the price of the bearing with appropriate material is 2 000 CZK. If we make the planned replacement, the machine will be down for 2 hours.
In case of enforced and unplanned replacement accompanied by serious damage of machine, the price of the bearing with appropriate material is 8 000 CZK. If we have to make unplanned replacement the machine will be

down for 14 hours, because we have to adjust and control machine. For comparison of the costs of maintenance, see table 1.

Table 1: Cost of maintenance

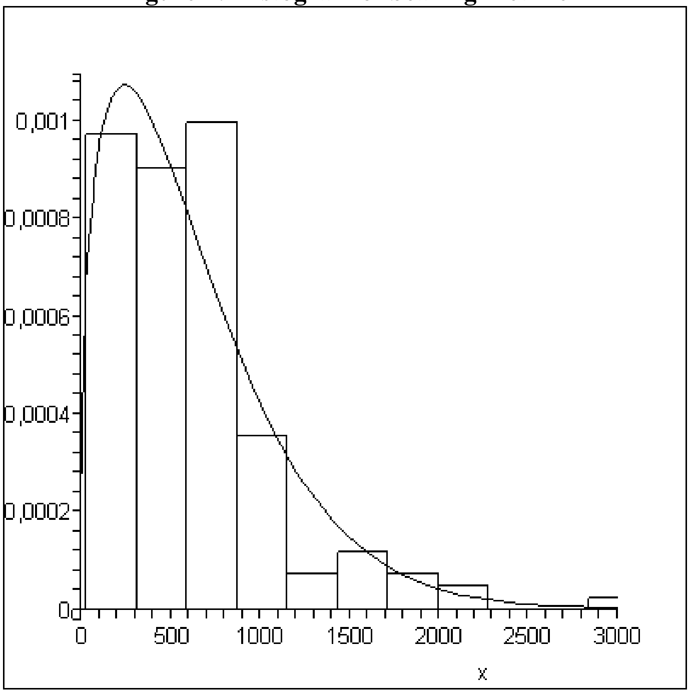
Maintenance form	Duration of downtime (hour)	Downtime cost (CZK/hour)	Material (CZK)	Total costs
Planned	2	500	2000	3000
Unplanned	14	500	8000	15000

There we have to decide how often we shall accomplish changing of bearing in order to minimize total maintenance costs. For finding the optimal maintenance period, we will use simulation. We will run the simulation of bearing lifetime with the program @RISK. The lifetime of bearing follows the Weibull distribution.
Parameters of Weibull distribution is computed according to equations (6) and (7) with program Maple:

$$\hat{c} = 1.332,\ \hat{\delta} = 685.137$$

Probability density function is depicted in figure 1.

Figure 1: Histogram of bearing lifetime



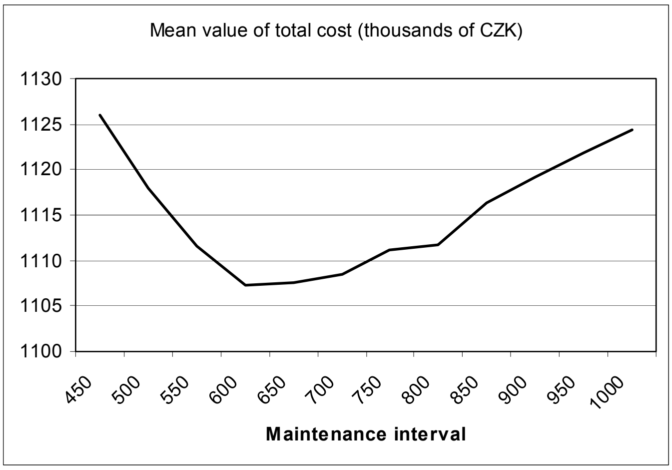
Now, we will simulate operation of the ball bearing in machine for substantial length of time, e.g. 50.000 hours.
For each maintenance period, we will simulate lifetime of bearing by sampling type Monte Carlo.
We run 300 iterations of twelve simulations (one simulation for one maintenance period) to determine which planned replacement interval minimizes total cost for 50 000 hours.

Table 2: Mean values of maintenance cost

Maintenance interval (hour)	Total costs	Maintenance interval (hour)	Total costs
450	1126,04	750	1111,13
500	1117,94	800	1111,77
550	1111,57	850	1116,35
600	1107,3	900	1119,22
650	1107,64	950	1121,87
700	1108,45	1000	1124,42

The optimal maintenance interval is 600 hours with the lowest total cost 1 107 300 CZK.

Figure 2: Total costs dependency on maintenance interval



Conclusions

Regarding the lifetime sample, we have estimated parameters of two-parameter Weibull distribution with the method based on moments. This method is mathematically intensive and can be successfully applied only with a special mathematical SW. The obtained distribution was used as a main input for lifetime simulation of a concrete device. A separate simulation, including 300 iterations, ran for each considered period. Those simulations were implemented in program @RISK. On the base of simulations outputs, we could determine the optimal maintenance period in order to minimize total costs.

References

[1] BARLOW, R.: Mathematical Theory of Reliability. John Wiley and Sons, 1967.
[2] KECECIOGLU, D.: Reliability & Life Testing Handbook. Prentice Hall, Inc., Englewood Cliffs, New Jersey, Vol. 1 and 2, 1993 and 1994.
[3] LAWLESS, J.,F.: Statistical models and methods for lifetime data. New York, Wiley,1982.
[4] LEHMAN, E., L.:Theory of point estimation. New York, Wiley,1983.
[5] LIEBLEIN, J., ZELEN, M: Statistical investigation of the fatigue life of deep-groove ball bearings. Journal of Research of the National Bureau of Standards, Vol. 57, 1956.
[6] MEEKER, W. Q., ESCOBAR, L. A.: Statistical Methods for Reliability Data. New York: John Wiley and Sons, Inc., 1998.
[7] http://www.weibull.com

Cognitive hierarchy process - an approach to decision making support*

Martin Gavalec, Karel Mls
University of Hradec Králové, Czech Republic
martin.gavalec@uhk.cz, karel.mls@uhk.cz

Abstract

A new method for computer support of decision processes in economics and management is described. The method uses cognitive maps, and is based on structuring the decision problem into a set of criteria, and on computing preferences for alternative solutions. The highest preference then determines the final decision. The suggested CHP method is similar to the previously known AHP method, but is more flexible and can be also used in situations with individual criteria evaluation for every alternative.

Keywords

Decision making support, fuzzy cognitive map, analytical hierarchy process, cognitive hierarchy process, evaluation of alternatives.

JEL: C51, C52, C63

1 Introduction

Making decisions belongs to the most important activities in economics and management. In this paper, a new method for computer support of decision processes is presented. The main tool used in the method are fuzzy cognitive maps. The basic idea of the suggested cognitive hierarchy process (CHP) method is similar to the well-known AHP method, however, CHP can also be used in situations with individual evaluation for the set of criteria in every alternative.

2 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

*This work was supported by Czech Science Foundation, # 402/06/1071

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. This binary representation of causality in CM makes difficulties in application to real problems. The representation of various causality strength is problematic, and also the application of several contradictory relationships to concrete concept leads to so called imbalanced CM. Thus the practical usage of CM in decision making process is very limited, without introducing some method for concept evaluation.

The notion of a fuzzy cognitive maps (FCM) was proposed by Kosko [9, 10] and later enriched by several authors [8, 13, 14, 16, 17]. Using fuzzy logic helps to solve the above mentioned problem 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 prediction as well, represent the most often cited domains of FCMs utilization, see [6, 7, 12, 15, 19].

3 Decision making support

Number of methods exist for formal support of decision processes. They are usually based on structuring the decision problem into smaller parts (alternatives, criteria, goals), and then on objective or subjective evaluations of the importance of various objectives and preferences for alternative solutions leading to the final decision.

One of the decision analysis techniques used for the formal support of the decision making process is the Analytical Hierarchy Process (AHP) method, invented by T. L. Saaty [18]. In its original form the technique is based on the Full Pairwise Comparison Method and relative normalization. Matrix of relative pair comparisons is created and used to determine the relative weights (relative importance) of every criterion to its subcriteria.

Full Pairwise Comparison Method involves comparing every element of a rating set to every other element of that set. Full pairwise comparisons are the most extensive and precise of the rating ranges. Normalization allows to handle different weight scales on an equal footing. All scales are converted to a common scale that takes a value between 0 and 1. Under the AHP methodology, a relative normalization approach is used where the weight of one subcriterion with respect to a given parent criterion is divided by the sum of the weights of all that parent criterion's subcriteria. For more about AHP see [2, 5], integration of multi-objective decision support and fuzzy logic is described in [11].

4 Cognitive hierarchy process

The present paper is a continuation of our previous work [3], where fuzzy cognitive maps were considered as a useful tool supporting the decision making process. In this section we describe a new decision support method emerging AHP and our evaluation method which was used for evaluation of dynamic changes in fuzzy cognitive maps [4]. The new method is called the Cognitive Hierarchy Process (CHP), and while the AHP method works with fixed evaluation of relative weights of different alternatives (cases), in the proposed CHP method every alternative generates its own specific system of weights.

We shall work with series of individual FCM's unified by a common template. We shall 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 $c \in C^*$, including the root node c_0, the notation $S(c)$ denotes the set of all successors of the node c.

For every individual FCM $M = M_s$ belonging to the system M^*, the evaluation mapping $e : C \to \langle 0, 1\rangle$ is defined recursively. If $S(c) = \emptyset$, i.e. if c is a leaf in the template tree M^*, then we put $e(c) = a_{p(c)c}$. If $S(c) \neq \emptyset$, then we denote

$$e^S(c) = \frac{1}{|S(c)|} \sum_{k \in S(c)} e(k)$$

and we put

$$e(c) = \begin{cases} e^S(c) & \text{for } c = c_0 \\ \sqrt{a_{p(c)c} \cdot e^S(c)} & \text{for } c \neq c_0 \end{cases}$$

5 Application

To illustrate basic ideas and properties of the proposed CHP method, some real world decisional situation will be presented. Let's suppose that there is a travel agency working on the new catalogue of destinations and that managers are looking for the best way to present the portfolio of destinations to different segments of customers. Data concerning any particular hotel, activity and route are well known, so their position in the decisional model should be considered as constant.

On the other hand, customer priorities can vary significantly and so they have strong impact to manager decisions.

In the presented case, ten alternatives $A_1, \ldots, A_{10}$ and five customer groups $G_1, \ldots, G_5$ are considered and analyzed with the help of the cognitive map of a common traveller. Characteristic properties of considered alternatives are shown in Table 1, and expectancy preferences of defined groups of possible customers are displayed in Table 2. Particular groups can be described as ‘culture focused, interested in history’ (G_1), ‘sportsmanlike looking for concrete activities at the concrete place’ (G_2), ‘adventurer without limits’ (G_3), ‘typical family with two little children’ (G_4), and ‘indecisive without any concrete idea’ (G_5).

The cognitive map of a common traveller can be described by enumerating his/her criteria and priorities. We consider the following criteria:

C_1 - place of the stay
C_2 - type of the stay
C_3 - activities
C_4 - stay in a city
C_5 - stay in a country
C_6 - stay by a sea
C_7 - stay in mountains
C_8 - stay on a fixed place
C_9 - travelling
C_{10} - combined stay
C_{11} - adventures
C_{12} - sport activities
C_{13} - culture events
C_{14} - historical monuments

	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}
A_1	1				1					1	1
A_2	1		1			1			1	1	
A_3		1		1		1		1	1		
A_4			1			1		1	1		
A_5	1				1						1
A_6		1		1			1	1	1		
A_7			1	1			1	1	1		
A_8	1	1					1		1	1	1
A_9				1	1				1		
A_{10}			1		1				1	1	

Table 1: Quality of particular alternatives

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}
G_1	1	0.5	0.8	1	0	0	0	1	0	0.2	0	0	0.3	0.8
G_2	1	0.5	1	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.9	0.4	0.1
G_3	0.2	0.6	1	0.1	0.2	0.4	0.6	0.2	0.8	0.3	1	0.8	0.1	0.1
G_4	1	1	0.5	0.2	0.3	0.8	0.4	0.8	0.3	0.5	0.1	0.4	0.3	0.2
G_5	0.3	0.3	0.3	0.5	0.5	0.3	0.2	0.8	0.2	0.4	0.2	0.3	0.4	0.4

Table 2: Subjective priorities of typical customers

The structure of the template cognitive map for the modeled example can be seen on Figure 1.

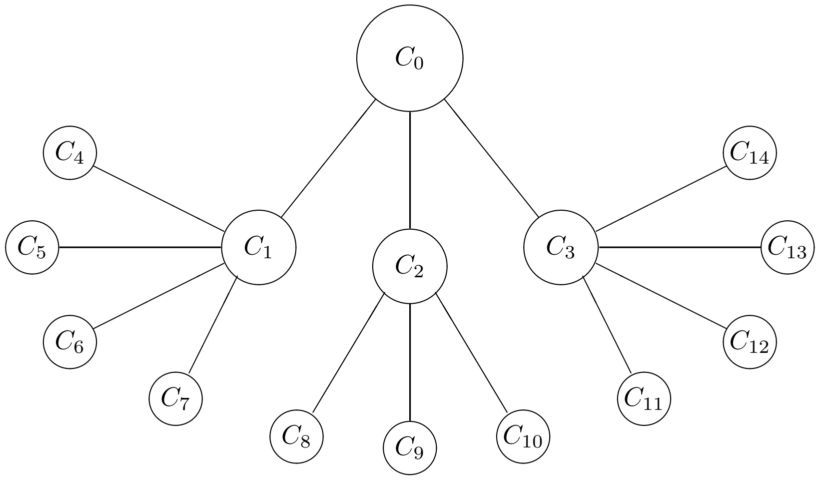


Figure 1: Structure of a template cognitive map

6 Results

The CHP method transforms the input values from tables 1 and 2 into evaluation vectors of individual cognitive maps. The core parts of these vectors consisting of values C_0, C_1, C_2, C_3 are shown in Table 3.

		A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
G_1	C_0	0.48	0.28	0	0	0.44	0.09	0.09	0.44	0.14	0.25
	C_1	0.50	0.50	0	0	0.50	0	0	0.50	0	0
	C_2	0.41	0	0	0	0.41	0.27	0.27	0.27	0.41	0.41
	C_3	0.54	0.33	0	0	0.42	0	0	0.54	0	0.33
G_2	C_0	0.38	0.52	0.52	0.46	0.31	0.51	0.51	0.52	0.40	0.44
	C_1	0.37	0.56	0.59	0.42	0.37	0.59	0.61	0.54	0.44	0.42
	C_2	0.27	0.36	0.36	0.36	0.27	0.32	0.32	0.32	0.27	0.27
	C_3	0.49	0.63	0.61	0.61	0.28	0.61	0.61	0.69	0.49	0.63
G_3	C_0	0.27	0.40	0.45	0.43	0.24	0.42	0.43	0.38	0.32	0.34
	C_1	0.13	0.22	0.25	0.18	0.13	0.25	0.27	0.20	0.20	0.18
	C_2	0.30	0.42	0.42	0.42	0.30	0.33	0.33	0.33	0.30	0.30
	C_3	0.40	0.55	0.69	0.69	0.28	0.69	0.69	0.62	0.47	0.55
G_4	C_0	0.41	0.46	0.44	0.41	0.37	0.46	0.48	0.48	0.41	0.47
	C_1	0.33	0.58	0.54	0.47	0.33	0.54	0.62	0.50	0.40	0.47
	C_2	0.55	0.43	0.43	0.43	0.55	0.49	0.49	0.49	0.55	0.55
	C_3	0.35	0.38	0.34	0.34	0.24	0.34	0.34	0.45	0.28	0.38
G_5	C_0	0.28	0.27	0.26	0.23	0.25	0.27	0.27	0.32	0.23	0.27
	C_1	0.23	0.31	0.29	0.20	0.23	0.29	0.27	0.33	0.18	0.20
	C_2	0.30	0.21	0.21	0.21	0.30	0.25	0.25	0.25	0.30	0.30
	C_3	0.31	0.30	0.27	0.27	0.22	0.27	0.27	0.37	0.20	0.30

Table 3: Evaluation of alternatives with individual preferences of customers

7 Conclusions

A new evaluation method CHP is suggested using individual fuzzy cognitive maps corresponding to given templates. The method enables to consider individual approach to evalution of criteria of the decision model. As a practical application, a case of a travel agency is studied, showing the possibility of involving the customer priorities into the managerial decision process.

References

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

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

[3] M. Gavalec, K. Mls, *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.

[4] M. Gavalec, K. Mls, *Processing of dynamic changes in linguistic fuzzy cognitive maps*, Abstracts of the 8th Inter. Conf. Fuzzy Sets Theory and its Applications, Liptovský Ján (2006), p. 47.

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

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

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

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

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

[10] B. Kosko, *Fuzzy engineering*, 1997, Upper Saddle River: Prentice-Hall. 549.

[11] M. Krapka, M. Zima, *NEFRIT - nástroj pro vícekriteriální hodnocení s použitím fuzzy*, in Inteligentní systémy pro praxi - Fuzzy logika, neuronové sítě, pravidlové expertní systémy. 2000. Luhačovice: AD&M Ostrava - Poruba.

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

[13] Y. Miao et al., *Dynamical Cognitive Network - an Extension of Fuzzy Cognitive Map*, IEEE Transactions on Fuzzy Systems, 2001. 9(5): p. 760-770.

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

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

[16] A.K. Tsadiras, K.G. Margaritis, *Introducing memory and decay factors in fuzzy cognitive maps*, in ISFL 95. 1995. Zurich, Switzerland.

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

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

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

STUR tests and their sensitivity for non-linear transformations and GARCH. A Monte Carlo analysis

Joanna Górka, Magdalena Osińska*
Nicolaus Copernicus University in Torun, Faculty of Economic Sciences and Management,
Ul. Gagarina 13a, 87-100 Torun, Poland
jgora@uni.torun.pl, emo@uni.torun.pl

Abstract

The problem of unit roots is widely known in econometrics, however not stochastic unit roots. The stochastic unit root model is a random coefficients model and can generate different series depending upon values of parameters, specifically the variance. In this paper we explore the problem of testing for stochastic unit roots. The STUR tests are examined for their robustness (or sensitivity) for logarithmic and Box Cox transformations of time series. We also study sensitivity of the test for the presence of GARCH. A Monte Carlo simulation is used to show possible limits of the tests.

Keywords

Stochastic unit roots, tests Z and E, Monet Carlo.
JEL: C2, C5

1 Introduction

The purpose of the paper is to analyze the performance of the stochastic unit root tests when the examined time series are transformed using standard transformation methods used widely in econometrics. Where series are supposed to be generated by non-linear mechanism they are often transformed to gain linearity and normality of the distribution. However these transformations often affect statistical inference made for original and transformed series. In the paper we ask the question whether Box-Cox and logarithmic transformations may be considered as “safe” for using stochastic unit roots tests and the Dickey Fuller test – for comparison. We also check the performance of the mentioned tests for two other cases: additive seasonal effect and GARCH disturbances in the stochastic unit root model.

The stochastic unit root model is an example on non-linear and non-stationary process which does not become stationary after taking differences of any order. This class of processes was considered in the articles by Leybourne, McCabe & Mills (1996), Leybourne, McCabe & Tremayne (1996) and Granger & Swanson (1997). The models describing stochastic unit root processes belong to a wide class of the time-varying parameters models and their state space representation can be easily written. In special cases different form of non-linear models - like threshold or bi-linear models - can be derived (Granger, Terasvirta (1993).

The STUR models are mainly applicable for financial time series. Sollis, Leybourne and Newbold (2000) have examined the STUR model for several stock indices. In 4 cases on 6 the stochastic unit root hypothesis were accepted. Osińska (2003) showed that the stochastic unit root model (3) - (4) fits well the exchange rates PLN/USD and EURO/PLN, observed daily.

The paper is organized as follows: in the second and third sections the model and tests procedures are shown. In the fourth part the Monte Carlo experiment results are presented. Conclusions close the paper in section five.

* Financial support of the Polish Committee for Scientific Research for the project 2 H02B 015 25 realisation in 2003-2006 is gratefully acknowledged.

2 The model

The most general representation of the STUR processes, given in Granger, Swanson (1997), is the following

$$y_t = \phi_t y_{t-1} + \varepsilon_t \qquad (1)$$

where $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

and

$$\phi_t = \exp(\alpha_t) \qquad (2)$$

where: α_t is a stationary process such that $\alpha_t \sim N(0, \sigma_\alpha^2)$, and its spectrum is equal to $g_\alpha(\omega)$.

From (1), after simple manipulations, we have

$$\phi_t = (y_t / y_{t-1})(1 - \varepsilon_t / y_t).$$

Taking (2) we may, clearly, write

$$\alpha_t = \log(y_t) - \log(y_{t-1}) + \log(1 - \varepsilon_t / y_t)$$

and after re-arranging

$$\alpha_t \approx \Delta \log(y_t) - \varepsilon_t / y_t.$$

So that if $\log(y_t)$ has an exact unit root, y_t posses a stochastic unit root. This property shows that taking logs instead of levels is not always a fully safe operation (Mills (1999)). This fact motivate us to check whether other related transformations may affect the performance of the STUR tests.

In the article concerning testing for the STUR, Leybourne, McCabe, Mills (1996) suggest the following simple random coefficient autoregressive model

$$y_t = \alpha_t y_{t-1} + \varepsilon_t \qquad (3)$$

where:

$$\begin{aligned} &\alpha_t = \alpha_0 + \delta_t,\ \alpha_0 = 1, \\ &\delta_t = \rho \delta_{t-1} + \eta_t, \\ &\delta_0 = 0 \text{ and } |\rho| \le 1. \end{aligned} \qquad (4)$$

Stochastic processes $\varepsilon_t \sim N(0, \sigma^2)$ and $\eta_t \sim N(0, \omega^2)$ are assumed to be independent.

If $|\rho| < 1$, then α_t constitutes the AR(1) process with mean equal to one, and for $\rho = 1$ it is a random walk. The latter happens also for $\alpha_0 = 1$ and $\omega^2 = 0$. If $\alpha_0 = 1$ and $\omega^2 > 0$ we have a process with a unit root in mean, called a stochastic unit root process.

3 Testing for the stochastic unit roots

Standard unit root tests (Dickey-Fuller, Philips-Perron, KPSS) are inappropriate in the case of the stochastic unit root. They tend to indicate that the process posses an exact unit root, and cannot distinguish between the two cases. The standard tests assume the unit root (I(1)) case under the null, and the stationary (I(0)) case under the alternative. These definitely do not concern the stochastic unit roots. The STUR processes do not become stationary after differencing of any order.

Leybourne, McCabe and Tremayne (1996) have proposed a testing procedure (LMT hereafter), where under the alternative the stochastic unit root is assumed (see also Leybourne. McCabe and Mills (1996)). The power of the tests was examined by Taylor and van Dijk (1999).

Hypotheses in the LMT test consider the variance ω^2 characteristics in the model (4). The null is

$$H_0 : \omega^2 = 0$$

that means the random walk process or ARIMA(p,1,0), while the alternative is as follows

$$H_1 : \omega^2 > 0 \ .$$

Interpretation of the alternative depends on the value of the ρ parameter in (4). When $|\rho|<1$, δ_t in (4) is a stationary process with a zero mean. For $\rho = 1$ it follows a random walk process.

To avoid the influence of the deterministic trend and the autocorrelation, the model can include the linear or quadratic time trend, and the autoregressive lags of the dependent variable as well, so it takes the following form:

$$y_t^* = \alpha_t y_{t-1}^* + \varepsilon_t \tag{5}$$

where

$$y_t^* = y_y - P_t - \sum_{i=1}^{p} \varphi_i y_{t-i} \tag{6}$$

where P_t means a deterministic component, for example trend

$$P_{1t} = \beta + \gamma t \quad \text{or} \quad P_{2t} = \beta + \gamma_1 t + \gamma_2 t^2$$

and the autoregressive part in (6) is stationary and its role is similar to the augmentation in the Augmented Dickey Fuller test.

If in H_1 $|\rho|<1$ then the Z statistics is computed in two steps:

1. estimate OLS the equation (7)

$$\Delta y_t = \Delta P_t + \sum_{i=1}^{p} \varphi_i \Delta y_{t-1} + \varepsilon_t \ . \tag{7}$$

2. compute the statistics

$$Z = T^{-\frac{3}{2}} \sigma^{-2} \kappa^{-1} \sum_{t=2}^{T} \left(\sum_{j=1}^{t-1} \varepsilon_j \right)^2 \left(\varepsilon_t^2 - \sigma^2 \right) \tag{8}$$

where: $\sigma^2 = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2$ and $\kappa^2 = T^{-1} \sum_{t=1}^{T} \left(\varepsilon_t^2 - \sigma^2 \right)^2$.

If in H_1 $\rho = 1$ then the following E statistics is strongly recommended (see Leybourne, McCabe, Mills (1996)):

$$\mathrm{E} = \mathrm{T}^{-3} \sigma^{-4} \sum_{\mathrm{i}=2}^{\mathrm{T}} \left\{ \left[\sum_{\mathrm{t}=\mathrm{i}}^{\mathrm{T}} \varepsilon_{\mathrm{t}} \left(\sum_{\mathrm{j}=1}^{\mathrm{t}-1} \varepsilon_{\mathrm{j}} \right) \right]^2 - \sigma^2 \sum_{\mathrm{t}=\mathrm{i}}^{\mathrm{T}} \left(\sum_{\mathrm{j}=1}^{\mathrm{t}-1} \varepsilon_{\mathrm{j}} \right)^2 \right\} . \tag{9}$$

Depending on the trend model choice P_{1t} or P_{2t} the statistics are denoted Z_1, Z_2 or E_1, E_2.

Granger and Swanson argue that the tests are robust against logarithmic transformation and GARCH effect, except of IGARCH (Granger, Swanson (1997)).

The test statistics do not converge to any standard distribution, so that the critical values have to be computed individually (see Leybourne et al. (1996a and b), Granger, Swanson (1997)).

We also propose to generalise the concept of choosing the deterministic trend term as well as the autoregression order in (6). It is due to the fact that arbitrary choice o trend polynomial and autoregressive lags may over/under-parametrize the model. Leybourne et al. (1996a) argue that it is better to take greater number of autoregressive lags in (6) than too small. When we however adjust these terms to the data it seems that the inference about the STUR may be more faithful. We check the mentioned procedure in the next section, denoting the tests simply Z and E (without numbering them).

4 Performance of the STUR tests. A Monte Carlo study

The experiment was designed to show the impact of log and Box-Cox transformation for the testing results for the presence of stochastic unit roots. In all cases the null was: series posses an exact unit root, while the alternative (which was true) was related to the stochastic unit root. We applied Z and E tests described above, and Dickey Fuller tests additionally.

The Box-Cox transformation is given by the formula:

$$\frac{y_t^{\lambda}-1}{\lambda}$$

where y_t is a time series generated here according to a STUR model and $\lambda \in [-3;3]$. For $\lambda = 0$ the Box-Cox transformation is equivalent to the logarithmic one, which results from

$$\lim_{\lambda \to 0} \frac{y_t^{\lambda}-1}{\lambda} = \ln y_t .$$

Additionally we verify the hypothesis that additive seasonal component may affect the performance o the test. As the STUR models are often applied to financial data, we assume the seasonal effect is equivalent to the Monday effect. That is why every fifth observation was reduced to the magnitude of 1% and 5% o its previous value, respectively.

In the last part of the experiment the STUR-GARCH models were generated in the following way:

$$y_t = \alpha_t y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = z_t h_t^{1/2}, \quad z_t \sim N(0,1)$$
$$h_t = \alpha_0 + \alpha_1 z_{t-1}^2 + \beta_1 h_{t-1}$$
$$\alpha_t = \alpha_0 + \delta_t$$
$$\delta_t = \rho \delta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \omega^2).$$

To show the effect of transformations, 16 types of STUR series were generated according to the following models:

STUR1	$\alpha_0 = 1$,	$\rho = 0.98$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR2	$\alpha_0 = 1$,	$\rho = 0.95$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR3	$\alpha_0 = 0.98$,	$\rho = 0.98$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR4	$\alpha_0 = 0.98$,	$\rho = 0.95$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR5	$\alpha_0 = 1$,	$\rho = 0.98$,	$\omega^2 = 0.05$	$\varepsilon_t \sim N(0,1)$
STUR6	$\alpha_0 = 1$,	$\rho = 0.95$,	$\omega^2 = 0.05$	$\varepsilon_t \sim N(0,1)$
STUR7	$\alpha_0 = 0.98$,	$\rho = 0.98$,	$\omega^2 = 0.05$	$\varepsilon_t \sim N(0,1)$
STUR8	$\alpha_0 = 0.98$,	$\rho = 0.95$,	$\omega^2 = 0.05$	$\varepsilon_t \sim N(0,1)$
STUR9	$\alpha_0 = 1$,	$\rho = 0.98$,	$\omega^2 = 0.1$	$\varepsilon_t \sim N(0,1)$
STUR10	$\alpha_0 = 1$,	$\rho = 0.95$,	$\omega^2 = 0.1$	$\varepsilon_t \sim N(0,1)$
STUR11	$\alpha_0 = 0.98$,	$\rho = 0.98$,	$\omega^2 = 0.1$	$\varepsilon_t \sim N(0,1)$
STUR12	$\alpha_0 = 0.98$,	$\rho = 0.95$,	$\omega^2 = 0.1$	$\varepsilon_t \sim N(0,1)$
STUR13	$\alpha_0 = 1.02$,	$\rho = 0.98$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR14	$\alpha_0 = 1.02$,	$\rho = 0.95$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR15	$\alpha_0 = 1.05$,	$\rho = 0.98$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$
STUR16	$\alpha_0 = 1.05$,	$\rho = 0.95$,	$\omega^2 = 0.01$	$\varepsilon_t \sim N(0,1)$

The experiment was taken within following panels, concerning the size o the sample: 250, 500 and 1000 observations respectively. Number of replications was equal to 1000 for each case. Some of the results are presented in tables 1-8. All results are available from the authors on request. The numerical results can be summarized as follows:

1. Logarithmic transformation

- tests Z1, Z2, E1, E2 are sensitive for logarithmic transformation;
- tests Z and E are more robust for the transformation;
- DF test is sensitive for the transformation.

2. Box-Cox transformation
   - tests Z1, Z2, E1, E2 are sensitive for Box-Cox transformation. This depends on the value of λ parameter. For $\lambda < 0$ and $\lambda \geq 2$, the ratio of rejection the null is relatively small (see figures 1-4). For $0 < \lambda < 2$ the ratio of rejection the null is similar to the results obtained for untransformed series (the tests are then robust; the transformation does not matter). For larger samples (500 and 1000 observations) the ratio of rejecting the null increases.
   - Tests Z and E seem to be more robust for the transformation, but only for $\lambda \leq 0$ (figures 5-6). The larger sample size, the greater ratio of rejection the null of exact unit root. So that for larger samples the stochastic unit root is more likely to be found. For $\lambda > 0$ tests Z and E perform in the same way as Z1, Z2, E1 and E2.
   - Concerning tests Z and E for $\lambda < 0$ we can observe that their results depend on the time trend choice. That can be explained by the fact, that we assumed 3 cases here, i.e. no trend, linear trend or quadratic one. In the case of Z1, Z2, E1 and E2 we assumed only two case: linear or quadratic time trend.
   - DF test is sensitive for the transformation, depending on the value of λ (figure 7).
   - For $\lambda = 0$ the results converge to that obtained for logarithmic transformation, which is obvious.
3. Additive seasonal effect
   - only DF is sensitive for the reducing every fifth value of the series of 1% or 5% of its previous value. The magnitude of the reaction is greatest for n=1000.
4. GARCH effect
   - only DF test is sensitive for GARCH in the residual process. It is especially visible for n=1000.

5 Conclusions

Recently developed concept of the stochastic unit roots has some advantages in comparison to the exact unit roots. First of all, the properties of the time series are more accurately investigated. There are some periods when process behaves like stationary, and some others when it is variance-non-stationary. However, the empirical identification of the STUR processes is not very popular yet. In the paper we examined some limits of the STUR test and their performance in the contexts of popular transformations. The STUR tests are indifferent when disturbances are generated by GARCH models and are robust for adding some seasonal effects. They are however sensitive when Box-Cox transformation is applied, but the size of reaction depends on the value of the parameter λ.

References

[1] Granger C.W.J., Swanson N.R. (1997), An introduction to stochastic unit-root process. *Journal of Econometrics* 80.

[2] Granger C.W.J., Teräsvirta T. (1993), *Modeling Nonlinear Economic Relationships*. Oxford University Press.

[3] Leybourne S.J., McCabe B.P.M., Mills T.C. (1996), Randomized unit root processes for modeling and forecasting financial time series: theory and applications. *Journal of Forecasting* 15.

[4] Leybourne S.J., McCabe B.P.M., Tremayne A.R. (1996) Can economic time series be differenced to stationarity? *Journal of .Business and Economic Statistics*, 14.

[5] Mills T.C. (1999), *The Econometric Modelling of Financial Time Series*. Second ed. Cambridge University Press.

[6] Osińska M. (2003) Stochastic Unit Root Processes Properties and Application. [In] *Macromodels'2003*. Eds. W. Welfe, A. Welfe. Wyd. Uniwersytetu Łódzkiego. Łódź 2004.

[7] Osińska M, Górka J. (2005) Identyfikacja nieliniowości w ekonomicznych szeregach czasowych. Analiza symulacyjna. [In] *Dynamiczne modele ekonometryczne*. Wyd. UMK Toruń, 2005.

[8] Taylor A.M.R, van Dijk D. (1999) Testing for Stochastic Unit Roots. Some Monte Carlo Evidence. *Econometric Institute Research Report EI-9922/A*.

[9] Sollis R., Leybourne S.J., Newbold P. (2000) Stochastic unit roots modeling of stock price indices. *Applied Financial Economics,* 10.

Appendix

Figure 1. The ratio of rejection the null relatively to the value of λ

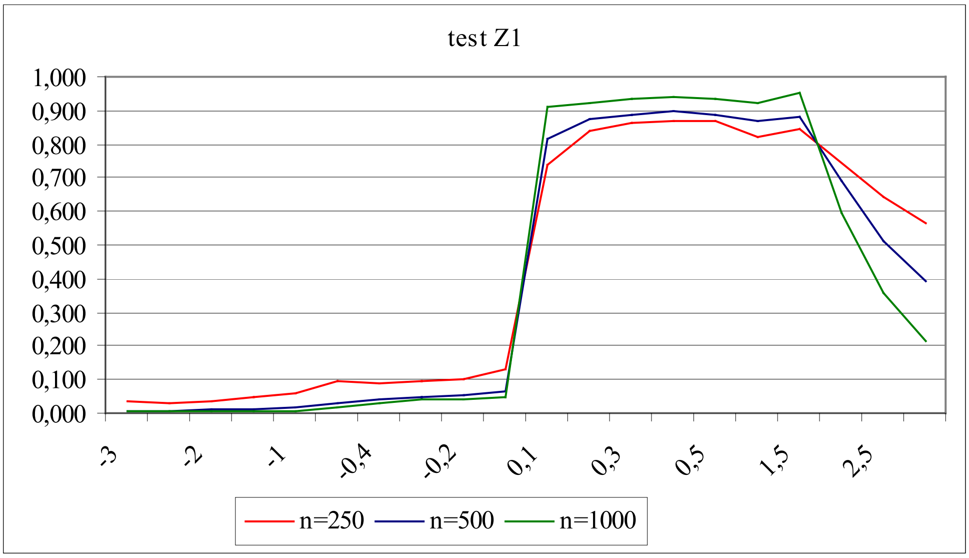


Figure 2. The ratio of rejection the null relatively to the value of λ

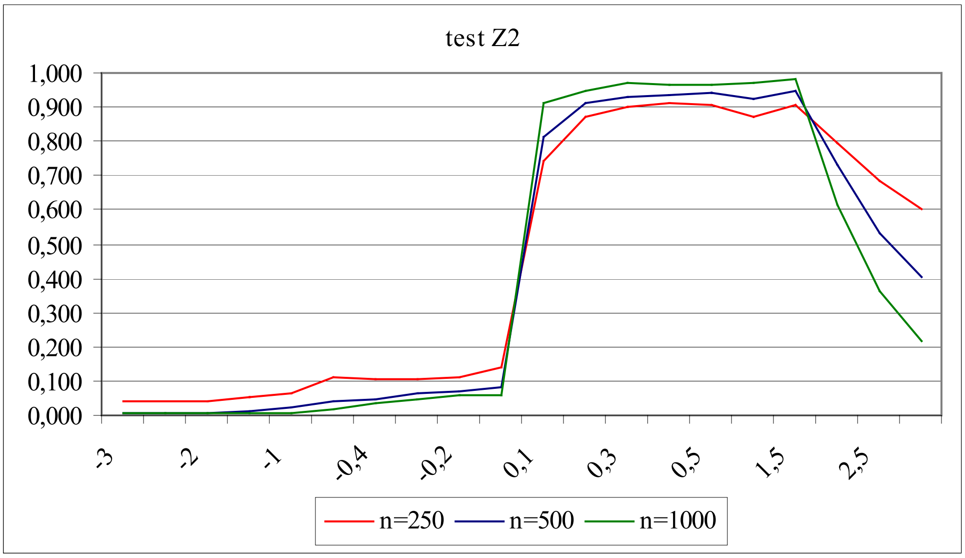


Figure 3. The ratio of rejection the null relatively to the value of λ

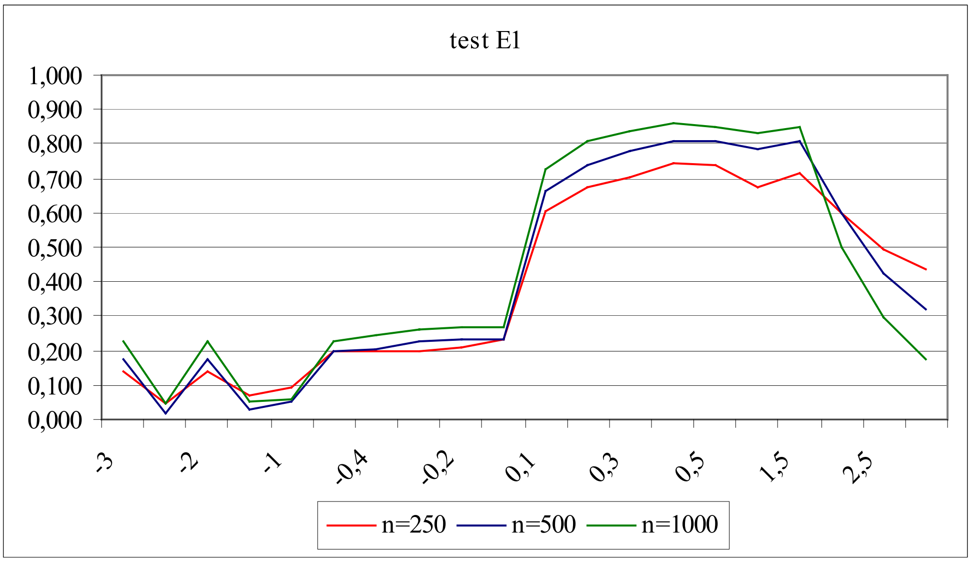


Figure 4. The ratio of rejection the null relatively to the value of λ

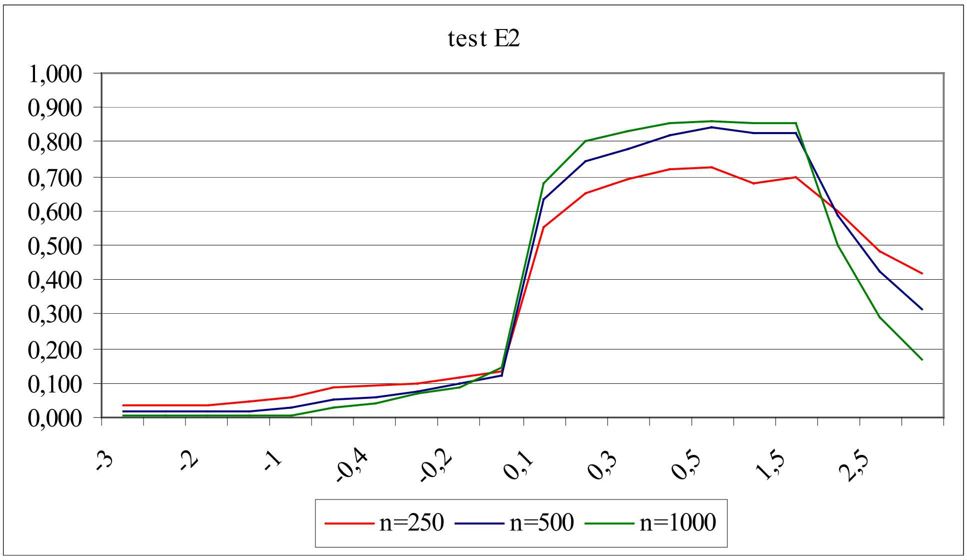


Figure 5. The ratio of rejection the null relatively to the value of λ

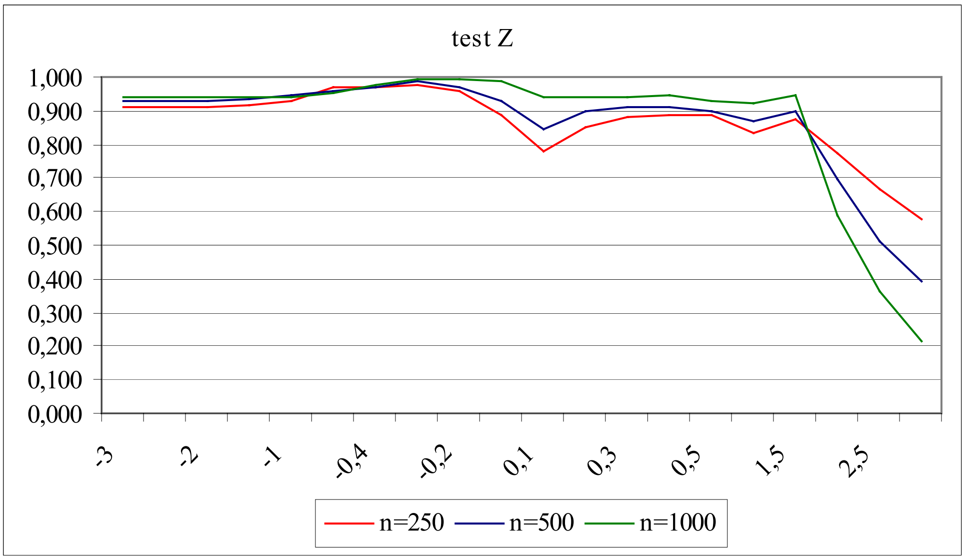


Figure 6. The ratio of rejection the null relatively to the value of λ

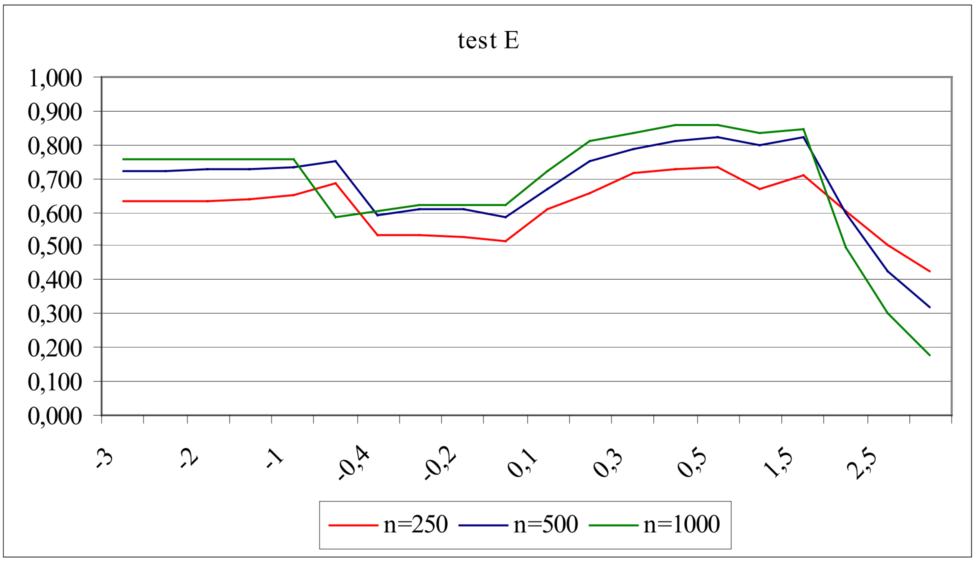


Figure 7. The ratio of rejection the null relatively to the value of λ

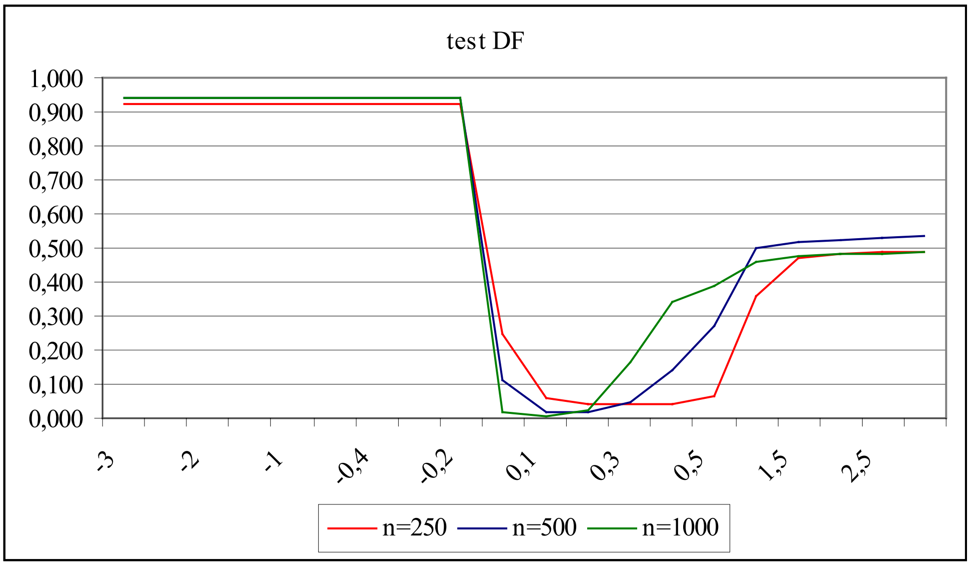


Table 1. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

STUR5: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR6: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR7: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR8: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

		λ=-3							lambda=-2,5						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,923	0,912	0,631	0,033	0,039	0,142	0,034	0,923	0,913	0,633	0,031	0,039	0,046	0,033
	STUR 6	0,891	0,877	0,677	0,032	0,035	0,150	0,032	0,891	0,886	0,681	0,037	0,047	0,050	0,035
	STUR 7	0,899	0,891	0,653	0,032	0,033	0,139	0,035	0,899	0,895	0,654	0,033	0,037	0,052	0,036
	STUR 8	0,912	0,898	0,678	0,031	0,036	0,137	0,034	0,912	0,902	0,681	0,032	0,038	0,052	0,035
n=500	STUR 5	0,940	0,929	0,723	0,006	0,006	0,173	0,015	0,940	0,929	0,723	0,007	0,007	0,016	0,015
	STUR 6	0,953	0,944	0,750	0,010	0,010	0,190	0,011	0,953	0,946	0,751	0,015	0,013	0,018	0,012
	STUR 7	0,942	0,930	0,721	0,007	0,007	0,173	0,017	0,942	0,932	0,723	0,009	0,009	0,023	0,019
	STUR 8	0,959	0,946	0,767	0,005	0,005	0,202	0,008	0,959	0,947	0,767	0,005	0,007	0,016	0,009
n=1000	STUR 5	0,943	0,940	0,755	0,003	0,003	0,224	0,006	0,943	0,940	0,755	0,003	0,003	0,044	0,006
	STUR 6	0,964	0,960	0,785	0,007	0,007	0,248	0,007	0,964	0,960	0,785	0,007	0,007	0,048	0,007
	STUR 7	0,942	0,940	0,744	0,003	0,003	0,217	0,003	0,942	0,941	0,745	0,004	0,004	0,030	0,004
	STUR 8	0,966	0,963	0,775	0,003	0,003	0,241	0,010	0,966	0,963	0,775	0,003	0,003	0,039	0,010

Table 2. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

		lambda=-2							lambda=-1,5						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,923	0,913	0,635	0,035	0,040	0,139	0,036	0,923	0,916	0,642	0,046	0,051	0,072	0,049
	STUR 6	0,891	0,891	0,689	0,046	0,051	0,163	0,043	0,891	0,901	0,701	0,057	0,066	0,078	0,058
	STUR 7	0,899	0,896	0,657	0,038	0,037	0,149	0,040	0,899	0,901	0,660	0,044	0,051	0,080	0,045
	STUR 8	0,912	0,907	0,684	0,036	0,042	0,145	0,038	0,912	0,914	0,689	0,047	0,055	0,077	0,043
n=500	STUR 5	0,940	0,931	0,725	0,010	0,008	0,176	0,015	0,940	0,933	0,728	0,011	0,012	0,031	0,020
	STUR 6	0,953	0,948	0,754	0,018	0,020	0,195	0,016	0,953	0,949	0,756	0,022	0,022	0,031	0,019
	STUR 7	0,942	0,934	0,724	0,010	0,011	0,174	0,019	0,942	0,936	0,727	0,016	0,017	0,041	0,024
	STUR 8	0,959	0,947	0,767	0,004	0,007	0,204	0,009	0,959	0,949	0,769	0,007	0,009	0,026	0,010
n=1000	STUR 5	0,943	0,940	0,755	0,003	0,003	0,224	0,006	0,943	0,941	0,756	0,004	0,004	0,053	0,007
	STUR 6	0,964	0,961	0,786	0,008	0,008	0,249	0,008	0,964	0,962	0,787	0,009	0,009	0,051	0,009
	STUR 7	0,942	0,941	0,745	0,004	0,004	0,218	0,004	0,942	0,942	0,746	0,005	0,005	0,031	0,005
	STUR 8	0,966	0,963	0,775	0,003	0,003	0,241	0,010	0,966	0,963	0,775	0,004	0,003	0,043	0,010
		lambda=-1							lambda=-0,5						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,923	0,931	0,650	0,062	0,067	0,093	0,056	0,923	0,971	0,686	0,096	0,112	0,199	0,089
	STUR 6	0,891	0,916	0,721	0,074	0,088	0,118	0,081	0,891	0,955	0,752	0,089	0,115	0,218	0,113
	STUR 7	0,899	0,923	0,683	0,068	0,075	0,114	0,071	0,899	0,962	0,718	0,095	0,125	0,209	0,108
	STUR 8	0,912	0,932	0,708	0,058	0,078	0,108	0,069	0,913	0,961	0,739	0,079	0,107	0,191	0,106
n=500	STUR 5	0,940	0,944	0,736	0,020	0,024	0,051	0,030	0,940	0,959	0,753	0,031	0,039	0,200	0,050
	STUR 6	0,953	0,961	0,764	0,024	0,036	0,065	0,030	0,953	0,985	0,789	0,038	0,057	0,212	0,054
	STUR 7	0,942	0,944	0,732	0,021	0,025	0,058	0,029	0,942	0,958	0,749	0,035	0,041	0,195	0,051
	STUR 8	0,959	0,960	0,779	0,014	0,021	0,068	0,023	0,959	0,981	0,798	0,026	0,044	0,221	0,041
n=1000	STUR 5	0,943	0,942	0,756	0,005	0,005	0,057	0,007	0,943	0,955	0,588	0,016	0,020	0,225	0,027
	STUR 6	0,964	0,967	0,791	0,013	0,014	0,063	0,013	0,964	0,985	0,610	0,026	0,030	0,263	0,032
	STUR 7	0,942	0,943	0,746	0,006	0,006	0,040	0,005	0,942	0,964	0,575	0,022	0,025	0,227	0,027
	STUR 8	0,966	0,969	0,781	0,008	0,009	0,057	0,016	0,966	0,985	0,624	0,019	0,022	0,248	0,030

Table 3. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

		lambda=-0,4							lambda=-0,3						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,923	0,971	0,532	0,090	0,108	0,200	0,092	0,923	0,974	0,532	0,096	0,108	0,200	0,100
	STUR 6	0,891	0,960	0,608	0,092	0,118	0,232	0,124	0,891	0,957	0,602	0,095	0,120	0,234	0,132
	STUR 7	0,899	0,966	0,555	0,096	0,125	0,210	0,117	0,899	0,967	0,556	0,096	0,127	0,213	0,129
	STUR 8	0,913	0,969	0,573	0,084	0,110	0,203	0,111	0,913	0,968	0,568	0,092	0,115	0,213	0,123
n=500	STUR 5	0,940	0,969	0,591	0,039	0,048	0,205	0,061	0,940	0,988	0,609	0,045	0,063	0,224	0,078
	STUR 6	0,953	0,979	0,587	0,037	0,052	0,215	0,063	0,953	0,982	0,587	0,034	0,052	0,215	0,067
	STUR 7	0,942	0,972	0,601	0,042	0,053	0,210	0,066	0,942	0,986	0,615	0,050	0,066	0,221	0,081
	STUR 8	0,959	0,984	0,616	0,030	0,047	0,228	0,052	0,959	0,986	0,618	0,031	0,051	0,235	0,068
n=1000	STUR 5	0,943	0,974	0,603	0,028	0,033	0,247	0,042	0,943	0,994	0,624	0,042	0,049	0,263	0,069
	STUR 6	0,964	0,994	0,618	0,033	0,038	0,269	0,046	0,964	0,995	0,621	0,032	0,042	0,269	0,054
	STUR 7	0,942	0,970	0,582	0,023	0,032	0,235	0,033	0,942	0,992	0,605	0,039	0,048	0,247	0,062
	STUR 8	0,966	0,989	0,628	0,024	0,027	0,260	0,039	0,966	0,994	0,633	0,027	0,034	0,265	0,053
		lambda=-0,2							lambda=-0,1						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,923	0,959	0,524	0,101	0,113	0,210	0,114	0,248	0,884	0,514	0,130	0,141	0,231	0,133
	STUR 6	0,891	0,937	0,609	0,123	0,147	0,248	0,151	0,256	0,875	0,591	0,159	0,188	0,271	0,206
	STUR 7	0,899	0,958	0,548	0,109	0,137	0,221	0,144	0,233	0,895	0,550	0,141	0,156	0,231	0,159
	STUR 8	0,913	0,946	0,562	0,115	0,133	0,218	0,142	0,276	0,867	0,545	0,151	0,187	0,234	0,177
n=500	STUR 5	0,940	0,971	0,610	0,051	0,070	0,232	0,098	0,114	0,931	0,586	0,067	0,080	0,233	0,125
	STUR 6	0,953	0,970	0,588	0,048	0,059	0,221	0,077	0,124	0,913	0,570	0,058	0,073	0,209	0,106
	STUR 7	0,942	0,977	0,616	0,057	0,063	0,223	0,094	0,115	0,936	0,587	0,073	0,078	0,232	0,121
	STUR 8	0,959	0,965	0,617	0,044	0,058	0,239	0,086	0,149	0,911	0,600	0,059	0,073	0,229	0,106
n=1000	STUR 5	0,943	0,994	0,624	0,043	0,056	0,269	0,088	0,015	0,987	0,621	0,045	0,057	0,266	0,145
	STUR 6	0,964	0,994	0,620	0,033	0,044	0,271	0,089	0,025	0,983	0,614	0,038	0,044	0,283	0,127
	STUR 7	0,942	0,993	0,608	0,042	0,053	0,257	0,086	0,013	0,992	0,610	0,041	0,054	0,252	0,119
	STUR 8	0,966	0,990	0,632	0,027	0,031	0,263	0,076	0,025	0,983	0,624	0,026	0,031	0,275	0,112

Table 4. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

		lambda=0,1							lambda=0,2						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,058	0,778	0,609	0,737	0,740	0,607	0,553	0,041	0,849	0,657	0,837	0,873	0,677	0,649
	STUR 6	0,023	0,804	0,683	0,798	0,806	0,686	0,659	0,019	0,876	0,727	0,875	0,905	0,733	0,730
	STUR 7	0,057	0,742	0,598	0,718	0,728	0,588	0,538	0,048	0,830	0,671	0,839	0,867	0,707	0,666
	STUR 8	0,033	0,812	0,635	0,812	0,810	0,655	0,612	0,031	0,884	0,715	0,886	0,921	0,734	0,708
n=500	STUR 5	0,015	0,844	0,669	0,815	0,813	0,665	0,635	0,019	0,900	0,753	0,877	0,911	0,741	0,743
	STUR 6	0,017	0,893	0,690	0,881	0,871	0,709	0,659	0,018	0,943	0,794	0,928	0,961	0,812	0,775
	STUR 7	0,020	0,865	0,658	0,825	0,819	0,675	0,620	0,023	0,922	0,751	0,891	0,933	0,750	0,727
	STUR 8	0,015	0,887	0,712	0,877	0,868	0,738	0,678	0,015	0,944	0,793	0,932	0,964	0,808	0,780
n=1000	STUR 5	0,005	0,939	0,722	0,911	0,915	0,729	0,681	0,025	0,941	0,813	0,924	0,946	0,809	0,801
	STUR 6	0,004	0,955	0,764	0,952	0,953	0,776	0,715	0,024	0,962	0,843	0,955	0,971	0,840	0,819
	STUR 7	0,007	0,942	0,719	0,910	0,918	0,721	0,673	0,035	0,946	0,797	0,928	0,960	0,803	0,788
	STUR 8	0,007	0,950	0,767	0,946	0,944	0,777	0,726	0,026	0,963	0,838	0,954	0,971	0,845	0,821
		lambda=0,3							lambda=0,4						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,041	0,879	0,716	0,864	0,899	0,706	0,690	0,039	0,887	0,729	0,867	0,914	0,743	0,723
	STUR 6	0,022	0,883	0,758	0,878	0,926	0,765	0,766	0,034	0,888	0,805	0,867	0,926	0,784	0,793
	STUR 7	0,050	0,852	0,723	0,847	0,886	0,730	0,726	0,053	0,851	0,743	0,851	0,905	0,754	0,758
	STUR 8	0,032	0,896	0,755	0,874	0,906	0,757	0,754	0,039	0,892	0,776	0,876	0,920	0,770	0,782
n=500	STUR 5	0,049	0,910	0,789	0,888	0,929	0,781	0,779	0,144	0,911	0,813	0,896	0,937	0,811	0,817
	STUR 6	0,037	0,942	0,824	0,927	0,959	0,822	0,809	0,133	0,940	0,850	0,926	0,960	0,832	0,833
	STUR 7	0,051	0,922	0,787	0,899	0,939	0,781	0,777	0,155	0,906	0,800	0,890	0,939	0,797	0,812
	STUR 8	0,037	0,943	0,831	0,926	0,961	0,827	0,821	0,136	0,949	0,859	0,930	0,962	0,849	0,849
n=1000	STUR 5	0,165	0,938	0,834	0,934	0,970	0,837	0,831	0,339	0,944	0,859	0,938	0,967	0,859	0,852
	STUR 6	0,166	0,958	0,868	0,947	0,978	0,857	0,853	0,321	0,964	0,883	0,952	0,982	0,874	0,885
	STUR 7	0,177	0,941	0,830	0,928	0,963	0,814	0,823	0,343	0,946	0,841	0,927	0,969	0,820	0,838
	STUR 8	0,157	0,962	0,867	0,953	0,980	0,866	0,863	0,312	0,962	0,879	0,954	0,977	0,875	0,874

Table 5. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

		lambda=0,5							lambda=1						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,065	0,886	0,731	0,870	0,904	0,737	0,725	0,359	0,836	0,670	0,822	0,869	0,674	0,683
	STUR 6	0,049	0,899	0,812	0,873	0,926	0,799	0,818	0,295	0,900	0,832	0,876	0,927	0,801	0,830
	STUR 7	0,074	0,866	0,766	0,854	0,901	0,764	0,773	0,362	0,818	0,688	0,813	0,867	0,689	0,718
	STUR 8	0,051	0,886	0,785	0,872	0,924	0,774	0,795	0,308	0,889	0,804	0,861	0,928	0,765	0,797
n=500	STUR 5	0,272	0,899	0,825	0,884	0,940	0,811	0,841	0,500	0,871	0,796	0,871	0,924	0,786	0,828
	STUR 6	0,278	0,941	0,850	0,932	0,969	0,849	0,850	0,526	0,932	0,862	0,922	0,961	0,864	0,864
	STUR 7	0,291	0,895	0,804	0,888	0,942	0,796	0,834	0,535	0,882	0,797	0,881	0,928	0,791	0,821
	STUR 8	0,270	0,946	0,874	0,928	0,956	0,864	0,867	0,533	0,932	0,880	0,926	0,964	0,874	0,881
n=1000	STUR 5	0,391	0,930	0,857	0,932	0,962	0,849	0,863	0,458	0,925	0,834	0,925	0,972	0,834	0,854
	STUR 6	0,390	0,962	0,892	0,950	0,984	0,883	0,894	0,458	0,953	0,891	0,952	0,978	0,886	0,885
	STUR 7	0,405	0,934	0,840	0,924	0,967	0,838	0,846	0,479	0,920	0,825	0,918	0,963	0,824	0,837
	STUR 8	0,384	0,957	0,890	0,949	0,973	0,871	0,885	0,436	0,948	0,889	0,952	0,973	0,887	0,890
		lambda=1,5							lambda=2						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,468	0,873	0,712	0,846	0,909	0,718	0,698	0,485	0,776	0,601	0,745	0,797	0,597	0,597
	STUR 6	0,391	0,891	0,794	0,858	0,927	0,782	0,803	0,411	0,841	0,738	0,812	0,883	0,718	0,736
	STUR 7	0,458	0,868	0,736	0,835	0,908	0,717	0,716	0,474	0,761	0,589	0,736	0,800	0,595	0,593
	STUR 8	0,402	0,887	0,763	0,854	0,924	0,733	0,765	0,423	0,864	0,714	0,827	0,896	0,697	0,699
n=500	STUR 5	0,520	0,898	0,820	0,882	0,950	0,807	0,823	0,526	0,699	0,600	0,692	0,731	0,601	0,589
	STUR 6	0,541	0,928	0,848	0,921	0,965	0,839	0,859	0,552	0,844	0,742	0,831	0,873	0,728	0,736
	STUR 7	0,555	0,922	0,825	0,902	0,952	0,810	0,832	0,563	0,712	0,610	0,706	0,733	0,609	0,611
	STUR 8	0,543	0,934	0,865	0,920	0,972	0,851	0,871	0,552	0,860	0,757	0,849	0,886	0,752	0,757
n=1000	STUR 5	0,475	0,946	0,849	0,950	0,982	0,851	0,853	0,483	0,592	0,498	0,597	0,613	0,501	0,500
	STUR 6	0,466	0,953	0,863	0,949	0,979	0,860	0,855	0,472	0,772	0,674	0,768	0,790	0,670	0,658
	STUR 7	0,495	0,936	0,837	0,945	0,976	0,846	0,843	0,496	0,593	0,499	0,598	0,613	0,507	0,499
	STUR 8	0,449	0,949	0,858	0,948	0,978	0,867	0,856	0,457	0,776	0,671	0,774	0,794	0,670	0,651

Table 6. The results of the DF, Z, E tests for Box-Cox transformation. The ratios of rejection of the null are reported at 0.05 significance level

		lambda=2,5							lambda=3						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,488	0,667	0,505	0,642	0,682	0,497	0,485	0,490	0,579	0,425	0,564	0,600	0,437	0,416
	STUR 6	0,418	0,774	0,656	0,743	0,802	0,649	0,658	0,424	0,715	0,596	0,682	0,738	0,588	0,593
	STUR 7	0,480	0,663	0,524	0,647	0,689	0,512	0,522	0,489	0,563	0,430	0,546	0,585	0,426	0,417
	STUR 8	0,429	0,812	0,647	0,769	0,829	0,640	0,647	0,432	0,732	0,581	0,707	0,745	0,574	0,568
n=500	STUR 5	0,532	0,514	0,425	0,511	0,535	0,422	0,422	0,535	0,395	0,322	0,390	0,406	0,320	0,316
	STUR 6	0,560	0,710	0,595	0,699	0,732	0,599	0,595	0,563	0,587	0,482	0,579	0,601	0,488	0,483
	STUR 7	0,566	0,512	0,436	0,513	0,537	0,428	0,433	0,568	0,396	0,322	0,396	0,402	0,326	0,318
	STUR 8	0,558	0,751	0,646	0,733	0,774	0,635	0,634	0,558	0,616	0,503	0,602	0,628	0,515	0,503
n=1000	STUR 5	0,485	0,361	0,301	0,360	0,363	0,295	0,291	0,487	0,217	0,178	0,215	0,217	0,175	0,171
	STUR 6	0,476	0,559	0,478	0,554	0,577	0,477	0,467	0,481	0,388	0,315	0,383	0,401	0,313	0,305
	STUR 7	0,498	0,366	0,306	0,363	0,368	0,301	0,288	0,499	0,219	0,181	0,215	0,218	0,181	0,180
	STUR 8	0,463	0,583	0,492	0,583	0,597	0,491	0,470	0,468	0,417	0,355	0,417	0,429	0,358	0,342

Table 7. The results of the DF, Z, E tests for additive seasonal effect. The ratios of rejection of the null are reported at 0.05 significance level

STUR5: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR6: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR7: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR8: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

		the reduction rate - 1% of its previous value							the reduction rate - 5% of its previous value						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,610	0,796	0,735	0,793	0,886	0,729	0,777	0,620	0,800	0,742	0,792	0,887	0,743	0,787
	STUR 6	0,630	0,875	0,805	0,850	0,920	0,783	0,819	0,647	0,887	0,807	0,854	0,929	0,798	0,819
	STUR 7	0,631	0,803	0,727	0,793	0,870	0,710	0,759	0,631	0,822	0,736	0,795	0,874	0,709	0,767
	STUR 8	0,615	0,874	0,814	0,840	0,908	0,794	0,821	0,619	0,861	0,812	0,838	0,900	0,793	0,823
n=500	STUR 5	0,842	0,868	0,796	0,867	0,928	0,789	0,825	0,842	0,867	0,798	0,864	0,934	0,786	0,822
	STUR 6	0,887	0,933	0,861	0,922	0,958	0,862	0,866	0,887	0,934	0,871	0,930	0,967	0,870	0,870
	STUR 7	0,856	0,881	0,796	0,880	0,927	0,787	0,821	0,856	0,880	0,798	0,880	0,929	0,783	0,819
	STUR 8	0,901	0,931	0,880	0,922	0,963	0,873	0,882	0,902	0,936	0,882	0,920	0,964	0,867	0,884
n=1000	STUR 5	0,893	0,925	0,839	0,925	0,969	0,832	0,855	0,893	0,926	0,837	0,921	0,964	0,832	0,853
	STUR 6	0,919	0,952	0,894	0,946	0,974	0,882	0,883	0,919	0,954	0,881	0,948	0,980	0,877	0,882
	STUR 7	0,905	0,922	0,825	0,922	0,963	0,821	0,842	0,905	0,930	0,830	0,922	0,957	0,827	0,837
	STUR 8	0,930	0,946	0,885	0,947	0,974	0,893	0,892	0,930	0,952	0,890	0,949	0,972	0,892	0,892

Table 8. The results of the DF, Z, E tests for logarithmic transformation and GARCH effect. The ratios of rejection of the null are reported at 0.05 significance level

STUR5: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR6: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 1$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR7: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.98\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

STUR8: $y_t = \alpha_t y_{t-1} + \varepsilon_t$, $\alpha_t = \alpha_0 + \delta_t$, $\alpha_0 = 0.98$, $\delta_t = 0.95\delta_{t-1} + \eta_t$, $\delta_0 = 0$, $\eta_t \sim N(0, 0.05)$, $\varepsilon_t \sim N(0,1)$

In the GARCH effect $\varepsilon_t = z_t\sqrt{h_t}$, $h_t = 0.1 + 0.1z_{t-1}^2 + 0.8h_{t-1}$, $z_t \sim N(0,1)$.

		Logarithmic transformation							GARCH effect						
		DF	test Z	test E	test Z		test E		DF	test Z	test E	test Z		test E	
					Z1	Z2	E1	E2				Z1	Z2	E1	E2
n=250	STUR 5	0,075	0,653	0,458	0,221	0,232	0,295	0,237	0,695	0,845	0,714	0,827	0,895	0,712	0,733
	STUR 6	0,033	0,643	0,513	0,276	0,288	0,359	0,319	0,642	0,894	0,842	0,879	0,929	0,824	0,844
	STUR 7	0,073	0,663	0,491	0,230	0,235	0,307	0,241	0,693	0,798	0,704	0,800	0,858	0,686	0,711
	STUR 8	0,036	0,657	0,488	0,284	0,288	0,345	0,299	0,620	0,900	0,823	0,879	0,930	0,791	0,822
n=500	STUR 5	0,014	0,664	0,466	0,141	0,135	0,267	0,196	0,908	0,919	0,802	0,899	0,937	0,785	0,813
	STUR 6	0,003	0,672	0,504	0,187	0,180	0,330	0,239	0,912	0,937	0,863	0,937	0,968	0,867	0,869
	STUR 7	0,009	0,693	0,491	0,133	0,126	0,282	0,188	0,907	0,921	0,811	0,919	0,949	0,807	0,828
	STUR 8	0,004	0,676	0,529	0,198	0,197	0,356	0,279	0,897	0,938	0,836	0,942	0,968	0,840	0,852
n=1000	STUR 5	0,000	0,749	0,532	0,094	0,094	0,308	0,210	0,892	0,933	0,828	0,927	0,974	0,821	0,834
	STUR 6	0,000	0,716	0,505	0,119	0,102	0,311	0,227	0,945	0,962	0,890	0,965	0,989	0,887	0,877
	STUR 7	0,000	0,737	0,514	0,087	0,082	0,294	0,195	0,910	0,926	0,834	0,938	0,972	0,838	0,840
	STUR 8	0,003	0,715	0,509	0,132	0,126	0,320	0,219	0,920	0,946	0,878	0,951	0,980	0,870	0,883

Minimum Wage Impact on Wage and Unemployment Distribution in the Czech Republic

Jana Hančlová
Faculty of Economics, Sokolská tř. 33, 701 21 Ostrava 1
jana.hanclova@vsb.cz

Abstract

The minimum wage exists in the Czech Republic since 1991. From the point of view of the conventional theory of competitive labour markets minimum wage increases cause a reduction in employment. The paper deals with the effects of minimum wage increases in whole wage or (un)employment distribution. We shortly introduce our econometric model and our results for 3 years period 1998 – 2000.

Keywords

Minimum Wage, Wage Distribution, Unemployment , Econometric Modeling,
JEL: AH, BB

1 Introduction

The minimum wage exists in the Czech Republic since 1991. During the first years of economic transformation the minimum wage has been kept at the very low levels not adjusted for an economic development in the country. The real minimum wage has consequently fallen and in 1998 it reached 61.6 per cent of its initial 1991 real value. At the same time, the relation of the minimum wage to the average wage dropped from the initial 52.8 per cent in 1991 to 22.7 per cent in 1998.

In 1998, the government announced its objective to increase the minimum wage up to its "stimulating" level of 15 per cent above the subsistence minimum, with the long-term goal of minimum wage level close to 50 per cent of the average wage. Thus, the minimum wage started to increase sharply during the years 1999 – 2004, see Table 1, lifting the minimum wage/average wage ratio up to 40 per cent. The ratio reached its 1991 level only in the end of 2003 and it does not bind seriously[1]. Nevertheless, such large increases in minimum wage level obviously gave a rise to a discussion on possible adverse effects of such increases on employment and effects on wages and hours worked.

Table 1: Czech minimum wage developments in 1993 -2003

Czech Republic	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**
Minimum wage	2200	2200	2200	2500	2500	2650	3,250 3,600 *	4000 4500 *	5000	5700	6200 6700[2]
Minimum wage, increase in %	0.0	0.0	0.0	13.6	0.0	6.0	22.6 10.8 *	11.1 12.5 *	11.1	14.0	8.8 8.1
Net MW	1.903	1.903	1.903	2.188	2.188	2.319	2.844 3,114 *	3.412 3.783 *	4.184	4715	5087
MW ratio in % of gross av. earns.	37.8	31.9	26.9	25.8	23.4	22.7	25.7 28.4 *	29.7 33.4 *	36.3	35.9	38.0 40.1[3]

[1] No more than 1 per cent of employed workers earn a minimum wage, see: Eriksson and Pytlikova (2004).

Source: Ministry of Labour, Czech Statistical Office, own calculations
* changes made in July 1 of given year, [2] MW from January 2004 [3] MW ratio of gross average earnings in 1Q of 2004

From the point of view of the conventional theory of competitive labour markets minimum wage increases cause a reduction in employment. Nevertheless, recent empirical work questioned the standard economic wisdom, showing that moderate minimum wage increases may, on contrary, have positive employment effects, see e.g. Katz and Krueger (1992), Card (1992), Card and Krueger (1995). Besides the effects on employment, the minimum wage increases may cause a shift in whole wage distribution, which can further blow up the employment effects; see Grossman (1984).
There has been a paradoxical relation occurring on the Czech labour market: the level of a net minimum wage did not exceed the level of the minimum subsistence wage for most of the years during the period 1991 – 2003. Thus there was little motivation for the unemployed without qualification to look for low-paid jobs *vis-à-vis* the existing level of social benefits, a fact contributing to the emergence of "unemployment trap" in the Czech Republic. Thus the minimum wage increases could play a positive role from the “motivation” point of view and thus lead to employment increases.
Thus, in this chapter we would like to shed some light on these issues and we study the effects of minimum wage increases on wages, employment and hours worked during the years 1998-2000. We conduct our econometric analyses using matched employee-employer data set.
In the next section, we shortly survey the existing literature on minimum wage effects. In Section 2 we describe the data used for analysis and introduce our model. Results from our econometric analyses are given in the 3rd section and the last section concludes.

2 Existing Research

There has been a heated debate on the effects of minimum wage increases on wages and employment during the past decades. The economists are divided as regards their view on what is the exact effect of the minimum wages on economic outcomes, one group supporting the prediction of standard economic theory and some the monopsony predictions. The vast majority of research was carried out in the United States.
Starting with earlier research, the empirical findings went well along the lines of conventional economic theory view, which claims that minimum wage increases lead to employment decrease; see Brown, Gilroy and Kohen’s (1982) and Brown (1999) for a survey of the earlier and later research respectively.
However, Card (1992), Katz and Krueger (1992) and especially Card and Krueger (1995) they all find it hard to identify any employment effect at all and they move the question from „how negative is the minimum wage employment effect?“ to „is there any impact on employment at all?“. Thus, monopsony explanation gained some new attention. Nevertheless, these results met strong criticism from Neumark and Wascher (1994), who doubted the „natural experiments“ method as a proper tool to analyse the consequences of minimum wage changes on employment. The discussion continue to develop further in similar fashion as above.
As far as Europe is concerned, we can also find similar debate, however their intensity is far from the academic debate in the United States. For example, in the case of UK there have been some findings of negative effect of minimum wage increases, e.g. Bazen (1990) and recently by Steward (2002), as well as no negative effect at all, e.g. Machin and Manning (1992), Dickens, Machin and Mannig (1999). Some studies have been carried for France as well. Especially interesting is a study by Abowd, Kramarz and Margolis (1999) that uses the

same data structure as we do. They find some negative effects on future employment probability as the consequence of minimum wage increases.
The employment effect of minimum wage increases may be further affected by increases in overall wages as a consequence of rising minimum wage. The mechanism behind is so called 'ripple effect' that causes increases in other parts of the wage distribution besides the most influenced lower part of distribution. One of the first empirical studies aimed at measuring the ripple effect was carried out by Gramlich (1976) and further by Grosman (1984), who find relatively strong ripple effect of minimum wage increases. Nevertheless, rather opposite evidence has been found in studies by DiNardo, Fortin and Lemieux (1996) and by Lee (1999).
Recently there have been few studies on the effect of minimum wage increases on wages in UK, which have found that the minimum wage increases cause a compression in the wage distribution, especially at its' lower part, see e.g. Machin, Manning and Rahman (2003). Also work (Machin, Manning and Rahman (2003)) studied the effects of introduction of a National Minimum wage in April 1999 on one heavily affected sector, the residential care homes industry, in the United Kingdom. This sector contains a large number of low paid workers and results of this study suggest that minimum wage raised the wages of a large number of care home workers, causing a very big wage compression of the lower end of the wage distribution, thereby strongly reducing wage inequality.
As well paper Bellmann (1996) presented an international analysis of the share of long-term unemployed of all unemployed for a number of industrial countries (OECD) in the period 1985-1991. The comparison reveals that a larger wage disparity (i.e. a lower first decile ratio of the distribution of incomes) reduces long-term find any negative employment effect after the increase of minimum wages in the late eighties and early nineties.
One particular study, which is interesting from our viewpoint, is Neumark, Schweitzer and Washer (2000). This study investigates the effect of minimum wage increases on wage levels, number of hours worked, employment and total income. It shows that employees whose wages are close to the minimum wage are generally adversely effected by its increase; their hours worked and employment decrease as a result of declining labour demand for this category of workers. By contrast, the income of employees from a higher income category is little affected. In the present chapter, we intend to test how the Czech labour market adapts to quick changes in minimum wage level, using the above approach.
As for the Czech Republic, there have been two studies analyzing the effects of minimum wage increases. Buchtikova (1995) carried out a simulation in an effort to answer the question as to what is the reaction of employment on an increase in wages in industry due to minimum wage growth. She simulated that e.g. the growth of hourly minimum wage from 12 to 14 Czech crowns declines employment for 0,53%. It is necessary, however, to take into account the short time series of the study (1991 to 1993) and that the research was based upon statistics from state companies which is why the results cannot be applied to the economy as a whole.
Eriksson and Pytlikova (2004) – the authors find that the impact on firm wages is rather large and that there are some, but not substantial job losses in reaction to minimum wage increases. They use the same data set as we do, however contrary to Eriksson and Pytlikova (2004) we use look at the changes from the employee point of view.

3 Data and Econometric Model

We use the data from an employer-employee matched data set, which comes from the labour cost survey "Average Earnings Information System".[2] Thus data set contains comprehensive information on both employers and employees in the Czech Republic, for further details on the data set see

[2] Provided by Trexima Ltd. Zlín for the purposes of the Czech Statistical Office.

Gottvald et al (2002) and Eriksson and Pytlikova (2004). We build our analysis on 1.049.582 and 1.056.724 employees for years 1998 and 2000 respectively. We use 3 years of the data for our analysis and we compare the situation in 1999 with 1998 (99/98) and further 2000 with 1999 (00/99). From these original databases are filtered individuals involved in all analyzed years 1998, 1999 and 2000 (it is 309 343 individuals) to obtain a panel dataset[3]. For individuals are observed selected characteristics like education (*EDUC*), age (*AGE*), gender (*SEX*), average hourly wage (w) and two institutional variables regional rate of unemployment (u), region of 13 administrative regions (r). Variable *EDUC* contents 5 groups according to ISCED 97 (1. primary education, 2. apprenticeship 3. secondary with GCE, 4. higher post-secondary schools + bachelors, 5. university). For gender validates that woman = 0, man = 1.

We suppose that a growth in minimum wages will have an influence on wage distribution across nine groups of employees, earning certain multiples of the minimum wage:

Table 2: The Definition of Wage Distribution in Nine Groups

j=	1	2	3	4	5	6	7	8	9
$w_{i,1}$	$\leq 1.3MW_1$	$\leq 1.4MW_1$	$\leq 1.5MW_1$	$\leq 2MW_1$	$\leq 3MW_1$	$\leq 4MW_1$	$\leq 5MW_1$	$\leq 6MW_1$	$\leq 8MW_1$

Notes: $w_{i,1}$... average hourly wage; MW_1... minimum wage level.

Following these groups there were defined dummy variables $R_j(w_{i,1};MW_1)$:

- $R_j(w_{i,1};MW_1)=1$, if wage $w_{i,1}$ of i-individual belongs to j's group of wage distribution with respect to minimum wage MW_1,
- $R_j(w_{i,1};MW_1)=0$ otherwise.

The ninth group of wage distribution $R_9(w_{i,1};MW_1)$ and region 81 – Moravia-Silesian region are used as reference variables hence they are omitted by reason of multicollinearity.

When dividing our sample into the nine wage groups, we obtain the following descriptive information on the basic features of wage distribution (Table 3):

Table 3: Characteristics of Wage Distribution According to Nine Groups

	Average Age			Average Education[1)]			Average Wages[2)]		
j =	**98**	**99**	**00**	**98**	**99**	**00**	**98**	**99**	**00**
1	31.5	34.3	38.2	2.45	2.27	2.54	17.177	22.578	27.829
2	32.1	34.7	38.7	2.37	2.31	2.50	20.749	26.702	33.406
3	33.2	36.0	38.5	2.31	2.28	2.51	26.882	34.149	42.113
4	34.8	37.0	38.6	2.46	2.51	2.74	38.371	49.141	60.348
5	37.4	39.3	40.0	2.65	2.69	2.86	52.147	66.353	82.394
6	39.1	40.4	40.7	2.67	2.71	2.96	66.396	84.813	104.947
7	40.2	41.1	42.2	2.72	2.86	3.05	80.847	103.665	128.431
8	41.1	42.5	43.6	2.87	3.09	3.10	101.207	129.026	159.055
9	43.4	44.6	44.8	3.17	3.44	3.29	157.189	204.537	260.404

Source: Own calculations based on Trexima data-sample.

[3] Procedure how to create panel database was as follows: All chosen individual characteristics of employees are taken the same in all periods and a sequence of characteristics - Company Identification Number, ISCO88, SEX, AGE, EDUCATION. ISCO88 = 4-digit level, AGE – filter +1 year in the next year, EDUCATION – if not stated, to shift into the next filtration. Where the characteristics fit one individual in a company for each period, it is assumed it is the same person and the same individual identifier is allocated to such a person. With an increase in the number of employees in a company, the number of persons in a group also increases, since the most significant classifying characteristic, the profession, contains a higher number of the same professions. Then a new average hourly wage for such employees is calculated and a so-called 'group individual' created but in our estimation figures as one individual. The filtration is based on the precondition that companies and institutions behave towards an individual according to 'regulations', which are in principle identical for all employees with the same characteristics (that is why any employee with a wage differing from the others within a particular group by more than 50% is excluded from that group).

Notes: [1)] ISCED 1-5 (classification see above)
[2)] Hourly annual average wage in nine groups of wage distributions in CZK.

Table 3 tells us that low-paid workers are typically those with lower education levels, a finding that is consistent with common-sense reasoning. Less obviously, we see that average age in low-paid groups of workers is increasing dramatically, as is their relative weight in the whole sample (see Graph 1). This alone would signal that minimum wage affects a growing fraction of the labour force.
To analyse the affect of minimum wage increases on wage distribution, we make use of the modified Neumark, Schweitzer and Wascher (2000) model in the following form:

$$\Delta w_{i,2/1} = \alpha + \sum_{j=1}^{8} \beta_j \cdot \Delta MW_{2/1} \cdot R_j(w_{i,j}; MW_1) + \delta_1 \cdot AGE_{i,1} + \delta_2 \cdot AGE_{i,1}^2 + \delta_3 \cdot EDUC_{i,1} +$$

$$+ \delta_4 \cdot SEX_{i,1} + \sum_{r=1}^{12} \lambda_r \cdot D_{i,1,r} + \varepsilon_i ,$$

where

$$\Delta w_{i,2/1} = 100 \cdot \frac{w_{i,2} - w_{i,1}}{w_{i,1}}, \quad \text{and} \quad \Delta MW_{2/1} = 100 \cdot \frac{MW_2 - MW_1}{MW_1}.$$

$w_{i,1}$ - average hourly wage of i-individual in year *1*,

MW_1 - level of minimum wage in year *1*,

$R_j(w_{i,1}; MW_1)$ - dummy variable for the relative position of the year *1* wage ($w_{i,1}$) relative to the year 1 minimum wage (MW_1), where group j = 1, 2, …, 9,

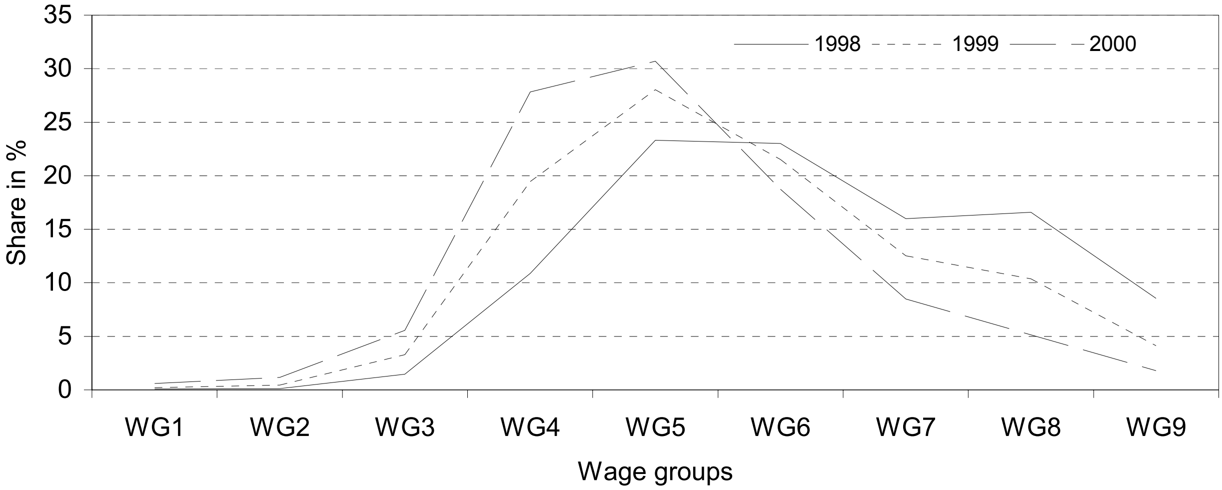
$D_{i,1,r} = 1$, if individual *i* is from region *r* in period 1,
= 0 otherwise (we include the fixed regional effects).

The subscripts *1* and *2* denote the year 1 and year 2, the subscript *i* means *individual i.* The parameters β_j in the first sum capture the effects of a given percent change in the minimum wage on each interval of the wage distribution defined by R_j.
Empirical analyses are carried out not only for dependent variable Δw, but alternatively also for Δu, which indicates the change in unemployment by region.

4 Empirical Results

As seen in Graph 1, the percentage share of individuals allocated in the wage groups near the minimum wage is increasing. This is shown by a movement of the curved lines to the left along with a significant increase in the number of individuals receiving wages which are near the minimum wage. With a growth in the minimum wage, the highest frequency of wage earners shifted from group 6 (four times the minimum wage) to group 5 (three times the minimum wage). The lowest wage group (1.3 to 1.5 of the minimum wage) increased its share in total employment from 0.13 % to 1.23 %, which is a very significant growth.
There is also a significant decrease in the number of individuals in the upper half of the wage distribution, in particular in the group with six times the minimum wage and more. Thus the overall nominal wage growth over the investigated period was relatively high, but clearly not high enough to compensate for the effect of the several times increased minimum wage.

Graph 1: Changes in Wage Distribution Related to Minimum Wage Levels, 1998-2000
(percentage share of employees in each of nine wage groups)



The increase in any wage of individual is influenced by numerous factors. The changing position of a certain individual in a particular group of wage distribution in a given period of time and in another one can be caused both by an increase in minimum wage and its effect on wage distribution, as well as by shifts depending on experience reached, education completed, or due to other reasons. The authors made no controls for such possible changes in career progress during the period investigated, and, therefore, the results presented combine all these influences. Thus the coefficients of elasticity are rather higher than those really corresponding to the actual effect of minimum wage increase. On the other hand, since individuals whose wages increased by more than 50% (presumably due to career progress) are excluded by filtration, the effects unrelated to minimum wage increase are limited. Below this limit, however, all factors are in fact demonstrated simultaneously.

Estimations, using OLS stepwise regressions method, in Table 4 indicate wage increases associated with an increase in minimum wage for a group of employed who are either paid by a minimum wage or wage just above its level (in the present case by its 1.3 multiple). The elasticity is higher than one, which means that the increase in wage evoked by the increase in minimum wage is even higher than its respective growth.

The estimates of β_j coefficients can be interpreted in a way that they measure the percentage change in wages, unemployment or hour worked resulting from a one percent increase in the minimum wage.

Wage elasticity regarding to wage distribution according to minimum wage in 99/98 and 00/99 decreases more or less exponentially however in 00/99 has weaker intensity. For example, in wage group just above minimum wage level the growth of minimum wage by 1% evokes growth of average hourly wage for 2,644% (99/98) and 1,567% (2000/99), ceteris paribus.

When the minimum wage increases, employees tend to improve their total wage by increasing the number of hours worked, i.e. offering an increased amount of work. Elasticity of number of hours worked regarding to wage distribution according to minimum wage in 99/98 and 00/99 has decreasing trend with exception of second group of wage distribution in 99/98. But we have to take care of very low level of coefficient of determination. This means that 1% growth of minimum wage evokes growth of average hours worked for 5,4 % in 99/98) and 2,2% in 2000/99 in the second group of probability distribution.

The growth in regional unemployment rates is a typical effect of minimum wage increase. The statistically indicated dependence for the first group of wage distribution in both periods could have been be predicted. The estimates are, however, low and/or insignificant (see Table 4).

Table 4: Effects of Minimum Wage Increase on Wages, Unemployment

Percent change in *MW* * dummy variable for:		Dependent variable = Δw		Dependent variable = Δu	
		99/98	**00/99**	**99/98**	**00/99**
$w \leq 1.3\ MW$	1	**2,644***** (0,048)	**1,567***** (0,028)	X	**0,033***** (0,012)
$1.3\ MW < w \leq 1.4\ MW$	2	**1,486***** (0,037)	**0,959***** (0,022)	**0,265***** (0,077)	**0,033***** (0,009)
$1.4\ MW < w \leq 1.5\ MW$	3	**0,985***** (0,013)	**0,554***** (0,010)	**0,084***** (0,024)	**0,029***** (0,004)
$1.5\ MW < w \leq 2\ MW$	4	**0,702***** (0,007)	**0,450***** (0,008)	**0,054***** (0,010)	**0,026***** (0,003)
$2\ MW < w \leq 3\ MW$	5	**0,482***** (0,006)	**0,341***** (0,007)	0,041*** (0,010)	**0,029***** (0,003)
$3\ MW < w \leq 4\ MW$	6	**0,368***** (0,006)	**0,259***** (0,007)	X	**0,018***** (0,003)
$4\ MW < w \leq 5\ MW$	7	**0,290***** (0.006)	**0,175***** (0,007)	**-0,013*** (0,008)	**0,009***** (0,003)
$5\ MW < w \leq 6\ MW$	8	**0,208***** (0,006)	**0,054***** (0,007)	X	**0,006**** (0,004)
R_{adj}^2		0.079	0.046	0.237	0.448
No. of observations		309,343	309,343	309,343	309,343

Notes: 10, 5 and 1 % significance levels are indicated by *, ** and ***, respectively. Standard errors are in parentheses. Dependent variable is annual percentage change in wages (average hourly wage per year; Δw), change in regional unemployment We include gender, education, age2 and regional dummies. The ninth minimum wage category has been excluded to avoid perfect multicollinearity. Estimates are OLS stepwise regressions. A panel is weighted to the original database structure in 1999.

5 Conclusion

Minimum wage could affect the wage structure and employment in many aspects, ranging from increasing unemployment among low-wage workers to compressing the entire wage structure, or the wage "ripple effect". In the Czech context, such studies are rather scarce. In this chapter we investigate the effect of minimum wage increases on wage distribution, number of hours worked and unemployment.

The growth in Czech minimum wage has confirmed its expected influence on wages and wage distribution. The entire wage distribution has started to slightly undulate due to the fact that wages of low-paid employees have grown significantly. This effect significantly weakens with a growing distance from the beginning of the wage distribution. This appears to confirm the assumption that the minimum wage was, after so many years of stagnation, so low and disconnected with the remaining relative wage structure that even its rapid growth has not prompted the ripple of the overall wage structure.

The effect of growing minimum wage on unemployment has only become slightly apparent for employees with the lowest wages and we cannot eliminate the effects unrelated to minimum wage increase.

References

[1] ABOWD, J., KRAMARZ, F., and MARGOLIS, D. (1999): Minimum Wages and Employment in France and United States. *NBER Working Paper,* No. 6996.

[2] BELLMANN, L. (1996): Wage differentiation and long-term unemployment – An International Comparison. NBER Working paper 19/1996.

[3] BROWN, CH., GILROY, C. and KOHEN, A. (1982): The Effect of the Minimum Wage on Employment and Uneployment. *Journal of Economic Literature,* 20, PP. 487-528.

[4] BROWN, CH. (1999): Minimum Wages: Employment and the Distribution of Income. In. Ashenfelder, O. and Card, D. (eds): *Handbook of Labour Economics*, North Holland.

[5] BUCHTIKOVA, A. (1995): Have minimum wages hit employment in the Czech transformation?, in Standing, G. and Vaughan-Whitehead, D., ed., Minimum wages in Central and Eastern Europe: From protection to destitution, *Central European University Press, ILO*.

[6] CARD, D. (1992): Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wages. *Industrial and Labor Relation Review,* 46, pp. 22-37.

[7] CARD, D., KRUEGER, A. (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey. *American Economic Review,* 84, pp. 772-793.

[8] CARD, D., KRUEGER, A. (1995): Myth and Measurement: The New Economics of the Minimum Wage. Princeton, New Jersey, Princeton University Press.

[9] DICKENS, R., MACHIN, S., MANNING, A. (1994): Estimating the Effect of Minimum Wages on Employment from the Distribution of Wages: A Critical View. *CEP Disscusion Paper no 02*03.

[10] DiNARDO, J., FORTIN, N., LEMIUX, T. (1996): Labour Market Institutions and the Distribution of Wages, 1973-1992: a Semiparametric Approach. *Econometrica 6*4, pp: 1001-1044.

[11] ERIKSSON, T., PYTLIKOVA, M. (2004): Firm-level consequences of large minimum-wage increases in the Czech and Slovak Republics. *Labour. Review of Labour Economics and Industrial Relations* Vol. 18, n° 1, pp: 75-103.

[12] GOTTVALD, J. et.al. (2002): Determinats of Individual Pay and Firms Pay Structures in the Czech and Slovak Republics. *VŠB-TU Ostrava.*

[13] GRAMLICH, E. (1976): Impact of Minimum Wages on Other Wages, Employment and Family Income. *Brooking Papers on Economic Activity* 7, pp: 409-451.

[14] GROSSMAN, J B. (1984): The Impact of the Minimum Wage on Other Wages. *Journal of Human Resources*, vol. 18 (Summer), pp. 359-378.

[15] KRUGMAN, P. (1994): Past and Prospective Causes of High Unemployment. Federal Reserve Bank of Kansas City Economic Review, 4th Q., pp: 23-43.

[16] LEE, CH., O¨ROURKE, B. (1999): The Impact of Minimum Wage Increases on Food and Kindred Products Prices: An Analysis of Price Pass-Through. *Technical Bulletin No. 1877.*

[17] MACHIN, S., MANNING, A. (1994): The effects of minimum wages on employment: theory and evidence from Britain. *NBER Working Paper* No. 4742.

[18] MACHIN, S., MANNING, A., Rahman, L. (2003): Where the Minimum Wage Bites Hard: Introduction of Minimum Wages ti a Low Wage Sector. *Journal of the European Economic Association 1*, pp. 156-180.

[19] NEUMARK, D., SCHWEITZER, M., WASCHER, W. (2000): The Effects of Minimum Wages throughout the Wage Distribution. NBER Working paper 7519.

Separation of convex polyhedral sets with uncertain data

Milan Hladík
Department of Applied Mathematics, Charles University,
Malostranské nám. 25, 118 00, Prague, Czech Republic
e-mail: milan.hladik@matfyz.cz

Abstract

This paper is a contribution to the interval analysis and separability of convex sets. Separation is a familiar principle and is often used not only in optimization theory, but in many economic applications as well. In real problems input data are usually not known exactly. For the purpose of this paper we assume that data can independently vary in given intervals. We study two cases when convex polyhedral sets are described by a system of linear inequalities or by the list of its vertices. For each case we propose a way how to check whether given convex polyhedral sets are separable for some or for all realizations of the interval data. Some of the proposed problems can be checked efficiently, while the others are NP-hard.

Keywords

Separating hyperplane, convex polyhedra, interval analysis, linear interval equation, linear interval inequalities.
JEL: C69

1 Introduction

In this paper we study separability of two convex polyhedral sets ($\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{C} \in \mathbb{R}^{l\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{d} \in \mathbb{R}^l$):

$$\mathcal{M}_1 \equiv \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \le \mathbf{b}\}, \tag{1}$$

$$\mathcal{M}_2 \equiv \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{C}\mathbf{x} \le \mathbf{d}\}. \tag{2}$$

There are various kinds of separability of convex sets (cf. [8]). We introduce so called weak and strong separation. Strong separation is dealt within the section 2 and this kind is especially convenient in order to utilize Theorem 2 and 3. Weak separation is dealt within the section 3.

Definition 1. Sets $X, Y \subset \mathbb{R}^n$ are called *weakly separable* if there exists a hyperplane $\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} = s\}$ such that $X \subseteq \overline{\mathcal{R}^-} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} \le s\}$, and $Y \subseteq \overline{\mathcal{R}^+} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} \ge s\}$ hold. Such a hyperplane $\mathcal{R}$ is called *the separating hyperplane* of the sets X, Y. Sets $X, Y \subset \mathbb{R}^n$ are called *strongly separable* if they are weakly separable and $dim\, X = dim\, Y = n$.

Let us remind the familiar separation theorem (see e.g. [2, 6]):

Theorem 1. *Convex sets X, $Y \subset \mathbb{R}^n$ are strongly separable if and only if $dim\, X = dim\, Y = n$, and $int\, X \cap int\, Y = \emptyset$.*

Let us introduce some notation. Vector $\mathbf{1}$ consists of ones, $diag(\mathbf{v})$ is a diagonal matrix with elements $v_1, \dots, v_n$. Given a matrix $\mathbf{M}$, the expressions $\mathbf{M}_{i,\cdot}$, $\mathbf{M}_{\cdot,j}$ denote i-th row and j-th column of the matrix $\mathbf{M}$, respectively. For vectors $\mathbf{a}, \mathbf{b}$ the inequalities $\mathbf{a} \le \mathbf{b}$ or $\mathbf{a} < \mathbf{b}$ are understood componentwise. For any set $\mathcal{X}$ let us denote by $\overline{\mathcal{X}}$, $int\, \mathcal{X}$, $dim\, \mathcal{X}$, and $conv\, \mathcal{X}$ the closer, the interior, the dimension, and the convex hull of $\mathcal{X}$, respectively. A sign of a real number $r \in \mathbb{R}$ is defined as follows: $sgn(r) = 0$ if $r = 0$, $sgn(r) = 1$ if $r > 0$ and $sgn(r) = -1$ if $r < 0$.

Let us introduce a convex polytope

$$\mathcal{Q}^* \equiv \left\{(\mathbf{u}, \mathbf{v}, v_{l+1}) \in \mathbb{R}^{m+l+1} \mid \begin{pmatrix} \mathbf{A}^T & \mathbf{C}^T & \mathbf{0} \\ \mathbf{b}^T & \mathbf{d}^T & 1 \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \\ 1 \end{pmatrix},\ (\mathbf{u}, \mathbf{v}, v_{l+1}) \ge \mathbf{0} \right\}. \tag{3}$$

With help of the set $\mathcal{Q}^*$ we can describe all separating hyperplanes of $\mathcal{M}_1,\ \mathcal{M}_2$ from $(1), (2)$. Theorems 2 and 3 comes from [4, 5].

Theorem 2. *Suppose that* $dim\,\mathcal{M}_1 = dim\,\mathcal{M}_2 = n$, $int\,\mathcal{M}_1 \cap int\,\mathcal{M}_2 = \emptyset$. *Let* $(\mathbf{u}, \mathbf{v}, v_{l+1}) \in \mathcal{Q}^*$, $\mathbf{u}^T\mathbf{A} \neq \mathbf{0}^T$, *and* $\eta \in \langle 0, v_{l+1}\rangle$ *is arbitrary. Then*

$$\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{u}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \eta\} \tag{4}$$

represents a separating hyperplane of the convex polyhedral sets $\mathcal{M}_1, \ \mathcal{M}_2$. *Conversely, any separating hyperplane* $\mathcal{R}$ *of* $\mathcal{M}_1, \ \mathcal{M}_2$ *we can express in the form of* (4) *for a certain* $(\mathbf{u}, \mathbf{v}, v_{l+1}) \in \mathcal{Q}^*$, $\mathbf{u}^T\mathbf{A} \neq \mathbf{0}^T$, *and* $\eta \in \langle 0, v_{l+1}\rangle$.

Theorem 3. *Let* $dim\,\mathcal{M}_1 = dim\,\mathcal{M}_2 = n$. *Then the convex sets* $\mathcal{M}_1$, $\mathcal{M}_2$ *are strongly separable if and only if* $\mathcal{Q}^* \neq \emptyset$.

1.1 Some results from interval analysis

Coefficients and right-hand sides of systems of linear equalities and inequalities are rarely known exactly. In interval analysis we suppose that these values vary independently in some real intervals. Let us introduce some notion. Interval matrix is defined as $\mathbf{M}^I = \{\mathbf{M} \in \mathbb{R}^{m\times n} \mid \underline{\mathbf{M}} \leq \mathbf{M} \leq \overline{\mathbf{M}}\}$. Next indroduce

$$\mathbf{M}^c \equiv \frac{1}{2}.(\overline{\mathbf{M}} + \underline{\mathbf{M}}), \quad \mathbf{M}^\Delta \equiv \frac{1}{2}.(\overline{\mathbf{M}} - \underline{\mathbf{M}}).$$

Then we can write

$$\mathbf{M}^I = \langle \underline{\mathbf{M}}, \overline{\mathbf{M}}\rangle = \langle \mathbf{M}^c - \mathbf{M}^\Delta, \mathbf{M}^c + \mathbf{M}^\Delta\rangle.$$

In the view of interval analysis there are two possibilities how to deal with the problem of finding a solution of interval linear system of equalities and inequalities. The system of interval linear inequalities

$$\mathbf{M}^I\mathbf{x} \leq \mathbf{m}^I \tag{5}$$

is *strongly solvable*, if every system $\mathbf{M}\mathbf{x} \leq \mathbf{m}$ is solvable for all $\mathbf{M} \in \mathbf{M}^I$, $\mathbf{m} \in \mathbf{m}^I$. Vector $\mathbf{x}^0$ is a *strong solution*, if $\mathbf{M}\mathbf{x}^0 \leq \mathbf{m}$ holds for all $\mathbf{M} \in \mathbf{M}^I$, $\mathbf{m} \in \mathbf{m}^I$. The interval system (5) is *weakly solvable*, if $\mathbf{M}\mathbf{x}^1 \leq \mathbf{m}$ holds for certain vector $\mathbf{x}^1$ and $\mathbf{M} \in \mathbf{M}^I$, $\mathbf{m} \in \mathbf{m}^I$ (such a vector $\mathbf{x}^1$ is called *a weak solution*). Similarly, we can define strong and weak solvability for other types of linear interval systems.

Theorem 4. *An interval system* $\mathbf{M}^I\mathbf{x} = \mathbf{m}^I, \ \mathbf{x} \geq \mathbf{0}$ *is weakly solvable if and only if the system*

$$\underline{\mathbf{M}}\mathbf{x} \leq \overline{\mathbf{m}}, \ \overline{\mathbf{M}}\mathbf{x} \geq \underline{\mathbf{m}}, \ \mathbf{x} \geq \mathbf{0} \tag{6}$$

is solvable. Moreover, a vector $\mathbf{x}$ *is a weak solution of the system* $\mathbf{M}^I\mathbf{x} = \mathbf{m}^I, \ \mathbf{x} \geq \mathbf{0}$ *if and only if it satisfies* (6).

Proof. See [10, Theorem 1.13]. □

Theorem 5. *An interval system* $\mathbf{M}^I\mathbf{x} = \mathbf{m}^I$, $\mathbf{x} \geq \mathbf{0}$ *is strongly solvable if and only the system*

$$(\mathbf{M}^c - diag(\mathbf{z})\mathbf{M}^\Delta)\mathbf{x} = \mathbf{m}^c + diag(\mathbf{z})\mathbf{m}^\Delta, \quad \mathbf{x} \geq \mathbf{0}$$

is solvable for each $\mathbf{z} \in \{\pm 1\}^m$.

Proof. See [10, Theorem 1.16]. □

Theorem 6. *An interval system* $\mathbf{M}^I\mathbf{x} \leq \mathbf{m}^I$ *is weakly solvable if and only if the system*

$$(\mathbf{M}^c - \mathbf{M}^\Delta diag(\mathbf{z}))\mathbf{x} \leq \overline{\mathbf{m}} \tag{7}$$

is solvable for some $\mathbf{z} \in \{\pm 1\}^n$. *Moreover, a vector* $\mathbf{x}$ *is a weak solution of the system* $\mathbf{M}^I\mathbf{x} \leq \mathbf{m}^I$ *if and only if it satisfies* (7).

Proof. See [10, Theorem 1.19]. □

Corollary 1. *An interval system* $\mathbf{M}^I\mathbf{x} \le \mathbf{0},\ \mathbf{x} \ne \mathbf{0}$ *is weakly solvable if and only if the system*

$$(\mathbf{M}^c - \mathbf{M}^{\Delta} diag(\mathbf{z}))\mathbf{x} \le \mathbf{0},\ \mathbf{x} \ne \mathbf{0}$$

is solvable for some $\mathbf{z} \in \{\pm 1\}^n$*.*

Theorem 7. *An interval system* $\mathbf{M}^I\mathbf{x} \le \mathbf{m}^I$ *is strongly solvable if and only if the system*

$$\overline{\mathbf{M}}\mathbf{x}^1 - \underline{\mathbf{M}}\mathbf{x}^2 \le \underline{\mathbf{m}}, \quad \mathbf{x}^1, \mathbf{x}^2 \ge \mathbf{0} \tag{8}$$

is solvable. Moreover, if $\mathbf{x}^1, \mathbf{x}^2$ *is a solution of* (8)*, then the vector* $\mathbf{x}^1 - \mathbf{x}^2$ *is a strong solution of the given interval system.*

Proof. See [10, Theorem 1.22 and 1.23]. □

2 Separation of interval convex polyhedral sets

In this section we deal with the strong separability of two convex polyhedral sets $\mathcal{M}_1$, $\mathcal{M}_2$ the input data of which can vary in given real intervals. Let us consider two families of convex polyhedral sets

$$\mathcal{M}_1^I \equiv \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}^I\mathbf{x} \le \mathbf{b}^I\}, \tag{9}$$

$$\mathcal{M}_2^I \equiv \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{C}^I\mathbf{x} \le \mathbf{d}^I\}, \tag{10}$$

where $\mathbf{A}^I = \{\mathbf{A} \in \mathbb{R}^{m\times n} \mid \underline{\mathbf{A}} \le \mathbf{A} \le \overline{\mathbf{A}}\}$, $\mathbf{C}^I = \{\mathbf{C} \in \mathbb{R}^{l\times n} \mid \underline{\mathbf{C}} \le \mathbf{C} \le \overline{\mathbf{C}}\}$, $\mathbf{b}^I = \{\mathbf{b} \in \mathbb{R}^m \mid \underline{\mathbf{b}} \le \mathbf{b} \le \overline{\mathbf{b}}\}$, $\mathbf{d}^I = \{\mathbf{d} \in \mathbb{R}^l \mid \underline{\mathbf{d}} \le \mathbf{d} \le \overline{\mathbf{d}}\}$. Matrices $\underline{\mathbf{A}}, \overline{\mathbf{A}} \in \mathbb{R}^{m\times n}$, $\underline{\mathbf{C}}, \overline{\mathbf{C}} \in \mathbb{R}^{l\times n}$ and vectors $\underline{\mathbf{b}}, \overline{\mathbf{b}} \in \mathbb{R}^m$, $\underline{\mathbf{d}}, \overline{\mathbf{d}} \in \mathbb{R}^l$ are fixed.

Let us assume that no matrix $\mathbf{A} \in \mathbf{A}^I$ contain the zero row and next assume that $dim\,\mathcal{M}_1 = n$ holds for all $\mathcal{M}_1 \in \mathcal{M}_1^I$ (i.e., the interval system $\mathbf{A}^I\mathbf{x} < \mathbf{b}^I$ is strongly solvable). Analogical assumptions for $\mathcal{M}_2^I$.

The former assumption can be verified easily, the latter assumption can be verified in the following way. The dimension of $\mathcal{M}_1$ is equal to n for all $\mathcal{M}_1 \in \mathcal{M}_1^I$ if and only if the interval system $\mathbf{A}^I\mathbf{x} \le \mathbf{b}^I - \varepsilon, \varepsilon > \mathbf{0}$ infinitesimal, is strongly solvable. According to Theorem 7 this happens if and only if the system

$$\overline{\mathbf{A}}\mathbf{x}^1 - \underline{\mathbf{A}}\mathbf{x}^2 \le \underline{\mathbf{b}} - \varepsilon,\ \mathbf{x}^1, \mathbf{x}^2 \ge \mathbf{0},$$

or, equivalently, the system

$$\overline{\mathbf{A}}\mathbf{x}^1 - \underline{\mathbf{A}}\mathbf{x}^2 < \underline{\mathbf{b}},\ \mathbf{x}^1, \mathbf{x}^2 \ge \mathbf{0} \tag{11}$$

is solvable. Moreover, if vectors $\tilde{\mathbf{x}}^1$, $\tilde{\mathbf{x}}^2$ form the solution of (11), then $\tilde{\mathbf{x}}^1 - \tilde{\mathbf{x}}^2 \in int\,\mathcal{M}_1$ holds for all $\mathcal{M}_1 \in \mathcal{M}_1^I$.

We are interested in two cases. We will study whether the convex polyhedral sets $\mathcal{M}_1$, $\mathcal{M}_2$ are strong separable either for some, or for all realization $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$.

2.1 Separability for some realization

The first case can be checked efficiently. According to Theorem 3, there exists two convex polyhedral sets $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$ which are strong separable if and only if the interval system

$$\begin{pmatrix} (\mathbf{A}^I)^T & (\mathbf{C}^I)^T & \mathbf{0} \\ (\mathbf{b}^I)^T & (\mathbf{d}^I)^T & 1 \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \\ 1 \end{pmatrix},\ (\mathbf{u}, \mathbf{v}, v_{l+1}) \ge \mathbf{0}. \tag{12}$$

is weakly solvable. From Theorem 4 we have that interval system (12) is weakly solvable if and only if the system

$$\begin{pmatrix} \underline{\mathbf{A}}^T & \underline{\mathbf{C}}^T & \mathbf{0} \\ \underline{\mathbf{b}}^T & \underline{\mathbf{d}}^T & 1 \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix} \le \begin{pmatrix} \mathbf{0} \\ 0 \\ 1 \end{pmatrix} \le \begin{pmatrix} \overline{\mathbf{A}}^T & \overline{\mathbf{C}}^T & \mathbf{0} \\ \overline{\mathbf{b}}^T & \overline{\mathbf{d}}^T & 1 \\ \mathbf{1}^T & \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix},\ \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix} \ge \mathbf{0}$$

is solvable.

2.2 Separability for all realizations

To verify whether all convex polyhedral sets $\mathcal{M}_1 \in \mathcal{M}_1^I, \mathcal{M}_2 \in \mathcal{M}_2^I$ are strongly separable, it is equivalent (see Theorem 3) to verify whether interval system (12) is strongly solvable. Theorem 5 enables us to check strong solvability of the interval system (12) with the exponential complexity. Polynomial complexity algorithm is not likely to exist; we will show that this problem is NP-hard.

Interval system (12) is strongly solvable if and only if the interval system

$$\begin{pmatrix} (\mathbf{A}^I)^T & (\mathbf{C}^I)^T & \mathbf{0} \\ (\mathbf{b}^I)^T & (\mathbf{d}^I)^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ v_{l+1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \ (\mathbf{u}, \mathbf{v}, v_{l+1}) \gneqq \mathbf{0} \tag{13}$$

is strongly solvable. It follows from Theorem 8 that checking the strong solvability of the interval system (13) is NP-hard problem.

Lemma 1. *Let $\mathbf{M} \in \mathbb{Q}^{n\times n}$ be a nonnegative positive definite matrix. Checking the solvability of the system*

$$|\mathbf{M}\mathbf{x}| \le \mathbf{1},\ \mathbf{1}^T|\mathbf{x}| > 1 \tag{14}$$

is a NP-hard problem.

Proof. It is a modification of the proof of Theorem 1.3 from [10] on the NP-hardness of testing the solvabiliy of a system $|\mathbf{M}\mathbf{x}| \le \mathbf{1},\ \mathbf{1}^T|\mathbf{x}| \ge 1$. □

Theorem 8. *Let $\mathbf{N}^I \subset \mathbb{R}^{n\times 2n}$. Checking strong solvability of an interval system*

$$\mathbf{N}^I\mathbf{x} = \mathbf{0},\ \mathbf{x} \gneqq \mathbf{0} \tag{15}$$

is NP-hard problem.

Proof. From Lemma 1 we have that checking solvability of the system (14) is NP-hard. Thus it is sufficient to prove that the system (14) has a solution if and only if an interval system

$$\langle \mathbf{M}^T - \mathbf{1}\mathbf{1}^T, \mathbf{M}^T + \mathbf{1}\mathbf{1}^T\rangle \mathbf{x}^1 + \langle -\mathbf{M}^T - \mathbf{1}\mathbf{1}^T, -\mathbf{M}^T + \mathbf{1}\mathbf{1}^T\rangle \mathbf{x}^2 = \mathbf{0},\ (\mathbf{x}^1, \mathbf{x}^2) \gneqq \mathbf{0}$$

is not strongly solvable, or equivalently an interval system

$$\begin{aligned} \langle \mathbf{M}^T - \mathbf{1}\mathbf{1}^T, \mathbf{M}^T + \mathbf{1}\mathbf{1}^T\rangle \mathbf{x}^1 + \langle -\mathbf{M}^T - \mathbf{1}\mathbf{1}^T, -\mathbf{M}^T + \mathbf{1}\mathbf{1}^T\rangle \mathbf{x}^2 &= \mathbf{0}, \\ \mathbf{1}^T\mathbf{x}^1 + \mathbf{1}^T\mathbf{x}^2 = 1,\ (\mathbf{x}^1, \mathbf{x}^2) &\ge \mathbf{0} \end{aligned} \tag{16}$$

is not strongly solvable (it is a special kind of the interval system (15)). The system (16) is not strongly solvable if and only if there exists $\mathbf{y} \in \{\pm 1\}^n$ such that a system

$$\begin{aligned} (\mathbf{M}^T - \mathbf{y}\mathbf{1}^T)\mathbf{x}^1 + (-\mathbf{M}^T - \mathbf{y}\mathbf{1}^T)\mathbf{x}^2 &= \mathbf{0}, \\ \mathbf{1}^T\mathbf{x}^1 + \mathbf{1}^T\mathbf{x}^2 = 1,\ (\mathbf{x}^1, \mathbf{x}^2) &\ge \mathbf{0} \end{aligned}$$

is not solvable (see [10, Theorem 1.16]). From the familiar Farkas Theorem it is sufficient and necessary that there exists a vector $(\mathbf{z}, z') \in \mathbb{R}^{n+1}$ satisfying the system

$$\begin{aligned} (\mathbf{M} - \mathbf{1}\mathbf{y}^T)\mathbf{z} + \mathbf{1}z' &\ge \mathbf{0}, \\ (-\mathbf{M} - \mathbf{1}\mathbf{y}^T)\mathbf{z} + \mathbf{1}z' &\ge \mathbf{0}, \\ z' &< 0, \end{aligned}$$

equivalently

$$\begin{aligned} (\mathbf{M} - \mathbf{1}\mathbf{y}^T)\mathbf{z} &> \mathbf{0}, \\ (-\mathbf{M} - \mathbf{1}\mathbf{y}^T)\mathbf{z} &> \mathbf{0}, \end{aligned}$$

or

$$|\mathbf{Mz}| < -\mathbf{1y}^T\mathbf{z}. \tag{17}$$

We claim that (17) has a solution if and only if the system (14) has a solution. If $\mathbf{x} \in \mathbb{R}^n$ solves (14), then it satisfies $|\mathbf{Mx}| < \mathbf{11}^T|\mathbf{x}|$ and vectors $\mathbf{z} = \mathbf{x}$, $\mathbf{y} = -sgn(\mathbf{x})$ forms a solution of (17). Conversely, if a certain $\mathbf{z} \in \mathbb{R}^n$ a $\mathbf{y} \in \{\pm 1\}^n$ satisfies (17), then

$$|\mathbf{Mz}| < |-\mathbf{1y}^T\mathbf{z}| \leq \mathbf{11}^T|\mathbf{z}|,$$

i.e.

$$|\mathbf{M}\frac{\mathbf{z}}{\mathbf{1}^T|\mathbf{z}|}| < \mathbf{1},$$

and a solution of (14) is a vector $\mathbf{x} = \frac{\mathbf{z}}{\mathbf{1}^T|\mathbf{z}|-\varepsilon}$, where $\varepsilon > 0$ is infinitesimal. □

Another possibility how to verify whether each couple $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$ is strongly separable is the following sufficient condition. Each two convex polyhedral sets $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$ are strongly separable, if convex hulls $conv\left(\cup_{\mathcal{M}_1 \in \mathcal{M}_1^I}\mathcal{M}_1\right)$, $conv\left(\cup_{\mathcal{M}_2 \in \mathcal{M}_2^I}\mathcal{M}_2\right)$ are strongly separable. Moreover, any separating hyperplane of these convex hulls is a separating hyperplane of convex polyhedral sets $\mathcal{M}_1$, $\mathcal{M}_2$ for all $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$. From Theorem 6 it follows that a vector $\mathbf{x} \in \mathbb{R}^n$ solves a system $\mathbf{Ax} \leq \mathbf{b}$ for certain $\mathbf{A} \in \mathbf{A}^I$, $\mathbf{b} \in \mathbf{b}^I$ if and only if the vector $\mathbf{x}$ solves

$$(\mathbf{A}^c - \mathbf{A}^\Delta diag(\mathbf{z}))\mathbf{x} \leq \overline{\mathbf{b}}$$

for some $\mathbf{z} \in \{\pm 1\}^n$. Hence

$$\bigcup_{\mathcal{M}_1 \in \mathcal{M}_1^I} \mathcal{M}_1 = \bigcup_{\mathbf{z} \in \{\pm 1\}^n} \left\{\mathbf{x} \in \mathbb{R}^n \mid (\mathbf{A}^c - \mathbf{A}^\Delta diag(\mathbf{z}))\mathbf{x} \leq \overline{\mathbf{b}}\right\}$$

and the problem is reduced to the problem of computing convex hull of finite (but exponential) number of convex polyhedral sets (for explicit description of convex hull of two convex polyhedral sets see [4]).

Note that the reverse implication generally does not hold, i.e. there can be all convex polyhedral sets $\mathcal{M}_1 \in \mathcal{M}_1^I$, $\mathcal{M}_2 \in \mathcal{M}_2^I$ strongly separable and convex hulls $conv\left(\cup_{\mathcal{M}_1 \in \mathcal{M}_1^I}\mathcal{M}_1\right)$, $conv\left(\cup_{\mathcal{M}_2 \in \mathcal{M}_2^I}\mathcal{M}_2\right)$ need not be strongly separable. The reason is that sets $\cup_{\mathcal{M}_1 \in \mathcal{M}_1^I}\mathcal{M}_1$, $\cup_{\mathcal{M}_2 \in \mathcal{M}_2^I}\mathcal{M}_2$ are not generally convex.

3 Convex polytopes

In this section we suppose that convex polytopes (bounded convex polyhedral sets) $\mathcal{M}_1$, $\mathcal{M}_2$ are described by the lists of their vertices as follows

$$\mathcal{M}_1 \text{ has vertices } \mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n,\ m \geq 1, \tag{18}$$

$$\mathcal{M}_2 \text{ has vertices } \mathbf{c}_1, \ldots, \mathbf{c}_l \in \mathbb{R}^n,\ l \geq 1. \tag{19}$$

Denote as $\mathbf{A} \in \mathbb{R}^{m \times n}$ such a matrix, for which $\mathbf{A}_{i,\cdot} = \mathbf{a}_i^T$, $i \in \{1, \ldots, m\}$ (i.e. rows of matrix $\mathbf{A}$ correspond to vectors $\mathbf{a}_i^T$) and analogically $\mathbf{C} \in \mathbb{R}^{l \times n}$ is such a matrix for which $\mathbf{C}_{j,\cdot} = \mathbf{c}_j^T$, $j \in \{1, \ldots, l\}$. We will also use the more transparent notion $\mathcal{M}_1 \equiv \mathcal{M}_1(\mathbf{A})$, $\mathcal{M}_2 \equiv \mathcal{M}_2(\mathbf{C})$.

For this situation it is convenient to study weak separability (Definition 1) of convex polytopes $\mathcal{M}_1$, $\mathcal{M}_2$, since nonemptiness of $\mathcal{M}_1$, $\mathcal{M}_2$ is guaranteed (against to full dimension).

Checking the existence of separating hyperplane of convex polytopes $\mathcal{M}_1$, $\mathcal{M}_2$ in spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ can be done in expected time $O(\sqrt{m+l})$, which is optimal (see [1]). But for the sake of this paper it is more convenient to use the standard linear programming problem: convex polytopes $\mathcal{M}_1$, $\mathcal{M}_2$ are weakly separable if and only if a convex polyhedral set

$$\mathcal{D} \equiv \left\{(\mathbf{r}, s) \in \mathbb{R}^{n+1} \mid \begin{pmatrix} \mathbf{A} & -\mathbf{1} \\ -\mathbf{C} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{r} \\ s \end{pmatrix} \leq \mathbf{0},\ \mathbf{r} \neq \mathbf{0}\right\} \tag{20}$$

is nonempty (whereas $\mathbf{r}^T\mathbf{x} = s$ represents a separating hyperplane).

In interval analysis we suppose that given values can vary is some intervals, i.e. the matrix $\mathbf{A}$ comes from the set $\mathbf{A}^I = \{\mathbf{A} \in \mathbb{R}^{m\times n} \mid \underline{\mathbf{A}} \leq \mathbf{A} \leq \overline{\mathbf{A}}\}$, where matrices $\underline{\mathbf{A}}$, $\overline{\mathbf{A}} \in \mathbb{R}^{m\times n}$ are fixed. Analogically for $\mathbf{C} \in \mathbf{C}^I$.

Like in the section 2 two natural questions arise: Are convex polytopes $\mathcal{M}_1(\mathbf{A})$, $\mathcal{M}_2(\mathbf{C})$ weakly separable for some realization $\mathbf{A} \in \mathbf{A}^I$, $\mathbf{C} \in \mathbf{C}^I$? Are $\mathcal{M}_1(\mathbf{A})$, $\mathcal{M}_2(\mathbf{C})$ weakly separable for all realizations $\mathbf{A} \in \mathbf{A}^I$, $\mathbf{C} \in \mathbf{C}^I$.

3.1 Separability for some realizations

Theorem 9. *Given an interval matrix $\mathbf{M}^I \subset \mathbb{R}^{m\times n}$. Checking weak solvability of an interval system*

$$\mathbf{M}^I\mathbf{x} \leq \mathbf{0},\ \mathbf{x} \neq \mathbf{0} \tag{21}$$

is NP-hard problem.

Proof. We proceed analogically as in [10, Theorem 1.18]. A vector $\mathbf{x} \in \mathbb{R}^n$ is a weak solution of the interval system (21) if and only if it is a solution of a system

$$\mathbf{M}^c\mathbf{x} \leq \mathbf{M}^\Delta|\mathbf{x}|,\ \mathbf{x} \neq \mathbf{0}. \tag{22}$$

It is sufficient to prove that checking solvability of (22) is NP-hard. We known (see [10, Theorem 1.3]) that checking solvability of system

$$|\mathbf{N}\mathbf{x}| \leq \mathbf{1},\ 1 \leq \mathbf{1}^T|\mathbf{x}| \tag{23}$$

is NP-hard. We claim that system (23) is solvable if and only if the system

$$|\mathbf{N}\mathbf{z}| \leq \mathbf{1}z',\ z' \leq \mathbf{1}^T|\mathbf{z}|,\ (\mathbf{z}, z') \neq \mathbf{0} \tag{24}$$

is solvable. When $\mathbf{x}$ solves (23), then $(\mathbf{z}, z') = (\mathbf{x}, 1)$ solves (24). Conversely, let $(\mathbf{z}, z')$ be a solution of (24). If $z' \neq 0$, then $z' > 0$ and a vector $\mathbf{x} = \frac{\mathbf{z}}{z'}$ solves the system (23). If $z' = 0$, then the system (23) is satisfied for a vector $\mathbf{x} = \frac{\mathbf{z}}{\mathbf{1}^T|\mathbf{z}|}$. The system (24) can be equivalently written down as

$$\mathbf{N}\mathbf{z} - \mathbf{1}z' \leq \mathbf{0},\ -\mathbf{N}\mathbf{z} - \mathbf{1}z' \leq \mathbf{0},\ z' \leq \mathbf{1}^T|\mathbf{z}|,\ (\mathbf{z}, z') \neq \mathbf{0}.$$

By choosing

$$\mathbf{M}^c = \begin{pmatrix} \mathbf{N} & -\mathbf{1} \\ -\mathbf{N} & -\mathbf{1} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad \mathbf{M}^\Delta = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{1}^T & 0 \end{pmatrix}$$

we reduce solvability of (24) to solvability of (22). Hence checking weak solvability of the interval system (21) is NP-hard. ☐

Remark 1. In contrary to interval systems of other types for the interval system (21) we have that for $m \leq n$ weak solvability can be checked in constant time and the system

$$\mathbf{M}\mathbf{x} \leq \mathbf{0},\ \mathbf{x} \neq \mathbf{0} \tag{25}$$

is solvable for all $\mathbf{M} \in \mathbf{M}^I$. The set of solutions of system $\mathbf{M}\mathbf{x} + \mathbf{1}x' = \mathbf{0}$ forms a vector space of dimension at least one. Hence there exists a vector $(\mathbf{x}, x') \neq (\mathbf{0}, 0)$ satisfying this system. If $x' = 0$, then $\mathbf{x}$ solves (25). If $x' \neq 0$, then the system (25) has a solution $\frac{\mathbf{x}}{x'}$.

The convex polytopes $\mathcal{M}_1(\mathbf{A})$, $\mathcal{M}_2(\mathbf{C})$ are weakly separable for some $\mathbf{A} \in \mathbf{A}^I$, $\mathbf{C} \in \mathbf{C}^I$ if and only if the interval system

$$\begin{pmatrix} \mathbf{A}^I & -\mathbf{1} \\ -\mathbf{C}^I & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{r} \\ s \end{pmatrix} \leq \mathbf{0},\ \mathbf{r} \neq \mathbf{0} \tag{26}$$

is weakly solvable. But according to Theorem 9 this is NP-hard problem, since interval system $\mathbf{M}^I\mathbf{x} \le \mathbf{0}$, $\mathbf{x} \ne \mathbf{0}$ is weakly solvable if and only if the following system is weakly solvable

$$\begin{pmatrix} \mathbf{M}^I & -\mathbf{1} \\ \mathbf{0}^T & -1 \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ x' \end{pmatrix} \le \mathbf{0},\ \mathbf{x} \ne \mathbf{0}.$$

This is a special kind of interval system (26). For checking (with exponential complexity) weak solvability of (26) we can use Corollary 1.

3.2 Separability for all realizations

The convex polytopes $\mathcal{M}_1(\mathbf{A})$, $\mathcal{M}_2(\mathbf{C})$ are weakly separable for all $\mathbf{A} \in \mathbf{A}^I$ a $\mathbf{C} \in \mathbf{C}^I$ if and only if the interval system (26) is strongly solvable. If the interval system (26) has a strong solution $(\mathbf{r}, s)$, then simply (26) is strongly solvable and for the hyperplane $\mathcal{R} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} = s\}$ we have

$$\mathcal{M}_1(\mathbf{A}) \subset \overline{\mathcal{R}^-} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} \le s\} \quad \forall \mathbf{A} \in \mathbf{A}^I,$$
$$\mathcal{M}_2(\mathbf{C}) \subset \overline{\mathcal{R}^+} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{r}^T\mathbf{x} \ge s\} \quad \forall \mathbf{C} \in \mathbf{C}^I.$$

The reverse implication holds only under some assumptions – see Theorem 10.

Assertion 1. *The set* $\cup_{\mathbf{A}\in\mathbf{A}^I}\mathcal{M}_1(\mathbf{A})$ *is convex.*

Proof. Let $\mathbf{A}^1, \mathbf{A}^2 \in \mathbf{A}^I$ and $\mathbf{x}^1 \in \mathcal{M}_1(\mathbf{A}^1)$, $\mathbf{x}^2 \in \mathcal{M}_1(\mathbf{A}^2)$. Denote by $\mathbf{a}_i^1$ and by $\mathbf{a}_i^2$, $i = 1, \ldots, m$, vertices of $\mathcal{M}_1(\mathbf{A}^1)$ and $\mathcal{M}_1(\mathbf{A}^2)$, respectively. Then vectors $\mathbf{x}^1,\ \mathbf{x}^2$ can be expressed as convex combinations

$$\mathbf{x}^1 = \sum_{i=1}^m \alpha_i^1\mathbf{a}_i^1, \quad \mathbf{x}^2 = \sum_{i=1}^m \alpha_i^2\mathbf{a}_i^2$$

for certain $\alpha_i^1, \alpha_i^2 \ge 0$, $\sum_{i=1}^m \alpha_i^1 = \sum_{i=1}^m \alpha_i^2 = 1$. An arbitrary convex combination of $\mathbf{x}^1,\ \mathbf{x}^2$ in the form $\mathbf{x}^c \equiv c^1\mathbf{x}^1 + c^2\mathbf{x}^2$ (where $c^1, c^2 \ge 0,\ c^1 + c^2 = 1$) is equal to

$$\mathbf{x}^c = c^1\sum_{i=1}^m \alpha_i^1\mathbf{a}_i^1 + c^2\sum_{i=1}^m \alpha_i^2\mathbf{a}_i^2 = \sum_{i=1}^m (c^1\alpha_i^1 + c^2\alpha_i^2)\left(\frac{c^1\alpha_i^1}{c^1\alpha_i^1 + c^2\alpha_i^2}\mathbf{a}_i^1 + \frac{c^2\alpha_i^2}{c^1\alpha_i^1 + c^2\alpha_i^2}\mathbf{a}_i^2\right).$$

Denote

$$\mathbf{a}_i^c \equiv \frac{c^1\alpha_i^1}{c^1\alpha_i^1 + c^2\alpha_i^2}\mathbf{a}_i^1 + \frac{c^2\alpha_i^2}{c^1\alpha_i^1 + c^2\alpha_i^2}\mathbf{a}_i^2$$

(a vector $\mathbf{a}_i^c$ is a convex combination of $\mathbf{a}_i^1$ and $\mathbf{a}_i^2$). Define a matrix $\mathbf{A}^c$ as follows $\mathbf{A}^c_{i,\cdot} \equiv (\mathbf{a}_i^c)^T$. Then $\mathbf{A}^c \in \mathbf{A}^I$. Since $\sum_{i=1}^m (c^1\alpha_i^1 + c^2\alpha_i^2) = c^1\sum_{i=1}^m \alpha_i^1 + c^2\sum_{i=1}^m \alpha_i^2 = 1$, the vector $\mathbf{x}^c = \sum_{i=1}^m (c^1\alpha_i^1 + c^2\alpha_i^2)\mathbf{a}_i^c$ is a convex combination of vectors $\mathbf{a}_i^c$. Thus $\mathbf{x}^c \in \mathcal{M}_1(\mathbf{A}^c)$. □

Theorem 10. *Let* $dim\,\mathcal{M}_1(\mathbf{A}) = dim\,\mathcal{M}_2(\mathbf{C}) = n$ *for all* $\mathbf{A} \in \mathbf{A}^I$, $\mathbf{C} \in \mathbf{C}^I$. *Then the interval system* (26) *has a strong solution if and only if* (26) *is strongly solvable.*

Proof. If (26) has a strong solution, then the interval system (26) is simply strongly solvable. The second implication we prove by contradiction. Suppose that (26) has not any strong solution, it means that the intersection $(\cup_{\mathbf{A}\in\mathbf{A}^I}\mathcal{M}_1(\mathbf{A})) \cap (\cup_{\mathbf{C}\in\mathbf{C}^I}\mathcal{M}_2(\mathbf{C}))$ is of full dimension. Hence there is a vector $\mathbf{x}^1$ belonging to the interior of $(\cup_{\mathbf{A}\in\mathbf{A}^I}\mathcal{M}_1(\mathbf{A})) \cap (\cup_{\mathbf{C}\in\mathbf{C}^I}\mathcal{M}_2(\mathbf{C}))$. This vector $\mathbf{x}^1$ belongs to $\mathcal{M}_1(\mathbf{A}^1)$ for certain $\mathbf{A} \in \mathbf{A}^I$. From assumptions of the theorem we have that $\mathcal{M}_1(\mathbf{A}^1) \cap (\cup_{\mathbf{C}\in\mathbf{C}^I}\mathcal{M}_2(\mathbf{C}))$ is of full dimension. Analogically there is $\mathbf{C} \in \mathbf{C}^I$ such that $int\,\mathcal{M}_1(\mathbf{A}^1) \cap int\,\mathcal{M}_2(\mathbf{C}^1) \ne \emptyset$. Therefore (for choice $\mathbf{A}^1,\ \mathbf{C}^1$) the interval system (26) is not strongly solvable. □

The existence of a strong solution of the interval system (26) can be check by two ways. First, we can compute $\cup_{\mathbf{A}\in\mathbf{A}^I}\mathcal{M}_1(\mathbf{A})$, $\cup_{\mathbf{C}\in\mathbf{C}^I}\mathcal{M}_2(\mathbf{C})$ and check weak separability of these convex polytopes. Denoting by $\mathbf{a}_i^j$, $i \in \{1, \ldots, m\},\ j \in J$ ($|J| = 2^n$), vertices of $\mathbf{A}^I_{i,\cdot}$ we have

$$\bigcup_{\mathbf{A}\in\mathbf{A}^I}\mathcal{M}_1(\mathbf{A}) = conv\left(\bigcup_{i\in\{1,\ldots,m\}}\bigcup_{j\in J}\{\mathbf{a}_i^j\}\right).$$

We reduced computing the union of infinitely many convex polytopes to computing the convex hull of finitely many points (concretely $m2^n$). Analogically we can compute $\cup_{\mathbf{C}\in\mathbf{C}^I}\mathcal{M}_2(\mathbf{C})$.

The second way is the following one. The interval system (26) has a strong solution if and only if there is $\mathbf{y}\in\{\pm 1\}^n$ such that an interval system

$$\begin{pmatrix} \mathbf{A}^I & -\mathbf{1} \\ -\mathbf{C}^I & \mathbf{1} \\ \mathbf{y}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{r} \\ s \end{pmatrix} \le \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -1 \end{pmatrix} \tag{27}$$

has a strong solution. The interval system (27) has a strong solution (see Theorem 7) if and only if the system

$$\begin{pmatrix} \overline{\mathbf{A}} & -\mathbf{1} \\ -\underline{\mathbf{C}} & \mathbf{1} \\ \mathbf{y}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{r}^1 \\ s^1 \end{pmatrix} - \begin{pmatrix} \underline{\mathbf{A}} & -\mathbf{1} \\ -\overline{\mathbf{C}} & \mathbf{1} \\ \mathbf{y}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{r}^2 \\ s^2 \end{pmatrix} \le \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -1 \end{pmatrix}, \quad (\mathbf{r}^1, s^1, \mathbf{r}^2, s^2) \ge \mathbf{0}$$

or, equivalently the system

$$\begin{pmatrix} \overline{\mathbf{A}} & -\mathbf{1} \\ -\underline{\mathbf{C}} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{r}^1 \\ s^1 \end{pmatrix} - \begin{pmatrix} \underline{\mathbf{A}} & -\mathbf{1} \\ -\overline{\mathbf{C}} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{r}^2 \\ s^2 \end{pmatrix} \le \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{y}^T(\mathbf{r}^1 - \mathbf{r}^2) < 0, \quad (\mathbf{r}^1, s^1, \mathbf{r}^2, s^2) \ge \mathbf{0}$$

is solvable for some $\mathbf{y}\in\{\pm 1\}^n$. On the whole we obtain that there exists a strong solution of the interval system (26) if and only if the system

$$\begin{pmatrix} \overline{\mathbf{A}} & -\underline{\mathbf{A}} & -\mathbf{1} \\ -\underline{\mathbf{C}} & \overline{\mathbf{C}} & \mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{r}^1 \\ \mathbf{r}^2 \\ s \end{pmatrix} \le \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{r}^1 \ne \mathbf{r}^2, \quad \mathbf{r}^1, \mathbf{r}^2 \ge \mathbf{0} \tag{28}$$

is solvable. Moreover, if $(\mathbf{r}^1, \mathbf{r}^2, s)$ solves (28), then the vector $(\mathbf{r}^1 - \mathbf{r}^2, s)$ is the required strong solution of the interval systems (27) and (26).

Acknowledgement

The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s.

References

[1] Chazelle, B., Liu, D., Magen, A. (2003): *Sublinear geometric algorithms*, STOC '03, 531–540, AMC, San Diego.

[2] Grünbaum, B. (2003): *Convex polytopes*, Springer, New York.

[3] Grygarová, L. (1997): *A calculation of all separating hyperplanes of two convex polytopes*, Optimization, **41**, 57–69.

[4] Grygarová, L. (1998): *On a calculation of an arbitrary separating hyperplane of convex polyhedral sets*, Optimization, **43**, 93–112.

[5] Hladík, M. (2004): *Explicit description of all separating hyperplanes of two convex polyhedral sets with RHS-parameters*, in: J. Šafránková, (Ed.), WDS'04: Part I, Mathematics and computer sciences, Matfyzpress, Prague, 63–70.

[6] Kemp, M. C. and Kimura, Y. (1978): *Introduction to mathematical economics*, Springer, New York.

[7] Martos, B. (1975): *Nonlinear programming. Theory and methods*, Akadémiai Kiadó, Budapest.

[8] Klee, V. (1969): *Separation and support properties of convex sets – a survey*, in: A. V. Balakrishnan, ed., Control theory and the calculus of variations, Academic Press, New York, 235–303.

[9] Rohn, J. (2000): *Computing the norm $\|A\|_{\infty,1}$ is NP-hard*, Linear and Multilinear Algebra **47**, No. 3, 195–204.

[10] Rohn, J. (2002): *Systems of interval linear equations and inequalities (rectangular case)*, Technical report No. 875, 1–69, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague.

Stylized Facts of Business Cycle in the Czech Republic*

Miroslav Hloušek
Faculty of Economics and Administration of Masaryk University,
Department of Applied Mathematics and Computer Science, Lipová 41a,
602 00 Brno, e-mail: hlousek@econ.muni.cz

Abstract
This paper deals with identification of stylized facts of Czech business cycle. Empirical time series are decomposed into trend and cyclical component using bandpass filter. Cross-correlations between cyclical component of GDP and of various time series are computed. Behaviour of the series over the cycle is determined, leading and lagging indicators are identified. The Granger causality between GDP and other aggregate variables is tested.

Keywords
Czech economy, stylized facts, bandpass filter, trend and cyclical components, Granger causality

JEL: C82, E32

1 Introduction

This paper tries to identify some stylized facts of business cycle in the Czech Republic. Comparison of model implications with stylized facts is useful for judgement of alternative economic theories and this paper contributes to it.

The paper is organized as follows. Section 2 introduces the reader into filtration of time series. Section 3 briefly describes data used for analysis and their transformation. Several characteristics describing business cycle fluctuations are presented in Section 4. Section 5 summarizes behaviour of real and nominal variables over the cycle and outlines differences among theories and facts in other countries. Final section discusses limitations of the analysis and suggests prospects for further research.

2 Decomposition procedure

The definition of business cycle fluctuations adopted in this paper is the deviation of actual time series from their long-run trends. These cyclical fluctuations are referred to as growth cycles. The linear filter is used to distinguish between the trend and cyclical components of economic time series. Here, I adopt the perspective of Baxter and King (1999) which draws on the theory of spectral analysis of time series data. The cyclical component can be thought of as those movements in the series associated with periodicities within a certain range of business cycle duration. I define this range of business cycle periodicities to be between six and thirty two quarters.[1] The ideal filter would preserve these fluctuations but would eliminate all other fluctuations, both the high frequency fluctuations (periods less than six quarters) associated

*This paper has been worked as a part of research activities at the grant project of GACR No. 402/05/2172 and with support of MŠMT project Research centers 1M0524.

[1]This setting is in accordance with Stock and Watson (1998) for better comparison of the results.

for example with measurement error and the low frequency fluctuations (periods exceeding eight years) associated with trend growth. This ideal filter cannot be implemented to finite data sets because it requires an infinite number of past and future values of the series. However, a feasible (finite-order) filter can be used to approximate this ideal filter. One widely used filter among macroeconomist is Hodrick-Prescott (1997) filter. However, this filter passes much of the high-frequency noise into the business cycle frequency band. The filter used in this study is bandpass filter designed by Christiano and Fitzgerald (1999), which mitigates this problem. For comparison, output gap estimated by these two filters can be seen in Figure 1 in Appendix.

Another problem of univariate filters generally is that data at the beginning and the end of time series are relatively poorly estimated and it is recommended to drop them out. Due to short data series of the Czech Republic I use all estimated values.

3 Data

I use data from the Czech National Bank and the Czech Statistical Office databases. The data series are seasonally adjusted using Kalman smoother. Frequency of data is quarterly; sample period is from 1995Q1 to 2005Q4. Most of the series were transformed by taking logarithm. Interest rate, unemployment rate and inflation rates are used without transformation. The cyclical components, usually referred to as gaps, are expressed as percentage deviation from trend.

The most important measure of business cycle are fluctuations in aggregate output. Thus real GDP is used as a benchmark for comparison of behaviour of time series within the cycle. The comovement between each series and real GDP is therefore examined.

4 Characteristics

The degree of comovement is quantitatively measured by cross-correlation of the cyclical component of each series with the cyclical component of real GDP. The magnitude of correlation coefficient indicates whether the variable is procyclical, countercyclical, or acyclical. The correlation is also calculated with phase shift up to five periods (forward and backward) which indicates if the variable is leading or lagging the cycle of GDP.

Specifically, the correlation is computed between y_t and x_{t+k} where y_t is the gap of GDP and x_{t+k} is the gap of relevant variable (both filtered by bandpass filter and expressed as deviation from trend value). A large positive correlation indicates procyclical behaviour of the series; a large negative correlation indicates countercyclical behaviour. A value of zero indicates absence of correlation: acyclical behaviour. For $k < 0$ the variable leads and for $k > 0$ the variable lags the cycle of output by k quarters. For example, if for some series the correlation is positive and maximum is at $k = -2$, it indicates that the variable is procyclical and tends to peak 2 quarters before real GDP. Further, the test of statistical significance of correlation coefficient is made.[2] The results are presented in Table 1. For better orientation, the largest absolute value of correlation coefficient is underlined, correlation coefficient that is statistically different from 0 is emphasized in bold. In next text I also distinguish if the variable is weakly ($|\rho| < 0.5$) or strongly ($|\rho| > 0.5$) correlated. This distinguishing is subjective and is not statistically tested.

Standard deviations of the cyclical component of each of the series are used as a measurement of variability. These values are also shown in Table 1.

Finally, the Granger (1969) causality between the cyclical component of GDP and of other variables is tested. The causality is examined in both directions. The test is based on adding of past values of explanatory variables into regression equation and testing if these variables improve explanatory power of

[2]The hypothesis of uncorrelation of the variables (which is equivalent to $\rho = 0$) is tested. The test statistic follows t-distribution; the test is two-tailed with 95 percent confidence intervals: $t_{stat} = \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{n-2}$, where ρ is the correlation coefficient and n is the number of observations.

the regression.[3] The number of lagged variables is set from one to five which corresponds to the phase shift calculated for correlations. Due to shortage of place the results are not presented in table but only mentioned in the text, if appropriate.

5 Stylized facts

The variables in Table 1 are sorted in the following way: GDP and its components, other real variables and nominal variables. Conclusions about behaviour of variables over the cycle are summarized and distinctions between stylized facts in the Czech Republic and other developed economies are mentioned.[4] The relationship between economic theory and facts is discussed.

5.1 Real facts

First, we look at the behaviour of GDP gap. Value of first order autocorrelation coefficient is 0.92 which is quite high compared to e.g. the United States where this value is about 0.85. It indicates certain persistence and rigid behaviour of this variable. For example, when shock hits the economy, it takes more time for GDP to return to its potential level.

Real consumption and investment are both strongly procyclical and lag output by one quarter. Investment is much more volatile than output. Consumption is usually more stable which is theoretically explained by smoothing behaviour of economic agents. In the case of the Czech Republic, consumption is also more volatile than output and it contradicts observations in other countries. Change of inventory stock (measured as share to GDP) is rather procyclical and lags output by one quarter. None of these variables has predictive content with respect to GDP.

Government expenditures are acyclical variable (the correlation coefficient is not statistically significant). It would indicate that government does not influence output and does not create political business cycle, at least on aggregate level. However, Granger causality test shows that government expenditures help to predict output for lags of two and four quarters. Thus, no clear conclusion can be made from these results.

Exports are weakly positively correlated and leads the cycle of GDP by two quarters. Imports are strongly procyclical and coincide with the cycle. The behaviour of imports is in accordance with economic theory – the volume of imports depends positively on domestic income.[5] Exports should depend on foreign income and moderate positive correlation with domestic output may come from synchronization of business cycles across countries. From the causality test is seen that output helps to predict imports; it was proved for lags from two to four quarters. Both exports and imports are much more volatile than output. Net exports (share to GDP) are weakly procyclical and leading variable.

Real wage is weakly procyclical and coincident variable. This is consistent with behaviour of the real wage in other countries. From the point of view of economic theory this stylized fact is in favour of real business cycle theory and monetary approach with sticky prices (rather than wages) assumptions.

Behaviour of real interest rate is tricky. The largest negative value of correlation coefficient is for shift of four quarters ahead, which indicates countercyclical and leading real interest rate. However, real interest rate is not contemporaneously correlated with the cycle of output and additionally there is rather important positive correlation four or five quarters behind the output. The real interest rate is usually assumed as

[3]The F statistic is used for this test:

$$F_{stat} = (n-k)\frac{\text{ESS}_R - \text{ESS}_{UR}}{q(\text{ESS}_{UR})}$$

where ESS_R and ESS_{UR} are the sums of squared residuals in the restricted and unrestricted regressions, respectively; n is the number of observations; k is the number of estimated parameters in the unrestricted regression; and q is the number of parameter restrictions. This statistic is distributed as $F(q, n-k)$. The significance level is 5 %. For more details about the test see e.g. Pindyck and Rubinfeld (1998)

[4]United States are usually used as reference economy.

[5]Beside dependence on real exchange rate.

Table 1: Cyclical behaviour of real and nominal variables

Variable x	Volatility (% Std.)	Cross Correlation of Real GDP with variable $x(t+k)$										
		$t-5$	$t-4$	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
Gross domestic product	1.26	-0.13	0.15	**0.45**	**0.72**	**0.92**	**1.00**	**0.92**	**0.72**	**0.45**	0.15	-0.13
Consumption	1.60	**-0.45**	-0.23	0.03	0.28	**0.48**	**0.59**	**0.59**	**0.51**	**0.38**	0.22	0.08
Investment	2.43	**-0.45**	-0.23	0.04	**0.31**	**0.51**	**0.63**	**0.64**	**0.58**	**0.49**	**0.40**	**0.32**
Stock change (share to GDP)	0.75	-0.17	-0.06	0.07	0.24	**0.42**	**0.57**	**0.63**	**0.55**	**0.36**	0.11	-0.10
Government expenditures	1.57	-0.09	-0.16	-0.20	-0.20	-0.14	-0.06	0.03	0.08	0.07	0.04	0.02
Exports	3.69	0.18	0.30	**0.37**	**0.40**	**0.39**	**0.34**	0.27	0.20	0.11	0.01	-0.12
Imports	2.88	-0.13	0.06	0.24	**0.39**	**0.50**	**0.56**	**0.56**	**0.49**	**0.36**	0.20	0.02
Net Exports (share to GDP)	1.77	**0.37**	**0.36**	**0.32**	0.24	0.13	0.03	-0.08	-0.15	-0.18	-0.19	-0.22
Real wage	1.60	-0.15	-0.00	0.15	0.28	**0.36**	**0.37**	**0.33**	0.23	0.10	-0.05	-0.17
Real interest rate (PRIBOR 3M)	1.41	**-0.59**	**-0.60**	**-0.55**	**-0.47**	**-0.34**	-0.17	0.02	0.21	**0.37**	**0.45**	**0.44**
Real money balances (M1/CPI)	5.19	0.03	0.18	0.29	**0.37**	**0.40**	**0.38**	**0.33**	0.25	0.15	0.06	-0.01
Employment	1.17	-0.30	-0.05	0.23	**0.48**	**0.65**	**0.70**	**0.62**	**0.49**	**0.36**	0.26	0.20
Labor productivity (Y/L)	0.95	0.19	0.26	**0.31**	**0.37**	**0.43**	**0.47**	**0.45**	**0.35**	0.15	-0.12	-0.42
Unemployment	0.62	**0.62**	**0.41**	0.14	-0.17	**-0.45**	**-0.68**	**-0.81**	**-0.86**	**-0.83**	**-0.74**	**-0.58**
Monetary aggregate M1	4.59	-0.11	0.03	0.15	0.25	**0.31**	**0.33**	**0.32**	0.28	0.23	0.18	0.15
Monetary aggregate M2	2.05	-0.29	-0.09	0.12	0.29	**0.40**	**0.45**	**0.45**	**0.43**	**0.40**	**0.37**	0.31
Consumer price index	1.28	**-0.54**	**-0.64**	**-0.65**	**-0.60**	**-0.50**	**-0.36**	-0.20	-0.01	0.20	**0.42**	**0.59**
Deflator GDP	1.75	**-0.32**	**-0.50**	**-0.63**	**-0.70**	**-0.72**	**-0.69**	**-0.58**	**-0.41**	-0.19	0.06	0.31
Inflation rate y/y (CPI)	1.83	**-0.39**	**-0.39**	**-0.32**	-0.22	-0.09	0.04	0.15	0.25	**0.34**	**0.42**	**0.44**
Inflation rate q/q (CPI)	2.36	**-0.32**	-0.24	-0.13	-0.01	0.10	0.19	0.26	**0.33**	**0.39**	**0.40**	**0.33**
Inflation rate (GDP deflator)	0.64	**-0.51**	**-0.49**	**-0.42**	**-0.31**	-0.18	-0.03	0.14	**0.32**	**0.49**	**0.60**	**0.61**
Nominal wage	1.41	**-0.66**	**-0.57**	**-0.42**	-0.23	-0.05	0.10	0.20	0.26	0.29	**0.32**	**0.34**
Nominal interest rate	1.90	**-0.79**	**-0.83**	**-0.78**	**-0.64**	**-0.43**	-0.17	0.10	**0.35**	**0.56**	**0.69**	**0.72**

leading indicator and that agrees. Monetary policy actions influence real interest rate (through controlling nominal interest rate) and subsequently influence real economic activity. However, the test of Granger causality fails the hypothesis that the real interest rate causes output.

Real money balances are moderately procyclical and leads the cycle by one quarter. The Granger causality test does not confirm monetarist view that money causes cycle. Real money balances are the most volatile variable.

Employment is strongly procyclical and coincident variable. Labor productivity is weakly procyclical and also coincide with the cycle of output. Both these stylized fact are consistent with real business cycle theory and are common in other countries. Unemployment is strongly countercyclical and lags output by two quarters. This high negative correlation with output supports well known Okun's law.

5.2 Nominal facts

Monetary aggregate M1 is weakly procyclical and coincide with output. Monetary aggregate M2 behaves in similar fashion but is more correlated and less volatile.[6] Similarly to real money balances the test of Granger causality did not proved causal relationship in any direction for both aggregates.

Price level expressed by consumer price index is strongly countercyclical and leads the cycle by three quarters. However, the correlation is changing with the phase shift; there is quite large positive correlation five quarters behind output. It can indicate that prices are rigid and adjust only slowly to clear the market. However, this observation is ruled out when you look at behaviour of GDP deflator. This measure of prices is strongly countercyclical and leads output by one quarter. These facts are similar also in other countries and contradict traditional Keynesian view of procyclical price level.

Inflation rate expressed by consumer price index (in year-on-year or quarter-on-quarter expression) is not contemporaneously correlated with the cycle. Considering phase shift, there is moderate positive correlation for lags of four or five quarters. It can be again explained by price stickiness, but the lag of one year is disputable. Rather important negative correlation five or four quarter before the cycle is hard to explain. Similar behaviour exhibit inflation rate expressed by deflator of GDP, only the correlation is stronger.

Nominal wage is countercyclical and leading indicator, slightly positive correlation occurs several quarters behind the cycle of GDP which can indicate certain wage stickiness. Nominal interest rate behaves similarly to the real interest rate, only the correlation is more striking. However, for economic subjects the real factors are more relevant; the nominal variables have only supplementary meaning. Additionally, the results of Granger causality are ambiguous for all nominal variables and do not prove any clear connections.

6 Conclusion

The analysis I made in this paper is empirical and is connected with some statistical difficulties. The results are influenced a lot by filtration method. Another setting of band pass filter or using of other filters can produce different results. Another problem is that the time span of Czech data is rather short and at most two cycles can be identified. These issues contribute to difficult interpretation of results. Further research will be aimed at examination of robustness of the results to filtration method.

References

[1] BAXTER, M., KING, R. G. Measuring business cycles: Approximate bandpass filters for economic time series. *The Review of Economics and Statistics,* Vol. 81, No. 4, November 1999, 575-593.

[6] M1 is the sum of currency in circulation and overnight deposits; M2 is the sum of M1, deposits with an agreed maturity of up to two years and deposits redeemable at notice of up to three months

[2] CHRISTIANO, L. J., FITZGERALD, T. J. *The band pass filer,* Cambridge, MA: NBER Working Paper 7257, July 1999.

[3] GRANGER, C. W. J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods, *Econometrica,* Vol. 37, No. 3, August 1969, 424-438.

[4] HODRICK, R. J., PRESCOTT, E. C. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking,* Vol. 29, No. 1, February 1997, 1-16.

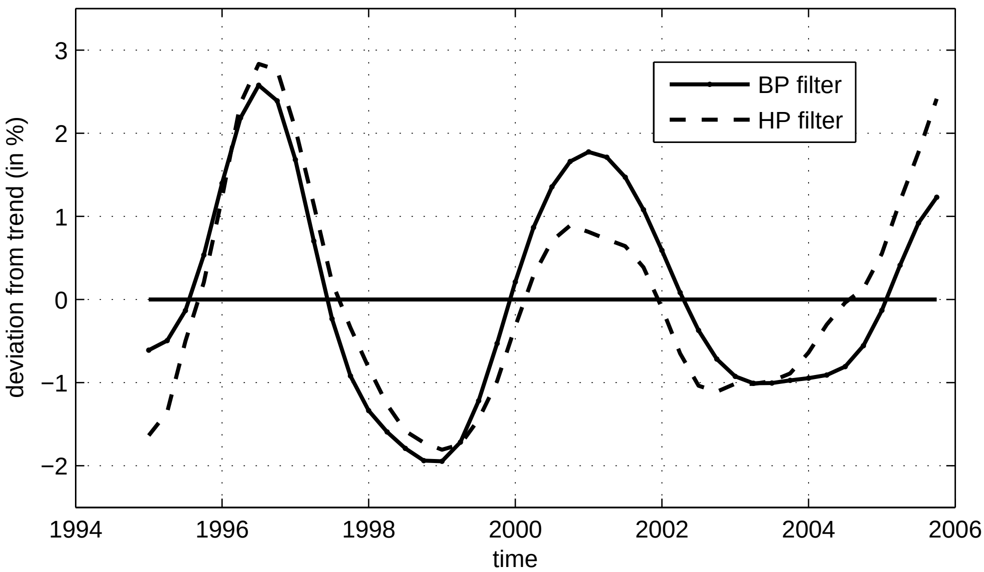
[5] KYDLAND, F. E., PRESCOTT, E. C. Business Cycles: Real Facts and a Monetary Myth. Federal Reserve Bank of Minneapolis: *Quarterly Review* Vol. 14, No. 2, spring 1990.

[6] PINDYCK, R. S., RUBINFELD, D. L. *Econometric Models and Economic Forecasts,* Irwin/McGraw-Hill, Fourth edition, 1998.

[7] STOCK, J. H., WATSON, M. W. *Business Cycle Fluctuations in U.S. Macroeconomic Time Series.* Cambridge, MA: NBER Working Paper 6528, April 1998.

Appendix

Figure 1: Estimated output gap



Source: Data CSO, author's filtration

Origin and Concentration: Corporate Ownership, Control and Performance

Jan Hanousek, Evžen Kočenda*, Jan Švejnar
* Charles University, CERGE, P.O.Box 882, Politických vězňů 7, 111 21 Prague
evzen.kocenda@cerge-ei.cz

Abstract

We analyze the effects of different types and concentration of ownership on performance using a population of firms in a model transition economy (Czech Republic) after mass privatization. Specifications based on first-differences and an unusual set of instrumental variables show that contrary to conventional wisdom, the effects of different types of ownership and ownership changes are limited. Concentrated ownership has a positive effect but only in some instances and a positive effect of foreign ownership is detectable primarily for majority ownership and for ownership by foreign industrial firms. The estimated effects of concentrated ownership support the agency theory and go against theories stressing the positive effects of managerial autonomy. Our results are also consistent with managers or stockholders "looting" the firms. The state as a holder of the golden share has a positive effect on employment and in some specifications also on output and profitability. Overall, our results suggest that the expectations and earlier findings of positive effects of privatization on performance were premature, with the effects of many types of ownership being indistinguishable from that of state ownership.

Keywords

ownership, performance, privatization, corporate governance, panel data, endogeneity, industrial organization

JEL: C33, D20, G32, G34, L20

1 Introduction

We analyze the effects of different types of ownership, changes in ownership and concentration of ownership on corporate performance using a population of firms in a model transition economy (the Czech Republic) during the period after mass privatization. In doing so, we take into account the effect of privatization and point out how our study avoids the key analytical problems that plague the large and controversial literature on privatization. In particular, our study has an important methodological dimension because many privatization schemes can be regarded as natural experiments that gave rise to endogenous ownership structures. Unlike other studies, we use an instrumental variables technique in a two-stage framework to tackle this issue.

Our analysis addresses one of the fundamental and most controversial economic questions, namely whether private firms perform better than state-owned enterprises (SOEs) and whether post-privatization ownership structures improve corporate performance. The issue has gained currency as extensive privatizations have taken place in most of the former command economies as well as in traditional developing countries, and as the most populous and rapidly growing countries -- China and India -- are in the process of privatizing their SOEs.

Apart from being diverse, the estimated performance effects found in much of the literature on transition economies are not firmly established. The credibility issue arises from three types of interrelated analytical problems that may be expected to be present in early studies. First, these studies rely on data covering short time periods immediately before and after privatization. They may hence at best capture the short–term effects of privatization rather than the medium and long-term effects of a switch from state to private or mixed ownership. Second, the early studies (a) use small and often unrepresentative samples of firms, (b) are frequently unable to identify accurately ownership because privatization is still ongoing or because the frequent post-privatization changes of ownership are hard to detect, and (c) often combine panel data from different accounting systems. Third, many of the early studies have not been able to control adequately for endogeneity of ownership (firms not being selected for privatization at random), and their estimates of the effects of privatization may hence be biased (Gupta, Ham and Svejnar, 2000).

Moreover, many of the early studies had access to limited data on firm ownership. As a result, they often treat ownership as a relatively simple categorical concept and are unable to distinguish the exact extent of ownership by individual owners or even relatively homogeneous groups of owners. As we discuss below, this also prevents many studies from providing evidence for a lively debate about the desirability of concentrated versus dispersed ownership on corporate performance.

In this paper we address systematically the three types of above-mentioned problems. In particular, in analyzing the performance effects of ownership, we (a) use panel data on a complete population of medium and large firms that went through the natural experiment of mass privatization in a model economy (Czech Republic) and that constitute the bulk of the country's economic activity, (b) cover a four-year period after privatization when accounting rules conforming to the international (IAP) standard were already in place and (c) control for endogeneity of ownership using a first-difference specification together with instrumental variables from rich data on pre-market initial conditions of these firms. Compared to earlier studies, we also develop a more systematic analytical framework for evaluating the performance effect of post-privatization ownership, distinguish between instantaneous and permanent effects of ownership changes, and use more detailed data on the extent of ownership by specific types of owners.

The fact that we use data from an economy that started almost completely state-owned and underwent virtually complete privatization means that we are analyzing a population of firms that experienced one of the most substantial changes in ownership. Unlike studies of partial privatization, we also benefit from a large variation in the values of the variables whose effect we analyze. Finally, by carrying out a detailed study of one model economy we are able to take into account specific legal and institutional features that relate to ownership and control, and we avoid the problem of not being able to control adequately for complex cross-country differences in the institutional and legal frameworks that confront comparative studies with a limited number of country-specific observations.

We find, contrary to expectations and results of many earlier studies, that the effects of different types and concentrations of ownership on firm performance are very limited and that many types of private owners do not bring about performance that is different from that of firms with substantial state ownership. We do detect, however, some significant effects of specific types of private ownership. In particular, a positive effect of concentrated ownership is discernible but only in some instances and for selected performance indicators, and a positive effect of foreign ownership is found primarily in the case of majority ownership and appears to be driven by the behavior of foreign industrial firms. The concentrated foreign owners (industrial companies) yield superior performance compared to all other types of owners in terms of growth of sales and in some specifications also profitability (strategic restructuring), and concentrated domestic owners (industrial companies and investment funds) reduce employment relative to others (defensive restructuring). Overall, our results highlight the benefits of deep privatization and restructuring accompanied by inflow of new capital and managerial culture. Thus, we provide microeconomic evidence that complements the works of Zines, Eilat and Sachs (2001) and Barrell and Pain (1997) who find that real, not just legal, privatization matters.

2 The Econometric Model

2.1 Model Specification

Our main goal is to analyze the performance effects of the principal types of ownership that we first observe after the large-scale privatization in 1996. In addition, we want to estimate the effects of the changes in ownership that took place in the 1996-99 post-privatization period. In the spirit of Ashenfelter and Card (1985) and Heckman and Hotz (1989), we specify a panel-data treatment evaluation procedure that fits our context and we supplement it with a rich set of instrumental variables. The variables that we use to provide an understanding of whether corporate restructuring associated with different types of ownership occurs more in terms of revenue or cost are companies sales revenues and labor cost. As profitability measures, we use operating profit on sales (profit/sales or return on sales) and the return on assets (ROA).

Let X_{ijt} be a given performance indicator, with subscripts denoting firm i with ownership type j, in year t. Moreover, let P_{ijt} denote ownership type j of firm i in year t. A logarithmic model of the level of performance may be specified as

$$\ln X_{ijt} = \alpha_i + \alpha\, t + (P_{ij1}t)\beta_j + (X_{ij1}t)\gamma_j + P_{ij\tau}\delta_j + [P_{ij\tau}(t-\tau)]\theta_j + (Dt)\varphi + \upsilon_{ijt}, \qquad (1)$$

where vector α_i controls for firm-specific (fixed effect) differences in the (level of) performance across firms, constant α captures the linearly time-varying performance effect of state single largest owner (SLO) or state majority ownership (depending on ownership categorization) in 1996-99, and all dummy variables in equation (1) are coded relative to α. Column vector β_j thus reflects the (linearly) time-varying effects on performance of all the other types of 1996 (initial post-privatization) ownership P_{ij1} relative to state SLO or state majority ownership.[1] Vector γ_j in turn captures the time-varying effect of the 1996 level of performance X_{ij1} on subsequent (1996-99) performance. Similarly, vector δ_j captures the time invariant (instantaneous) effect on the level of performance of a firm changing its 1996 ownership to a new ownership category $P_{ij\tau}$ in a given year τ after 1996. Complementing δ_j, vector θ_j reflects the time-varying effect on performance brought about by the new type of ownership $P_{ij\tau}$ established in the firm at time τ. Finally, vector φ represents the time-varying effects of D, industry and annual dummy variables as well as dummy variables reflecting the form of privatization of the firm (first or second wave, both waves, or outside of the voucher scheme), and υ_{ijt} is the error term.

For estimating purposes, it is useful to express equation (1) in the form of the annual rate of change specification. In particular, letting y_{ijt} be the percentage change of X_{ijt} from t - 1 to t, equation (1) may be expressed in a first-difference specification as an estimating equation

$$y_{ijt} = \alpha + P_{ij1}\beta_j + X_{ij1}\gamma_j + \Delta P_{ij\tau}\delta_j + P_{ij\tau}\theta_j + D\varphi + \varepsilon_{ijt}, \qquad (2)$$

where $\varepsilon_{ijt} = \upsilon_{ijt} - \upsilon_{ijt-1}$ is the error term. Equation (2) permits us to estimate all the parameters of interest.

Like other studies in this area, the three key econometric issues that we have to account for are: omitted variables bias, measurement error, and endogeneity of ownership. We address omitted variables bias by including a number of important control variables. In dealing with measurement error in ownership, performance and other variables, we note that the error can induce standard attenuation as well as more complicated biases in estimated coefficients. As discussed above, the earlier studies of privatization often suffer from mis-measurement of the ownership variables and performance indicators, including outliers that may seriously affect the estimated coefficients. In collecting the present data set, we have placed particular emphasis on identifying precisely individual owners and changes in ownership, as well as collecting several indicators of performance from a period when the IAP accounting system was in place. We have also tested for and eliminated outliers that affect the estimates.

As to endogeneity of ownership we address this problem as follows. First, we use the first-difference specification in equation (2) with the aforementioned covariates as a panel data treatment evaluation procedure to control for the possibility that firms are not assigned to different ownership categories at random and that certain types of owners (e.g., foreigners) may acquire firms that are inherently superior or inferior performers. Second, since first-differencing does not fully address all types of endogeneity, especially those where the effect is time-varying, we also employ an instrumental variable strategy.

[1] Coding the ownership dummy variables so that the effects of non-state ownership forms is measured relative to the effect of state ownership is useful because firms in which the state retains ownership are the ones that are least privatized and under the null hypothesis also least restructured. More importantly, the approach also accords with our desire to investigate change in performance as firms switch from state to private ownership.

2.2 Instrumental Variables

First, we use the Hausman (1978) specification test for assessing endogeneity of the initial post-privatization ownership. We employ the first-difference IV method in which we treat ownership as potentially endogenous and instrument it by IVs that we describe presently. The test is carried out by differencing the two sets of parameter estimates and standardizing the vector of differences by the difference in the covariance matrices of the two sets of estimates. The resulting quadratic form is asymptotically chi-squared with degrees of freedom equal to the number of parameters being tested. Results of the Hausman test confirm that 1996 ownership should be treated as endogenous.

We proceed by using a unique set of firm-specific instrumental variables from the pre-privatization (pre-1992) period. The instrumental variables reflect economic, institutional, industry, and geographic characteristics of the SOEs in the pre-market period, and we use them to instrument the initial post-privatization ownership that we observe in the market economy in 1996. All the instrumental variables pass a standard adequacy test.

For each firm we have collected detailed information from all the proposed privatization projects that were submitted to the government before privatization. We use the number of privatization projects per se as an important IV since many SOEs attracted several privatization project proposals, reflecting the degree of investor interest and expected future performance of the firm. Moreover, for each privatized firm we use as IVs the pre-privatization data on registered (share) capital, net asset value, total number of shares, number of shares entering voucher privatization, number of shares allocated through voucher privatization, value of shares allocated through voucher privatization in voucher points, geographic and industry location of the firm, and the structure of share ownership among various domestic and foreign parties as proposed in the winning privatization project. The share ownership variables include the share that the government intended to keep for the short or long term. Finally, our set of IVs contains annual observations on the SOE's sales, profit, debt, and employment during the three consecutive years preceding privatization. The three-year panel permits us to capture the evolution of enterprise performance before privatization. For the sake of comparability across firms, we scale these indicators by the total number of shares.

In addition to controlling for endogeneity of the ownership structure resulting from privatization, we control for possible endogeneity problems associated with changes in ownership in the 1996-99 period by including in equation (1) ownership group fixed effects δ_j for firms undergoing ownership changes. These δ_j effects may be interpreted as proxying unobserved performance characteristics of the acquired firms (i.e., new owners cherry picking winners or taking over losers) or reflecting the time invariant effects of new ownership on the level of performance. In order to check the robustness of our results, we have also estimated models that, analogously to including X_{ij1} as a regressor, include $X_{ij\tau}$ -- the performance achieved by the previous owner at the time τ when there is a change of ownership in 1996-99. This specification does not produce materially different results from those of equation (2).

3 Empirical Results of the Effects of Ownership on Performance

Our estimates are generated by the Huber (1967)--White (1982) procedure yielding heteroskedasticity-adjusted residuals in the presence of instrumental variables. We have also checked that the residuals are free from serial correlation. We employ a two-stage least squares procedure in which we instrument all variables related to ownership. The approach provides consistent estimates that are not affected by potential model misspecification.

In examining the results, we note the extent to which different types of ownership result in defensive restructuring (reducing labor cost and possibly also sales) versus strategic restructuring (increasing sales revenues, labor productivity and/or profits). Since the latter outcomes are inferred from the relative effects on sales, labor cost and profitability (e.g., increased sale and/or reduced labor costs not being accompanied by higher profits), these findings are also consistent with other phenomena such as changes in non-labor costs, and non-sales income.

The estimated coefficients make it clear that in the first four years after privatization the performance effects of different types of ownership are surprisingly limited and that many types of private ownership do not generate effects that are different from those of majority or SLO state ownership. Moreover, the overall fit of these regressions suggests that ownership explains a very small part of total variation in the rate of change of corporate performance after privatization.

3.1 The Single Largest Owner

According to our estimates, the only initial post-privatization SLO that has a positive, time-varying effect on sales is foreign industrial company. In terms of labor costs (employment), only firms with domestic industrial companies and investment funds as SLOs show a negative effect relative to the state. Finally, only firms with foreign industrial companies as SLOs have a positive effect on profit/sales and no SLO type generates a significant effect on ROA. The post-privatization foreign industrial owners thus increase profitability by enhancing the rate of growth of sales, without having a differential effect from the state firms on the rate of growth of labor cost (employment). Their domestic counterparts and investment fund SLOs reduce the rate of growth of labor cost, but do not display a corresponding positive effect on profit. The restructuring carried out by foreign industrial firms is of a strategic nature, while that performed by the domestic industrial company and investment fund SLOs is of a defensive type.

The time-varying performance effects of the SLOs that come into existence after 1996 display a number of similarities to, but also more statistical significance than, the effects of the immediate post-privatization SLOs. The basic pattern persists in that (a) most types of private owners do not show significant deviations from the sales, labor cost and profitability effects given by the base category of state SLOs, (b) foreign industrial firms raise sales and (c) domestic industrial and investment fund owners reduce labor cost. The new patterns are that firms acquired after 1996 by investment funds and portfolio companies experience a reduction in sales, foreign industrial SLOs increase not only sales but also labor costs and they no longer have a positive effect on profitability, bank SLOs have a positive effect on profit/sales and ROA, and non-industrial foreign SLOs have a negative effect on profit/sales. These results suggest that the more recent foreign industrial owners acquire firms to expand production but they no longer hold back the rate of growth of labor cost (employment), investment funds reduce the scale of operations, bank and portfolio company SLOs increase efficiency by reducing non-labor costs and/or increasing non-sales income, and domestic industrial and foreign non-industrial SLOs may suffer from transfer pricing.

3.2 Extent of Ownership

Majority and minority post-privatization ownerships by most types of private owners do not generate effects that are statistically different from the base effect of majority state ownership. The notable exception is majority ownership by foreign companies which has a strong positive effect on the rate of change of sales, thus generating an effect that parallels that of foreign industrial SLOs. The difference is that majority foreign-owned firms, unlike foreign industrial SLOs, do not produce a positive effect on profitability. This difference may be brought about by the different composition of the majority and SLO foreign groups, rising non-labor costs or falling non-sale income in the majority foreign owned firms, or dissipation of profit by majority foreign owners through transfer pricing. Firms with majority and blocking minority domestic private ownership are the only ones that significantly reduce labor costs (employment).

Overall, the effects of initial post privatization ownership indicate that concentrated foreign ownership raises the rate of increase in sales revenue, while highly as well as moderately concentrated domestic owners reduce the rate of increase in labor cost (employment) relative to others. These asymmetric findings with respect to sales and labor cost effects of concentrated domestic and foreign owners are provocative because it has been widely presumed that both domestic and foreign private ownership, especially in highly concentrated forms, would lead to substantial strategic restructuring and increases in sales -- domestically and/or on the world markets.

Firms in which the state retains a golden share register positive time-varying effects on sales, labor cost and ROA. These effects complement the estimates from the SLO specification and suggest that the state pursues an objective of increasing employment and output (revenue), while also inducing profit-oriented restructuring relative to assets. Since the state retains golden shares primarily in state-owned and domestic private firms, the effect of a golden share moderates the tendency in some of these firms to reduce output (sales) and/or employment.

4 Concluding Observations

With the former Soviet bloc and other developing countries having privatized state-owned enterprises, and the economies of China, India and Vietnam being in the process of privatization, it is important to have a solid understanding of the post-privatization effects of evolving different forms of ownership on performance. While theory generates conflicting predictions, most surveys of the early empirical literature suggest that a shift from state to private ownership tends to improve economic performance. However, much of the early literature suffers from serious data problems and inadequate treatment of endogeneity of ownership, thus leaving most results in doubt. In this paper, we analyze this issue using rich panel data covering an entire population of firms that went through mass privatization in a model transition economy -- the Czech Republic. In doing so, we have the benefit of sizable variation in key variables during a large natural experiment and we address carefully the principal data issues, including omitted variables bias, measurement error and endogeneity of ownership.

Overall, our econometric estimates present a much less sanguine picture than the generally accepted stylized facts, suggesting that the expectations and early findings of positive effects of immediate post-privatization ownership structures on corporate performance were premature. Contrary to many earlier studies, our results indicate that the performance effects of privatization and different types of ownership are on the whole surprisingly limited and that many types of private owners do not generate performance that is different from that of firms with state ownership. This lack of difference in performance is provocative because it has often been assumed that private owners would perform better than the state and the extent of inefficiency associated with various types of private ownership has been underestimated. There are two key exceptions to this overall result. First, concentrated foreign owners (foreign industrial companies) yield superior performance in terms of growth of sales and in some specifications also profit – thus reflecting the presence of strategic restructuring. Second, concentrated domestic owners (industrial companies and investment funds) reduce employment – thus engaging in defensive restructuring. These findings are consistent with the agency theory prediction that concentrated ownership results in superior corporate performance and they go against theories stressing the positive effects of managerial autonomy. Overall, our results highlight the benefits of deep privatization and restructuring accompanied by inflow of new capital and managerial culture.

References

[1] Ashenfelter, O., and Card, D. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs". *Review of Economics and Statistics* 67(4), 648-60.

[2] Barrell, Ray and Pain, Nigel 1997. Foreign Direct Investment, Technological Change, and Economic Growth Within Europe. *The Economic Journal*, 107(445), 1770-1786.

[3] Gupta, N., J. C. Ham, and J. Svejnar. 2000. Priorities and Sequencing in Privatization: Theory and Evidence from the Czech Republic. *William Davidson Institute WP No. 323*.

[4] Hausman, J. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6):1251-1271.

[5] Heckman, J., and Hotz, V. 1989. "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training." *Journal of the American Statistical Association* 84(408), 862-74.

[6] Huber, P. J. 1967. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 221-223. Berkeley, CA: University of California Press.

[7] White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50: 1-25.

[8] Zinnes C., Eilat Y., Sachs J. (2001): "The Gains from Privatization in Transition Economies: Is 'Change of Ownership' Enough?" *IMF Staff Papers*, 48, Special Issue.

Measurement of supplier-customer system complexity based upon entropy

Jiří Hofman [a], Ladislav Lukáš [b]
University of West Bohemia in Pilsen, Faculty of Economics,
[a] Dept. of Business Economics and Accounting, [b] Dept. of Statistics and Operations Research,
Husova 11, 30614 Plzeň
[a] jiri.hofman@fek.zcu.cz, [b] lukasl@kso.zcu.cz

Abstract

Paper discusses a systematic approach for measuring supplier-customer relations operational complexity based upon entropy. There is well-known that, as the operational complexity of the system increases, there is an associated increase in the amount of information required to describe that system. Basically, business economics knows two types of complexity of supplier-customer systems – a structural one, and operational one, respectively. At first, some basic notions from information theory are introduced in order to understand entropy as a theoretic measure of quantity of information. A unique feature of this measure is that it captures, in relative terms, the amount of information required to describe the state of the system. Further, derivation of well-adopted measure for operational complexity of supplier-customer system concerning variations in quantities and time is given in detail. On the base of a prototype supplier-customer system using analytic approach of inventory control methods the necessary data requirements are presented. The basic idea consists in introduction of corresponding set of states monitoring various flow variations, e.g. order – forecast, delivery – order, and actual production – scheduled production. These states of variation are defined with respect to control decisions, by considering the severity of such variation. The measure of operational complexity from an information-theoretic perspective provides a detailed and flexible analysis of supplier-customer systems.

Keywords

Business economics, supplier-customer systems, complexity measures, information and entropy.

JEL: C63, C81, L25, M21.

1 Introduction

Basically, business economics knows two types of complexity of supplier-customer systems, a structural complexity and an operational one, in particular. The structural complexity is usually defined as that associated with the static variety of a system and their design dimensions. On the contrary, the operational complexity can be defined as the uncertainty associated with the dynamic system. Hence, a measure of operational complexity should express behavioural uncertainty of the system with respect to a specified level of its control. The operational complexity of supplier-customer system is associated with specific data provided by inventory management. Such data regard both the uncertainty of information and material flows within and across organizations.

2 Theoretical background

Information theory provides a means of quantifying complexity. Of the complexity measures available, Shannon's information-theoretic measure and corresponding entropy are well-known quantities, which measure the expected amount of information required to describe the state of a system.

Shannon introduced the concept of measuring the quantity of information, by means of entropy, in his work on a mathematical theory of information and general theory of communication. In general, the complexity of a system increases with increasing levels of disorder and uncertainty of its states.

Basic mathematical model of information complexity assumes that N objects are given. Since any object has to be identify uniquely a unique binary code $(a_1,..., a_d)$ is assigned to each object, where a_i , $i=1,...,d$ belongs to $\{0,1\}$, and d is the lowest exponent satisfying the relation $2^d \leq N$, or, in another words, an integer satisfying $0 \leq d - \log_2 N < 1$. Hence, the quantity $I = \log_2 N$ gives the length of most effective binary coding for unique identification of N objects.

In probabilistic framework, one assumes a trial which results 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 $p_1+...+p_N = 1$.

Making a large number of independent trials n, we get ratios $n(A_i)/n$ approaching p_i, $i=1,...,N$, where $n(A_i)$ denotes the number of occurrences of event A_i within such n independent trials. There is also evident that $n(A_1)+...+ n(A_N) = n$ holds. The total number of possible outcomes, where events A_i, $i=1,...,N$, appear $n(A_i)$ times each, is $N_n = n!/(n_1!... n_N!)$, where $n_i \approx np_i$.

In order to express $\log_2(N_n)$ in analytic form, for $n \to \infty$, one may use the Stirling formula, i.e. $m! \approx m^m e^{-m} \sqrt{(2\pi m)}$, for large integer m.

Hence, we get after technical manipulation

$$\log_2(N_n) \approx n\log_2(n) - \sum_{i=1}^{n} np_i \log_2(np_i) + (\log_2(\sqrt{(2\pi n)}) - \sum_{i=1}^{n} \log_2(\sqrt{(2\pi n_i)})),$$

$$\log_2(N_n) \approx -n\sum_{i=1}^{n} p_i \log_2(p_i). \tag{1}$$

Now, coupling both results together we are able to express the length, denoted d_n, of the most effective binary coding of any outcome of all possible N_n ones in following form

$$d_n \approx \log_2(N_n) \approx -n\sum_{i=1}^{n} p_i \log_2(p_i). \tag{2}$$

Keeping in mind we have n independent trials, the equation (2) yields in average a quantity, denoted I, for each individual trial

$$I = -\sum_{i=1}^{n} p_i \log_2(p_i). \tag{3}$$

The expression (3) gives a motivation to introduce the quantity $I(p_1,...,p_N)$, which measures an average quantity of information relating an appearance of one event of $\{A_1,...,A_N\}$ carried in any individual trial, in the following form

$$I(p_1,...,p_N) = -\sum_{i=1}^{N} p_i \log_2(p_i). \tag{4}$$

In particular, if the events $\{A_1,...,A_N\}$ obey an uniform distribution, we get immediately

$$I_u = -\sum_{i=1}^{N} (1/N) \log_2(1/N) = \log_2(N). \tag{5}$$

which is the greatest value possible obtained. It reflects the most uncertain situation, as all events are equally likely, and each observation is equally unpredictable.

The alternative way how to get the formula (4) is based upon functional approach. Since it is natural to assume $I(p_1,...,p_N)$ to be a continuous function, two basic properties are to be fulfilled by $I(p_1,...,p_N)$:

i) to remain invariant against any permutation of its arguments $p_1,...,p_N$ as the set of events $\{A_1,...,A_N\}$ is always the same,

ii) to take the following specific functional form

$$I(p_1,...,p_N) = q_1 I(p_1/q_1, ..., p_m/q_1, 0,...,0) + q_2 I(0,...,0, p_{m+1}/q_2, ..., p_N/q_2),$$

which actually defines behaviour of $I(p_1,...,p_N)$ depending upon quantities I_1, I_2. These ones express information quantities related to conditional probabilities of conjunctions of events.

More precisely, if we know that a compound event $B_1 = \bigcup_{k=1}^{m} A_k$ has occurred, i.e. if any event from $\{A_1,...,A_m\}$ has already appeared, then the corresponding quantity of information I_1 should be calculated by expression

$$I_1 = I(p_1/q_1, ..., p_m/q_1, 0,...,0),$$

where $q_1 = \sum_{k=1}^{m} p_k = \mathrm{P}(B_1)$, is probability of a compound event B_1, and $p_k/q_1 = \mathrm{P}(A_k|B_1)$, are conditional probabilities of events A_k conditioned by the compound event B_1, for k=1,...,m.
The quantity I_2 is defined in a similar way assuming the complement event to B_1, denoted B_2, and expressed by $B_2 = \bigcup_{k=m+1}^{N} A_k$, has occurred, i.e. if any event from $\{A_{m+1},...,A_N\}$ has already appeared. In particular, I_2 should be calculated by expression

$$I_2 = I(0,..., 0, p_{m+1}/q_2, ..., p_N/q_2),$$

where $q_2 = \sum_{k=m+1}^{N} p_k = \mathrm{P}(B_2)$, is probability of a compound event B_2, and $p_k/q_2 = \mathrm{P}(A_k|B_2)$, are conditional probabilities of events A_k conditioned by the compound event B_2, for k=m+1,...,N.
The functional solution reads as follows

$$I(p_1,...,p_N) = -\, c \sum_{k=1}^{N} p_k \log(p_k), \tag{6}$$

where c is a positive constant, $c > 0$, and base of logarithms is arbitrary $b > 1$. For, $c = 1$, and $b = 2$, the expression (6) takes the form of (4). For more details, see book [4].

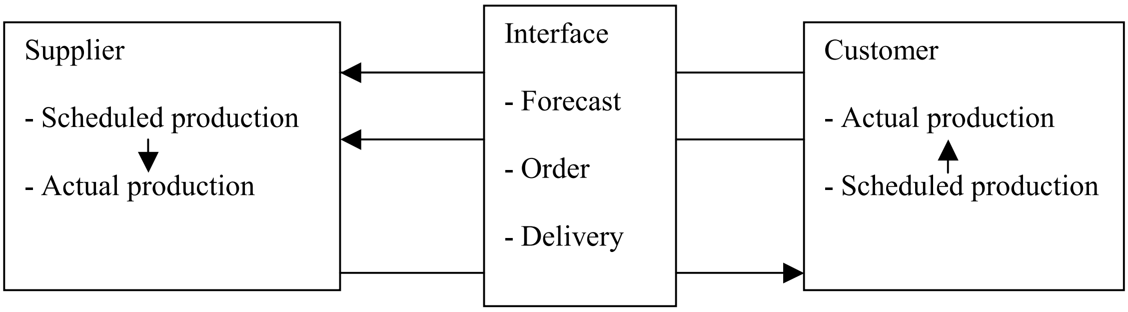
The information-theoretic measure defined by expression (4) of a system, which states are described by $\{A_1,...,A_N\}$ having probabilities $(p_1,...,p_N)$, is called an *entropy* of the system.

3 Operational complexity of supplier-customer system

From a point of view of management science, the supplier-customer system belongs to theory of inventory control. However, available analytic results and commonly used numerical procedures for solving optimal cost inventory problems are usually to restrictive as regards their theoretic assumptions. Hence, they are used as approximations in practical supplier-customer systems, only. Such reason stems from fact that analytic models are mainly uni-commodity ones, and moreover usually assume constant rate of demand. Some well-known and suitable optimal cost inventory models are given in [2]. Nevertheless in practice, a management of any supplier-customer system needs quantity of goods and their delivery times including lead times to be determined precisely.

There is evident that for effective supply system management it is critical that goods arrive in the right quantity at the specified time. Since in practice, there are deviations in quantity and delivery times, as well, there is very reasonable to define an theoretic interface in any supplier-customer system, where such deviations are to be measured.

Operational complexity of supplier-customer system results at this interface where the actual deliveries of goods deviate in quantity and/or time from that expected. Following [5], we are able to sketch a basic scheme introducing the interface within a supplier-customer system.



Definition of variables.
Let us consider a set of products $\{P_1, ..., P_n\}$, which is handled within a supplier-customer system.
In general, there are two types of variables relating quantity and time to be considered for a particular product P_i, i=1,...,n, and at both supplier and customer side, and at the interface, particularly.

	Quantity	Time
A) Supplier side:		
- scheduled production:	${}_{s,s}Q_i$, i=1,...,n,	${}_{s,s}T_i$, i=1,...,n,
- actual production:	${}_{s,p}Q_i$, i=1,...,n,	${}_{s,p}T_i$, i=1,...,n,
B) Interface:		
- forecast:	${}_{i,f}Q_i$, i=1,...,n,	${}_{i,f}T_i$, i=1,...,n,
- order:	${}_{i,o}Q_i$, i=1,...,n,	${}_{i,o}T_i$, i=1,...,n,
- delivery:	${}_{i,d}Q_i$, i=1,...,n,	${}_{i,d}T_i$, i=1,...,n,
C) Customer side:		
- scheduled production:	${}_{c,s}Q_i$, i=1,...,n,	${}_{c,s}T_i$, i=1,...,n,
- actual production:	${}_{c,p}Q_i$, i=1,...,n,	${}_{c,p}T_i$, i=1,...,n,

which gives 14n variables describing supplier-customer system, in total.

In order to get a structure of supplier-customer system to fit into a framework of information theory discussed in the first part, and thus enabling an entropy to be considered as a information-theoretic measure, we have to recast the crude logical structure defined by the variables introduced, into a set of events $\{A_1,...,A_N\}$ describing the states of system having corresponding probabilities $(p_1,...,p_N)$.
This is a very crucial point, and in general, it depends on various aspects of supplier-customer system investigated, in particular upon its structural complexity, too. It means, that even structurally simple systems can posses high operational complexity, and if approaching its upper bound, i.e. the value $\log_2(N)$ for uniformly distributed events, to be almost unpredictable.

Usually, the quantities ${}_{(.,.)}Q_i$, and ${}_{(.,.)}T_i$, i=1,...,n, are continuous variables over specific ranges. In that case, we have to discretize them introducing non-overlapping proper quantity and time bands, which cover the specific ranges. In case of discrete variables, we may use their values directly.

Operational complexity in supplier-customer system is defined as the amount of information required to describe the state of system in terms of the quantity and time variations across material flows and time-information flows that exist.
Roughly speaking, each event A_k is to be defined on the base of flow variation considered, and quantities available for that could take the following form, e.g. (Order – Forecast), (Delivery – Order), (Actual production – Scheduled production), etc.

Hence, there are quantities of following structure:

$({}_{i,o}Q_i - {}_{i,f}Q_i)$, $({}_{i,o}T_i - {}_{i,f}T_i)$, $({}_{i,d}Q_i - {}_{i,o}Q_i)$, $({}_{i,d}T_i - {}_{i,o}T_i)$,
$({}_{s,p}Q_i - {}_{s,s}Q_i)$, $({}_{s,p}T_i - {}_{s,s}T_i)$, $({}_{c,p}Q_i - {}_{c,s}Q_i)$, $({}_{c,p}T_i - {}_{c,s}T_i)$, etc.

Across each flow variation the corresponding quantity and/or time variations should be monitored. We note that both positive and negative differences, which give mathematical values of these variations, should be tackled separately. That means, they are generating different states of supplier-customer system, in general.
Very important for setting bounds of the variations is also a managerial reasoning and decision making as to a severity of such variations. Frequency of monitoring plays an important role, too, which is linked to cycle time of delivery goods, in general. The paper [5] assumes monitoring of variations will be issued by controllers, key people within the organization who are responsible for managing the monitored flow variations. However, it could be provided automatically by the special report issues generated from advanced managerial information system, too.

Basically, it is recommendable to mollify the amount of information required to describe the quantity variation or time variation into separate states of supplier-customer system across one flow variation for one product P_i considered. Each of them will be described by a scalar variable, and the lower and upper bounds of the states can be defined. General question arises, how many states to define.
There is evident, that the basic state across particular flow variation should reflect either no variation of the inspected scalar variable, or just a tolerable one captured between basic acceptable bounds. Such state defines a desired or expected state as being judged by the management of organization, and it is called *in-control state*.
On the contrary, related to one in-control state it is evidently necessary to define other states by setting their bounds applied to the same inspected scalar variable, which mainly depend upon severity of such variations from managerial point of view, again. They are called *out-of-control states*.

Now, we have to discuss probabilities of both types of states, i.e. in-control state and out-of-control states, too. At the first stage and for a purpose to keep expressions simple as possible, we temporally neglect an index of the product considered P_i, and we do not specify if a scalar variable describing the flow variation considered would be either a quantity variation or a time variation. We call it just a common scalar variable.
Let us denote p_1 the probability of in-control state of a common scalar variable, further s will denote a total number of states considered for that scalar variable, i.e i=1 denotes the in-control state and out-of control states are indexed i=2,...,s, consecutively, and finally, p_i, i=2,...,s will denote probabilities of all out-of-states considered. Since these s states form a set of complete events as to the flow variation considered, we get directly

$$\sum_{i=1}^{s} p_i = 1, \text{ hence } \sum_{i=2}^{s} p_i = 1 - p_1 \tag{7}$$

Usually, within information-theoretic literature the entropy of a system defined by expression (4) is denoted by H, and we know it is an additive quantity as to the set of events $\{A_1,...,A_N\}$, in general. Since our analysis of supplier-customer system has concerned just one flow variations with s mutually disjunctive states, we may adopt the following expression for calculating an entropy of that flow variation

$$h(p_1,...,p_s) = -\,p_1 \log_2(p_1) - \sum_{i=2}^{s} p_i \log_2(p_i). \tag{8}$$

Now, we shall consider a general supplier-customer system with its interface included, which consists of set of products $\{P_1, ..., P_n\}$, each P_i having r_i flow variations, i=1,...,n, and each flow variation being represented by s_{r_i} mutually disjunctive states, i.e one desired in-control state and the others out-of-control states, as discussed earlier.

On the base of formula (8), we may express the entropy of that supplier-customer system in the following way

$$H = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i}\left(p_{ij1}\log_2(p_{ij1}) - \sum_{k=2}^{s\_r\_i} p_{ijk}\log_2(p_{ijk})\right), \tag{9}$$

where p_{ij1} stands for probability of desired in-control state adhering to j-th flow variation from r_i ones considered with i-th product P_i. Whereas p_{ijk} are corresponding out-of-control states introduced.

However, it could be usefull in some cases to express out-of-control states and their probabilities p_{ijk} as conditional ones, in particular conditioned by their compound event to the corresponding in-control state. Thus, on the base of (7), and using property that all states introduced are mutual disjunctive ones, we may write

$$\sum_{k=2}^{s\_r\_i} p_{ijk} = 1 - p_{ij1}\text{, or equivalently } (1 - p_{ij1})^{-1}\sum_{k=2}^{s\_r\_i} p_{ijk} = 1,$$

which yields

$$p_{ijk} = (1 - p_{ij1})\, q_{ijk}\,,\quad\text{and}\quad \sum_{k=2}^{s\_r\_i} q_{ijk} = 1, \tag{10}$$

where q_{ijk} are conditional probabilities introduced.
Substituting (10) into (9) gives

$$H = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i}\left(p_{ij1}\log_2(p_{ij1}) - \sum_{k=2}^{s\_r\_i}(1 - p_{ij1})q_{ijk}\log_2((1 - p_{ij1})q_{ijk})\right),$$

and finally, we obtain

$$H = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i}\left(p_{ij1}\log_2(p_{ij1}) - (1 - p_{ij1})\log_2(1 - p_{ij1}) - (1 - p_{ij1})\sum_{k=2}^{s\_r\_i} q_{ijk}\log_2(q_{ijk})\right). \tag{11}$$

The operational complexity of supplier-customer system measured by entropy (11) depends upon the set of mutually disjunctive states introduced with all flow variations considered and their corresponding probabilities. Inspecting its structure we notice that it can be divided into three additive terms H_1, H_2 and H_3, respectively

$$H = H_1 + H_2 + H_3\,,$$

$$H_1 = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i} p_{ij1}\log_2(p_{ij1}),\qquad H_2 = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i}(1 - p_{ij1})\log_2((1 - p_{ij1}),$$

$$H_3 = -\sum_{i=1}^{n}\sum_{j=1}^{r\_i}(1 - p_{ij1})\sum_{k=2}^{s\_r\_i} q_{ijk}\log_2(q_{ijk}), \tag{12}$$

where H_1 represents the amount of information required to describe the system is in desired in-control states, H_2 represents the amount of information required to describe the system is out of desired in-control states, and H_3 represents the additional amount of information required to describe the system is in all out-of-control states considered.

There is evident that these quantities provide more useful information than H only. At first, the ratio H_2/H_1 describes information-theoretic fraction showing how much is the system out of desired in-control states related to being in all desired in-control states. At second, the quantity H_3 could be tackled as stronger information-theoretic measure of operational complexity of the supplier-customer system, because it gives an expected amount of information to describe the extent to which the system occurs in all out-of-control states considered.

Now, a natural and crucial question arises how to get all probabilities introduced. On the contrary to information-theoretic framework using entropy as a measure of expected amount of information describing state of system, in general, the estimation of probabilities depends heavily on the specific supplier-customer system particularly investigated. So far, we conclude simply probabilities p_{ijk} should be estimated from data collected by monitoring the supplier-customer system, and stored in a proper data base. More detailed analysis of that topic and description of procedure how to do it goes beyond the paper concerns.

4 Conclusion

The measure of operational complexity from an information-theoretic perspective provides a detailed and flexible analysis of supplier-customer systems. We have discussed a necessary background from information theory, which provides a platform for derivation of formulas giving appropriate instruments for calculating such measures. In general, operational complexity of supplier-customer system is formulated as expected amount of information required to describe the system state being defined by set of flow variations covered by appropriate desired in-control states and out-of-control states introduced, respectively. The outlined framework of operational complexity is linked to static description of system. However, it could be generally applied for calculating the corresponding entropy measures in different time periods providing the relevant amount of monitored data is at disposal, thus giving a time series of interested quantities in time.

Evidently, from managerial point of view how to apply this analysis in practical cases of supplier-customer systems, there is still a lot of work to elaborate. In particular, the amount of data necessary for application of proposed operational complexity analysis, the collection of data, data processing procedures, the estimation of probabilities of all mutually disjunctive states within the specific supplier-customer system considered, there are just some basic and important features, which should be further elaborated. The research thereon is ongoing.

Acknowledgement: This work was supported by grant No. 402/05/2392 of the Grant Agency of Czech Republic. The second author acknowledges gratefully also funding by the project LC06075 of the Ministry of Education of Czech Republic.

References

[1] 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.

[2] Lukáš, L. *Pravděpodobnostní modely některých manažerských úloh* (*Probabilistic models of some managerial problems*, in czech). Západočeská univerzita v Plzni, 2005, ISBN 80-7043-402-3, 220 str.

[3] Nagurney, A., Cruz, J., Matsypura, D. Dynamics of Global Supply Chain Supernetworks. Math. & Computer Modelling 37 (2003), 963-983.

[4] Prochorov, Ju. B., Rozanov, Ju., A. *Teorija verojatnostej* (*Theory of probability*, in russian). Nauka, Moskva 1967.

[5] 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.

Approximations in stochastic and robust programming problems

Michal Houda
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic,
Pod Vodárenskou věží 4, 182 08 Praha, Czech Republic, houda@karlin.mff.cuni.cz

Abstract
Optimization procedures are very useful tools in many economic decision-making problems. We deal with the case of optimization problems where uncertainties in parameters occur. The stochastic programming approach considers the probability distribution of uncertain parameters and seeks for a solution that is feasible up to a certain level of probability (chance-constrained programming). Robust programming techniques search for such a solution that satisfies simultaneously all possible realizations of the parameters. Both methods require some kind of approximation because of computational difficulties. The paper deals with such approximations and illustrates the essential difference between the two above-mentioned methods. Even if a variety of economic problems lead to the same optimization program, one is required to choose a correct method to solve it; the economic background of the problem is crucial for such decision.

Keywords
Chance-constrained problem, empirical distribution function, robust optimization problem, sampled problem.
JEL: C44, C61

1 Introduction: uncertainty of the data

The traditional optimization research became an integral part of the post-war science and many advances and applications in various fields of the area were obtained, including applications in finance, engineering, management, control, etc. The real world carries the uncertainty of the data as a generic property of all the models of mathematical programming. There are many ways to handle the uncertainty and to give applicable results of the optimization procedures.

Consider an optimization problem of the form

$$\text{minimize } c(x;\xi) \text{ subject to } x \in X,\ f(x;\xi) \leq 0 \tag{1}$$

where $\xi \subset \mathbf{R}^s$ is a data element of the problem, $x \in X \subset \mathbf{R}^n$ is a decision vector, the dimensions n, s, M and the mappings $c : \mathbf{R}^n \times \mathbf{R}^s \to \mathbf{R}$ and $f : \mathbf{R}^n \times \mathbf{R}^s \to \mathbf{R}^M$ are structural elements of the problem. This is a general framework for a large class of optimization problems which we characterize further by

- insufficient knowledge of the data; all that is known about the data vector ξ is that it belongs to a given *uncertainty set* $\Xi \in \mathbf{R}^s$;
- the constraints of problem (1) are required to be satisfied as much as possible given the actual realization of $\xi \in \Xi$.

If a realization of ξ is known and fixed in advance, standard deterministic optimization algorithms can be used to solve problem (1). This is rarely the case; in practice, uncertainty of the data is typical in the modelling framework, for example:

- the data ξ is not known at the time when the decision (value of x) have to be made, and will be realized in the future (the data can represent future demands and prices in economy, loads to the bridge in truss construction, weather conditions, etc.);
- the data ξ cannot be measured or estimated exactly even if it is realized before a concrete decision is taken (material properties, measuring errors, etc.);
- the data is certain and the optimal solution of the problem can be computed exactly, but such solution cannot be implemented exactly due to physical characteristics of the solution (e. g. uncertain production of some commodity, properties of construction, etc.). The last can be easily modeled via uncertainty in the parameters of the model, not in the decision vector x;

- the model itself is an approximation of a complicated real-world phenomenon and uncertainty comes directly from the modeling process.

Dealing with uncertainty is a kind of bread-and-butter problems that classical optimization try to solve. Several approaches were developed. First, the uncertainty is simply *ignored* at the stage of building the model and/or finding an optimal solution of it. The data is replaced by some nominal values (e. g. averages, expected values) and the accuracy of the optimal solution is (or should be) inspected ex-post by *sensitivity analysis*. This is a traditional way to control the stability of the model but it is limited only to an already generated solution. There are examples where the "ignoring uncertainty" approach leads to a solution that is not acceptable in practice (see e. g. Kall's linear programming example, [15]).

Stochastic programming handles the uncertainty of stochastic nature. More precisely, we consider ξ to be a random vector and assume that we are able to identify its underlying probability distribution. The idea of stochastic programming approach is to incorporate available information about data through its probability distribution and solve the new model by means of deterministic optimization (the new model was said to be a "deterministic equivalent" in early works on stochastic programming). There are various ways of doing that and there are many papers and books dealing with particular branches of stochastic programming. The stochastic programming community recognizes Dantzig's paper [7] as the initial work in the area; there are also a large number of books devoted to stochastic programming and its applications ([4], [16], [19], [23], [24], and others).

The concept of so-called *robust optimization* does not have such a long history. It introduces an alternative way to handle uncertainty in the model by the so-called "worst-case" analysis: we look for such a solution that satisfies the constraints for all possible realizations of ξ and we optimize the worst-case objective function among all robust solutions. Even if such paradigm is classical in statistical decision theory, the real development in this area of optimization dates only to the last decade starting with the work [2]. On the other hand, robust optimization problems are not new (they are a part of semi-infinite programming problems); also the influence of the robust control theory is evident and not negligible.

2 Mathematical model of uncertainty

2.1 Uncertain convex program

An *uncertain convex program* (UCP) is a *family* of convex optimization programs (1) parameterized by $\xi \in \Xi$. Without loss of generality we consider the following form of (UCP):

$$\text{minimize } c'x \text{ subject to } x \in X,\ f(x;\xi) \le 0, \tag{2}$$

where $\xi \in \Xi \subset \mathbf{R}^s$, $X \subset \mathbf{R}^n$ is convex and closed set, the objective is linear and the scalar-valued function $f : X \times \Xi \to \mathbf{R}$ is convex in x for all $\xi \in \Xi$. In fact, the linearity of the objective can be imposed by considering the problem minimize t subject to $x \in X, c(x;\xi) \le t, f(x;\xi) \le 0$ instead of problem (1); and multiple valued convex constraint functions $f_i(x;\xi)$ can be converted into a single scalar-valued function of the form $f(x;\xi) := \max_{i=1,\dots,M} f_i(x;\xi)$. If the realization of ξ is known and fixed, we use the deterministic optimization to solve problem (2). This corresponds to the approach of ignoring uncertainty as described above. In many cases, such solution is very sensitive to perturbations of ξ and one of the following methods must be used.

2.2 Chance (probability) constrained program

A *chance* (or *probability*) *constrained program* (PCP) is a particular variant of stochastic programming problem. We assume that ξ is a random vector defined on some probability space $(\mathbf{\Omega}, \mathcal{A}, \mathbf{P})$ with known probability distribution $P \in \mathcal{P}(\Xi)$ where $\mathcal{P}(\Xi)$ is the space of all probability measures defined on Ξ. Next, we require the constraints of (UCP) to be fulfilled with a prescribed level of probability ε. The problem reads

$$\text{minimize } c'x \text{ subject to } x \in X_\varepsilon := \left\{x \in X;\ P\{f(x;\xi) > 0\} \le \varepsilon\right\}. \tag{3}$$

The feasible solution of this problem allows to violate the original constraints with a small level of "risk"; it is sometimes called ε-feasible solution. The problem (PCP) need not be convex even if f is convex in x for all ξ; another issue arises when we try to evaluate the probability in the definition of X_ε because it usually involves multidimensional integrals. One of the possible approaches overcoming this issue is the subject of Section 3.

The term of "chance-constrained" optimization is coined with an early work of Charnes and Cooper [6]; most of the relevant literature is resumed in [19]. Main results of the topic concern conditions under which (3) is a convex program and how to convert the probability constraints into an explicit deterministic form.

2.3 Robust convex program

In the *robust convex program* (RCP) we look for a solution to optimization program (2) that satisfies the constraints for all possible realizations of ξ, i. e., that is feasible for any member of problems belonging to the family (UCP). The problem can be rewritten as

$$\text{minimize } c'x \text{ subject to } x \in X,\ f(x;\xi) \leq 0 \text{ for all } \xi \in \Xi. \tag{4}$$

The symbol Ξ is overloaded here: first, in (2), it represents the uncertainty set which the (unknown) parameter ξ belongs to, and second, in (4), it is the set of parameters for which the constraints must be fulfilled. In fact, this overloading is not too much important: the two sets usually coincide for the reason that information we have about the uncertain parameter is the same as the risk we want to hedge against.

(PCP) is a convex program but it is numerically hard to solve because of an infinite number of constraints. There are several relaxation techniques to deal with this issue, see e. g. [2], [3], and references therein. We describe a so-called “randomized” approach [5] in the following.

3 Approximation to stochastic and robust optimization programs

The probability distribution P of ξ is rarely known precisely as required by the chance-constrained methods solving (3). Similarly, numerically intractable problems arising from (4) are common in practice. *Sampling techniques* are useful in this context for both approaches dealing with the uncertain convex program, stochastic and robust programming. Let us introduce ideas of sampling for both presented approaches.

3.1 Chance-constrained “sampled” problem

Let $\xi_1, \ldots, \xi_N$ be a set of independent samples distributed according to P (distribution of the random vector ξ). We define the *empirical distribution function* as a discrete random vector of the form

$$P_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i}$$

where δ_ξ denotes the Dirac measure placing the unit mass at ξ. We approximate, for the given sample, the problem (PCP) by replacing the original distribution P in (3) by P_N. We denote the resulting problem as (PCP$_N$):

$$\text{minimize } c'x \text{ subject to } x \in X[\varepsilon, N] := \Big\{x \in X;\ \frac{1}{N} \operatorname{card}\{i;\ f(x;\xi_i) > 0\} \leq \varepsilon\Big\}. \tag{5}$$

The main idea behind this program is that the relative frequency of constraint violations is approximately the desired level ε of infeasibility in (PCP). An extensive literature on stability in stochastic programming deals with the question how far the resulting optimal solution of (PCP$_N$) is from the optimal solution of (PCP), see e. g. [12], [17], [18], [21], [22] and others. If the assumptions of the general stability theorem (Theorem 1 in [12]) are fulfilled, the distance between optimal solutions of the two problems are expected to converge to zero.

3.2 Robust sampled convex problem

Consider again $\xi_1, \ldots, \xi_N$ a set of independent samples from a given probability distribution P. Note that the original (RCP) problem does not involve any information about stochastic nature of parameter ξ. This is why this approach adopted the name *randomized program* or *robust sampled convex program* (SCP$_N$):

$$\text{minimize } c'x \text{ subject to } X[N] := \Big\{x \in X;\ f(x;\xi_i) \leq 0 \text{ for all } i = 1, \ldots, N\Big\}. \tag{6}$$

This is a relaxation of the original robust problem: we does not require the original constraints to be satisfied for all realizations of $\xi \in \Xi$ but only for a certain finite but sufficiently large number of samples. This approach has several favorable impacts:

- the problem is convex, it has a finite number of constraints and it is effectively computable;
- it incorporates weights to the individual parameter instances of ξ – their absence was also the subject of criticism of the common robust framework;
- in addition, realizations of ξ used in (SCP$_N$) are that which are most probably to happen.

The randomized approach to (RCP) was proposed in [5] and [8]; in [9] the idea was extended to the case of the so-called ambiguous chance-constrained programming where the distribution P is known only approximately. For a survey of these results see [13].

The solution of (6) approximates the solution of (RCP): the higher the number of samples, the closer the solutions are. In order to find an optimal solution of (SCP_N) that is sufficiently close to the optimal solution of (RCP) one need a rather high number of samples to be generated. The authors of above-cited papers have sought for a rule on the sample size N that assures the optimal solution of (SCP_N) to be ε-feasible, i. e. feasible in the chance-constrained problem (PCP). But one cannot expect that this solution is near to the optimal solution of (PCP). In fact, in [14] we gave a comparative numerical study on a simple optimization problem where this conclusion is approved. We develop this idea in detail from the economical and practical point of view in the following section.

4 Application issues of stochastic and robust programming problems

4.1 Simple example

The following example is taken from [14] where it is examined in more details. Consider a simple uncertain convex program

$$\text{minimize } x \text{ subject to } x \geq \xi, x \in X \tag{7}$$

where $\xi \in \mathbf{R}$ is distributed according to the standard normal distribution. The solution to the chance-constrained program (PCP) related to (7) with $\varepsilon = 0.05$ is 0.95 quantile of normal distribution (approximately 1.64). The number of samples assuring that the optimal solution of (SCP_N) is ε-feasible is about 240. Figure 1 shows how optimal values of both (PCP_N) (dotted line) and (SCP_N) (solid line) behave for sequences of 240 and 3000 samples. On the one hand you should note the convergence property of solutions of (PCP_N) to the optimal value of (PCP); on the other hand robust sampled solutions are getting away from this point as far as the number of samples increases. We are going to discuss this feature in a practical point of view.

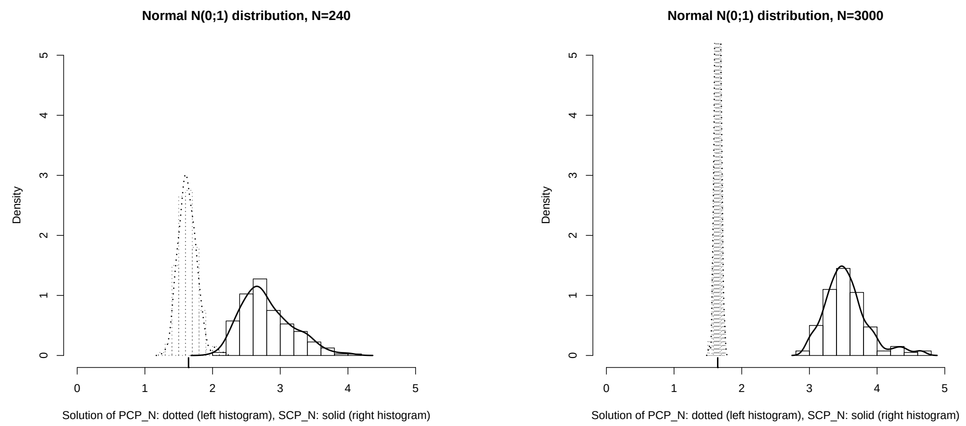


Figure 1: Optimal solutions to sampled problems.

4.2 Applications of the chance-constrained problems

There is a huge number of applications in stochastic programming due to the long history of the subject. A collection of the most important ones is given in the Wallace and Ziemba's book [24]. We give here only a short overview of selected particular tasks that was solved in real-world applications. Other items include applications in agriculture, power generation and electricity distribution, military, production control, telecommunications, transportation and many others.

- Chemical engineering ([11]). A continuous distillation process is frequently very dependent on a controlled rate of its inflow; if the last is of stochastic nature, it cannot be processed immediately but has to be stored in a feed tank. The objective is to find the optimal feed control with the prescribed lower and upper level of the inflow preventing the feed tank to be empty or full, together with the fact that costs compensating possible level violations are difficult to model.

- Finance – portfolio selection. The objective is to select the optimal portfolio of bonds in order to maximize final amount of money and to cover necessary payments in all years. The last is modeled via liquidity constraints we want to satisfy with some high level of probability.
- Water management ([20]): one of the very beginning application of chance-constrained problems. A number of reservoirs must be designed in order to control flooding due to random stream inflows.

These models have in common that we estimate the probability distribution (needed to solve the optimization problem) by means of observations from the past. The resulting solution in (PCP_N) is an approximation to the (unknown) solution of (PCP) and the approximation is better as the number of samples (observations) is higher. Furthermore, our solution of the chance-constrained "sampled" problem is only approximatively ε-feasible for a given level ε. On the other hand, this level is usually not crucial for real applications if our preferences are pointed towards costs saving solutions – this is also the case of all mentioned applications.

4.3 Applications of robust sampled problem

The number of real-world applications of the robust programming is a little more sparse. The most important are the robust truss topology design and the robust portfolio selection problems; there are also other applications, especially in finance (e. g. option modeling), management (supply chain management), or engineering (power supply).

- Robust Truss Topology Design ([1]). The problem is to select the optimal configuration of a structural system (mechanical, aerospace, ...) that is subjected to one ore more given loads (nominal loads) and an unspecified set of small uncertain loads. The goal is to find such configuration that the construction is rigid to all of the prescribed loads.
- Robust portfolio selection ([10]). Here, the uncertain parameters are the modeling errors in the estimates of market parameters and they are assumed to lie in a known and bounded uncertainty set. The robust portfolio is the solution to an optimization problem where the worst-case behaviour of parameters is assumed.

The optimal solutions to robust problems hedge against the worst-case realization of uncertain parameters regardless their "importance". The randomized (sampled) approach incorporate the information about importance to the model via probability distribution of samples so that the optimal solution of the sampled problem does not have to satisfy the constraints for all possible realizations of parameter. At the same time, the probability of such violation is small and for a given ε one could easily compute the number of samples to generate in order to have an ε-feasible solution. Indeed, if the number of samples is significantly greater than required, the optimal solution of the sampled problem also hedges against the parameters with the smaller probability of occurrence. This could be the task if the risk of constraint violation has to be minimized as much as possible and costs of doing that are of smaller importance. This is usually the case of the truss constructions mentioned above.

5 Conclusion

Many economic and engineering applications lead to the same optimization model that incorporate uncertainty parameter. One can handle this parameter in a variety of ways: including ignoring uncertainty, using stochastic or robust programming. But the decision about how to deal with uncertainty cannot be left out of an economical analysis – the desired optimal solution of the problem is closely related to the economical and managerial background of the problem.

Acknowledgements. This research was supported by the Czech Science Foundation under the projects No. 402/03/H057, 402/04/1294, and No. 402/05/0115.

References

[1] A. Ben-Tal and A. Nemirovski. Robust truss topology desing via semidefinite programming. *SIAM Journal on Optimization*, 7(4):991–1016, 1997.

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[3] D. Bertsimas and M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical Programming*, 107(1-2):5–36, 2006.

[4] J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer-Verlag, New York, 1997.

[5] G. Calafiore and M. C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming, Ser. A*, 102(1):25–46, 2005.

[6] A. Charnes and W. W. Cooper. Chance constrained programming. *Management Science*, 6:73–79, 1959.

[7] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[8] D. P. de Farias and B. Van Roy. On constraint sampling in the linear programming approach to approximate dynamic programming. *Mathematics of Operations Research*, 29(3):462–478, 2004.

[9] E. Erdoğan and G. Iyengar. Ambiguous chance constrained problems and robust optimization. *Mathematical Programming*, 107(1-2):37–61, 2006.

[10] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[11] R. Henrion and A. Möller. Optimization of a continuous distillation process under random inflow rate. *Computers & Mathematics with Applications*, 45(1-3):247–262, 2003.

[12] R. Henrion and W. Römisch. Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. *Mathematical Programming*, 84(1):792–818, 1999.

[13] M. Houda. Estimation in chance-constrained problem. In H. Skalská, editor, *Proceedings of the 23rd International Conference Mathematical Methods in Economics 2005*, pages 134–139, Hradec Králové, 2005. Gaudeamus.

[14] M. Houda. Comparison of approximation in stochastic and robust programming. In *Proceedings of the 7th Prague Symposium on Asymptotic Statistics and 15th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes Prague Stochastics 2006*, Prague, 2006. MatfyzPress. Accepted for publication.

[15] P. Kall. *Stochastic Linear Programming*. Springer-Verlag, Berlin, 1976.

[16] P. Kall and S. W. Wallace. *Stochastic Programming*. Wiley, Chichester, 1994.

[17] V. Kaňková. On the convergence rate of empirical estimates in chance constrained stochastic programming. *Kybernetika*, 26:449–461, 1990.

[18] V. Kaňková. A note on relationship between distribution function estimation and estimation in stochastic programming. In P. Lachout and J. Á. Víšek, editors, *Transactions of the 12th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 122–125, Prague, 1994. Academy of Sciences of the Czech Republic.

[19] A. Prékopa. *Stochastic Programming*. Akadémiai Kiadó, Budapest, 1995.

[20] A. Prékopa and T. Szántai. Flood control reservoir system design using stochastic programming. *Mathematical Programming Study*, 9:138–151, 1978.

[21] S. T. Rachev and W. Römisch. Quantitative stabilisy in stochastic programming: The method of probability metrics. *Mathematics of Operations Research*, 27(4):792–818, 2002.

[22] W. Römisch. Stability of stochastic programming problems. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming. Handbooks in Operations Research and Management Science*, volume 10, pages 483–554. Elsevier, Amsterdam, 2003.

[23] A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. Elsevier, Amsterdam, 2003.

[24] S. W. Wallace and W. T. Ziemba, editors. *Applications of Stochastic Programming*, volume 5 of *MPS-SIAM Series on Optimization*. SIAM, Philadelphia, 2005.

Mathematical Methods in Accounting

Milan Hrdý
University of West Bohemia in Pilsen, Faculty of Economics,
Husova 11, 30614 Plzeň
hrdy@fek.zcu.cz

Abstract

The modern international accounting notwithstanding if it is based on international accounting standards, on US GAAP or on national standards requires more and more assets to be evaluated in their real or fair value. It must reflect the real economic situation in the firm as precisely as possible and so it can improve its predicability ability, mainly for the purpose of the financial management of the firm. One of the most suitable approach how to evaluate the accounting assets is the use of the modern mathematical methods. These methods are used mainly in the process of the evaluation of different types of securities, for example when we are calculating the fair (real) value of the bonds with the use of the present value of the future interests and principals. The mathematical methods in accounting have an extra important position in the sphere of the accounting standards for financial institutions. Very often is necessary to identify the amortised cost of a financial assets or financial liability or accured interests. Sometimes is necessary to overrate the value of securities. Particularly important are these methods in the process of evaluation of the derivates. This process is very complicated, mainly if we are calculating the fair (real) value of the monetary forward. The mathematical methods can help us also to identify of the expected losses of some special derivates. This contribution describes the usage of the mathematical methods in accounting generally and also on the concrete examples.

Keywords

Methods, Mathematics, Accounting
JEL: G0

1 Introduction

The modern accounting methods require more and more assets to be evaluated in their real or fair value. This concerns mainly IAS (International accounting standards) or GAAP (Generally Accepted Accounting Principles), but also these methods are gradually projected into national accounting standards including the czech accounting standards. But identifying of the real or fair value of assets is sometimes very complicated. Sometimes the mathematical methods can help the accountants mainly if they are trying to evaluate the financial assets.

The principles of evaluation methods are stated in the International Valuation standard 39 which concens the financial instruments. There is a lot of different cases where the financial instruments must be evaluated in their fair value, some instruments (bonds) are based on the principle that the fair value equals the present value of the future incomes which the owner of the instrument can obtain within some time period. Sometimes are necessary the special methods, for example in case of option pricing or derivates valuations. This contribution concerns mainly the options´and derivates´ valuation.

2 International Accounting Standard 39

IAS 39 *Financial Instruments: Recognition and Measurement* was issued in December 2003 and is applicable for annual periods beginning on or after 1 January 2005. IAS 39 prescribes principles for recognising and measuring all types of financial instruments besides some speciál exceptions. IAS 39 applies to those contracts to buy or sell a non-financial item that can be settled net in cash or another financial instrument, or by exchanging financial instruments, as if the contracts were financial instruments. However, IAS 39 does not apply to any such contracts that were entered into and continue to be held for the purpose of the receipt or delivery of a non-financial item in accordance with the entity's expected purchase, sale or usage requirements.

A financial asset or liability is recognised when the entity becomes a party to the instrument contract. A financial liability is derecognised when the liability is extinguished. A financial asset is derecognised when, and only when the contractual rights to the cash flows from the asset empire, or the entity transfers substantially all the risks and rewards of ownership of the asset, or the entity transfers the asset, while retaining some of the risks and rewards of ownership, but no longer has control of the asset (ie the transferee has the ability to sell the asset). The risks and rewards retained are recognised as an asset.

Financial assets and liabilities are initially recognised at fair value. Subsequent measurement depends on how the financial instrument is categorised:

- At amortised cost using the effective interest metod.
- Held-to-maturity investments: non-derivative financial assets with fixed or determinable payments and maturity that the entity has the positive intention and ability to hold to maturity.
- Loans and receivables: non-derivative financial assets with fixed or determinable payments that are not quoted in an active market.
- Financial liabilities that are not held for trading and not designated at fair value through profit or loss.

Concerning fair value financial asset or liability that is classified as held for trading, is a derivative or has been designated by the entity at inception as at fair value through profit or loss. Non-derivative financial assets that do not fall within any of the other categories. The unrealised movements in fair value are recognised in equity until disposal or sale, at which time, those unrealised movements from prior periods are recognised in profit or loss. If there is objective evidence that a financial asset is impaired, the carrying amount of the asset is reduced and an impairment loss is recognised. A financial asset carried at amortised cost is not carried at more than the present value of estimated future cash flows. An impairment loss on an available-for-sale asset that reduces the carrying amount below acquisition cost is recognised in profit or loss.

IAS 39 provides also for two kinds of hedge accounting, recognising that entities commonly hedge both the possibility of changes in cash flows (i.e. a cash flow hedge) and the possibility of changes in fair value (ie a fair value hedge). Strict conditions must be met before hedge accounting is applied:

- There is formal designation and documentation of a hedge at inception.
- The hedge is expected to be highly effective (ie the hedging instrument is expected to almost fully offset changes in fair value or cash flows of the hedged item that are attributable to the hedged risk).
- Any forecast transaction being hedged is highly probable.
- Hedge effectiveness is reliably measurable (ie the fair value or cash flows of the hedged item and the fair value of the hedging instrument can be reliably measured).
- The hedge must be assessed on an ongoing basis and be highly effective.

When a fair value hedge exists, the fair value movements on the hedging instrument and the corresponding fair value movements on the hedged item are recognised in profit or loss. When a cash flow hedge exists, the fair value movements, on the part of the hedging instrument that is effective, are recognised in equity until such time as the hedged item affects profit or loss. Any ineffective portion of the fair value movement on the hedging instrument is recognised in profit or loss.

IAS 39 requires derivatives that are embedded in non-derivative contracts to be accounted for separately at fair value through profit or loss.

3 Derivates Valuation for Accounting Purpose

Financial derivates have grown rapidly during the past decade primarily because of fundamental changes in global financial markets, advancements in computer technology, and fluctuations in interest and currency exchange rates. Derivates, which are defined as financial products such as swaps, options, futures, forwards and unstructed receivables deriving their value from underlying financial instrument, have become increasingly important and widely used in global business. IAS 39 establishes, as was has been mentioned in previous chapter, accounting and reporting steps for derivate instruments and hedging activities by requiring that affected entities regognize all derivates as either assets or liabilities in financial statements nad measure them at fair value. Derivates has been used for managing financial risks, speculating on the price of the financial instruments, for reducing the cost of raising capital, earning higher investment returns, for adjusting investment portfolios to take advantage of miss-pricing between stock baskets and stock index futures, and for combining derivates with other financial instruments to create new and more powerful financial products.

The increasing use of derivates contracts available over the counter, on exchanges, and through private placements has raised serious concerns regarding their proper valuations. A number of mathematical valuation models for different types of derivates have been developed based on the premise that if the suggested model accurately determines the value of a derivate, its marketmarket price should equal its theoretical fair value. The quantitative models range from the relatively simple models (for example Binomical Option Pricing Model) to more complex and sophisticated models (for example Black-Scholes model, Digital Contracts).These models help accountants to state the fair value of these derivates.

As some examples it can be mentioned here the binomical model for one – period **Binomical Option Pricing Formula**, **Black** and **Scholes Pricing Model and Digital Contracts for Valuation of Derivates**. Binomical Option Pricing Formula determines the option price as a weighted average of the two possible option prices at expiration, discounted at the risk-free rate. Mathematically the option price is calculated as follows:

$$C = (P*C(u) + (1 - P)*C(d)) / (1 + r))$$

where:
C is theoretical fair value of call option,
$C(u)$ = Max [0, $S * (1 + u) - E$], is the price of call option when it goes up, and S is stock price, E is the exercise price of call, and u is the percentage increase in value of stock,
$C(d)$ = Max [0, $S * (1 + d) - E$], is the price of call option when it goes down, and d is the percentage decrease in value of stock,
$P = (r - d) / (u - d)$, where r is the risk-free rate.

This model can be demonstrated on the following example. Assume that a stock is currently priced at 300 Euro and can go up to 354 Euro (an increase of 18 %), or down to 240 Euro (a decrease of 20 percent). Furthermore the exercise price of a call option is is 250 Euro, and the risk-free rate is 12 %. The theoretical fair value for the accounting purpose is calculated as follows:

$C(u)$ = Max [0, $S * (1 + u) - E$] = Max [0, 300 * (1 + 0,18) – 250 = 104

$C(d)$ = Max [0, $S * (1 + d) - E$] = Max [0, 300 * (1 – 0, 20) – 250] = 0

$P = (r - d) / (u - d)$ = (0,12 – (- 0,20)) : (0,18 – (- 0,20)) = 0,842

$C = (P*C(u) + (1 - P)*C(d)) / (1 + r))$ = ((0,842*104) + (0,158*0)) / (1 + 0,12) = **78 Euro**

The problems with the Black-Scholes Option Model is analysed in the next chapter. Digital contracts are simple bulding blocks that provide a unified approach for determining formulas for a wide variety of financial instruments. Digital contacts are simple because their payoffs are either “on” or “off”, indicating that a digital option pays at maturity either one dollar (on) or nothing (off), depending on its payoff event. Rezaee (2001) suggested a three-step valuation process with digital contracts.

The first step is the determination of the risk neutral probability of a particular payoff event. The second step involves the development of formulas for the digital contracts.
The third step is to use these formulas to value financial derivate contracts. To simplify, mathematically, a pure European style call option can be valued according to the Rezaee (2001) as follows:

$$P(c) = \sum a_i * k(S_i, t_i, T_i, M_i) + \sum b_j * L (S_j, t_j, T_j, M_j),$$

where $k(S_i, t_i , T_i, M_i)$ is the value at time t_i of receiving 1 dollar at time T_i, the maturity date, if and only if the event M_i occurs, and $L (S_j, t_j, T_j, M_j)$ is the value at a time t_j of receiving one share of stock at time t_j of receiving one share of stock at time T_j (without dividends), if and only if the event M_i occurs.
Generally, the probability that any event $M_{(.)}$ happens depends on the current stock price S, so the values of k and L are determined based on stock price S. Thus, the value of call option, is determined based on the value of stock.

The above mentioned models are used especially in the accounting practises of the financial institutions. The identification of the fair value of the financial asssets is there very complicated and so these mathematical models can be very usefull.

Some special methods are used also of valuation of the american options. American options are derivative securities for which the holder of the security can choose the time of exercise. In an American put, for example, the option holder has the right to sell an underlying security for a specified price K (the strike price) at any time between the initiation of the agreement (t = 0) and the expiration date ($t = T$). The exercise time τ can be represented as a stopping time; so that American options are an example of optimal stopping time problems.

Valuation of American options presents at least two difficulties. First, there is a singularity in the option characteristics at the expiration time. For American puts and calls on equities with dividends, a thorough analysis of this singularity was performed by Evans, Kuske, and Keller (2002).
A second difficulty occurs for Monte Carlo valuation of American options. Monte Carlo methods are required for options that depend on multiple underlying securities or that involve path dependent features. Since determination of the optimal exercise time depends on an average over future events, Monte Carlo simulation for an American option has a “Monte Carlo on Monte Carlo" feature that makes it computationally complex. There are several methods for overcoming this difficulty withAmerican options.

The first, developed by Broadie and Glasserman (2004) involves two branching processes, the first of which provides an upper bound and the second a lower bound on option price. The second method, is a Martingale optimization formula developed in Rogers (2002) that provides a dual formulation of the Monte Carlo valuation formula and leads naturally to an upper bound on the option price. The third is the Least Squares Monte Carlo (LSM) derived by Longstaff and Schwartz (2001). Finally it is the work by the authors on use of quasi-random sequences in LSM Chaudhary (2004).

4 Option Valuation for Accounting Purpose

The history of the development of option valuation approaches are very good describes at Breadley Education Institution [10]. Modern option pricing techniques, with roots in stochastic calculus, are often considered among the most mathematically complex of all applied areas of finance. These modern techniques derive their impetus from a formal history dating back to 1877, when Charles Castelli wrote a book entitled The Theory of Options in Stocks and Shares. Castelli's book introduced the public to the hedging and speculation aspects of options, but lacked any monumental theoretical base.

Twenty three years later, Louis Bachelier offered the earliest known analytical valuation for options in his mathematics dissertation "Theorie de la Speculation" at the Sorbonne. He was on the right track, but he used a process to generate share price that allowed both negative security prices and option prices that exceeded the price of the underlying asset. Bachelier's work interested a professor at MIT named Paul Samuelson, who in 1955, wrote an unpublished paper entitled "Brownian Motion in the Stock Market".

During that same year, Richard Kruizenga, one of Samuelson's students, cited Bachelier's work in his dissertation entitled "Put and Call Options: A Theoretical and Market Analysis". In 1962, another dissertation, this time by A. James Boness, focused on options. In his work, entitled "A Theory and Measurement of Stock Option Value", Boness developed a pricing model that made a significant theoretical jump from that of his predecessors. More significantly, his work served as a precursor to that of Fischer Black and Myron Scholes, who in 1973 introduced their landmark option pricing model.

The Black and Scholes Option Pricing Model didn't appear overnight, in fact, Fisher Black started out working to create a valuation model for stock warrants. This work involved calculating a derivative to measure how the discount rate of a warrant varies with time and stock price. The result of this calculation held a striking resemblance to a well-known heat transfer equation. Soon after this discovery, Myron Scholes joined Black and the result of their work is a startlingly accurate option pricing model. Black and Scholes can't take all credit for their work, in fact their model is actually an improved version of a previous model developed by A. James Boness in his Ph.D. dissertation at the University of Chicago. Black and Scholes' improvements on the Boness model come in the form of a proof that the risk-free interest rate is the correct discount factor, and with the absence of assumptions regarding investor´s risk.

The Black-Scoles model takes the well-known following form:

$$C(t) = S*N(d_1) - K*\exp(-rt)*N(d_2),$$

where C is theoretical call premium, S is current stock price, t is time until option respiration, K is option striking price, r is risk-free interest rate, and $N(.)$ denotes cumulative standart mornal distrinution.

The quantities d_1 and d_2 take the following form

$$d_1 = (\ln(S/K) + (r+ s^2/2)/(s\sqrt{t}) \quad , \quad d_2 = d_1 - s\sqrt{t} \, ,$$

where s denotes standard deviation of stock returns.

In order to understand the model itself, it is useful to divide it into two parts. The first part, $SN(d1)$, derives the expected benefit from acquiring a stock outright. This is found by multiplying stock price S by the change in the call premium with respect to a change in the underlying stock price $N(d1)$. The second part of the model, $K*\exp(-rt)N(d2)$, gives the present value of paying the exercise price on the expiration day.

The fair market value of the call option is then calculated by taking the difference between these two parts. There are some limitations of Black-Scholes Model which are the following:
The stock pays no dividends during the option's life.Most companies pay dividends to their share holders, so this might seem a serious limitation to the model considering the observation that higher dividend yields elicit lower call premiums. A common way of adjusting the model for this situation is to subtract the discounted value of a future dividend from the stock price.

European exercise terms are used. European exercise terms dictate that the option can only be exercised on the expiration date. American exercise term allow the option to be exercised at any time during the life of the option, making american options more valuable due to their greater flexibility. This limitation is not a major concern because very few calls are ever exercised before the last few days of their life. This is true because when you exercise a call early, you forfeit the remaining time value on the call and collect the intrinsic value. Towards the end of the life of a call, the remaining time value is very small, but the intrinsic value is the same.

Markets are efficient. This assumption suggests that people cannot consistently predict the direction of the market or an individual stock. The market operates continuously with share prices following a continuous Itô process. To understand what a continuous Itô process is, you must first know that a Markov process is "one where the observation in time period t depends only on the preceding observation." An Itô process is

simply a Markov process in continuous time. If you were to draw a continuous process you would do so without picking the pen up from the piece of paper.
No commissions are charged. Usually market participants do have to pay a commission to buy or sell options. Even floor traders pay some kind of fee, but it is usually very small. The fees that Individual investor's pay is more substantial and can often distort the output of the model.

Interest rates remain constant and knot. The Black and Scholes model uses the risk-free rate to represent this constant and known rate. In reality there is no such thing as the risk-free rate, but the discount rate on U.S. Government Treasury Bills with 30 days left until maturity is usually used to represent it. During periods of rapidly changing interest rates, these 30 day rates are often subject to change, thereby violating one of the assumptions of the model.Returns are lognormally distributed. This assumption suggests, returns on the underlying stock are normally distributed, which is reasonable for most assets that offer options.

There are example of valuations in different situations, which are based on the well-known greeks frequently used in financial analysis:

Delta – measures sensitivity of calculated option value to small changes in the share price:

$$\Delta = \partial C/\partial S = N(d_1),$$

Gamma – measures the second sensitivity to price, i.e. the sensitivity of calculated delta sensitivity to small changes in share price:

$$\Gamma = \partial C^2/\partial S^2 = \varphi(d_1)/(S\sigma\sqrt{t}),$$

Theta – measures the calculated option value sensitivity to small changes in time till maturity:

$$\theta = - \partial C/\partial t = S\varphi(d_1)/(2\sqrt{t}) - rK \exp(-rt)N(d_2),$$

Vega, sometimes *kappa* – measures the calculated option value sensitivity to small changes in volatility:

$$\kappa = \partial C/\partial\sigma = S\sigma\varphi(d_1)\sqrt{t},$$

Rho – measures sensitivity to risk-free interest rate:

$$\rho = \partial C/\partial r = KT \exp(-rt)N(d_2),$$

where $\varphi(.)$ denotes probability density function of standard normal distribution.

Since 1973, the original Black and Scholes Option Pricing Model has been the subject of much attention. Many financial scholars have expanded upon the original work. In 1973, Robert Merton relaxed the assumption of no dividends. In 1976, Jonathan Ingerson went one step further and relaxed the the assumption of no taxes or transaction costs. In 1976, Merton responded by removing the restriction of constant interest rates. The results of all of this attention, that originated in the autumn of 1969, are alarmingly accurate valuation models for stock options.

5 Conclusion

The mathematical methods are very often used in accounting practise, mainly if the accountants must identify the value of financial assets such as bonds, options and the other types of derivates. According to the IAS 39 the financial assets must be mostly evaluated in their fair value. The problem is how to identify this value. Here the mathematical and numerical methods can be very useful.
In option valuation there are Binomic models, Black and Scholes models and Digital Model available. The situation is more complicated if we need to evaluate the american option or to evaluate corporate bond with stochastic interest rate.

References

[1] Caflish, R.E., Chaudhary, S.: Monte Carlo Method for American Option. Winter Simulation Conference, University of California, USA, 2004

[2] Black, M., Scholes, F.: The pricing of options and corporate liabilities. Journal Political Economy, 1973, p.637 – 654

[3] Boness, J.: A theory and measurement of a stock option value. Dissertation Work, USA, 1962

[4] Longstaff, F., Schwarz, E.: Valuing american options by simulation. The review of financial studies, 2002, p. 113 - 147

[5] Merton, R.: The theory of rational option pricing. Bell Journal of Economics and Management Science, 1973, p.141-183

[6] Broadie, M., Glasserman, P.: Pricing american style securities using simulation. Journal of Economics Dynamics and Control, 1997, p. 1323 – 1352

[7] Rezaee, Z.: Financial Institutions Valuations, Mergers, and acquisitions. John Wiley and Sons, USA, 2001, ISBN 0-471-39449

[8] Rogers, L.C.G.: Monte Carlo Valuation of american option. Mathematical Finance, 2002, p. 271-286

[9] IAS 39 – Financial Instruments

[10] http://bradley.bradley.edu/~arr/bsm/pg04.html

Change in the mean versus random walk: a simulation study

Marie Hušková, Miriam Marušiaková
Charles University of Prague, Department of Statistics, Sokolovská 83, CZ – 186 75 Prague 8, Czech Republic;
marie.huskova@karlin.mff.cuni.cz, marusiak@csmat.karlin.mff.cuni.cz

Abstract

The paper concerns statistical procedures for distinguishing between a change in the mean in a location model with stationary error terms and the situation when the error terms in the location model form a random walk. A simple two step test procedure is described, possible modifications are discussed and its behaviour is checked on a simulation study. The related theoretical results are derived in Aue et al. (2006).

Keywords

CUSUM statistic, Bartlett estimator, change in mean, random walk
JEL: C15

Acknowledgement

The work was supported by grants GAČR 201/06/0186 , GAČR 201/05/H007 and MSM 0021620839. The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s..

1 Introduction

We consider the time series model:

$$Y_i = \mu_i + e_i, \quad i = 1, \dots, n, \tag{1}$$

where $\mu_1, \dots, \mu_n$ are location parameters and $e_1, \dots, e_n$ are random errors fulfilling some regularity conditions.

One of the basic problems considered in the change point analysis is to detect a change in the location in the location model (1). It can be formulated as hypothesis testing problem $H_0 : \mu_1 = \dots = \mu_n$ versus $H_1 : \mu_1 = \dots = \mu_m \neq \mu_{m+1} = \dots = \mu_n$ for some $m < n$, where the error terms form a stationary sequence. The test statistics are usually constructed in a way that their large values indicate a change. In probability terms it means that as the number of observations tends to infinity the test statistic is bounded in probability when there is no change while it tends to infinity under a change. However, the important role plays also dependence among the observations. Particularly, if the error terms form a random walk (sometimes called unit root alternatives) then the test statistics used for our testing problem become also large even for no change in the location. The problem is to propose test procedures that would distinguish between a change in location and the unit root alternatives. In the present paper we are interested in test procedures that would distinguish among the following 3 scenarios:

(A) no change in the location ($\mu_1 = \dots = \mu_n$) and the error terms form a stationary sequence;

(B) there is a change in the location (for some $m < n$ $\mu_1 = \dots = \mu_m \neq \mu_{m+1} = \dots = \mu_n$) and the error terms form a stationary sequence;

(C) the error terms form a random walk (i.e., unit root alternatives).

Basic information on change points can be found, e.g., in Csörgő and Horváth (1997), Antoch et al. (2001). Related results to the above formulated problems are in the papers Giraitis et al. (2001), Hurst (1951), Kwiatkovski et al. (1992), Perron (1990) and Perron and Vogelsang (1992). We can meet such problems in modelling various financial time series as well as in modelling structural changes in econometrics. More details can be found, e.g., in Perron (1990, 2006) and Perron and Vogelsang (1992).

2 CUSUM-type procedure

Aue et al. (2006) have proposed test procedure based on versions of the CUSUM (Page, 1954), R/S (Lo, 1991) and V/S (Giraitis et al., 2001) statistics. They have studied the limit behaviour for all three above mentioned scenarios. The main purpose of the present paper is to discuss various modifications of the CUSUM type procedure developed in Aue et al. (2006) and to present the results of simulation study.
Next we formulate more precisely three considered scenarios:

(A) $\mu_1 = \ldots = \mu_n$ and the partial sums of the error terms $\{e_i\}$ satisfy the functional central limit theorem, i.e., with some positive σ, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} e_i \to^{\mathcal{D}[0,1]} \sigma W(t), \tag{2}$$

where $\lfloor . \rfloor$ is the integer part, $\{W(t); t \in [0, 1]\}$ denotes a Brownian motion and $\to^{\mathcal{D}[0,1]}$ denotes for weak convergence in the Skorohod space $\mathcal{D}[0, 1]$. For definition and basic assertion, see, e.g., Davidson (2002).

(B) for some $\theta \in (0, 1)$ $\mu_1 = \ldots = \mu_m \neq \mu_{m+1} = \ldots = \mu_n$ with $m = \lfloor n\theta \rfloor$ and the partial sums of error terms $\{e_i\}$ satisfy the functional central limit theorem, i.e., (2).

(C) $|\mu_i| < c$ for some $c > 0$, $i = 1, \ldots, n$ and the error terms $\{e_i\}$ itself satisfy the functional central limit theorem, i.e., with some positive σ, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} e_{\lfloor nt \rfloor} \to^{\mathcal{D}[0,1]} \sigma W(t). \tag{3}$$

Notice that (C) admits changes in the location.

The CUSUM test statistic for testing (A) versus (B) has the form:

$$T_n = \frac{1}{s_n \sqrt{n}} \max_{1 \le k < n} \left| \sum_{i=1}^{k} (Y_i - \overline{Y}_n) \right|, \tag{4}$$

where s_n is an estimator of the scale σ in (2). One can use the so called Bartlett estimator that takes into account possible change in the location defined as follows:

$$s_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \overline{Y}_{in}(\widehat{m})\right)^2 + \frac{2}{n} \sum_{j=1}^{q(n)} \left(1 - \frac{j}{q(n)+1}\right) \sum_{i=1}^{n-j} \left(Y_i - \overline{Y}_{in}(\widehat{m})\right) \left(Y_{i+j} - \overline{Y}_{i+j,n}(\widehat{m})\right) \tag{5}$$

with

$$q(0) > 0, \quad q(n) \to \infty, \quad \frac{q(n) \log n}{n} \to 0 \quad n \to \infty, \tag{6}$$

$$\overline{Y}_{in}(\widehat{m}) = I\{1 \le i \le \widehat{m}\} \overline{Y}(0, \widehat{m}) + I\{\widehat{m} < i \le n\} \overline{Y}(\widehat{m}, n),$$

$$\overline{Y}(j, k) = \frac{1}{k - j} \sum_{i=j+1}^{k} Y_i, \quad 0 \le j < k \le n.$$

Large values of T_n indicate that the scenario (A) is not true. If this is the case then we calculate an estimator of the change point m, e.g., we can use

$$\widehat{m} = \min \left\{ k = 1, \ldots, n : \left| \sum_{i=1}^{k} \left(Y_i - \overline{Y}_n\right) \right| = \max_{1 \le k < n} \left| \sum_{i=1}^{k} (Y_i - \overline{Y}_n) \right| \right\} \tag{7}$$

and split the observations into two parts $(Y_1, \ldots, Y_{\widehat{m}})$ and $(Y_{\widehat{m}+1}, \ldots, Y_n)$ and calculate the analogs of T_n based on the single parts. For this step we have to calculate $T_n(0, \widehat{m})/s_n(0, \widehat{m})$, $T_n(\widehat{m}, n)/s_n(\widehat{m}, n)$ defined by

$$T_n(j, k) = \frac{1}{\sqrt{k - j}} \max_{j \le v \le k} \left| \sum_{i=j+1}^{v} \left(Y_i - \overline{Y}(j, k)\right) \right|, \quad 0 \le j < k \le n \tag{8}$$

$$s_n^2(j,k) = \frac{1}{k-j}\sum_{i=j+1}^{k}\left(Y_i-\overline{Y}(j,k)\right)^2+\frac{2}{k-j}\sum_{v=1}^{q(k-j)}\left(1-\frac{v}{q(k-j)+1}\right) \cdot \sum_{i=j+1}^{k-v}\left(Y_i-\overline{Y}(j,k)\right)\left(Y_{i+j}-\overline{Y}(j,k)\right), \quad 0\le j<k\le n. \tag{9}$$

The test procedure is the following:

(i) Calculate T_n according to (4) and for chosen level α find the $1-\alpha$ quantile $x_{1-\alpha}$ of the distribution of $\sup_{0<t<1}|B(t)|$, where $\{B(t);\ t\in[0,1]\}$ is a Brownian bridge;

(ii) if $T_n < x_{1-\alpha}$, our decision: there is no evidence that the scenario (A) is violated;

(iii) if $T_n \ge x_{1-\alpha}$ we calculate $\widehat{m}$ and

$$T_{n,1} = \max\left\{\frac{T_n(0,\widehat{m})}{s_n(0,\widehat{m})},\ \frac{T_n(\widehat{m},n)}{s_n(\widehat{m},n)}\right\}$$

and find the quantile $x_{\sqrt{1-\alpha}}$;

(iv) if $T_{n,1} < x_{\sqrt{1-\alpha}}$ then our decision is (B), i.e., there is a change in the mean (there can be eventually more than one change);

(v) if $T_{n,1} \ge x_{\sqrt{1-\alpha}}$, then our decision is (C), i.e., the unit alternative holds true.

Here are the crucial asymptotic results proved in Aue et al. (2006):

(I) Under scenario (A), as $n\to\infty$, $T_n/s_n \to^D \sup_{0<t<1}|B(t)|$ and $\widehat{m}/n$ has the same limit distribution as $\arg\max_{0<t<1}|B(t)|$.

(II) Under scenario (B), as $n\to\infty$, $T_n \to^P \infty$, $\widehat{m}/n \to^P \theta$ and

$$T_{n,1} \to^D \max\left\{\sup_{0<t<1}|B_1(t)|,\ \sup_{0<t<1}|B_2(t)|\right\},$$

where $\{B_1(t),\ t\in(0,1)\}$ and $\{B_2(t),\ t\in(0,1)\}$ are independent Brownian bridges.

(III) Under scenario (C), as $n\to\infty$, $T_n \to^P \infty$ and $T_{n,1}\to^P \infty$.

These asymptotic results immediately imply that

$$\lim_{n\to\infty} P_{(A)}\left(\text{reject}\,(A)\right)=\alpha, \quad \limsup_{n\to\infty} P_{(B)}\left(\text{reject}\,(B)\right)\le\alpha, \quad \lim_{n\to\infty} P_{(C)}\left(\text{reject}\,(B)\right)=1$$

Modifications: Quite analogously we can develop test procedures based on other maxtype test statistics.

3 Simulation results

In this section, we present the results of our simulation study examining the empirical size and power of the CUSUM type procedure described in the previous sections. All computations are performed in R (R Development Core Team, 2005).

We consider the model (1), where

$$\mu_i = \begin{cases} 0 & i=1,\dots,m \\ \Delta & i=m+1,\dots,n \end{cases}$$

and the error terms form an autoregressive sequence of order 1:

$$e_i = \rho e_{i-1} + v_i, \quad i=2,\dots,n$$

with v_i independent and normally distributed with zero mean and unit variance. For our simulations we use

(i) sample size $n\in\{100,200\}$;

(ii) location of the change point $m\in\{n/4,n/2\}$;

(iii) size of the change $\Delta \in \{0, 1, 1.5, 2\}$;

(iv) autoregressive coefficient $\rho \in \{0.3, 0.5, 0.7, 0.9, 1\}$.

For each combination of the parameters we made 10^5 replications. The results are summarised in Table 1. Columns denoted by A show proportions of replications (out of 10^5) for which the scenario (A) was rejected. Similarly, columns B give proportions of replications (out of the replications for which $T_n \geq x_{1-\alpha}$ hold) rejecting the scenario (B) in the second step of the procedure. For the significance level $\alpha = 0.05$ the critical values $x_{1-\alpha}$ and $x_{\sqrt{1-\alpha}}$ are equal to 1.358 and 1.48, respectively (Kiefer, 1959, Table 1-2). We chose $q(n) = \lfloor n/\log(n)^2 \rfloor$ in the formula of the Bartlett estimator - it satisfies all conditions in (6). For calculation of the test statistic T_n we use the standard Bartlett estimator $s_n^2(0, n)$, cf (9) and Bartlett estimator with correction for the possible change point (5).

In case of no change ($\Delta = 0$) and stationary errors ($0 < \rho < 1$) the CUSUM procedure with the standard Bartlett estimator performs much better than the procedure with the corrected Bartlett estimator. For the latter one we can see substantial size distortions. The test statistic T_n, cf (4), rejects the null hypothesis, i.e., the scenario (A), too frequently for all values of the parameters. For $\rho = 0.9$ and sample size $n = 100$ the rejection rate is even 59 %. The bias can be caused by the fact that the change point estimator $\hat{m}$, cf (7), can be either small or close to n under no change situation. In that case the mean estimator used in (5) is calculated from a small number of observations and is not very reliable.

As the dependence among errors increases (ρ increases), the tendency of both procedures to reject the null hypothesis is higher. Especially for $\rho = 0.9$ the procedures perform poorly. The procedure with the standard Bartlett estimator has a reasonable empirical size only if the dependence among errors is weaker ($\rho \leq 0.5$).

For larger sample size n the empirical size of the test T_n is improved for all considered values of the parameters.

If the scenario (B) holds, i.e., there is a change in the location ($\Delta > 0$) and the errors are stationary ($0 < \rho < 1$), the power of the procedures is quite high for weaker dependence among errors (small ρ). As ρ increases, the test procedures become less sensitive.

As the sample size n increases, the power of the test to reject no change situation is considerably higher. The exception is the situation when the parameter ρ is near 1 ($\rho = 0.9$). Here the power declines surprisingly with larger sample size n.

As expected, the power of the test increases with the size of the change Δ.

The results also depend on the location of the change point. The test has a higher power when the change point is located in the middle of the time series ($m = n/2$) than when it is in the first quarter ($m = n/4$).

The procedure is more powerful if we use the Bartlett estimator with correction for a possible change point instead of the standard Bartlett estimator.

If the error terms form a random walk ($\rho = 1$), the test statistic T_n, cf (4), is large also under no change situation ($\Delta = 0$). The procedure behaves in the same way for different Δ - it can not distinguish between a situation with no change and a situation with a change if the errors are nonstationary. The power of the procedure does not depend on the location of the change point.

In sample sizes of 100 observations we decide correctly for the scenario (C) in approximately $0.77 \cdot 0.48 \approx$ 37 % of all simulated data. But when the sample size reaches 200, the power of the procedure increases dramatically - almost $0.83 \cdot 0.78 \approx 68$ % of all replications indicate that the scenario (C) is true. We can even improve the power if we use the Bartlett estimator that takes into account possible change in the location.

The procedure behaves very similarly for normal errors v_i and for Laplace ones (not shown in Table 1).

4 Summary

We studied the performance of the two step CUSUM-type procedure which is able to distinguish between change in the location and random walk alternatives. We used the test with standard Bartlett estimator and Bartlett estimator that takes into account possible change in the location. The latter one seemed to be more powerful. On the other hand, it suffered from size distortions. In order to obtain reasonable size, the procedure could be modified in the following way: if the change point estimator lies near the boundaries of the sample, one should rather use the standard Bartlett estimator as an estimator of a scale.

The power of the CUSUM procedure generally declines as the dependence among the error terms is stronger. When the errors are close to nonstationary behaviour, the procedure performs very poorly and its power is even lower for increased sample size. Possibly different results can be obtained if we choose, e.g., $q(n) = \sqrt{n}$ in the Bartlett estimator.

Values		T_n with standard Bartlett estimator								T_n with Bartlett estimator with correction							
		$n=100$				$n=200$				$n=100$				$n=200$			
		$m=n/2$		$m=n/4$		$m=n/2$		$m=n/4$		$m=n/2$		$m=n/4$		$m=n/2$		$m=n/4$	
ρ	Δ	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
0.3	0.0	0.035	0.006	0.035	0.006	0.036	0.016	0.037	0.011	0.120	0.006	0.119	0.009	0.101	0.014	0.100	0.017
0.3	1.0	0.821	0.007	0.494	0.009	0.988	0.020	0.866	0.027	0.931	0.007	0.776	0.010	0.997	0.022	0.960	0.025
0.3	1.5	0.993	0.008	0.880	0.014	1.000	0.023	0.999	0.040	0.999	0.008	0.980	0.014	1.000	0.025	1.000	0.040
0.3	2.0	1.000	0.008	0.992	0.017	1.000	0.025	1.000	0.042	1.000	0.008	1.000	0.018	1.000	0.025	1.000	0.041
0.5	0.0	0.055	0.011	0.054	0.012	0.047	0.021	0.048	0.020	0.169	0.015	0.167	0.017	0.123	0.026	0.124	0.026
0.5	1.0	0.599	0.011	0.325	0.014	0.868	0.032	0.599	0.032	0.780	0.012	0.600	0.017	0.939	0.033	0.795	0.032
0.5	1.5	0.913	0.013	0.645	0.020	0.997	0.038	0.935	0.050	0.971	0.014	0.874	0.020	0.999	0.039	0.985	0.049
0.5	2.0	0.993	0.015	0.896	0.026	1.000	0.042	0.998	0.062	0.999	0.015	0.981	0.027	1.000	0.040	1.000	0.062
0.7	0.0	0.112	0.032	0.111	0.035	0.080	0.059	0.079	0.060	0.277	0.042	0.277	0.042	0.186	0.067	0.186	0.066
0.7	1.0	0.399	0.032	0.251	0.032	0.552	0.068	0.327	0.059	0.600	0.036	0.477	0.042	0.700	0.073	0.525	0.063
0.7	1.5	0.661	0.033	0.416	0.043	0.857	0.079	0.603	0.077	0.817	0.037	0.668	0.044	0.929	0.081	0.793	0.079
0.7	2.0	0.863	0.039	0.617	0.052	0.979	0.091	0.849	0.097	0.942	0.039	0.834	0.052	0.993	0.090	0.948	0.096
0.9	0.0	0.374	0.161	0.376	0.160	0.292	0.294	0.291	0.293	0.590	0.184	0.593	0.187	0.471	0.313	0.471	0.310
0.9	1.0	0.434	0.160	0.401	0.164	0.388	0.301	0.338	0.287	0.636	0.177	0.618	0.190	0.557	0.309	0.523	0.302
0.9	1.5	0.500	0.161	0.435	0.170	0.493	0.306	0.400	0.280	0.686	0.177	0.651	0.188	0.643	0.313	0.581	0.295
0.9	2.0	0.583	0.161	0.486	0.175	0.613	0.321	0.476	0.290	0.746	0.176	0.692	0.197	0.739	0.325	0.655	0.303
1.0	0.0	0.769	0.482	0.769	0.481	0.832	0.790	0.831	0.789	0.884	0.477	0.883	0.478	0.915	0.786	0.915	0.788
1.0	1.0	0.769	0.482	0.768	0.487	0.834	0.791	0.832	0.799	0.886	0.476	0.884	0.477	0.916	0.785	0.913	0.788
1.0	1.5	0.772	0.480	0.772	0.485	0.833	0.789	0.834	0.790	0.886	0.473	0.885	0.482	0.916	0.788	0.915	0.786
1.0	2.0	0.774	0.482	0.772	0.490	0.831	0.789	0.832	0.790	0.885	0.475	0.886	0.487	0.916	0.787	0.915	0.785

Table 1: Simulated probabilities of rejection of the scenarios (A) and (B), i.e. simulated probability $P\left(T_n \geq x_{1-\alpha}\right)$ and conditional probability $P\left(T_{n,1} \geq x_{\sqrt{1-\alpha}} | T_n \geq x_{1-\alpha}\right)$, in columns A and columns B, respectively. The test statistic T_n, cf (4), is calculated with standard Bartlett estimator $s_n^2(0, n)$, cf (9) or Bartlett estimator with the correction for a possible change point (5). The significance level is $\alpha = 0.05$.

References

Antoch, J., Hušková, M., and Jarušková, D. (2001). Off-line quality control. In *Multivariate Total Quality Control: Foundation and Recent Advances*, pages 1–86. Springer-Verlag.

Aue, A., Horváth, L., Hušková, M., and Ling, S. (2006). On distinguishing between random walk and changes in the mean alternatives. in preparation.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. J. Wiley.

Davidson, J. (2002). *Stochastic Limit Theory*. Oxford University Press, New York.

Giraitis, L., Kokoszka, P., and Leipus, R. (2001). Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, 10:1002–1024.

Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–799.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramer-V.Mises tests. *The Annals of Mathematical Statistics*, 30:420–447.

Kwiatkovski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178.

Lo, A. (1991). Long-term memory in stock market prices. *Econometrica*, 59:1279–1313.

Page, E. (1954). Continuous inspection schemes. *Biometrika*, 57:100–105.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business and Economic Statistics*, 8:153–162.

Perron, P. (2006). Dealing with structural breaks. In *Palgrave Handbook of Econometrics: Econometric Theory Vol.1*.

Perron, P. and Vogelsang, T. (1992). Testing for a unit root in a time series with a changing mean: corrections and extensions. *Journal of Business and Economic Statistics*, 10:467–470.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Fuzzy logic as liquidity solution in a bank

Jaroslav Charouz
Československá obchodní banka – CSOB, Karviná
jcharouz@csob.cz

Abstract

Banks as financial institutions are obliged to hold for their customers enough liquidity at every point of time. This is rather problematic because of not clear maturity of some passive products. There is a common approach to this problem. A Bank has to determine costumer behavioural connected with a product. It can apply some time series analysis as technical approach or make some fundamental inferences from costumer information e.g. from financial statements. The inferences couldn't be done when analyzing retail costumers. This is not easy to get from the analysis sound and robust information. As an alternative, fuzzy logic could be applied. This allows for the bank to make distribution of liquidity according to chosen significant parameters and to employ expert's opinions at the same time.

Keywords

Bank, liquidity, fuzzy logic,
JEL: G21

1 Liquidity requirements from Basel I to banks (14 principles)

Liquidity framework is given implicitly by Basel committee and 14 rules proposed by the committee are significant for the whole banking system. Developing a structure for managing liquidity the following rules should be satisfied:

1.1 Each bank should have an agreed strategy for the day-to-day management of liquidity. This strategy should be communicated throughout the organization.

1.2 A Bank's board of directors should approve the strategy_and significant policies related to the management of liquidity. The board should also ensure that senior management takes the steps necessary to monitor and control liquidity situation of the bank and immediately if there are any material changes in the bank's current or prospective liquidity position.

1.3 Each bank should have a management structure in place to execute effectively the liquidity strategy. This structure should include the ongoing involvement of members of senior management. Senior management must ensure that liquidity is effectively managed, and that appropriate policies and producers are established to control and limit liquidity risk. Bank should set and regularly review limits on the size of their liquidity position over particular time horizons.

1.4 A bank must have adequate information system for measuring, monitoring, controlling and reporting liquidity risk. Reports should be provided on a timely basis to the bank's board of directors, senior management and other appropriate personnel.

Measuring and Monitoring Net Funding Requirements - Each bank should establish a process for the ongoing measurement and monitoring of net funding requirements:

1.5 Each bank should establish a process for the ongoing measurement and monitoring of net funding requirements.

1.6 A bank should analyse liquidity utilising a variety of "what if" scenarios.

1.7 A bank should review frequently the assumptions utilised in managing liquidity to determine that they continue to be valid.

Managing market access, the following rule should be satisfied:

1.8 Each bank should periodically review its efforts to establish and maintain relationships with liability holders, to maintain the diversification of liabilities, and aim to ensure its capacity to sell assets.

Contingency planning should be as follows:

1.9 A bank should have contingency plans in place that address the strategy for handling liquidity crises and include procedures for making up cash flow shortfalls in emergency situations.

Foreign currency liquidity management:

1.10 Each bank should have a measurement, monitoring and control system for its liquidity positions in the major currencies in which it is active. In addition to assessing its aggregate foreign currency liquidity needs and the acceptable mismatch in combination with its domestic currency commitments, a bank should also undertake separate analysis of its strategy for each currency individually.

1.11 Subject to the analysis undertaken according to Principle 10, a bank should, where appropriate, set and regularly review limits on the size off its cash flow mismatches over particular time horizons for foreign currencies in aggregate and for each significant individual currency in which the bank operates.

Internal controls for liquidity risk management:

1.12 Each bank must have an adequate system of internal controls over its liquidity risk management process. A fundamental component of the internal control system involves regular independent reviews and evaluations of the effectiveness of the system and, where necessary, ensuring that appropriate revisions should be available to supervisory authorities.

Role of public disclosure in improving liquidity is as follows:

1.13 Each bank should have in place a mechanism for ensuring that there is an adequate level of disclosure of information about the bank in order to manage public perception of the organisation and its soundness.

1.14 Supervisors should conduct an independent evaluation of a bank's strategies, policies, procedures and practices related to the management of liquidity. Supervisors should require that a bank has an effective system in place to measure, monitor and control liquidity risk. Supervisors should obtain sufficient and timely information from each bank with which to evaluate its level of liquidity risk and should ensure that the bank has adequate liquidity contingency plans.

The above stated 14 rules are suggestions of BASEL I, which are valid recently. They are reflecting supervisor's viewpoint and at the same time they represent the system of liquidity management. The rules 1- 4 define a structure for liquidity management. A management of every financial institution (e.g. banks) has to be aware of a crucial position among all financial risks. The management of liquidity is really in the heart of risk management because the bank has to obtain sources and only after that it can wisely manage the sources.

The rules 5-7 are tailored to measure net funding requirements. The bank has to take into consideration volatility of assets and liabilities and also off-balance items. The off-balance items are volatile from its' own nature. But some prediction has to be taken from past development and thereafter take some extrapolations. In this part there is a need to construct liquidity ladder, which reflects maturities of assets and liabilities in chosen time intervals. Generally a guide line for fulfilling such report is a contractual maturity. If there is a product with no explicit maturity, the consequent assumptions have to be taken.

The rule 8 defines the market access. A bank has to consider its' possibility for liquidity generation under both common market conditions and worsen market conditions.

The rule 9 defines the general planning. So called contingence plan for liquidity crisis has to be developed e.g. in common, pessimistic and optimistic variation and has to contain approaches of handling liquidity crisis.

The rules 10-11 are focussed on current liquidity management with the stress on FX risk (foreign exchange rate). A bank has to thoroughly reconsider the liquidity risk, which arises from changes of FX rates. The changes can increase an amount of expected inflows. However, bigger changes can for example arise from a psychological influence on depositors of the currency. A bank could be located in two states of the world, firstly it could have only exposures in the foreign currency and secondly it could have also sources in the currency. Generally, the second situation is less risky.

The rule 12 is oriented on internal control. The process of auditing the liquidity management has to be convincing. The processes has to be clear, consistent and has to be tested for its' integrity. Also the working force has to be adequately experienced and educated.

The rule 13 there are needed rules about public disclosure and the rule. A regular, understandable and appropriate information flow about an institution could reduce instability deposits.

The rule 14 generally expects supervisor independent supervision.

2 Foreseen application from Basel II

There are no liquidity requirements in Basel II at first sight. This could be confusing in reality, however it is not fully true. In fact liquidity risk and its potential effects on the financial system is a major concern of regulators. From regulatory perspective a liquidity requirement (coverage) for liquidity risk could be established explicitly. There could be pronounced Best practices for liquidity risk management thereafter or standards for sound payment system could take place.

As explicit reference of liquidity in Basel guidance lines we can mention § 741 in Revised Framework (June 2004): Liquidity risk: "Liquidity is crucial to the ongoing viability of any banking organization. Banks' capital positions can have an effect on their ability to obtain liquidity, especially in a crisis. Each bank must have adequate systems for measuring, monitoring and controlling liquidity risk. Banks should evaluate the adequacy of capital given their own liquidity profile and the liquidity of the markets in which they operate"

The architecture of Basel II consists of three pillars: First one focuses on explicit risk charge for coverage of explicit known risk that is the Credit risk, Market risk and Operational risk. The second one is oriented to the supervisory review and the third deals with the market discipline. A virtual application of this third point of view is firstly determined by the proposal of Basel committee but by the law process this proposal is after many cycles of suggestions from commercial banks and central banks standardised in European directions. National law system of every member of EU has to reflect EU directions.

From this perspective we can say that there are three basic risks: credit risk, market risk and operational risk. But there are other risks, too, for example: settlement risk, residual risk, securitization risk, concentration risk, interest rate risk (general interest rate risk), reputation risk and liquidity risk. When dealing with these type of risks we are using stress tests (to prove significance of assumptions), scenario analysis, economic and regulatory environment and finally capital planning. The Supervisory review process is oriented on these activities. General idea is to compare results with a peer group, which takes into consideration the so called individual capital guidance, processing, application of control and system improvement, but it can also bring business restrictions.

3 Practical liquidity problems

Practical liquidity problems arise from classical loan theory, shiftability theory and anticipated income theory. This generally concerns assets. Liability management is oriented on funding costs. A consistent ALM view tries to match liability to assets; at least from liquidity point of view it is important. When quantifying liquidity it is important to think about alternative interest rates, which could influence our liquidity position. When assessing liquidity situation of firm it is also important to figure out prepayment scenarios, which influence the liquidity. Our goal is to optimize risk/return ratio. We have to be aware, that the asset allocation in context of liquidity management and liability management gives conclusions in profitability.

We need to realize that consumption of liquidity is not only on the liability side but also on the assets side. At the same time we need to recognise that a development of liquidity is closely connected to economic fundaments. Liquidity problems have to be recognised from point of view of both system and specific needs.

But not only the above mentioned arguments are significant when dealing with liquidity in financial institutions. A complex approach has to include even more. There is also need to quantify a volume and timing of cash flow ex-ante, anticipate the cash flow. We also have to think about contingent cash flow. Timing of future cash flow is situation specific. Sometimes the nature of funding is clear and understandable but not always (look at the case study). Sometimes a significant inflow or outflow can occur suddenly.

Common situation is to generate liquidity by shifting claims from one party to another (directly or indirectly). For contingent funding appropriate but in times of systematic crisis it would hardly work.

The starting point to measure liquidity is the decision, which factors are important, e.g. the amount of shiftable assets, amount of self-liquidate loans, volatility of wholesale funds, stickiness of core deposits, marketability of investments (pledged or not saleable). The simplest approach is to construct some liquidity ratio. There are many liquidity ratios e.g. liquid assets/all assets, liquid assets/short term liabilities, loan-to-deposits ration and so on. This is, however, not a appropriate method.

How to measure liquidity better? - certainly, by complex description of liquidity characteristics of assets and liabilities. What are better liquidity criteria? Maybe the measurement has to be *comprehensive*, which means that it includes all assets, liabilities and off-balance items. It has to be also *flexible*, which means that it can react to changes, e.g. new product development. Other characteristics could be *prospective,* and *scenario specific*. Because only a measurement which takes into account future assumptions and generates scenarios could be cold complex.

We have to reconsider every asset and liability from liquidity point of view. We have to reconsider its volatility and also identify the trend (drift) of the item. In banks there is a commonly used model for liquidity modelling. We need, when constructing the model, to find which assets are liquid and illiquid and which liabilities are volatile and stable. The net liquid assets are found if the liquid assets are higher than the volatile liabilities.

A more extensive approach is so called the Basic Surplus Deficit (BSD). BSD is adoption of liquid assets to purchased funds and the ratio of liquid assets to short-term liabilities. Like the liquidity coverage ratio the "bottom line" for BSD is the size of liquidity cushion. An improvement of BSD is thanks to its inclusion of cash flow from potential loan sale. It includes also potential losses. And we take into the consideration also some time horizon e.g. 30 days.

When taking into consideration a more complex future development, we divide time into the time buckets. The width of the buckets is proportional to exactness of measurement. This concept is called the liquidity gap. This "gap approach" was historically extended also to the interest rate risk. There is also possibility, when using this approach, to take into consideration also new businesses. Our intent is to get a positively covered liquidity coverage ratio, which is the ratio of total projected inflows and total projected outflows. When dealing with cash flow projection there is a need to recognize the static vs. dynamic gap, option risk (call risk – asset side, prepayment risk – asset side, withdrawal risk – liability side), market driven cash flow (margin payments call).

In principle, we can measure cash flow uncertainties by 1. combining the measure of mismatch liquidity with measures of contingency/stand-by/prudential risk. 2. scenario based liquidity cash flow projections. There must be at least three scenarios: ordinary course of business, bank-specific crisis scenario, systemic crisis scenario (e.g. bank run). 3. cash flow modelling. That means the Monte Carlo simulation or other numerical techniques to generate future in- and outflows or direct calculation of liquidity at risk.

The liquidity on asset liability side could be observed from the following viewpoints: Prepayments and Curtailments. The prepayment is an excess of required scheduled payments. It will reduce principals and subsequent interest payments. Prepayments usually accelerate when interest decrease that is motivated by possibility of the new funding at a lower rate. Some products (e.g. mortgage bonds) have a prepayment penalty. The curtailments are lower payments than required. They are connected with the credit risk, client's bonity and economical development.

The problem of prepayments is more standardised in the USA. There are two industry conventions for benchmark prepayment rates: Conditional Prepayment Rate (CPR) and Public Security Association (PSA). The PSA takes into consideration monthly series of CPR, and assumes prepayment rate

increases as pool ages. The standard benchmark is 100% PSA (e.g. CPR = 0.2% for the first month, increasing by 0.2% per month up to the 30th month. For conversion of CPR to the single month mortality, we have $SMM = 1-(1-CPR)^{1/12}$.
For simulation of optional cash flows it is important to realize the cash flow causality as follows:

Derivative instruments
 -Cash flow = $f(market\_risk)$ [1]

Pre-payment options
 - Cash flow = $f(refinancing\_rate)$

Increase/decrease of low-cost funds
 - Cash flow = $f(alternative\_rates)$

Credit risk
 - Credit risk = short put position

But the bottom line of all above mentioned considerations is a realization of a strong integration of the market, credit and liquidity risk.

4 Case study – a fuzzy logic approach

Current accounts are common example of a product that has not explicit maturity. Consequently we are also not able to determine residual maturity. In this example our intent is to suggest possible approach to make theoretical division of the liability, to get some positions, from time structure point of view, which will serve as a benchmark for assets allocation. This assets allocation must follow fundamental financial logic, firstly at least possible outflow (liquidity risk) a secondly possible products allocations (duration of products, point risk, credit profile of assets). We can also use riskless yield to construct FTP (inter bank rate) to derive cost of product.

The intention of the case study is to use Mamdani fuzzy regulator for duration determination of liabilities, which could lead us to relevant investment activities. Inputs are, as descried below in detail form, two characteristics of one kind of bank's primary deposit. The fuzzy regulator is able to take into account an expert opinion thanks to table of rules, which serves as inferential tool. The table of rules is describing signification of the state of the world according the two inputs. Consequently, the system allows us to change investment policy when monitored characteristics change their magnitude.

Products, which are in the centre, of our attention are Current Accounts (CA), Time Accounts (TA), Saving Accounts (SA) and Credit Lines (CL). Firstly we have to determine main characteristics, which are influencing liquidity behaviour of the products. In line with [1] we have chosen two main determinants: volatility σ and trend d as inputs and liquidity outflow w as output. The states of the world are denoted by the help of tools and notation of fuzzy set theory as follows:

[1] Cash flow from derivatives are mainly determined by market risk

$$\sigma \hat{=} A\{A_1, A_2, A_3\} = \{L, M, H\} \tag{1}$$

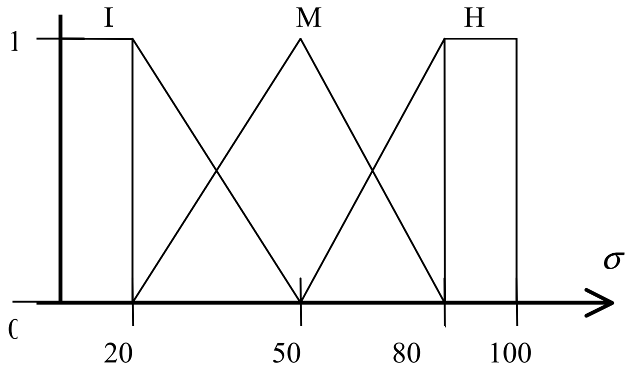


Fig. 1 Input variable volatility σ

The states of the world of σ are determined by historical approach. That means historical time series of CA (5 year history with daily observations), the deviations from 6 month moving averages are transformed to [0,100] interval. Derived value can take linguistic value L (low), M (middle) and H (high) see Fig 1.

$$d \hat{=} B\{B_1, B_2, B_3\} = \{D, S, I\} \tag{2}$$

The second input variable is given by (2). It presents the next significant information about position in the product. It reflects the slope of a trend. The linear trend d is defined on interval $\lfloor -45^0, +45^0 \rfloor$ and deviation from zero (S - middle of the interval of the fuzzy set) is described by membership functions. The states of the world described by this fuzzy set reflects a possible situation development of d, D (decreasing), S (stable), I (increasing) see Fig 2.

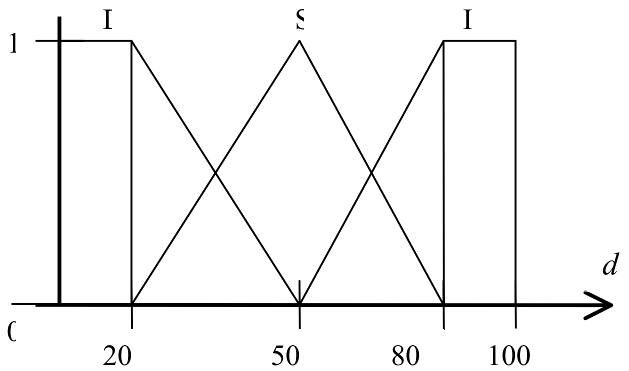


Fig. 2. Input variable trend d

$$w \hat{=} C\{C_1, C_2, C_3\} = \{L, M, H\} \tag{3}$$

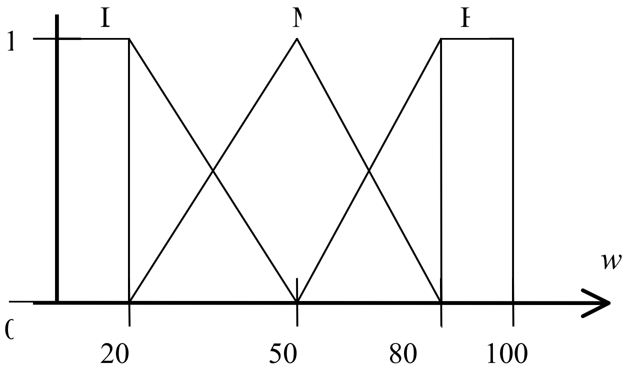


Fig. 3 Output variable liquidity outflow w

The output generated by (1) and (2) is described by (3). The liquidity outflow is understood as approximation of liquidity position in the product derived from chosen characteristic of product. The states of the world are described by L (low), M (middle) and H (high). The interval $[0,100]$ is perceived as a possibility the liquidity position will remain without change in next 10 years.
As all input characteristics (σ, d) and output function (w) are defined on the same interval their membership functions could be described jointly (substitute σ, d, w for v).

$$\mu_L(v) = \begin{cases} 1 & for \quad 0 \le v \le 20 \\ \dfrac{50-v}{30} & for \quad 20 \le v \le 50 \end{cases} \tag{4}$$

$$\mu_M(v) = \begin{cases} \dfrac{v-20}{50} & for \quad 20 \le v \le 50 \\ \dfrac{50-v}{30} & for \quad 50 \le v \le 80 \end{cases} \tag{5}$$

$$\mu_H(v) = \begin{cases} \dfrac{v-50}{30} & for \quad 50 \le v \le 80 \\ 1 & for \quad 20 \le v \le 50 \end{cases} \tag{6}$$

Our next task is to define, how foreseen states of the world (input variables) will influence the result (output variable). This is the way we can employ both expert opinion and non-crisp vision of the world. This can be done by the table of rules

d/σ	L	M	H
D	M	M	H
S	L	L	H
I	L	L	M

Tab. 1 Table of rules

The Tab. 1 contents IF-THEN rules as follows:

rule number 1: IF σ is *L* and *d* is *D* than *w* will stay M
rule number 2: IF σ is *M* and *d* is *D* than *w* will stay M
rule number 3: IF σ is *H* and *d* is *D* than *w* will stay H
….
rule number 9: IF σ is *H* and *d* is *I* than *w* will stay M

From real bank position we have found σ=40 and *d=25.*

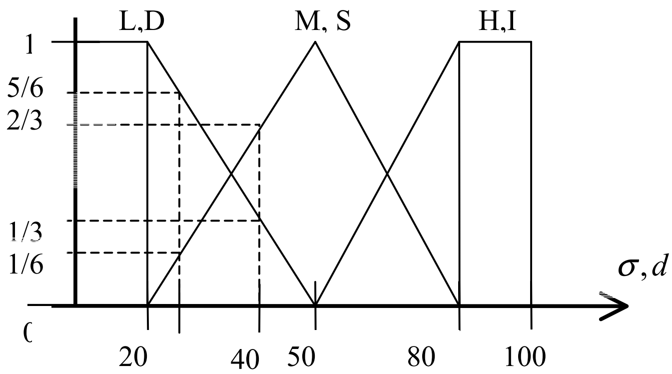


Fig. 4 Inputs variables – membership functions derivation

Control output (CO) of each rule is defined by operation conjunction applied on its strength and conclusion as follows:

$$\begin{aligned}
&\text{CO of rule 1: } \alpha_{ij} \wedge \mu_{Cij}(z) = \min(\alpha_{ij}, \mu_{Cij}(z)) \\
&\text{CO of rule 2: } \alpha_{ij+1} \wedge \mu_{Cij+1}(z) = \min(\alpha_{ij+1}, \mu_{Cij+1}(z)) \\
&\text{CO of rule 3: } \alpha_{i+1j} \wedge \mu_{Ci+1j}(z) = \min(\alpha_{i+1j}, \mu_{Ci+1j}(z)) \\
&\text{CO of rule 4: } \alpha_{i+1j+1} \wedge \mu_{Ci+1j+1}(z) = \min(\alpha_{+1ij+1}, \mu_{Ci+1j+1}(z))
\end{aligned} \tag{7}$$

According our number from the two characteristic we can substitute (7):

$$
\begin{aligned}
\alpha_{11} &= \eta_L(40) \wedge \eta_D(25) = \min(1/3, 5/6) = 1/3 \\
\alpha_{12} &= \eta_M(40) \wedge \eta_D(25) = \min(2/3, 5/6) = 2/3 \\
\alpha_{21} &= \eta_L(40) \wedge \eta_S(25) = \min(1/3, 1/6) = 1/6 \\
\alpha_{22} &= \eta_M(40) \wedge \eta_S(25) = \min(1/3, 2/3) = 1/1
\end{aligned}
\tag{8}
$$

Following table show as how we get from relationships of input to relation sheep of output

d/σ	$\mu_L(40)\ 1/3$	$\mu_M(40)\ 2/3$
$\mu_D(25)\ 5/6$	$\mu_M(z)$	$\mu_M(z)$
$\mu_S(25)\ 5/6$	$\mu_L(z)$	$\mu_L(z)$

Tab. 2 Table of membership functions

Next thing is to aggregate sub-products to get one fuzzy logic controller (Mamdani´s fuzzy controller). It is defined as follows:

$$\mu_{agg}(z) = (\alpha_{ij} \wedge \mu_{Cij}(z)) \vee (\alpha_{i+1,j} \wedge \mu_{Ci+1,j}(z)) \vee (\alpha_{i,j+1} \wedge \mu_{Ci,j+1}(z)) \vee (\alpha_{i+1,j+1} \wedge \mu_{Ci+1,j+1}(z))$$

$$\mu_{agg}(z) = \max\{(\alpha_{ij} \wedge \mu_{Cij}(z)) \vee (\alpha_{i+1,j} \wedge \mu_{Ci+1,j}(z)) \vee (\alpha_{i,j+1} \wedge \mu_{Ci,j+1}(z)) \vee (\alpha_{i+1,j+1} \wedge \mu_{Ci+1,j+1}(z))\}$$

In our example:

$$\mu_{agg}(z) = \max\{\min(1/3 \wedge \mu_M(z)) \vee (2/3 \wedge \mu_M(z)) \vee (1/6 \wedge \mu_L(z)) \vee (1/6 \wedge \mu_L(z))\}$$

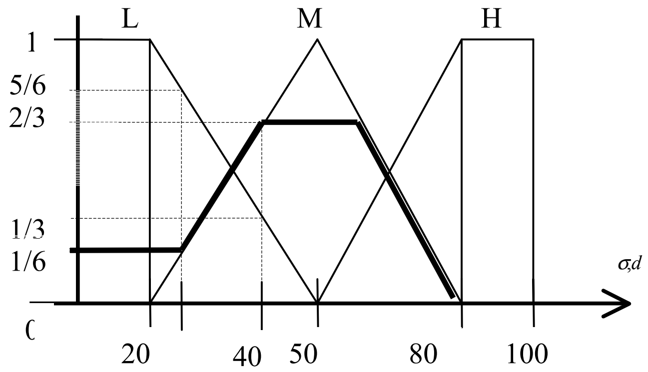


Fig. 5 Inputs variables – membership functions derivation

Fuzzy space showed by Fig. 5 could be understood as liability division, which is reflecting both development of the product d and stability of the product σ. The definition of fuzzy output was made in a way, to get output, which could be utilized for constructing a theoretical product division. Because the so called product is in fact a financial source for the firm, applying fuzzy logic tools allows us to allocate this means in a logical way.

Fig. 6 shows the division of sources. The final fuzzy set determines the division of sources up to 8 years, while the average duration of the theoretical sources is approximately 5 years. The centre of gravity is close to 4.5 year.

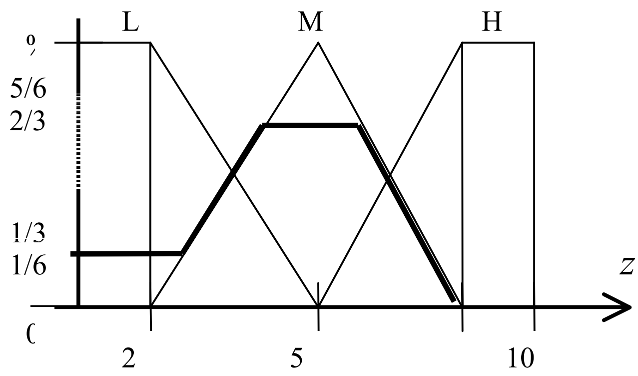


Fig. 6 FLC output – Theoretical division of sources

Following the determined structure of sources we can adjust the present structure of assets to be in line with foreseen outflow. We have to be aware of the fact that in reality financial statements follow accountant implications. There is a possibility to construct financial positions from cashflow reports for the best approximation of the real cash. For liquidity purpose it is the best approach.

In fact, the reality is much more complex and the liquidity viewpoint is also influenced by a need of liquidity buffer (many firms cannot imagine to exist without Buy-Sell or Repo operations, which serve for utilization of short term liquidity surplus. But there are also credits implications, which influence the inflow.

5 Conclusion

A liquidity framework, which was given in part 1, is defined by 14 rules. Generally we can say, that bank should have its strategy for steering liquidity, that senior management should take decisions in liquidity place, that bank should have proper system for managing liquidity, the system should be able to deal with every day liquidity. Banks should imply a variety of IF –THEN scenarios. Banks have to take some assumptions when dealing with liquidity, the assumptions have to be reconsidered frequently. Bank liquidity system should be oriented not only to structural liquidity of institution but also on market liquidity aspects. The banking liquidity system should also deal with contingency planning. There should exist an explicit system for dealing with FX risk connected with liquidity. The communication of a bank is also important in a way that the bank disclosure approach should be standardised. All mentioned requirements should be revised by the supervisor. As foreseen development of Basel suggestions could take conclusion in liquidity risk requirement, so it is good to implement, when dealing with liquidity, all practises which are commonly used in risk management of market risk, to steer liquidity in a qualified way. As practical problems when dealing with liquidity are connected with projection of the cash flow rather than to solve ex-ante problems, the fuzzy logic solution is proposed for no maturity products. A development of a product and its volatility, the characteristics extracted from past, could through expert opinions matrix give the liquidity structure of the product. This approach is simple and follows the common sense, which could be also considered as an advantage.

References

[1] Bojadziev,G a Bojadziev, M: Fuzzy Logic for Business, Finance and Management – World Scientific Publishing, 1997, ISBN 9810228945

[2] Sound Practices for Managing Liquidity in Banking Organisations, February 2000, on www.bis.org/publ/bcbs69.pdf

[3] Stephen Morris and Hyun Song Shin: Liquidity Black Holes, Cowles Foundation discussion paper no. 1464

[4] Michael J. Fleming: Measuring Treasury Market Liquidity, FRBNY Economic Policy Review, 2003

[5] Federal Reserve Bank of San Francisco: What determines the Credit Spread?, Economic research, 2004

DEA models with random inputs and outputs*

Josef Jablonský

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Praha 3
jablon@vse.cz, http://nb.vse.cz/~jablon

Abstract

Data Envelopment Analysis (DEA) models are general tool for evaluation the relative efficiency of production units described by multiple inputs and outputs. The standard DEA models are based on deterministic inputs and outputs. The paper formulates basic DEA models supposing that inputs and outputs are random variables with given continuous probabilistic distributions. Under this assumption the efficiency scores for evaluated production units are random variables. The aim of the paper is to try to describe the properties of the random efficiency scores by means of Monte Carlo experiments. The experiments are realised within the *MS Excel* environment together with *LINGO* optimisation solver and *Crystal Ball* add-in application for Monte Carlo analysis in spreadsheets. The computations were realised on the real set of almost 200 bank branches of one of the Czech commercial banks. The paper shows that the simulation experiments can offer more information that standard linear programming approaches, e.g. by this way it is possible to define new concepts of super-efficiency in DEA. Their disadvantage is the time consumption for calculations connected with simulation compared to standard optimisation approaches. Nevertheless, the time consumption is compensated by possibilities to consider different probabilistic distributions for inputs and outputs and available detailed probabilistic characteristics of efficiency scores. The experiments show that this approach can be an appropriate instrument for analysing DEA problems with imprecise data and can produce interesting results in comparison with other approaches.

1 Introduction

Data envelopment analysis (DEA) is a tool for measuring the relative efficiency and comparison of decision making units (DMU). The DMUs are usually characterised by several inputs that are spent for production of several outputs. Let us consider set E of *n* decision making units E = {DMU_1, DMU_2, …, DMU_n}. Each of the units produces *r* outputs and for their production spent *m* inputs. Let us denote $X_j = \{x_{ij}, i=1,2,\dots,m\}$ the inputs and $Y_j = \{y_{ij}, i=1,2,\dots,r\}$ the outputs for the DMU_j. Then X is the (m,n) matrix of inputs and Y (r,n) matrix of outputs.

The basic principle of the DEA in evaluation of efficiency of the DMU_q, $q\in\{1,2.,,,.n\}$ consists in looking for a virtual unit with inputs and outputs defined as the weighted sum of inputs and outputs of the other units in the decision set - $X\lambda$ a $Y\lambda$, where $\lambda=(\lambda_1, \lambda_2,\dots, \lambda_n)$, $\lambda>0$ is the vector of weights of the DMUs. The virtual unit should be better (or at least not worse) than the analysed unit DMU_q. The problem of looking for a virtual unit can generally be formulated as standard linear programming problem:

* The research is supported by the Grant Agency of Czech Republic – grant no. 402/06/0150.

$$\begin{aligned}
\text{minimise} \quad & z = \theta, \\
\text{s.t.} \quad & Y\lambda \geq Y_q, \qquad\qquad (1) \\
& X\lambda \leq \theta X_q, \\
& \lambda \geq 0.
\end{aligned}$$

The DMU_q is to be considered as efficient if the virtual unit is identical with analysed unit (does not exist the virtual unit with better inputs and outputs). In this case $Y\lambda = Y_q$, $X\lambda = X_q$ and minimum value of $z = \theta = 1$. Otherwise the DMU_q is not efficient and minimum value of $\theta<1$ can be interpreted as the need of proportional reduction of inputs in order to reach the efficient frontier. The presented model is so called input oriented model because its objective is to find reduction of inputs in order to reach the efficiency. Analogously can be formulated output oriented model.

Model (1) presents just the basic philosophy of the DEA models. The input oriented form of the Charnes, Cooper, Rhodes model (CCR-I) is formulated as follows:

$$\begin{aligned}
\text{minimise} \quad & z = \theta - \varepsilon(e^T s^+ + e^T s^-), \\
\text{s.t.} \quad & Y\lambda - s^+ = Y_q, \qquad\qquad (2) \\
& X\lambda + s^- = \theta X_q, \\
& \lambda, s^+, s^- \geq 0,
\end{aligned}$$

where $e^T = (1,1,\ldots,1)$ and ε is a very small constant (usually 10^{-6} or 10^{-8}). Presented formulations (1) and (2) are very close each other. The variables s^+, s^- are just surplus and slack variables expressing the difference between virtual inputs/outputs and appropriate inputs/outputs of the DMU_q. Obviously, the virtual inputs/outputs can be computed with optimal values of variables of the model (2) as follows:

$$X'_q = X_q\theta^* - s^-,$$
$$Y'_q = Y_q + s^+.$$

The modification of the CCR-I model is the BCC-I model. The constraints of the model (2) are extended by convexity condition $e^T\lambda = 1$ in this model.

For evaluation of efficiency of all DMUs of the decision set it is necessary to solve *n* linear programming optimisation problems (2) or its modifications. Each of the LP program contains (*n*+*m*+*r*+1) variables and (*m*+*r*) constraints. These LP programs are relatively small but their repeated solution for all units can be too time consuming. It is more useful to formulate one bigger LP program that can find out the efficiency of all units by one optimisation run. This model can be written for the CCR-I model as follows:

$$\begin{aligned}
\text{minimise} \quad & \sum_{q=1}^{n}\left(\theta_q - \varepsilon\left(\sum_{i=1}^{r} s_{iq}^+ + \sum_{i=1}^{m} s_{iq}^-\right)\right) \\
\text{s.t.} \quad & \sum_{j=1}^{n} y_{ij}\lambda_{jq} - s_{iq}^+ = y_{iq}, \quad i = 1,2,\ldots,r,\ q = 1,2,\ldots,n, \qquad (3) \\
& \sum_{j=1}^{n} x_{ij}\lambda_{jq} - s_{iq}^- = \theta_q x_{iq}, \quad i = 1,2,\ldots,m,\ q = 1,2,\ldots,n, \\
& \lambda_{jq} \geq 0, s_{iq}^+ \geq 0, s_{iq}^- \geq 0, \theta_q \geq 0.
\end{aligned}$$

The formulation for the BCC-I model is extended by the following constraints:

$$\sum_{j=1}^{n} \lambda_{jq} = 1, \qquad q = 1,2,\ldots,n.$$

The model (3) contains $n(n+m+r+1)$ variables and $n(m+r)$ constraints. The appropriate LP program for the DEA model with 69 DMUs and 3 inputs and 3 outputs (computational experiments presented in the last section of the paper are realised by means of the CCR-I model on the problem of this size) contains 5244 variables and 414 constraints. The problem of this size can be simply solved in several seconds.

2 Data envelopment analysis with random inputs and outputs

Inputs and outputs usually reflect past values of the DMU's characteristics. That is why the standard DEA models suppose that the inputs and outputs of evaluated units are given by fixed values. For evaluation and estimation of the future efficiency of the DMUs it is useful to consider imprecise values of the inputs and outputs. These can be given as interval values or more generally as random variables with defined probabilistic distribution. The elements of the matrix of inputs $X = (x_{ij})$ can be taken as random variables with probabilistic density function $f(x_{ij})$ and cumulative distribution function $F(x_{ij})$. Similarly, the elements of the matrix of outputs $Y = (y_{ij})$ are random variables with probabilistic density function $f(y_{ij})$ and cumulative distribution function $F(y_{ij})$. The following probabilistic distributions will be taken into account for our experiments:

- general distribution with probabilistic density function defined by piecewise linear function with parameters $p_{ij} \le c_{ij} \le d_{ij} \le q_{ij}$ (fig.1) – special cases of this distribution are uniform or triangular distribution,
- normal distribution with mean μ_{ij} and standard deviation σ_{ij}.

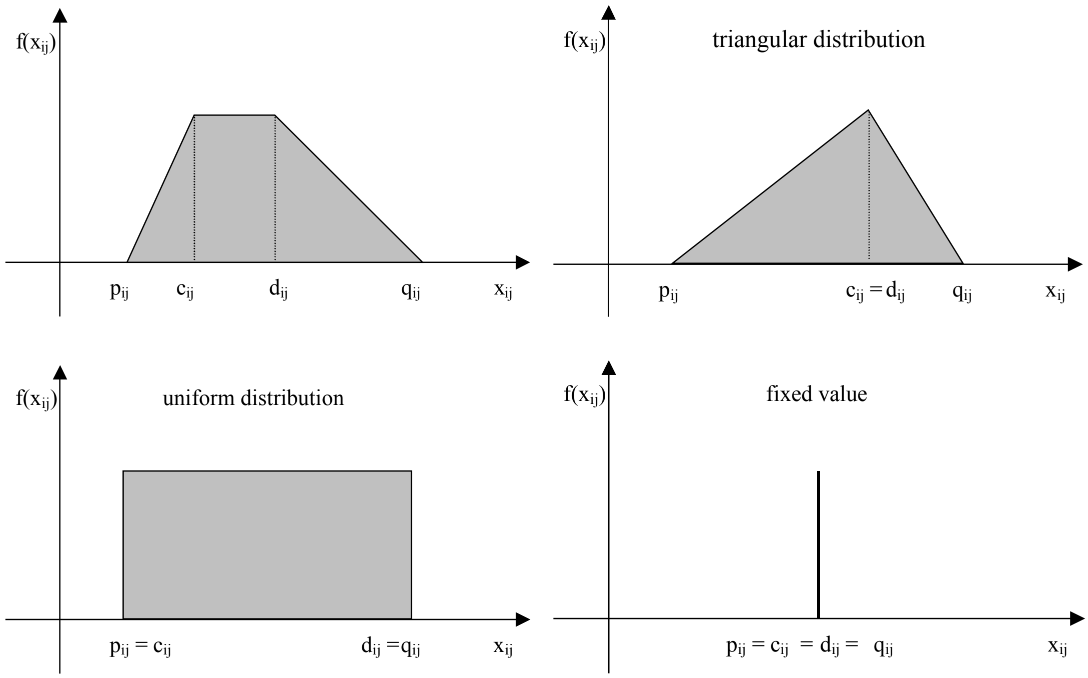


Fig. 1: Probabilistic density function of general distribution.

There are several approaches for dealing with imprecise data in DEA. The paper [3] supposes uniform distribution for inputs and outputs and makes it possible to split the set of DMUs into three subsets $E = E^{++} \cup E^{+} \cup E^{-}$:

1. DMUs always efficient (subset E^{++}) – this class contains units that are efficient even their inputs and outputs are set into worst values and the inputs and outputs of other units are on the best border.
2. DMUs never efficient (subset E^{-}) – units not efficient by considering their best input and output values and worst values for rest of the units.
3. DMUs efficient just by suitable setting up of inputs and outputs within the given interval for all the DMUs (subset E^{+}).

The approach presented in paper [3] is based on solving several LP problems. The results can be obtained very quickly but the information given by this approach are limited. Another possibility for analysis of DEA models with random inputs and outputs is simulation approach. It is more time consuming but offers more information about the distribution of efficiency scores of DMUs.

Detailed information about the probabilistic distribution of efficiency scores for DEA models with inputs and outputs defined as random variables can be probably received only by realisation of random trials. Each of the trials contains several steps:

1. Generation of all the random variables of the model. This step was realised within MS Excel environment by means of built-in functions or VBA procedures in our experiments. The random variables are described by above mentioned probabilistic distributions.
2. The LP problem (3) is solved with input values generated in the previous step. Solving of problem (3) supposes high quality optimisation solver. Our application co-operates with LINGO modelling system. This system can read input data prepared in spreadsheets, perform optimisation by means of powerful solver and return optimal results back to the spreadsheet.
3. Information from the random trials are processed and evaluated by means of MS Excel add-in application called Crystal Ball. This application collects information from the random trials and offers comfort possibilities for their presentation in both the numerical and graphical form (frequency chart on fig. 1). As the final result it is possible to receive basic statistics of the efficiency scores – lower and upper bounds, mean value, standard deviation, etc. The efficiency scores are random variables with general distribution with given mean value and standard deviation.

Based on the simulation steps the DMUs can be divided, analogously to the mentioned LP approach, to the subsets E^{++}, E^{+} and E^{-}. The elements of the first subset cannot be probably further ordered. For the DMUs belonging to the subset E^{+} the probability of their efficiency can be derived as the ratio of the number of trials with the efficiency of the appropriate DMU to the number of all trials. The elements of the last subset can be ordered e.g. by mean value of their efficiency score.

3 Evaluation of bank branches – simulation experiments

One of the most often described applications of DEA models is evaluation of efficiency of bank branches. In our experiments we had the possibility to work with the real data set of 81 bank branches of a Czech bank. According to the size of the branches the data set was divided into two subsets with 69 small and 12 large branches. Our experiments were performed on the set of small branches. Based on the discussion with the bank management we selected for the analysis three inputs and three outputs. The available past exact data were taken from financial sheets of the bank of September 2000. The following inputs and outputs were taken into account (minimal, mean and maximal value of the characteristics are given in parenthesis):

Inputs:
- average number of employees of the branch (min 4, mean 10.2, max 28),
- operational costs in thousand of CZK (5534, 10997, 36261),
- floor space in square meters (40, 231, 649).

Outputs:
- total number of accounts (276, 2943, 16632),
- number of transactions per month (42840, 202421, 724338),
- value of savings in millions of CZK (66, 1533, 8573).

We suppose that the past exact data will go on as the random variables in the future. In order to compare the results of simulation experiments with the LP based results of paper [3] we use for the random variables:

- uniform distribution over the interval $<p_{ij}, q_{ij}>$, where p_{ij}, q_{ij} values can be received by percentage deviations from exact values specified for input and output characteristics (Table 1 – x denotes past exact values of inputs/outputs),
- normal distribution with mean value μ_{ij} and standard deviation σ_{ij} derived from past exact values (Table 1).

	Uniform		Normal	
	p_{ij}	q_{ij}	μ_{ij}	σ_{ij}
Inputs				
number of employees	0.95x	1.05x	1.00x	0.05x
operational costs	0.975x	1.075x	1.025x	0.03x
floor space	1.00x	1.00x	1.00x	0
Outputs				
# of accounts	0.975x	1.10x	1.0375x	0.04x
# of transactions	1.00x	1.15x	1.075x	0.05x
value of savings	0.95x	1.05x	1.00x	0.03x

Table 1: Definition of random variables for simulation experiments.

Each simulation step consists in solving an optimisation problem (3) with several thousands of variables and hundreds of constraints. The total time of simulation run takes, depending on the number of trials, relatively long time, even the finding the optimal solution of problem (3) takes just few second. That is why we use as maximal number of trials 1000 in our experiments.

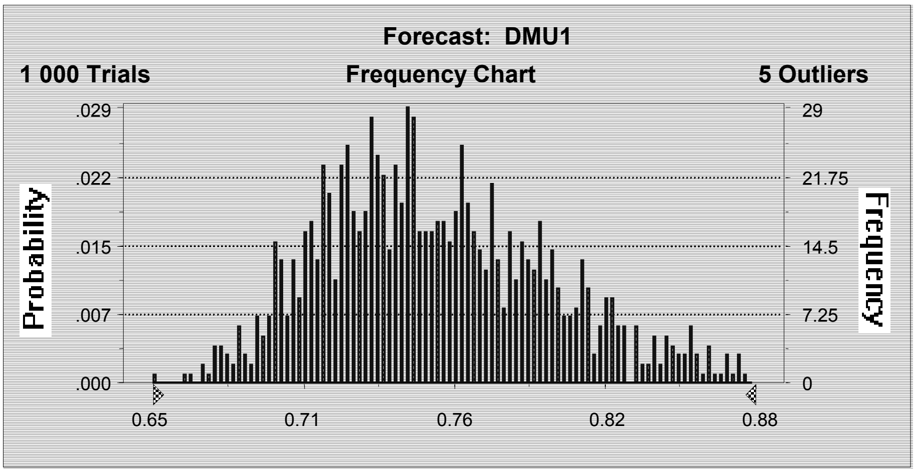


Fig. 1: Frequency chart.

As mentioned above, experiments are realised within MS Excel spreadsheet with co-operation with modelling and optimisation system LINGO and system Crystal Ball for support of simulation experiments in spreadsheets. Based on the results of simulation the DMUs can be divided, similarly to the LP approach [3], into three subsets: E^{++} (always efficient), E^{-} (never efficient) and E^{+} (efficient in case of suitable setting of inputs and outputs within given ranges). The units of subset E^{++} cannot be further ordered and the units of subset E^{-} can be compared by mean values of their efficiency scores. The most interesting is the analysis of the units of subset E^{+}. The LP approach [3] compares these units by values of endurance indices computed by especially formulated LP program. In the simulation approach it is possible to estimate the probability that the unit is efficient as the ratio of successful trials (unit is efficient) and all the trials (1000 in our experiments). This probability can be very good base for the ordering of units of this subset.

The comparison of results of both the LP approach [3] and simulation approach are quite interesting and will be presented in the table below. In simulation experiments we used probabilistic distributions as mentioned in Table 1. The results given by both the distributions are very close each other. It is not necessary to discuss units of the subset E^{++}. All the approaches must give same results in this subset. In our example this subset contains 6 elements: units #7, #36, #40, #56, #68 and #76. The results for the subset E^{+} are presented in Table 2. First two columns of this table contain endurance indices given by LP approach and order of the DMUs according to these indices (the branches are ordered by the results of LP approach). Next four columns are information about results of simulation experiments by

	LP approach		Uniform distribution				Normdist
	Endurance	Order	Mean	st.dev.	$P(\theta = 1)$	Order	$P(\theta = 1)$
Branch #18	0.818	1	0.926	0.077	0.365	9	0.307
Branch #03	0.818	2	0.966	0.042	0.458	8	0.442
Branch #39	0.817	3	0.999	0.003	0.969	2	0.938
Branch #64	0.812	4	0.860	0.057	0.011	14	0.019
Branch #21	0.809	5	0.979	0.029	0.507	6	0.511
Branch #77	0.805	6	0.997	0.009	0.799	4	0.809
Branch #49	0.803	7	0.952	0.038	0.173	10	0.173
Branch #74	0.795	8	0.960	0.030	0.137	11	0.142

Branch #62	0.793	9	0.999	0.003	0.986	1	0.957
Branch #47	0.791	10	0.932	0.037	0.058	12	0.056
Branch #09	0.790	11	0.824	0.049	xxx	18	0.012
Branch #13	0.787	12	0.914	0.046	0.047	13	0.052
Branch #38	0.782	13	0.788	0.041	xxx	20	xxx
Branch #32	0.750	14	0.803	0.026	xxx	19	xxx
Branch #52	0.652	15	0.992	0.019	0.757	5	0.774
Branch #65	0.644	16	0.998	0.006	0.879	3	0.809
Branch #54	0.642	17	0.851	0.042	xxx	17	0.004
Branch #20	0.641	18	0.870	0.043	xxx	16	0.017
Branch #78	0.637	19	0.974	0.037	0.493	7	0.513
Branch #71	0.631	20	0.890	0.047	0.005	15	0.031
Branch #79	0.626	21	0.723	0.052	xxx	22	xxx
Branch #1	0.623	22	0.756	0.043	xxx	21	xxx

Table 2: Simulation results – subset E^+.

using of uniform distribution – mean value and standard deviation of efficiency scores, probability of efficiency and ordering of DMUs. It is possible to remember that just 15 of 22 units indicated as possibly efficient by LP approach are indicated in the same way in the simulation approach and moreover, units #64, #13 and #47, #71 are efficient with very low probability (less than 0.1). The similar results are given by using of normal distribution (last column of Table 2) – just 11 units are indicated as possibly efficient with probability greater than 0.1. The results for the subset E^- are more or less the same in all the approaches. The advantage of the simulation approach is the conclusion given that the DMUs of this subset are random variables with normal distribution and mean and standard deviation received by simulation trials.

4 Conclusions

The paper shows that the simulation techniques can offer more information that standard LP approach. Their disadvantage is the time consumption for calculations connected with simulation compared with standard optimisation approaches. Nevertheless, the time consumption is compensated by possibilities to consider different probabilistic distributions for inputs and outputs, available detailed probabilistic characteristics of efficiency scores, etc. The experiments show that this approach can be an appropriate instrument for analysing DEA problems with imprecise data and can produce interesting results in comparison with other approaches. Advantages of this approach are modesty, flexibility and visualisation. The results by comparison of bank branches of a Czech commercial bank were helpful for bank management.

References

1. Charnes,A., Cooper,W.W., Lewin,A., Seiford,L.: Data Envelopment Analysis: Theory, Methodology and Applications. Kluwer Publ., Boston 1994.
2. Crystal Ball – User Manual. Decisioneering Inc., Denver 1999.

3. Fiala,P., Jablonsky,J., Despotis,D., Smirlis,Y.: DEA with Interval Data: An Application to the Evaluation of Branches of a Czech Bank. Submitted to the proceedings book of a special IFORS conference "New Trends in Bank Management", Athens 2001.
4. Green,R.H.: DIY DEA: Implementing Data Envelopment Analysis in the Mathematical Programming Language AMPL. Omega 24 (1996), 4, pp.489-495.
5. Jablonský,J., Dlouhý,M.: Modely hodnocení efektivnosti produkčních jednotek (Models for evaluation of efficiency of production units) Professional Publishing, Praha 2004.
6. Jablonský,J.: Modely hodnocení efektivnosti produkčních jednotek (Models for evaluation of efficiency of production units). Politická ekonomie, 2 (2004), pp.206-220.
7. Schrage,L.: Optimization modelling with *LINGO*. Lindo Systems Inc., Chicago 1999.

IZAR – multicriteria decision support system

Jana Kalčevová, Petr Fiala

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Prague 3

kalcevov@vse.cz, pfiala@vse.cz

Abstract

Many of real decision making problems are evaluated by multiple criteria. For applying of appropriate multicriteria methods it is necessary a software support. The paper presents a universal multicriteria decision support system IZAR. The main component of the system IZAR is an expert system helping the user with choosing the method most suitable for available information about the problem. IZAR not only suggests the most suitable method but also applies it immediately to the problem solving. By means of properly specified questions the expert system selects the right procedure for solving the problem, when it takes into account its peculiarities regarding entries and additional information, which user can assign to the system. An appropriate classification of multicriteria models and methods is needed because of universality of the system. The system solves discrete and continuous problems. Basic models for multiattribute evaluation and multi-objective optimization problems are included. Methods for multiattribute evaluation problems are classified by types of preference information and by calculation principles. Preference information is given as aspiration levels, ordinal information, and cardinal information. Basic calculation principles are utility maximization, minimization of a distance from the ideal alternative, and evaluation by preference relation. Methods for multiobjective optimization problems are classified by means of setting user information, as a priori information, a posteriori information, and progressive information. This project is supported by FRVŠ – grant no. 2949/2006. The system will be available on web pages for all interested users and it will be also distributed on CD for users without internet connection.

Keywords

Multiple criteria, Multiattribute evaluation, Multiobjective optimization, Models, Methods, Decision support system

1 Introduction

IZAR is non-commercial software set for students of decision theory. The final version of IZAR should be the universal system for solving of discrete and continuous single objective and also multiobjective optimization problems. The first part of this software is common background with an expert system of IZAR. This part is implemented and at present the authors are focused on the second part of IZAR, continuous problems. The idea is focused on linear programming models but in the future we probably expand IZAR about some nonlinear methods. The third part – discrete problems – should be implemented in the end of the year 2006.[1]

Whole IZAR system is implemented in Smalltalk/X, a dialect Smalltalk-80. Smalltalk/X virtual machine and runtime is available for both MS Windows and Linux and also for FreeBSD.

IZAR system is divided into three main parts:

- a core mathematical library consisting of a set of basic mathematical types and operations (matrixes and matrix operations like matrix multiplication for example),
- user interface part for communication with user in a convenient way,

[1] These three parts are supported by FRVŠ grant no. 2949/2006

- set of implemented methods.

All methods are implemented as external independent program units those can be dynamically loaded or unload to or from running system. Methods are implemented in slightly modified version of Pascal [4] which is interpreted by specialized build-in interpreter. Pascal is also used as a scripting language, so the user can work with IZAR system non-interactively.

This design gives a possibility to study and extend the IZAR system, especially about the implemented methods, by the user – this is one of the most important features of the system.

2 Single objective continuous problems

The simplest problem that can be solved by IZAR is continuous problem with the set of feasible solutions given by linear constraints in the form of inequations and with only one objective function. The problem can be written as:

$$\begin{array}{ll} z = f(x) \rightarrow \min \textit{ or } \max & \\ \text{subject to} & \\ g_j(x) \le b_j & \text{for } j = 1,\dots,p \\ g_k(x) \ge b_k & \text{for } k = p+1,\dots,q \\ g_l(x) = b_l & \text{for } l = q+1,\dots,m \\ x_i \ge 0 & \text{for } i = 1,\dots,n \end{array}$$

where *f(x)* and *g(x)* are linear functions, *n* is a number of variables and *m* is a number of constraints. Constraints can be written in the equation or inequation form. All constraints with non-negativity conditions construct the feasible set.

The single objective problems can be solved by simplex method. This method is implemented in IZAR and is used for search optimum of minimum or maximum objective functions on the feasible set. For constraints of type "less or equal than" (with non-negative right hand sides) is used one-phase simplex method, for feasible set with at least one constraint of type "equal" or "greater or equal than" two-phase simplex method is used. Algorithms for both methods are described in [3] and they are implemented in the modified simplex method form. For now, a multiplicative version of simplex algorithm is used, but it is subject to change in order to achieve better performance, since simplex algorithm is basic building block for most implemented methods.

IZAR solves the problem by this method and in the result window represents results and simplex tables also. All non-integer coefficients are presented in a fraction form.

3 Multiobjective continuous problems

Multiobjective continuous problems differ from single objective ones in a number of objective functions (see [1]). These functions can be minimal or maximal and the model can be written in the form:

$$\begin{array}{ll} z_r = f_r(x) \rightarrow \min \text{ for } r = 1,\dots,t & \\ z_s = f_s(x) \rightarrow \max \text{ for } s = t+1,\dots,v & \\ \text{subject to} & \\ g_j(x) \le b_j & \text{for } j = 1,\dots,p \\ g_k(x) \ge b_k & \text{for } k = p+1,\dots,q \\ g_l(x) = b_l & \text{for } l = q+1,\dots,m \\ x_i \ge 0 & \text{for } i = 1,\dots,n \end{array}$$

where *f(x)* and *g(x)* are linear functions, *n* is a number of variables, *m* is a number of constraints and *v* is a number of objective functions. Constraints can be written in the equation or inequation form as well as in previous case.

3.1 Data sources

IZAR should be a very friendly software consequently we focused on easy data input. The data can be loaded from data file which is in CPLEX format (for single criteria problems) and in augmented CPLEX format for multicriteria problems. The second way how to enter data is a manual work. The window for data dimension has four parts. The first one is about name of problem that is default named New problem. Figure 1 displays problem of furniture factory that products tables and chairs and for production are needed time, wood and chrome. The factory maximizes profit and minimizes total costs at the same time. The second part of this window is about variables. User can change number of variables and their names. Default names of variables are x1, x2, etc. The third part makes possible to type in a number of constraints and their names (default named Constraint 1, Constraint 2, etc.). The last part is given for objectives functions. The user can change their number and names of course. Note that the number of variables, constraints and objective functions are unbounded. They are limited only by computational power of CPU an amount of available.

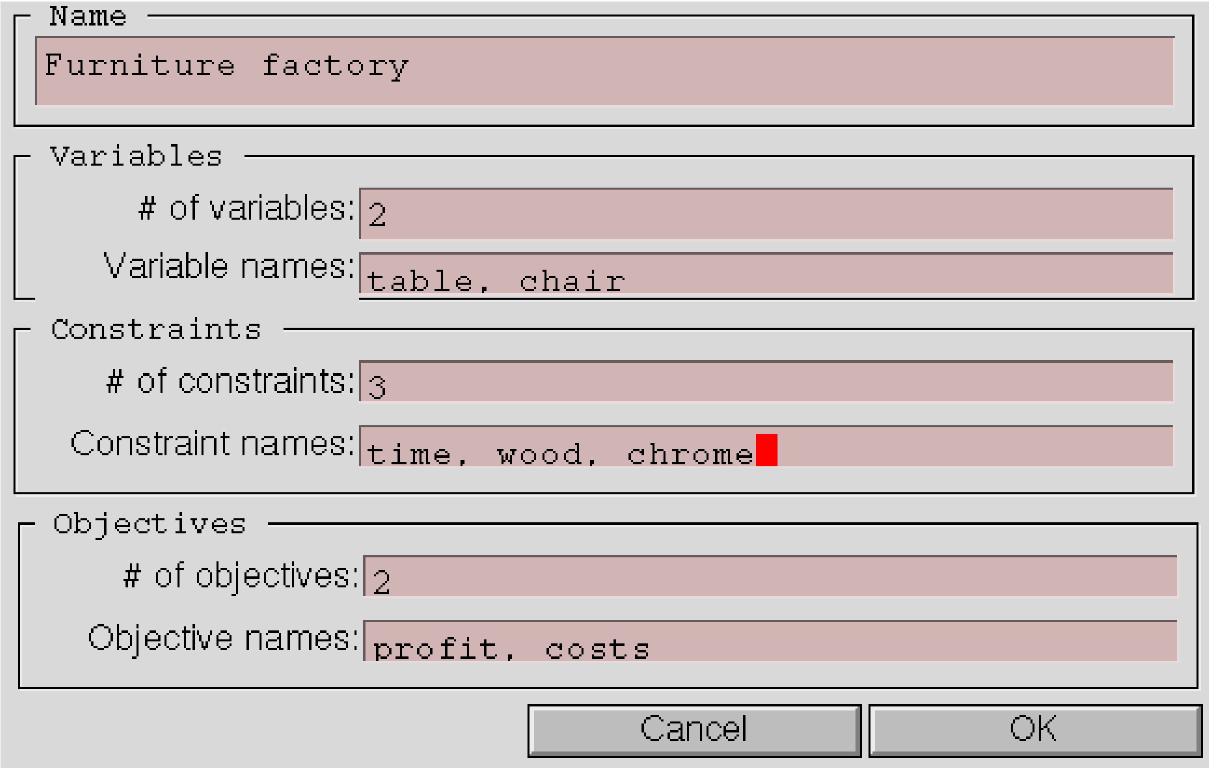


Figure 1: Problem dimension

Next Window (Figure 2) shows part for data submission. The upper part of window is for objective functions. User should enter prices and types of objective functions extremes. The following part of window is intended for structural coefficients and right hand sides of constraints. And here the user can change the sign of inequality of course. Below is a place for method selection. The methods are line in the menu together with possibility to pick up an expert system for automatic selection of method. After this method selection IZAR finds compromise solution and displays the results.

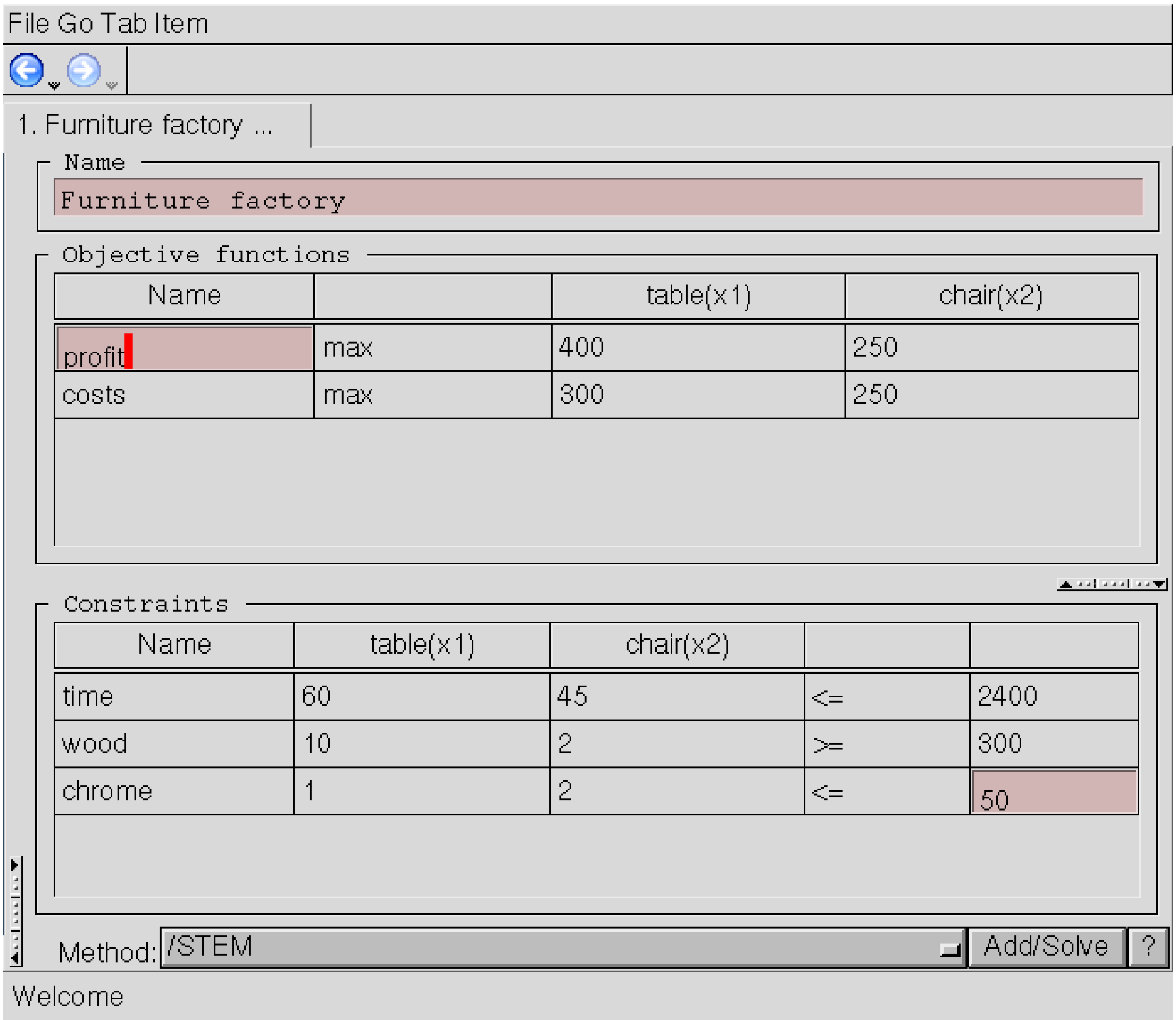


Figure 2: Functions and method selection

3.2 Expert system of IZAR

An expert system can be very useful. The expert system for discrete problems was proposed in [2]. We propose a similar expert system for multiobjective optimization problems. The main menu for method selection offers eleven methods for these problems:

1. maximal utility method,
2. minimal distance from the ideal solution method,
3. lexicographic method,
4. goal programming,
5. maximal probability method,
6. minimal component method,
7. multicriteria simplex method.
8. GDF method,
9. Zionts-Wallenius method,
10. STEM,
11. Steuer method,

The main component of the IZAR system is an expert system helping the user with choosing the method most suitable for available information about the problem. By means of properly specified questions the expert system selects the right procedure for solving the problem, when it takes into account its peculiarities regarding entries and additional information, which user can assign to the system. The choosing of the suitable method is based on following classifications and questions.

Methods are classified by means of setting user information:

- a priori information (methods 1-6),
- a posteriori information (method 7),
- a progressive information (interactive methods 8-11).

The user chooses the way of specification of the importance of each criterion:

- weights,
- order of the criteria,
- goal (ideal) values.

Weights can be assessed:

- directly,
- by ordinal ordering of the criteria,
- by cardinal evaluation of criteria.

The information which can affect the choice of the method is a calculation principle:

- maximization of the utility,
- minimization of the distance from goal (ideal) values,

The form of substitution information:

- explicitly expressed value of substitution (rates of substitution),
- implicitly expressed value of substitution.

IZAR not only suggests the most suitable method but also applies it immediately to the problem solving.

3.3 Compromise solution

Methods for multiobjective optimization problems are classified by means of setting user information. A priori information is given before starting of problem solving, a posteriori information is given after computation, and progressive information is given during calculation process.

After choosing of asked method the calculation process is started and there is required additional information. For maximal utility method and minimal component method weights are needed, goal values have to be given for goal programming, minimal distance from the ideal solution method works with both (weights and ideal values) but ideal value is calculated automatically. Order of the criteria is required for lexicographic method and maximal probability method needs no information as well as multicriteria simplex method. Progressive information is given for all four interactive methods – GDF method, Zionts-Wallenius method, STEM and Steuer method.

Each method provides a compromise solution and the user has a possibility to solve the same problem by the other method or by the same method with different additional information. The user also can change the model or exit IZAR. The results of all methods are saved in table for easy comparison.

4 Conclusion

The IZAR system is in a development stage. The continuous problems are implemented and in the following semester this part of IZAR system will be tested at the Economic University in Prague, Department of Econometrics, in the course of Decision Theory. The discrete problems are prepared for implementation.

The IZAR system has several main advantages. The first is the design of the system that gives a possibility to study and extend the IZAR system by the user. The user with elementary knowledge of

Pascal language can read the program code for each method and study what the method does and how it works. The user experienced in Pascal programming can implemented his own methods or improve the existing parts of the IZAR system.

The second advantage is unlimited number precision. Most of currently available softwares represent numbers (i.e. value of variables, coefficients, constants etc.) as floating-point numbers, because floating-point arithmetic is quite fast (majority of operations can be done in hardware). Although usage of floating-point arithmetic is fast and quite easy, it introduces almost unsolvable problem with correctness of results, because floating-point arithmetic is subject to rounding errors. Since there are problems, IZAR represents all numbers as fractions, i.e. as pair of two (potentially unlimited) integers – nominator and denominator, so it is possible to represent any rational number within the IZAR system without any precision lost. Usage of fractions is transparent to user. The main disadvantage of fraction number representation approach is computational and memory complexity – solving of large problems is quite slow. But the CPUs became faster and faster and the authors of this article believe that it is much more important to obtain correct result after (potentially) long time than to obtain (potentially) imperfect result in a while. However problem with solving of large sets of large problems is solvable. The IZAR system provides non-interactive (batch) mode, in which no user interaction is needed. It is one solution of this problem. On the other side since all methods are written in Pascal language, it is possible to compile one specialized program using ordinary Pascal compiler (for example) that produces highly optimized code for target platform using floating-point arithmetic. The resulting specialized program will be much faster, but may produce incorrect results (because of rounding errors in floating-point arithmetic).

The third advantage is unbounded dimension of solved problem. Programs for solving of linear programming models are usually limited by the number of variables and the number of constraints. This program has no limits. The user can solve the problems of arbitrary number of variables, constraint and objective functions with respect to computer capacity.

The fourth advantage is a number of implemented methods. IZAR knows eleven methods for multiobjective optimization problems those are named in section 3.2. As it is written above, the user can extend this system about additional methods. In the final version will be implemented methods for multiattribute evaluation problems and in the event of interest the system can be extend for example about data envelopment analysis (DEA models) and so on.

All methods are inserted in the expert system. It chooses the method most suitable for the problem. By specified questions the expert system selects the right procedure for solving the problem and so the system can be used by users without decision theory knowledge.

The last but not least advantage of the IZAR system is the fact that such system was created. There exist a lot of systems for linear programming problems but all are focused on one objective function. The IZAR system is the first one for multiobjective optimization problems that is accessible to students of Czech universities and not only to them. The final version placed on web sides will be open to all people interested in decision problems.

Bibliography

[1] Fiala Petr: Modely a metody rozhodování. 1. vyd. Praha : Oeconomica, 2003. ISBN 80-245-0622-X.

[2] Fiala Petr, Fábry Jan: SYMCA-System for Multicriteria Analysis. Central European Journal for Operations Research and Economics, 1993, roč. 2, č. 2, s. 177–184.

[3] Lagová Milada, Jablonský Josef: Lineární modely. 1. vyd. Praha : VŠE, 1999. ISBN 80-7079-959-5.

[4] Wirth Niklaus, Kathy Jensen: PASCAL – User Manual and Report. Springer-Verlag, 1974.

Stochastic Programming Programs with Linear Recourse; Application to Problems of Two Managers

Vlasta Kaňková

Department of Econometrics
Institute of Information Theory and Automation, AS ČR
Pod Vodárenskou věží 4, 182 08 Praha 8
kankova @utia.cas.cz

Abstract

[1] Stochastic programming problems with recourse are a composition of two (outer and inner) optimization problems. A solution of the outer problem depends on the "underlying" probability measure while a solution of the inner problem depends on the solution of the outer problem and on the random element realization. Evidently, a position and optimal behaviour of two managers can be (in many cases) described by this type of the model in which the optimal behaviour of the main manager is determined by the outer problem while the optimal behaviour of the second manager is described by the inner problem. We focus on an investigation of properties of the inner problem.

Keywords

Stochastic programming problems with linear recourse, stability, empirical estimates, Lipschitz function, strongly convex function

JEL classification: C 44
AMS classification: 90 C 15

1 Introduction

First, let us consider an example of a "classical" production planning problem in which quantity b of raw materials is random. In the case of a deterministic technology matrix T and a deterministic cost vector c we obtain (under some additional assumptions) the "underlying" linear programming problem with a random element b in the form.

Find

$$\max\{c^{'}x \,|\, Tx \leq b, \; x \geq 0\}.$$

Furthermore, let us complete the above mentioned "classical" example to a special situation under which the unutilized raw materials can be employed for a next production and, moreover, this second production can be organized by a relatively independent manager (maybe in another locality). Evidently, the aim of the second manager is to maximize profit from this additional production. We can suppose that the additional problem can be written in the form.

Find

$$\max_{\{y \in R^{n_1} : \, Wy = b - Tx, \, y \geq 0\}} q^{'}y.$$

[1]This research was supported by the grant agency of the Czech Republic under Grants 402/04/1294, 402/05/0115 and the Grant Agency of AS CR under Grant A 7075202.

Supposing that the main manager has also a profit from the inner problem, his or her decision is determined by the problem.

Find

$$\max_{x \in C} \mathsf{E}_{F^b}\{c^{'}x + K^{*} \max_{\{y \in R^{n_1}:\, Wy = b - Tx,\, y \geq 0\}} q^{'}y\} \quad \text{for some} \quad K^{*} > 0. \tag{1}$$

The matrices T, W and the vector q are supposed to be (in this special case) deterministic of suitable types; $c^{'}$, $q^{'}$ denotes a transposition of the vector c, q; F^b denotes the distribution function of b; $C \subset R^n$ is a nonempty (maybe polyhedral) set; the symbol E_{F^b} is reserved for the mathematical expectation corresponding to F^b.

Under the assumption of a complete information about P_{F^b} the decision of the second manager is determined by the problem.

Find

$$\max_{\{y \in R^{n_1}:\, Wy = b - Tx(F^b),\, y \geq 0\}} q^{'}y, \tag{2}$$

where $x(F^b)$ is a solution of the outer problem (1). Evidently, the decision of the inner problem (in the situation mentioned above) depends also on the underlying probability distribution. In details, if the underlying probability measure P_{F^b} is replaced by another P_{G^b}, then the solution $x(F^b)$ and the solution $x(G^b)$ can be different and consequently the solutions of the inner problem $y(x(F^b),\, b)$, $y(x(G^b),\, b)$ can be also different. Moreover, if F^b in (1) is replaced by an empirical distribution function, then the inner problem depends also on this estimate. Consequently, surely, it is reasonable to investigate the inner problem from the point of view of a stability (considered with respect to a probability measures space) and from the point of view of empirical estimates. We shall consider this type of problems in a rather more general setting.

2 Mathematical Definition of the Problem

To define the above mentioned problem in a rather more general setting, let $\xi := \xi(\omega)$ $(m \times 1)$ be a random vector defined on a probability space $(\Omega,\, \mathcal{S},\, P)$; $q\,(n_1 \times 1)$, $W(m \times n_1)$, $T(m \times n)$, $m \leq n$ be a deterministic vector and matrices. We denote by F^{ξ}, $P_{F^{\xi}}$ the distribution function and the probability measure corresponding to the random vector ξ; $Z_{F^{\xi}}$ the support of $P_{F^{\xi}}$. Let, moreover, $g_0(x,\, z)$ be a function defined on $R^n \times R^m$; $C \subset R^n$ be a nonempty, closed convex set. Symbols x, y denote n–dimensional decision vector and n_1–dimensional decision vector depending on the decision x and the realization of ξ. (R^n denotes the n–dimensional Euclidean space.)

Stochastic programming problems with linear recourse (in a rather general setting) can be introduced as a problem.

Find

$$\varphi(F^{\xi}) = \min_{x \in C} \mathsf{E}_{F^{\xi}}\{g_0(x,\, \xi) + \min_{\{y \in R^{n_1}:\, Wy = \xi - Tx,\, y \geq 0\}} q^{'}y\}. \tag{3}$$

Evidently, the problem (3) is a composition of the outer and inner problems. The solution $x(F^{\xi})$ of the outer problem depends on the probability measure $P_{F^{\xi}}$, the solution $y(:= y(x(F^{\xi}),\, \xi))$ of the inner problem depends on the solution $x(F^{\xi})$ of the outer problem and on the realization of the random element ξ. Consequently, the solution of the inner problem depends (in the case of the optimal solution $x(F^{\xi})$ of the outer problem) also on the probability measure $P_{F^{\xi}}$. In details, if the underlying probability measure $P_{F^{\xi}}$ is replaced by another $P_{G^{\xi}}$, then the solution $x(F^{\xi})$ and the solution $x(G^{\xi})$ can be different and, consequently, the solutions of the inner problem $y(x(F^{\xi}),\, \xi)$, $y(x(G^{\xi}),\, \xi)$ can be also different. The investigation of the stability of the problem (2) (considered with respect to the probability measure space equipped with a suitable metric) has been already studied in [3]. The aim of this contribution is to generalize and complete one stability assertion and, furthermore, to investigate the case when F^{ξ} is replaced by an empirical distribution function F^{ξ}_N, $N = 1, 2, \ldots$ determined by an independent random sample $\{\xi^i\}_{i=1}^N$ corresponding to the distribution function F^{ξ}. We denote

$$
\begin{aligned}
Q(x,\,\xi) &= \min_{\{y \in R^{n_1}:\, Wy = \xi - Tx,\, y \geq 0\}} q^{'} y, \qquad Q^1_{F^{\xi}}(x) \;=\; \mathsf{E}_{F^{\xi}} Q(x,\,\xi),\\
\mathcal{Y}(x,\,\xi) &= \{y \in R^{n_1} :\, Q(x,\,\xi) = q^{'} y;\; Wy = \xi - Tx,\; y \geq 0\},\\
Q_{F^{\xi}}(x) &= \mathsf{E}_{F^{\xi}}\{g_0(x,\,\xi) + \min_{\{y \in R^{n_1}:\, Wy = \xi - Tx,\, y \geq 0\}} q^{'} y\},\\
\mathcal{X}(F^{\xi}) &= \{x \in C : Q_{F^{\xi}}(x) = \varphi(F^{\xi})\}.
\end{aligned}
\tag{4}
$$

It follows from the relations (4) that $Q(x,\,\xi)$, $\mathcal{Y}(x,\,\xi)$ depend on the vector x only through the value Tx. Consequently, there exists a function $Q_T(t,\,z)$ defined for $t \in R^m$, $t = Tx$, $x \in C$ such that

$$
Q_T(t,\,z) = Q(x,\,z) \quad \text{for} \quad x \in C, \quad t = Tx, \quad z \in R^m. \tag{5}
$$

Furthermore, it follows from the above discussed analysis that in the case when the both managers know F^{ξ}, then the inner problem is a problem of linear programming and, consequently, stochastic properties of $Q(x,\,\xi)$ can be determined. However in the case when the complete knowledge of $P_{F^{\xi}}$ is not guaranteed, then for every realization ξ the value $Q(x,\,\xi)$ depends on the decision approximation of $x(F^{\xi})$. We try to generalize a stability result of the inner problem (with respect to the probability measures space) [3] and, furthermore, to investigate statistical behaviour of $Q(x(F^{\xi}_N),\,\xi)$.

3 Some Definitions and Auxiliary Assertions

First, we introduce a system of the assumptions.

- A.1 W is a complete recourse matrix (for every $z \in R^m$ there exists $y \geq 0$ such that $Wy = z$),
- A.2 there exists $u \in R^m$ such that $u^{'} W \leq q$,
- A.2' there exists a vector $\bar{u} \in R^m$ such that $\bar{u}^{'} W < q$ componentwise,
- A.3 there exists $\int\limits_{R^m} \|z\| dP_{F^{\xi}} < +\infty$ ($\|\cdot\|$ denotes the Euclidean norm in R^m),
- A.4 the probability measure $P_{F^{\xi}}$ is absolutely continuous with respect to the Lebesgue measure on R^m,
- A.4' A.4 is fulfilled and, moreover, there exists a convex open set $V \subset R^m$, constants $r > 0$, $\rho > 0$ and a density f^{ξ} of $P_{F^{\xi}}$ such that

$$
f^{\xi}(t^{'}) \geq r \quad \text{for all} \quad t^{'} \in R^m \quad \text{with} \quad \mathrm{dist}(t^{'},\,V) \leq \rho,
$$

- A.5
  - a. for every $x \in C$ there exists a finite $\int\limits_{R^m} g_0(x,\,\xi) dP_{F^{\xi}}$,
  - b. for every $z \in Z_{F^{\xi}}$, $g_0(x,\,z)$ is a strongly convex function on C,
  - c. $g_0(x,\,z)$ is a uniformly continuous function on $C \times Z_{F^{\xi}}$,
- A.6 $Z_{F^{\xi}}$ is a convex set, $f^{\xi}(t^{'}) \geq r$ for all $t^{'} \in Z_{F^{\xi}}$.

Proposition 1. [7] If

i. A.1, A.2', A.3 and A.4 are fulfilled, then $\mathsf{E}_{F^\xi} Q(x, \xi) = \mathsf{E}_{F^\xi} Q_T(t, \xi)$, $t = Tx$ is a strictly convex function of Tx on any open convex subset $V_1 \subset R^m$ of a support of P_{F^ξ}.

ii. A.1, A.2', A.3 and A.4' are fulfilled, then $\mathsf{E}_{F^\xi} Q_T(t, \xi)$ is a strongly convex function on V (given by the assumption A.4').

Remark 1. Evidently, if we add the assumption A.6, then $\mathsf{E}_{F^\xi} Q_T(t, \xi)$ is strongly convex on every open convex $V \subset Z_{F^\xi}$ such that $V(\varepsilon) \subset Z_{F^\xi}$ for some $\varepsilon > 0$. ($V(\varepsilon)$ denotes ε–neighbourhood of V.)

To recall the next assertion we introduce the Wasserstein metric $d_{W_1}(\cdot, \cdot)$ (see also [5]) .To this end let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures on R^s, $s \geq 1$. If $\mathcal{M}_1(R^s) = \{\nu \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\| \nu(dz) < \infty\}$ and $\mathcal{D}(\nu, \mu)$ denotes the set of measures in $\mathcal{P}(R^s \times R^s)$ whose marginal measures are ν, μ, then

$$d_{W_1}(\nu, \mu) = \inf\{ \int\limits_{R^s \times R^s} \|z - \bar{z}\| \kappa(dz \times d\bar{z}) : \kappa \in \mathcal{D}(\nu, \mu)\}, \quad \nu, \mu \in \mathcal{M}_1(R^s).$$

Proposition 2. [6] Let C be a polyhedral set. Let, moreover, A.1, A.2 and A.3 be fulfilled. If

1. $\mathsf{E}_{F^\xi} Q_T(t, \xi)$ is a strongly convex function on a convex open set U_P containing $T\mathcal{X}(F^\xi)$ and, moreover, $\mathcal{X}(F^\xi)$ is a nonempty, bounded set,
2. there exists n–dimensional deterministic vector c such that $g_0(x, z) = c^{'} x$,

then there exist constants $L^* > 0$, $\delta^* > 0$ such that

$$\Delta[\mathcal{X}(F^\xi), \mathcal{X}(G^\xi)] \leq L^* [d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{whenever} \quad P_{G^\xi} \in \mathcal{M}_1(R^m), d_{W_1}(P_{F^\xi}, P_{G^\xi}) \leq \delta^*.$$

($\Delta[\cdot, \cdot]$ denotes the Hausdorff distance of closed subsets of R^n; for the definition see e.g. [1]).

Lemma 1. [3] (for details see e.g. [4]) If A.1 and A.2 are fulfilled, $C \subset R^n$ is a nonempty, convex, compact set, then

1. for $z \in R^m$, $Q(x, z)$ is a piecewise linear, convex, continuous and Lipschitz function of x on C (consequently also of Tx on TC) with the Lipschitz constant not depending on $z \in R^m$,
2. for every $x \in R^n$, $Q(x, z)$ is a convex and Lipschitz function of $z \in R^m$ with the Lipschitz constant not depending on $x \in C$ (consequently also on Tx),

If, moreover,

3. $g_0(x, z)$ is for every $z \in Z_{F^\xi}$ a strongly convex function on C, then $Q_{F^\xi}(x)$ is a strongly convex function on C.

Proposition 3. [2] Let C be a nonempty, compact set. If $\bar{g}(x, z)$ is a uniformly continuous, bounded function on $C \times Z_{F^\xi}$, then

$$P\{|\min_{x \in C} \mathsf{E}_{F^\xi_N} \bar{g}(x, \xi) - \min_{x \in C} \mathsf{E}_{F^\xi} \bar{g}(x, \xi)| \longrightarrow_{N \longrightarrow +\infty} 0\} = 1.$$

If, moreover, C is a convex set and $\bar{g}(x, z)$ is for every $z \in Z_{F^\xi}$ a strictly convex function on C, then also

$$P\{\|\bar{x}_N(F^\xi_N) - \bar{x}(F^\xi)\| \longrightarrow_{N \longrightarrow +\infty} 0\} = 1,$$

where

$$\min_{x \in C} \mathsf{E}_{F^\xi} \bar{g}(x, \xi) = \mathsf{E}_{F^\xi} \bar{g}(\bar{x}(F^\xi), \xi), \quad \min_{x \in C} \mathsf{E}_{F^\xi_N} \bar{g}(x, \xi) = \mathsf{E}_{F^\xi_N} \bar{g}(\bar{x}_N(F^\xi_N), \xi).$$

4 Main Results

In this section we try to present two assertions dealing with stability and empirical estimates of the inner problem. The aim of these new results is, first, to complete and generalize one result of [3] (including the special case of the example presented in the introduction), furthermore, we shall deal with empirical estimates.

4.1 Stability Results

Theorem 1. Let the assumptions A.1, A.2', A.3 and A.4' be fulfilled, C be a polyhedral set, $\mathcal{X}(F^{\xi})$ be a nonempty bounded set. If

1. there exists an n–dimensional deterministic vector c such that $g_0(x, z) = c^{'}x$,
2. $x(F^{\xi}) \in \mathcal{X}(F^{\xi}), \quad T\mathcal{X}(F^{\xi}) \subset V \quad$ for some V fulfilling A.4',

then there exist constants $L_Q^1 > 0$, $d_Q^1 > 0$ such that for $P_{G^{\xi}} \in \mathcal{M}_1(R^m)$, $d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}}) \le d_Q^1$ there exists $x(G^{\xi}) \in \mathcal{X}(G^{\xi})$ fulfilling the relation

$$|\, Q(x(F^{\xi}), \xi) - Q(x(G^{\xi}), \xi)\,| \le L_Q^1 [d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})]^{\frac{1}{2}}, \quad \xi \in R^m. \tag{6}$$

Proof. First, it follows from the assertion of Proposition 1 that $\mathsf{E}_{F^{\xi}} Q_T(t, \xi)$ is a strongly convex function on V. Consequently, the assumptions of Proposition 2 are fulfilled. According to the assertion of Proposition 2 there exist constants $L^* > 0$, $\delta^* > 0$ such that

$$\Delta[\mathcal{X}(F^{\xi}), \mathcal{X}(G^{\xi})] \le L^*[d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})]^{\frac{1}{2}}, \quad \text{whenever} \quad P_{G^{\xi}} \in \mathcal{M}_1(R^m),\, d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}}) \le \delta^*.$$

Furthermore, let $x(F^{\xi}) \in \mathcal{X}(F^{\xi})$ be arbitrary, then according to the last inequality there exists $x(G^{\xi}) \in \mathcal{X}(G^{\xi})$ such that

$$\|x(F^{\xi}) - x(G^{\xi})\| \le L^*[d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})]^{\frac{1}{2}} \quad \text{whenever} \quad P_{G^{\xi}} \in \mathcal{M}_1(R^m),\, d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}}) \le \delta^*.$$

However since it follows from Lemma 1 that there exists a constant $L^{'} > 0$ such that

$$|Q(x, \xi) - Q(x^{'}, \xi)| \le L^{'}\|x - x^{'}\| \quad \text{for every } x,\, x^{'} \in C,\ \xi \in R^m,$$

we can see that there exists $x(G^{\xi}) \in \mathcal{X}(G^{\xi})$ such that

$$|Q(x(F^{\xi}), \xi) - Q(x(G^{\xi}), \xi)| \le L^{'}\|x(F^{\xi}) - x(G^{\xi})\| \le L^{'}L^*[d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})]^{\frac{1}{2}},$$

whenever $P_{G^{\xi}} \in \mathcal{M}_1(R^m)$, $d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}}) \le \delta^*$. Consequently setting $L_Q^1 = L^{'}L^*$ we obtain the assertion of Theorem 1.

4.2 Empirical Estimates

Theorem 2. Let $Z_{F^{\xi}}$ be a compact set, C be convex and compact set. If

1. the assumptions A.1, A.2, A.3, A.4 and A.5 are fulfilled,
2. $\xi, \xi^1, \ldots, \xi^N, \ldots$ are stochastically independent, corresponding to the distribution function F^{ξ}; F_N^{ξ} is an empirical distribution function determined by $\{\xi^k\}_{k=1}^N$, $N = 1, 2, \ldots$,

then

$$P\{Q(x(F_N^{\xi}), \xi) \longrightarrow_{(N \longrightarrow \infty)} Q(x(F^{\xi}), \xi)\} = 1.$$

Proof. First, it follows from the assumptions and from the assertion 1 of Proposition 3 that the solution of $x(F^{\xi})$ of the problem (1) is a singleton and, moreover,

$$P\{x(F^{\xi}_N) \longrightarrow_{(N \longrightarrow \infty)} x(F^{\xi})\} = 1.$$

Furthermore, it follows from the assertion 1 of Lemma 1 that $Q(x, z)$, $Q_T(t, z)$, are bounded and uniformly continuous functions on $C \times Z_{F^{\xi}}$ and $TC \times Z_{F^{\xi}}$. Consequently, for every $z \in Z_{F^{\xi}}$

$$P\{Q(x(F^{\xi}_N), z) \longrightarrow_{(N \longrightarrow \infty)} Q(x(F^{\xi}), z)\} = 1.$$

Employing the properties of the conditional probability measure we can see that the assertion of Theorem 2 is valid.

5 Conclusion

In the contribution we have considered stochastic programming problems with linear recourse and random only right hand side (in the constraints of the "underlying" programming problem). The aim of the contribution has been to investigate the stability and empirical estimates from the point of view of the inner problem. Some stability results presented in [3] has been by this generalized.

The investigation has been restricted to the case of deterministic technology matrix T. However, employing the results of the paper [7] we can obtain assumptions under which very similar assertions can be proven also in the case of random T. The introduced results can be useful in some economic problems including the problem of two managers as well as they can be also employed in the case of the multistage stochastic programs.

References

[1] J -B. Hiriart-Urruty and C. Lemaréchal: Fundamentals of Convex Analysis. Springer, Berlin 2001.

[2] V. Kaňková: Optimal Solution of Stochastic Optimization Problem with Unknown Parameters. In: Trans. 7th Prague Conf. 1974, Academia, Prague 1977. 349–353.

[3] V. Kaňková: On Stability of Stochastic Programming Problems with Linear Recourse. In: Mathematical Methods in Economics, Hradec Králové 2005(H. Skalská, ed.), Faculty of Informatics and Management University of Hradec Králové, Hradec Králové 2005. 188–195.

[4] J. Mayer: Stochastic Linear Programming Algorithms (A Comparison Based on a Model Management System). Gordon and Breach Science Publishers, Amsterdam 1998.

[5] S. T. Rachev. Probability Metrics and the Stability of Stochastic Models. Wiley, Chichester 1991.

[6] W. Römisch and R. Schulz: Stability of Solutions for Stochastic Programs with Complete Recourse. Math. Oper. Res. *18* (1993), 590–609.

[7] R. Schulz: Strong Convexity in Stochastic Programs with Complete Recourse. J. Comput. Math. *56* (1994), 3–22.

Application of martingale approximations to AR, MA and ARMA processes

Jana Klicnarová
Faculty of Agriculture, University of South Bohemia, Studentská 13, České Budějovice, janaklic@zf.jcu.cz

Abstract
In many economic applications central limit theorems or invariance principles are used. The conditions for possibility of using CLTs or IPs are usually stated by any mixing conditions. We use new approach to this problem and derived these conditions in the way of martingale approximation.

Keywords
AR, MA, ARMA processes, martingale approximation, central limit theorem, inveriance principle.
JEL: C19

1 Introduction

For many economic applications, we need to verify for the stochastic process (X_k) satisfying of conditions for central limit theorems or invariant principles. It means if we put

$$S_n := \sum_{k=1}^{n} X_k,$$

then we find conditions under which the sequence

$$\left(\frac{S_n}{s_n}, n \in \mathbb{N}\right)$$

converges in distribution to a mixture of normal distribution (we say **the sequence** (S_n) **satisfies CLT** for short) or that

$$\left(\frac{S_{\lfloor nt \rfloor}}{s_n}, t \in [0,1]\right)$$

converges in distribution to a mixture of Wiener measures in the space of cadlag functions – $D([0,1])$ (**the sequence** (S_n) **satisfies IP**) for some sequence of positive numbers $(s_n)_{n\in\mathbb{N}}$.

We focus on stationary sequences, especially on AR, MA and ARMA processes.

The conditions for central limit theorems or invariance principles are usually based on so called mixing conditions. But modern methods use for verifying of conditions for central limit theorems or invariance principles conditions for existence of martingale approximation. These methods are less restrictive and easier to be verified. The principle is to divide the sequence (X_k) into the sum of sequences (M_k) and (R_k) such that

$$X_k = M_k + R_k,$$

where (M_k) is a martingale and (R_k) is a rest. The R_k must be small enough and for martingale (M_k) we use some martingale central limit theorem or martingale invariance principle, see for instance [2].

Conditions for existence of martingale approximation for adapted stationary sequence we can find in [8]. There is proved that the martingale approximation exists if

$$||E(S_n|\mathcal{F}_0)|| = o(\sigma_n),$$

where $(\mathcal{F}_k)$ are σ-fields generated by the sequence of (X_k), more precisely $\mathcal{F}_k = \sigma(X_i, i \leq k)$, and by σ_n we denote $||S_n||$, where $||\cdot||$ is L_2-norm, more precisely $||S_n||^2 = ES_n^2$. There is also proved that the adapted stationary process (X_k) such that $EX_k^p < +\infty$ for some $p > 2$ satisfy the invariance principle if

$$||E(S_n|\mathcal{F}_0)|| = o(\sqrt{n}\log^{-q} n)$$

for $q \geq 2$.

Volný in [5] extended this result for non adapted stationary processes (in that case the σ-field $(\mathcal{F}_i)$ is arbitrary, it is not generated by the process) and showed that the martingale approximation for non adapted processes exists if

$$||E(S_n|\mathcal{F}_0)|| = o(\sigma_n) \quad \text{and} \quad ||S_n - E(S_n|\mathcal{F}_n)|| = o(\sigma_n).$$

In [6], there is proved that the non adapted stationary sequence of X_k such that $EX_k^p < +\infty$ for some $p > 2$ satisfies the invariance principle if

$$||E(S_n|\mathcal{F}_0)|| = o(\sqrt{n}\log^{-q} n) \text{ and } ||S_n - E(S_n|\mathcal{F}_n)|| = o(\sqrt{n}\log^{-q} n)$$

for some $q \geq 2$.

The other conditions for the invariance principle are also studied for example in [3].

Many authors study the conditions for martingale approximation and moreover for central limit theorems or invariance principles only for linear processes. These results can be found for example in [7], [4], [1].

2 Notation

Let $(\Omega, \mathcal{A}, \mathrm{P})$ be a probability space and $T : \Omega \to \Omega$ be a bijective bimeasurable measure preserving transformation. By $\mathcal{F}_0$ we will denote a σ-field of $\mathcal{A}$ such that $\mathcal{F}_0 \subset T^{-1}(\mathcal{F}_0)$. We do not assume that X_0 is $\mathcal{F}_0$-measurable (then we speak about non adapted sequences). And by $\mathcal{F}_i$ we denote $T^{-i}(\mathcal{F}_0)$.

We denote by $\mathcal{I}$ the σ-field of all T-invariant sets. (Recall that the probability P is ergodic if every elements of the σ-field $\mathcal{I}$ have measure 0 or 1.)

And let H_i denote the space of $\mathcal{F}_i$-measurable and square integrable random variables.

For building of AR, MA and ARMA processes we need an underlying sequence of non correlated zero-mean random variables. We will denote the sequence by $(\xi_i)_{i\in\mathbb{Z}}$ and then, we can write for arbitrary linear process

$$X_k = \sum_{i=-\infty}^{+\infty} a_i \xi_{k-i},$$

where a_i are constants such that $\sum_{i=-\infty}^{+\infty} a_i^2 < +\infty$. The sequence $(\xi_i)_{i\in\mathbb{Z}}$, which generates linear process, we can expressed as $\xi_i = \xi_0 \circ T^i$, where ξ_0 is zero mean random variable from the space $H_0 \ominus H_{-1}$. Then we can write $X_0 = \sum_{i\in\mathbb{Z}} a_i \xi_0 \circ T^{-i}$ and $X_k = X_0 \circ T^k$.

Recall that we say that the sequence (X_k) is a MA(n) sequence, or sequence of moving averages of degree n, if we can write

$$X_k = \sum_{i=0}^{n} a_i \xi_{k-i}.$$

The sequence (X_k) is so-called AR(p) if

$$X_k - \sum_{i=1}^{p} b_i X_{k-i} = \xi_k.$$

It is known that the process AR(p) is stationary, we can expressed it as a linear process, if roots of the following equation

$$1 - \sum_{i=1}^{p} b_i x^i = 0 \tag{1}$$

lie outside of the unit circle.

And the ARMA(p,m) process (X_k) is defined by

$$X_k - \sum_{i=1}^{p} b_i X_{k-i} = \xi_k - \sum_{j=1}^{m} c_j \xi_{k-j}.$$

The process ARMA(1,1) is stationary if $|b_1| < 1$. And generally, the process is stationary if the condition (1) holds.

3 Results

Our aim is to show which results we can apply for AR, MA and ARMA processes. It was mentioned above that one of the results for the invariance principle can be derived from the theorem in [6]. We can obtain the following proposition.

Proposition 1 *If*

$$||E(S_n|\mathcal{F}_0)|| = o(\sqrt{n}\log^{-q} n) \quad \textit{and} \quad ||S_n - E(S_n|\mathcal{F}_n)|| = o(\sqrt{n}\log^{-q} n). \tag{2}$$

holds for some $q \geq 2$, $\mathrm{E}|\xi_0|^p < +\infty$ *for some* $p > 2$ *and* $\sigma^2 := \lim_{n\to+\infty} \frac{\sigma_n^2}{n}$ *satisfies* $0 < \sigma^2 < +\infty$, *then the process*

$$\left(\frac{S_{\lfloor nt\rfloor}}{\sigma_n}, t \in [0,1]\right)$$

converges in distribution to Brownian motion in the space $D([0,1])$.

Let us recall that in case of linear processes we have

$$||E(S_n|\mathcal{F}_0)||^2 = \sum_{j=-\infty}^{-n}\left(\sum_{i=j+1}^{j+n} a_i\right)^2$$

and

$$||S_n - E(S_n|\mathcal{F}_n)||^2 = \sum_{j=0}^{+\infty}\left(\sum_{i=j+1}^{j+n} a_i\right)^2.$$

So, we can the condition (2) expressed as

$$\sum_{j=-\infty}^{-n}\left(\sum_{i=j+1}^{j+n} a_i\right)^2 = o(n\log^{-2q} n) \quad \text{and} \quad \sum_{j=0}^{+\infty}\left(\sum_{i=j+1}^{j+n} a_i\right)^2 = o(n\log^{-2q} n).$$

In case when the condition $\mathrm{E}|\xi_0|^p < +\infty$ is satisfied for no $p > 2$ and we have finite only the second moment we can use the following proposition, see [1, Cor 4]:

Proposition 2 *Let* $s_n = \sqrt{n}|a_{-n} + \ldots + a_n|$. *If the following two conditions hold,*

1. $\limsup_{n\to+\infty} \frac{\sum_{i=-n}^{n}|a_i|}{|\sum_{i=-n}^{n} a_i|} < +\infty$,
2. *Either* $\sum_{k=1}^{n}\sqrt{\sum_{|i|\geq k} a_i^2} = o(s_n)$, *or* $\sum_{i\in\mathbb{Z}}|a_i| < +\infty$,

then the sequence satisfies IP. More precisely $\{s_n^{-1}S_{\lfloor nt\rfloor}, t \in [0,1]\}$ *converges in distribution in (D([0,1]),d) (the space of cadlag functions with Skorohod metric) to* $\sqrt{\mathrm{E}(\xi_0^2|\mathcal{I})}W$, *where* W *is a standard Brownian motion independent od* $\mathcal{I}$.

If we focus on AR, MA and ARMA processes we can easily seen that processes MA(n) satisfies CLT and IP for each coefficients a_i. Other processes we have to express as linear processes, at first, and then we can try to apply previous results. It is easily seen that the process AR(p) we can express as the linear process with coefficients

$$\begin{aligned} a_k &= 1,\ k = 0, \\ &= \sum_{i=1}^{k} b_i a_{k-i},\ k \leq p, \\ &= \sum_{i=1}^{p} b_i a_{k-i},\ k > p. \end{aligned}$$

Especially, the sequence AR(1) we can express as

$$X_k = \sum_{i=0}^{n} b_1^i \xi_{k-i}.$$

And this process is stationary if $|b_1| < 1$.

The sequence AR (2) we can write as

$$X_k = \sum_{i \geq 0} a_i \xi_{k-i},$$

where $a_0 = 1, a_1 = b_1$ and $a_j = b_1 a_{j-1} + b_2 a_{j-2}$ for $j \geq 2$. And the process is stationary if $b_1 + b_2 < 1$, $-b_1 + b_2 < 1$ and $|b_2| < 1$.

So, it is seen that the stationary AR(1) processes satisfies CLT and IP for every coefficients, too. (In fact, both of these cases (MA(n) and AR(1)) are so-called short-memory processes). Recall that we say that the process (X_k) is **short-memory** if the covariances of X_k are summable. In our case the process is short-memory if the sum $\sum_{i \in \mathbb{Z}} |a_i|$ is finite. In case of short-memory processes we can also use $s_n = \sqrt{n}$. In case of stationary AR(2) process we can see that it is short-memory if coefficients b_1, b_2 are positive. And similarly result there is also for AR(p) processes.

If we express the ARMA(p,m) process as the linear process, then coefficients a_i are in the form as follows. Let us denote $q := \max(p, m)$ and without loss of generality suppose that $p = m = q$ (we allow zero coefficients).

$$\begin{aligned} a_k &= 1,\ k = 0, \\ &= -c_k + \sum_{i=1}^{k} b_i a_{k-i},\ k \leq q, \\ &= \sum_{i=1}^{p} b_i a_{k-i},\ k > q. \end{aligned}$$

Now we can easily verify if some process is short-memory and the CLT and IP are satisfied and in another case, in case of so-called long-memory processes, we must for IP verify conditions of one of previous propositions.

References

[1] Dedecker, J., Merlevéde, F., Volný, D. (2006): On the weak invariance principle for non adapted sequences under projective criteria. *In preparation.*

[2] Hall, P., Heyde, C.C. (1980): *Martingale limit theory and its application*. Acamedic Press, New York.

[3] Peligrad, M., Utev, S. (2005): A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.*, **33**, p. 798-815.

[4] Peligrad, M., Utev, S. (2006): Central limit theorem for stationary linear processes. *To appear in The Annals of Probability.*

[5] Volný, D. (2005): Martingale Approximations of non Adapted Stochastic Processes with Nonlinear Growth of Variance. *To appear in Dependence in Probability and Statistics, Lecture Notes in Statistics, Vol. 187, 2006.*

[6] Volný, D., Klicnarová, J.(2006): *in preparation.*

[7] Wu, W.B., Min, W. (2005): On linear processes with dependent innovations for sums of stationary processes. *Stoch. Processes Appl.*, **115**, p. 939-958.

[8] Wu, W.B., Woodroofe, M. (2004): Martingale Approximations for Sums of Stationary Processes. *The Annals of Probability,* **2**, p. 1674-1690.

Goodwin's Predator-Prey Model with Endogenous Technological Progress

Jan Kodera, Miloslav Vosvrda*
Centre for Economic Dynamics and Econometrics in Prague
vosvrda@utia.cas.cz

Abstract:
Contemporary economics contains mainly two approaches for an explanation of fluctuations of economic activity indicators. The first approach expresses fluctuations as the expression of an environment that is fundamentally uncertain and subject to random external shocks. The second approach expresses fluctuations as the expression of a deterministic dynamic process producing more complex behaviour of the economic system. A purpose of this paper is to derive from traditional Goodwin's model the predator-prey model with the specific function for technological progress. This approach gives a system of differential equations and stochastic differential equations. A base of this system contains as variables in interest a rate of employment, a share of labour, and different forms of a variable rate of the technological progress.

Keywords:
non-linear three-equation dynamic model, predator-prey model, limit cycle
JEL *classification:* C6, E44

1 Introduction

Three versions of the non-linear form of Goodwin's accelerator are introduced as a models allowing for technological progress. These demonstrate inherent and self-sustained oscillations in the economic system as well as an explosive nature of the accelerator. The non-linear accelerator with three versions of technological progress generates different types of inherent oscillations. The version with embodied exogenous technological progress gives nonlinear oscillations. This system lacks any attractor. The second version assumes the technological progress as endogenous variable describing by specific differential equation. Introducing this equation for technological progress we get the enlarged Goodwin model and even more complex dynamics. The third version introduces stochastic technological progress. This model generates dynamics disturbed by external random shocks.

Our task is presentation of the traditional Goodwin model at first and then formulation of Goodwin model with deterministic endogenous technological progress and with stochastic technological progress and comparison of their behaviour. Systems behaviour will be illustrated by numerical examples where system parameters respect economic nature. Such illustrative tools as phase portraits of each system and evolutions of its variables are used in the article.

2 Deterministic Version

Let us assume that the Leontieff technology (Goodwin (1951))

$$y_t = \min\left\{\frac{k_t}{\sigma}, a_t l_t\right\}, \text{ with } a_t = a_0 e^{\alpha t} \tag{1}$$

governs the production, that all labour income is consumed and all capital income (profits) is invested, that the labour supply increases exponentially by the following form $n_t = n_0 e^{\beta t}$, and the growth rate of wages, $\dot{w}_t / w_t$, increases with the employment rate, υ_t, given in a linear Phillips curve-type relationship in real wages by the expression

* The support from GA CR under the grants 402/06/0990, and from MSM0021620841, and from GA AS CR under the grant A7075202 is gratefully acknowledged.

$$\frac{\dot{w}_t}{w_t} = -\gamma + \rho \upsilon_t, \text{ with } \upsilon_t = \frac{l_t}{n_t} \quad (2)$$

The real wage growth is bounded even at full employment. Employment is driven by capital accumulation since $l_t = y_t / a_t = k_t / (\sigma a_t)$, where the σ is interpreted as the capital-output ratio. Hence

$$\frac{\dot{l}_t}{l_t} = \frac{\dot{k}_t}{k_t} - \frac{\dot{a}_t}{a_t} = \frac{\dot{k}_t}{k_t} - \alpha, \quad (3)$$

where we are taking the rate of growth of labour productivity $\frac{\dot{a}_t}{a_t}$ as a Harrod-neutral constant α. Since $y_t = a_t l_t$ and the wage is $w_t l_t$, the share of labour, u_t, is

$$u_t = \frac{w_t l_t}{y_t} = \frac{w_t}{a_t} \quad (4)$$

and the share of capital is $1 - u_t$. By assumption, all capital income is invested, so

$$\dot{k}_t = (1 - u_t) y_t = (1 - u_t) k_t / \sigma, \quad (5)$$

hence the growth rate of capital is

$$\frac{\dot{k}_t}{k_t} = \frac{1 - u_t}{\sigma}. \quad (6)$$

By using expression (3), we obtain

$$\frac{\dot{l}_t}{l_t} = \frac{1 - u_t}{\sigma} - \alpha. \quad (7)$$

The dynamic equation for the employment rate is therefore given by

$$\dot{\upsilon}_t = \left(\frac{\dot{l}_t}{l_t} - \frac{\dot{n}_t}{n_t} \right) \cdot \upsilon_t = \left(\frac{1 - u_t}{\sigma} - (\alpha + \beta) \right) \cdot \upsilon_t. \quad (8)$$

An increase in the share of labour u_t squeezes the profit rate bringing about a decrease of a capital accumulation. With lower capital accumulation, the demand for labour forces increases more slowly while the labour supply increases at the exogenous rate β. The change in the employment rate depends negatively on the share of labour. An equation of motion for the share of labour is

$$\dot{u}_t = \left(\frac{\dot{w}_t}{w_t} - \frac{\dot{a}_t}{a_t} \right) \cdot u_t = (-\gamma + \rho \upsilon_t - \alpha) \cdot u_t \quad (9)$$

A higher employment rate causes the growth rate of wages to increase faster, which has a positive effect on the share of labour. Together, (8) and (9) form the Volterra predator-prey dynamic system (Borrelli R. L. and Coleman C. S. (1996), and Desai M., Henry B., Mosley A., and Pemberton M. (2004)). Equation (9) states that the employment rate, υ, the 'prey', attracts the share of labour, *u*, the 'predator'. Equation (8) states that 'prey' υ tries to escape the 'predator' *u*. The steady states $\bar{u}$ and $\bar{\upsilon}$ of this dynamic system are

$$0 = \frac{1 - \bar{u}}{\sigma} - (\alpha + \beta) \qquad \bar{u} = \frac{\eta_1}{\theta_1} \text{ and } \bar{v} = \frac{\eta_2}{\theta_2} \quad (10)$$
$$0 = -\gamma + \rho \bar{v} - \alpha$$

where $\eta_1 \equiv \frac{1}{\sigma} - (\alpha + \beta), \eta_2 \equiv \alpha + \gamma, \theta_1 \equiv \frac{1}{\sigma}, \theta_2 \equiv \rho$. A higher rate of technological progress,

α, is associated with a lower share of labour $\bar{u}$ and a higher rate of employment $\bar{\upsilon}$.

Example 1

Let us introduce a numerical example with the following parameters

$\alpha = 0.04$ $\beta = 0.003$ $\gamma = 2.54$ $\sigma = 2$ $k0 = 0.01$ $k = 1.05$ $g = 2.58$

Initial conditions for the system differential equations (8), (9) are $v_0 = 0.4$, $u_0 = 0.6$. The steady states $\bar{u}$ and $\bar{\upsilon}$ of this system are

$$\text{vsteady} := \frac{\alpha + \gamma}{\rho} \qquad \text{usteady} := \frac{\frac{1}{\sigma} - (\alpha + \beta)}{\frac{1}{\sigma}} \qquad \rho = \left(\sqrt{k} \Big/ 1 + g \right)^2$$

or numerically it is follows as vsteady=4.935, usteady=0.794. Figure 1a and Figure 1b show dynamics of v and u separately. Figure 2a shows dynamics of v and u simultaneously and Figure 2b shows a limit cycle in the dynamics of u against v and equilibrium point (usteady,vsteady). This Figure shows that if the rate of employment as prey significantly increases then rapidly increasing the share of labor as predator causes decreasing of the rate of employment. Figure 3a shows a cyclic character of the growth rate of wages and Figure 3b shows an explosive behavior of the dynamics of the output.

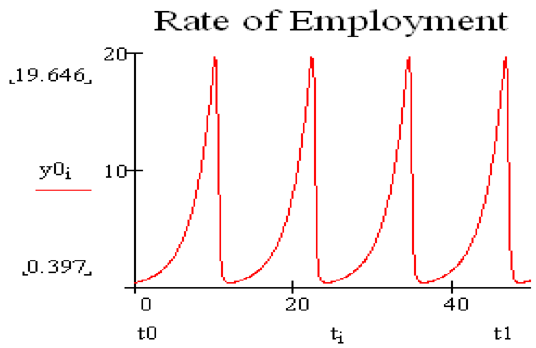


Figure 1a

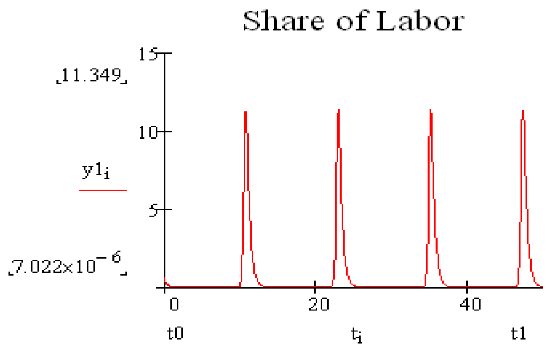


Figure 1b

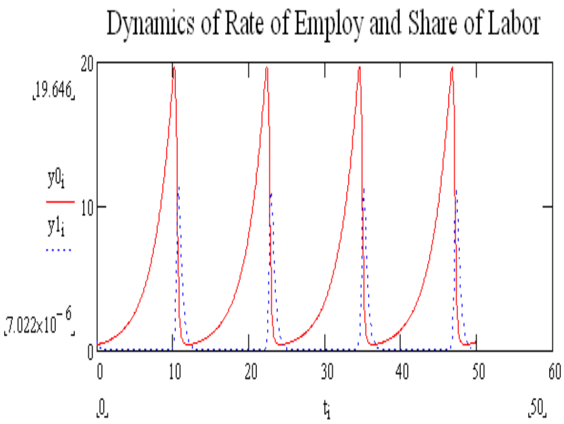


Figure 2a

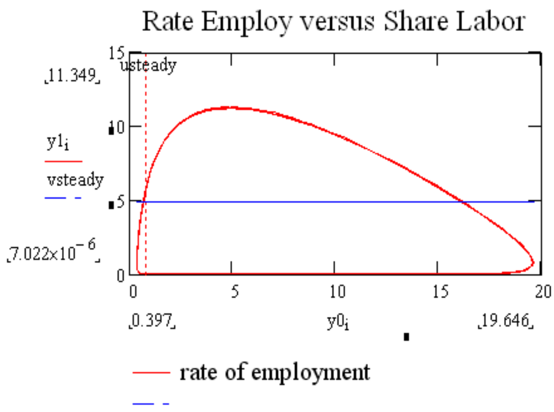


Figure 2b

There are moderate modifications. The first modification is about a profit rate. In Goodwin's model, the expression $\frac{1-u_t}{\sigma}$ is the profit rate labeled by θ_t. Equation (6) says that growth rate of capital equals the profit rate. Then an investment function as (6) assumes to invest all profits independently of their profitability. It would be more appropriate to form the investment function of the gap among the actual rate of profit θ_t and its tolerate rate, θ_t^*. Now we will implement this gap into the share of labor as follows: the profit rate θ_t falls to θ_t^* then

u_t increases to u_t^*. If u_t goes to u_t^* then $\dot{k}_t$ goes to $-\infty$. Then the investment function can acquire the following form

$$\frac{\dot{k}_t}{k_t} = \lambda \ln\left[\frac{u_t * - u_t}{1 - u_t}\right] \tag{11}$$

where λ>0 is the speed of adjustment. This investment function embodies fact that the difference among the actual and tolerated shares of labor determines investment (Desai M., Henry B., Mosley A., and Pemberton M. (2004)). Then equation (8) could have the following form

$$\dot{\upsilon}_t = \left(\lambda \ln\left[u_t^* - u_t\right] - (\alpha + \beta)\right) \cdot \upsilon_t - \lambda \ln\left[1 - u_t\right] \cdot \upsilon_t \tag{12}$$

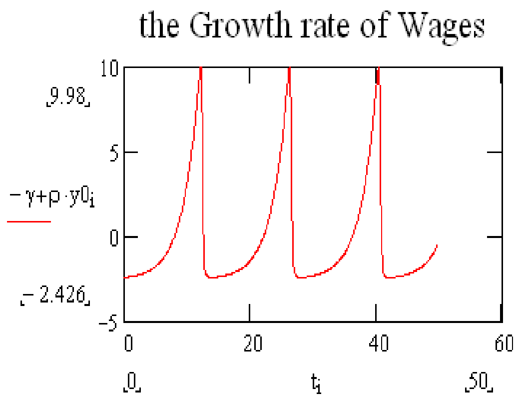


Figure 3a

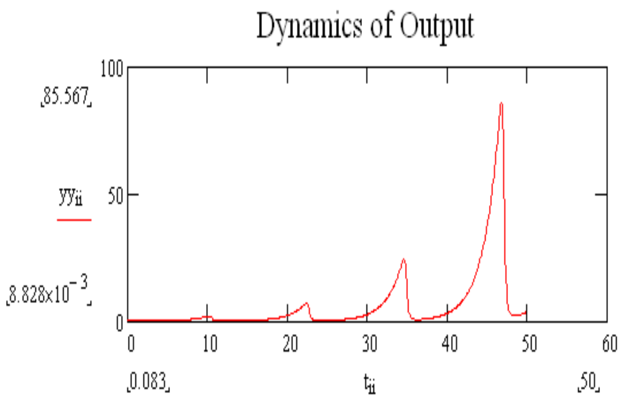


Figure 3b

The second modification is about a growth wages. Let us return to Phillips's original nonlinear form for money wages representing by the following equation

$$\frac{\dot{w}_t}{w_t} = -\gamma 1 + \rho 1 \cdot (1 - \upsilon_t)^{-\eta}, \text{ with } \eta > 0 \tag{13}$$

If $v_t \to 1$ then $\frac{\dot{w}_t}{w_t} \to \infty$. Now the equation of motion for the share of labour (9) could have the following form

$$\dot{u}_t = (-\gamma 1 - \alpha) \cdot u_t + \rho 1 \cdot (1 - v_t)^{-\eta} \cdot u_t \tag{14}$$

3 Deterministic Model with an Endogenous Technological Progress

In the preceding section the system of differential equations (8) and (9) has been derived from the traditional Goodwin model. The steady state variables $\overline{u}$ and $\overline{v}$ are given by the relations (10). Assuming that technological process does not act i.e. α=0, we get $\overline{u} = 1 - \sigma\beta$ and $\overline{v} = \gamma / \rho$ for the steady state variables. Now we introduce endogenous technological progress. It is natural to assume that the stationary level of technological progress which does not reduce or expand the labour force in the Leontieff technology is equal one. Technological progress is negatively influenced by the gap between technological progress and its stationary non-influencing level a-1. On the other side it is positively influenced by the difference between actual and steady state level of labour share given by the expression $u - (1 - \sigma\beta)$. Expressing the common activity of both mentioned factors we get the following equation describing the dynamics of technological progress

$$\dot{a}_t = -\lambda \cdot (a_t - 1) + \kappa \cdot (u_t - (1 - \sigma \cdot \beta)) \tag{15}$$

In this way the dynamic of technological progress gets more complex form than the expression for exogenous technological progress stated in the original model i.e. $a_0 e^{\alpha t}$. So it is necessary to use to keep the original notation $\dot{a}_t / a_t$ for the rate of growth of endogenous technological progress instead of constant α used for exogenous technological progress. So the equations (8) and (9) will get more general form

$$\begin{aligned} \dot{u}_t &= (\rho \cdot v_t - \gamma - \frac{\dot{a}_t}{a_t}) \cdot u_t \\ \dot{v}_t &= \frac{1 - u_t}{\sigma} v_t - (\frac{\dot{a}_t}{a_t} + \beta) \cdot v_t \end{aligned} \tag{16}$$

The equations (15) and (16) give modified Goodwin model with endogenous technological progress. But the system is not presented in canonical form which is the most appropriate. Such presentation could be made after simple adaptation

$$\begin{aligned} \dot{v}_t &= \frac{1 - u_t}{\sigma} \cdot v_t - (\frac{A_t}{a_t} + \beta) \cdot v_t \\ \dot{u}_t &= (\rho \cdot v_t - \gamma - \frac{A_t}{a_t}) \cdot u_t \\ \dot{a}_t &= -\lambda \cdot (a_t - 1) + \kappa \cdot (u_t - (1 - \sigma \cdot \beta)) \end{aligned} \tag{17}$$

where

$$A_t = -\lambda \cdot (a_t - 1) + \kappa \cdot (u_t - (1 - \sigma \cdot \beta)). \tag{18}$$

Example 2

Let us introduce a numerical example with the following parameters

$\lambda = 0.04$ $\beta = 0.05$ $\gamma = 1.9$ $\sigma = 10$ $k = 0.5$ $\rho = 2.27$

The initial conditions for the system differential equations (17), (18) are $v_0 = 0.835$, $u_0 = 0.5$, and $a_0 = 1$. Having solved this numerical example we get the following phase portrait of three differential equations and further three Figures showing the evolution of technological progress a, labour share u and the rate of employment v. Figures 4a, and 4b show dynamics of v, u and Figure 5a shows dynamics of a. Figure 5b shows a phase portrait of the system (17).

The evolution of employment rate

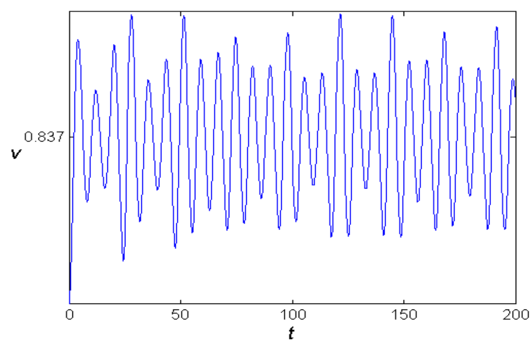


Figure 4a

The evolution of labour share

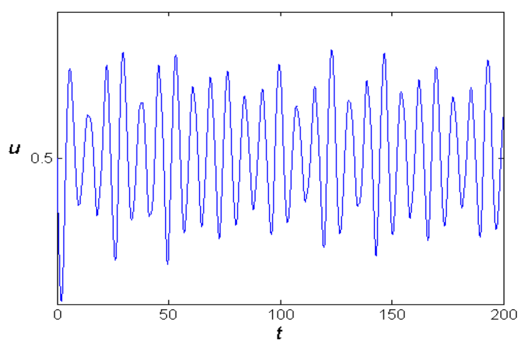


Figure 4b

The evolution of labour productivity

Phase portrait of the system

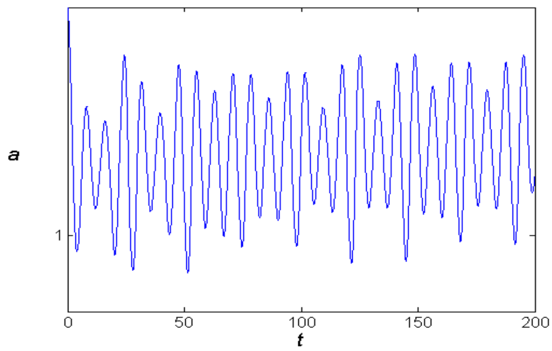


Figure 5a

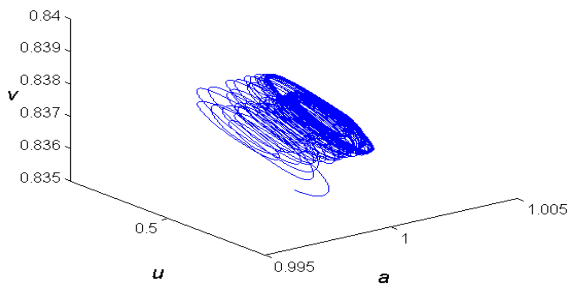


Figure 5b

4 Stochastic Setting

Let us assume the Leontieff technology again

$$y_t = \min\left\{\frac{k_t}{\sigma}, a_t l_t\right\}, \text{ but with } a_t = a_0 e^{\alpha\varepsilon_t} \tag{19}$$

where ε_t reflects the effects of the shocks. We assume that ε_t is a first order autoregressive process in the following form

$$\varepsilon_t = \alpha_1 \varepsilon_{t-1} + \eta_t \tag{20}$$

where η_t are white-noise disturbances and $-1 < \alpha_1 < 1$. Employment is driven by capital accumulation since $l_t = y_t / a_t = k_t / (\sigma a_t)$. Hence

$$\frac{\dot{l}_t}{l_t} = \frac{\dot{k}_t}{k_t} - \frac{\dot{a}_t}{a_t} = \frac{\dot{k}_t}{k_t} - \alpha - \dot{\varepsilon}_t.. \tag{21}$$

and α is again the rate of technological process $\dot{\varepsilon}_t$ is the Brownian motion. Technology is a subject to random disturbances as well. Thus,

$$a_t = a_0 \exp(\alpha t + \varepsilon_t) \tag{22}$$

Since $y_t = a_t l_t$ and the wage is $w_t l_t$, the share of labour, u_t, is

$$u_t = \frac{w_t l_t}{y_t} = \frac{w_t}{a_t} \tag{23}$$

and the share of capital is $1-u_t$. Since all capital income is invested, we have

$$k_t = (1-u_t)\, y_t = (1-u_t)\, k_t / \sigma,$$

hence

$$\frac{\dot{k}_t}{k_t} = \frac{1-u_t}{\sigma} \Rightarrow \frac{\dot{l}_t}{l_t} = \frac{1-u_t}{\sigma} - \alpha - \dot{\varepsilon}_t. \tag{24}$$

The equation of motion for the employment rate is therefore given by

$$\dot{\upsilon}_t = \left(\frac{\dot{l}_t}{l_t} - \frac{\dot{n}_t}{n_t}\right) \cdot \upsilon_t = \left(\frac{1-u_t}{\sigma} - (\alpha + \dot{\varepsilon} + \beta)\right) \cdot \upsilon_t \tag{25}$$

The equation of motion for the share of labour is

$$\dot{u}_t = \left(\frac{\dot{w}_t}{w_t} - \frac{\dot{a}_t}{a_t}\right) \cdot u_t = \left(-\gamma + \rho\upsilon_t - (\alpha + \dot{\varepsilon}_t)\right) \cdot u_t \tag{26}$$

A higher employment rate causes the wage rate to increase faster, which has a positive effect on the share of labour. Together, (25) and (22) form the Volterra predator-prey dynamic system. Equation (22) states that the employment rate, υ, the 'prey', attracts the share of labour, u, the 'predator'. Equation (25) states that 'prey' υ tries to escape the 'predator' u.

$$\dot{a}_t = (\alpha + \dot{\varepsilon}_t) \cdot a_t . \tag{27}$$

The steady states $\bar{u}$ and $\bar{\upsilon}$ are

$$\bar{u} = \frac{\eta_1}{\theta_1} \text{ and } \bar{\tau} = \frac{\eta_2}{\theta_2} \quad \eta_1 \equiv \frac{1}{\sigma} - (\alpha + \beta), \eta_2 \equiv \alpha + \gamma, \theta_1 \equiv \frac{1}{\sigma}, \theta_2 \equiv \rho \tag{28}$$

A higher rate of technological progress, α , is associated with a lower share of labour $\bar{u}$ and a higher rate of employment $\bar{\upsilon}$.

Example 3

Let us introduce a numerical example with the following parameters

$\alpha_0 = 0.001$	$\alpha_1 = 0.9$	$\alpha = 0.1$	$\beta = 0.003$	$\gamma = 2.54$	k=1.05
$\sigma = 1$	$\sigma_1 = 0.1$	$\sigma_2 = 0.3$	$\sigma_3 = 0.1$	g=0.02	

Initial conditions for the system differential equations (25), (26), and (27) are $v_0 = \sigma_1 \cdot \varepsilon_0 + 0.5$, $u_0 = \sigma_2 \cdot \varepsilon_0 + 0.2$, and $a_0 = \sigma_3 \cdot \varepsilon_0$. The steady states $\bar{u}$ and $\bar{\upsilon}$ of this system are

vsteady= $(\alpha + \gamma)/\rho$ usteady= $1 - \sigma \cdot (\alpha + \beta)$ $\rho = \left(\sqrt{k} \big/ 1+g \right)^2$

or numerically it is follows as $\text{vsteady} = 2.616$, $\text{usteady} = 0.897$. Figure 6a, and Figure 6b shows dynamics of v and u separately. Figure 7a, and Figure 7b shows the dynamics of v and u simultaneously and dynamics of u against v. Figure 8a, and Figure 8b shows the growth rate of wages and dynamics of output.

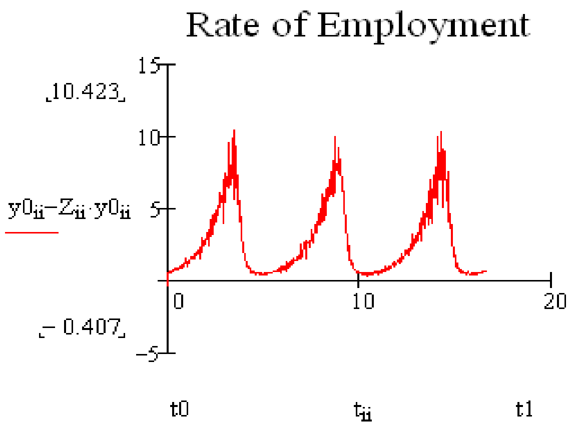


Figure 6a

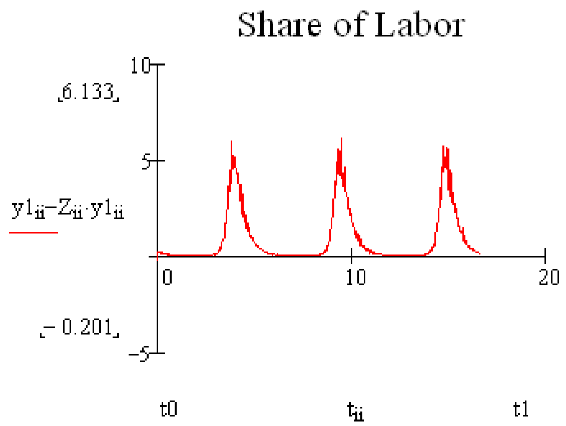
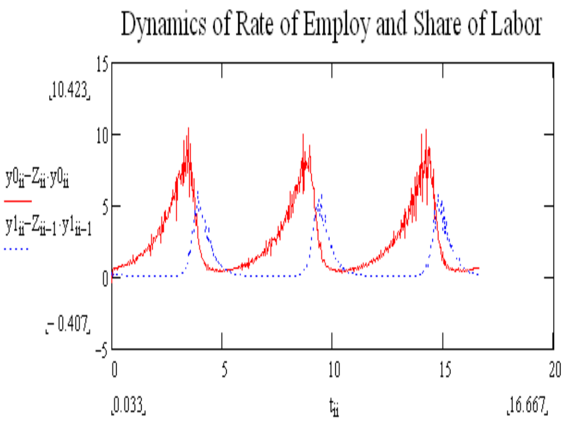


Figure 6b



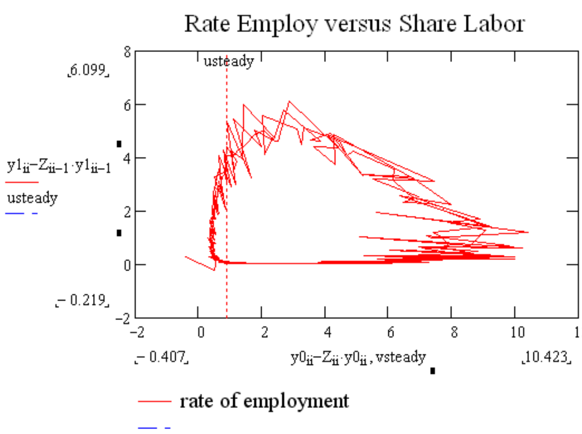


Figure 7a **Figure 7b**

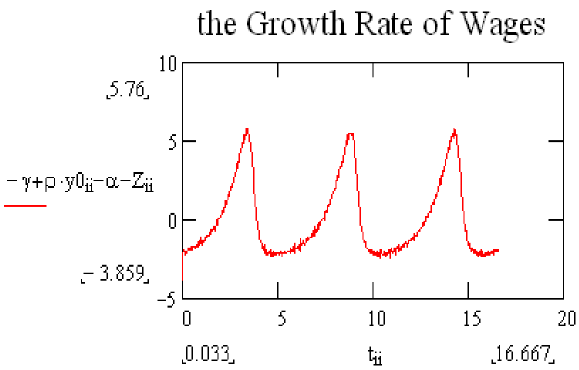


Figure 8a

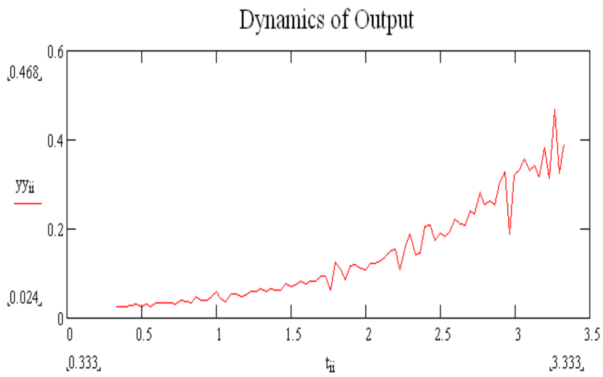


Figure 8b

4 Conclusions

In our opinion the deterministic way to technological shocks generation is very promising what was shown in the section 3. The problem is relatively small amplitude of the oscillations. On the other hand, these oscillations seem to be immanent to the system and are not the consequence of mistakes in computations or of the mistakes generated by rounding. The oscillations of an embodied technological progress, a share of labor and a rate of employment are non-periodical. So we can see that the change in the approach to technological progress changes the dynamics of the system considerably. Remember that the original Goodwin system exhibits non-linear periodical oscillation while the system with the endogenous deterministic technological progress outputs more complex dynamic. Interesting features are presented by the Goodwin model with a stochastic technological progress as has been shown in the section 4. Such representation through a time series introduces technological shocks into all components of this system. This spreading of the stochastic technological progress all through by the Brownian motion promotes chilling out of the dynamics of output (Figure 3b, and Figure 8b)

5 References

Borrelli R. L. and Coleman C. S. (1996): *Differential Equations: A Modeling Perspective,* Wiley

Desai M., Henry B., Mosley A., and Pemberton M. (2004): *A Clarification of the Goodwin Model of the Growth Cycle*, Discussion Paper 04-04, University College London, ISSN 1350-6722

Gandolfo G.: *Economic Dynamics*, Springer Verlag, Berlin, 1997

Goodwin R.M. (1951): The Nonlinear Accelerator and the Persistence of Business Cycles, *Econometrica*, vol. 19, nu. 1, pp.1-17

Oksendal B. (1998): *Stochastic Differential Equations*, Springer

Veneziani R. (2001): *Structural Stability and Goodwin's „A Growth Cycle*". Survey, Discussion Paper no.24, Bank of Italy

Stability of optimal portfolios: non-smooth utility approach

Miloš Kopa
Charles University Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Prague, Czech Republic, kopa@karlin.mff.cuni.cz

Abstract

This paper deals with utility functions and their application in a portfolio selection problem. The stability of expected utility and investment strategy of optimal portfolios in dependence on the choice of utility function is analyzed. Contrary to [2] and [3] the assumption of differentiability is not employed. Applying the theory of variational analysis, see Rockafellar & Wets [4], under assumption of hypoconvergence of utility functions, the limit set of optimal portfolios is analyzed. When solving the portfolio selection problem with non-smooth or discontinuous utility functions a computational problems may appear. This stability result allows us to approximate an intractable utility functions.

Keywords

Stability, optimal portfolio, expected utility, variational analysis
JEL: G11, D81

1 Introduction

In this paper we use utility functions, so that when solving portfolio selection problem, the optimal portfolio has the maximal expected utility. Utility functions are very useful for modeling the investor's behavior, e.g. risk aversion (or seeking). On the other hand it can be difficult to solve the portfolio selection problem for some types of utility functions. Therefore some approximations can be employed. We explore the stability of the set of optimal portfolios due to changes in a utility function. In Kopa [2] and Kopa [3], under assumption of twice differentiability of utility functions, a qualitative stability result was derived. In this paper we relax this assumption. Under assumption of hypoconvergence of utility functions, the limit set of optimal portfolios is analyzed. We follow Rockafellar & Wets [4] in defining the basic terms of variational analysis. Contrary to Rőmisch [6], we assume that probability distributions of returns are known.

The remainder of this text is structured as follows. In section 2, we recall the portfolio selection problem. In section 3, we present a stability result under assumption of convergence of Arrow-Pratt absolute risk averse measure. In section 4, we define variational analysis approach. In section 5, stability results are derived and required assumptions are discussed.

2 Portfolio selection problem

Suppose that an investor wishes to allocate his wealth among assets $i = 1, ..., n$ and he chooses $\mathbf{x} = (x_1, ..., x_n)'$ to maximize the expected utility of final wealth. This model will be formulated as:

$$\begin{aligned} \max Eu(x_0 &+ \mathbf{r}'\mathbf{x}) \\ \text{subject to}: \mathbf{1}'\mathbf{x} &= x_0 \\ x_i &\geq 0 \end{aligned} \tag{1}$$

x_0 . . . the initial wealth
$\mathbf{r}$ . . . the random vector of returns per unit of wealth
$\mathbf{x}$ . . . the investment strategy
u . . . the utility function

Assuming multiplicative approach, we could also formulate the problem as:

$$\begin{aligned} \max Eu(\mathbf{r}'\mathbf{x}x_0) \\ \text{subject to}: \mathbf{1}'\mathbf{x} = 1 \\ x_i \geq 0, \end{aligned} \tag{2}$$

Of course, it is assumed that all considered expected values exist.

3 Stability of optimal portfolio due to changes in ARA measure

Kallberg & Ziemba [1] proved that investors with the same Rubinstein measure of global risk aversion, defined as:

$$r_g(x_0) = -\frac{x_0 E\left[u''(w)\right]}{E\left[u'(w)\right]}$$

where $w = x_0\mathbf{r}'\mathbf{x}$, have the same optimal investment strategies, i.e. the same optimal solutions of (2), under the additional assumption that $\mathbf{r}'\mathbf{x}$ is normally distributed. In Pratt [5], another measure of risk aversion was suggested. Arrow-Pratt absolute risk aversion (ARA) measures have "similar" optimal portfolios. The Arrow-Pratt absolute risk aversion measure is defined as

$$r(x) = -\frac{u''(x)}{u'(x)} \tag{3}$$

for an increasing, twice differentiable utility function $u : I \to R$ and for every $x \in I$. Kallberg & Ziemba [1] also empirically examined the extent to which investors with "similar" ARA measures have "similar" optimal portfolios. The more precisely formulations of this result for portfolio selection problem (1) were derived in Kopa [2] and Kopa [3]. To present the most general result of this theory, we recall the definition of the *Hausdorf distance* between two sets, A and B:

$$d_h(A, B) = \max\{\max_{a\in A} d(a, B), \max_{b\in B} d(b, A)\} \quad \text{where } \ d(p, Q) = \min_{q\in Q} d(p, q)$$

and $d(p, q)$ is the Euclidean distance from p to q.

Theorem 1:

Let the following assumptions hold:

(i) There exists an interval $\langle a, b\rangle \subseteq I$ such that $P(x_0 + \mathbf{r}'\mathbf{x} \in \langle a, b\rangle) = 1$.
for any choice of $x_i \geq 0, i = 1, ..., n$, satisfying: $\mathbf{1}'\mathbf{x} = x_0$.

(ii) Utility functions $u : I \to R$, $u_k : I \to R$, $k = 1, 2\ldots$ are increasing and twice differentiable in the interval $I \subseteq R$ and $r(x), r_1(x), r_2(x), \ldots$ are their ARA measures.

(iii) $\lim_{k\to\infty} r_k(x) = r(x) \quad \forall x \in \langle a, b\rangle$,

(iv) $u''(x), u''_k(x), k = 1, 2, ...$ are continuous and negative in interval $\langle a, b\rangle$.

Let

$$\begin{aligned} X &= \{\mathbf{x} = (x_1, x_2, ..., x_n) : \mathbf{1}'\mathbf{x} = x_0, \ \ x_i \geq 0, \ \ i = 1, 2, .., n\} \\ X^k &= \arg\max_{\mathbf{x}\in X} Eu_k(x_0 + \mathbf{r}'\mathbf{x}) \quad k = 1, 2, ... \\ X^* &= \arg\max_{\mathbf{x}\in X} Eu(x_0 + \mathbf{r}'\mathbf{x}) \end{aligned}$$

then

$$\limsup_{k\to\infty} d_h(X^k, X^*) = 0.$$

We refer to Kopa [3] for proof and more details.

4 Variational analysis

In this section, following Rockafellar & Wets [4], we recall the basic terms of variational analysis. We consider expected utility as a function of investment strategy i.e.

$$f(\mathbf{x}) = -Eu(x_0 + \mathbf{r}'\mathbf{x}).$$

Definition 2:

(i) The function $f : R^n \to R$ is *lower semicontinuous* (lsc) at $\overline{\mathbf{x}}$ if

$$\liminf_{\mathbf{x}\to\overline{\mathbf{x}}} f(\mathbf{x}) \geq f(\overline{\mathbf{x}})$$

and lower semicontinuous on R^n if this holds for every $\overline{\mathbf{x}} \in R^n$. The function $f : R^n \to R$ is *upper semicontinuous* (usc) at $\overline{\mathbf{x}}$ if $-f$ is lsc at $\overline{\mathbf{x}}$ and upper semicontinuous on R^n if $-f$ is lower semicontinuous on R^n.

(ii) For $f : R^n \to R$, the *epigraph* of f is the set

$$\text{epi} f = \{(x, a) \in R^n \times R | a \geq f(\mathbf{x})\}.$$

(iii) For $f : R^n \to R$, the *level set* of f is the set

$$\text{lev}_\alpha f = \{\mathbf{x} \in R^n | f(\mathbf{x}) \leq \alpha\}.$$

The epigraph consists of all the points of R^{n+1} lying on or above the graph of f. For α finite, the level sets correspond to the "horizontal cross section" of the epigraph. According to Rockafellar & Wets [4], Th. 1.6 the following properties of a function f are equivalent:

(a) f is lower semicontinuous on R^n;

(b) epi$f$ is closed in R^{n+1};

(c) $\text{lev}_\alpha f$ is a closed set in R^n for all α.

The basic tool for epiconvergence approach is definition of a limit of a sequence of sets $\{C^k\}_{k\in\mathcal{N}}$ and eventually level-bounded sequence using the following notation of index sets:

$$\begin{aligned} N_\infty &= \{N \subset \mathcal{N} | \mathcal{N}\backslash N \text{ is finite}\} \\ N_\infty^\sharp &= \{N \subset \mathcal{N} | N \text{ is infinite}\} \end{aligned}$$

where $\mathcal{N}$ represents the set of natural numbers. Since $N_\infty^\sharp$ consists of all subsequences of $\mathcal{N}$ it is easily seen that $N_\infty \subset N_\infty^\sharp$.

Definition 3:

(i) For a sequence $\{C^k\}_{k\in\mathcal{N}}$ of subsets of R^n, the *outer limit* is the set :

$$\limsup_{k\to\infty} C^k = \{x \,|\, \exists N \in N_\infty^\sharp, \exists x^k \in C^k, k \in N \text{ with } x^k \xrightarrow{N} x\}.$$

while the *inner limit* of $\{C^k\}_{k\in\mathcal{N}}$ is the set:

$$\liminf_{k\to\infty} C^k = \{x \,|\, \exists N \in N_\infty, \exists x^k \in C^k, \ k \in N \text{ with } x^k \xrightarrow{N} x\}.$$

The *limit* of the sequence $\{C^k\}_{k\in\mathcal{N}}$ exists, if the outer and inner limit sets are equal:

$$\lim_{k\to\infty} C^k := \limsup_{k\to\infty} C^k = \liminf_{k\to\infty} C^k.$$

(ii) For any sequence $\{f_k\}_{k\in\mathcal{N}}$ of functions on R^n, the *lower epi-limit* ($e - \liminf_k f_k$) is the function having as its epigraph the outer limit of the sequence of sets epi f_k:

$$\text{epi}(e - \liminf_k f_k) = \limsup_k(\text{epi}(f_k)).$$

The *upper epi-limit* ($e - \limsup_k f_k$) is the function having as its epigraph the inner limit of the sequence of sets epi f_k:

$$\text{epi}(e - \limsup_k f_k) = \liminf_k(\text{epi}(f_k)).$$

When upper and inner limit coincide, the *epi-limit* ($e - \lim_k f_k$) is said to exist: $e - \lim_k f_k = e - \liminf_k f_k = e - \limsup_k f_k$. In this event the functions f_k are said to *epi-converge* to f ($f_k \xrightarrow{e} f$).

(iii) A sequence $\{f_k\}_{k\in\mathcal{N}}$ of functions on R^n is *eventually level-bounded* if for each $\alpha \in R$ the sequence of level sets ($\text{lev}_\alpha f_k$) is eventually bounded, i.e. for some index set $N \in N_\infty$ the set $\bigcup_{k\in N} \text{lev}_\alpha f_k$ is bounded.

Directly from the definition of epi-limit and from the definition of the limit of sets (epigraphs) we can see that: $e - \liminf_k f_k \leq e - \limsup_k f_k$ and $f_k \xrightarrow{e} f \Leftrightarrow \text{epi} f_k \longrightarrow \text{epi} f$.

5 Stability in variational analysis approach

Applying Rockafellar & Wets [4], Th. 7.33 in the context of the portfolio selection problem we can conclude the following stability result.

Theorem 4:

Let $f_k(\mathbf{x}) = -Eu_k(x_0 + \mathbf{r}'\mathbf{x})$ and $f(\mathbf{x}) = -Eu(x_0 + \mathbf{r}'\mathbf{x})$. Suppose the sequence $\{f_k\}_{k\in\mathcal{N}}$ is eventually level-bounded, and $f_k \xrightarrow{e} f$ with f_k and f lsc. Then

(i) $\limsup_k X^k \subset X^*$

(ii) $Eu_k(x_0 + \mathbf{r}'\mathbf{x}^k) \longrightarrow Eu(x_0 + \mathbf{r}'\mathbf{x}^*)$ for any $\mathbf{x}^k \in X^k$ and $\mathbf{x}^* \in X^*$.

Reformulating the assumptions of Theorem 4 in terms of utility functions we obtain the following result.

Corollary 5:

Suppose that the interval I is bounded. Let $u : I \longrightarrow R$ and $u_k : I \longrightarrow R$, $k = 1, 2, \ldots$, be usc utility functions with $-u_k \xrightarrow{e} -u$. Let $\mathbf{r}$ be a random vector with bounded support. Then

(i) $\limsup_k X^k \subset X^*$

(ii) $Eu_k(x_0 + \mathbf{r}'\mathbf{x}^k) \longrightarrow Eu(x_0 + \mathbf{r}'\mathbf{x}^*)$ for any $\mathbf{x}^k \in X^k$ and $\mathbf{x}^* \in X^*$.

Proof:

Since the union of domains of u, u_k, $k = 1, 2, \ldots$ is bounded and the support of $\mathbf{r}$ is bounded the union of all level sets of expected utility functions ($\bigcup_{k\in\mathcal{N}} \mathrm{lev}_\alpha[-Eu_k(x_0 + \mathbf{r}'\mathbf{x})]$) is bounded for any choice of $\alpha \in R$, i.e. the sequence $\{-Eu_k(x_0 + \mathbf{r}'\mathbf{x})\}_{k\in\mathcal{N}}$ is eventually level-bounded.

To show that $-u_k \xrightarrow{e} -u$ implies $-Eu_k(x_0 + \mathbf{r}'\mathbf{x}) \xrightarrow{e} -Eu(x_0 + \mathbf{r}'\mathbf{x})$, we apply Rockafellar & Wets [4], Th. 7.2. dealing with sufficient and necessary condition of epiconvergence: $f_k \xrightarrow{e} f$ if and only if at each point $\mathbf{x}$ both following statements hold true:

(a) $\liminf_k f_k(\mathbf{x}^k) \geq f(\mathbf{x})$ for every sequence $\mathbf{x}_k \longrightarrow \mathbf{x}$

(b) $\limsup_k f_k(\mathbf{x}^k) \leq f(\mathbf{x})$ for some sequence $\mathbf{x}_k \longrightarrow \mathbf{x}$.

Using Fatou's lemma and assumption $-u_k \xrightarrow{e} -u$, especially (a), we obtain:

$$\begin{aligned} \liminf_k \int_{R^n} -u_k(x_0 + \mathbf{r}'\mathbf{x}_k)dP(\varrho) &\geq \int_{R^n} \liminf_k - u_k(x_0 + \mathbf{r}'\mathbf{x}_k)dP(\mathbf{r}) \\ &\geq \int_{R^n} -u(x_0 + \mathbf{r}'\mathbf{x})dP(\mathbf{r}) \end{aligned}$$

for every sequence $\mathbf{x}_k \longrightarrow \mathbf{x}$ which proves (a) with $f_k(\mathbf{x}^k) = -Eu_k(x_0 + \mathbf{r}'\mathbf{x}_k)$ and $f(\mathbf{x}) = -Eu(x_0 + \mathbf{r}'\mathbf{x})$. In the same manner, for some sequence $\mathbf{x}_k \longrightarrow \mathbf{x}$ we have:

$$\begin{aligned} \limsup_k \int_{R^n} -u_k(x_0 + \mathbf{r}'\mathbf{x}_k)dP(\mathbf{r}) &\leq \int_{R^n} \limsup_k - u_k(x_0 + \mathbf{r}'\mathbf{x}_k)dP(\mathbf{r}) \\ &\leq \int_{R^n} -u(x_0 + \mathbf{r}'\mathbf{x})dP(\mathbf{r}), \end{aligned}$$

i.e. (b) holds true and the proof of epiconvergence of sequence $\{-Eu_k(x_0 + \mathbf{r}'\mathbf{x})\}_{k\in\mathcal{N}}$ is complete.

Finally, lower semicontinuity of $-Eu_k(x_0 + \mathbf{r}'\mathbf{x})$, $k = 1, 2, \ldots$, and $-Eu(x_0 + \mathbf{r}'\mathbf{x})$ will be derived. From the assumption of upper semicontinuity of u and u_k, $k = 1, 2, \ldots$ and Fatou's lemma we conclude:

$$\begin{aligned} \liminf_l \int_{R^n} -u_k(x_0 + \mathbf{r}'\mathbf{x}_l)dP(\mathbf{r}) &\geq \int_{R^n} \liminf_l - u_k(x_0 + \mathbf{r}'\mathbf{x}_l)dP(\mathbf{r}) \\ &\geq \int_{R^n} -u_k(x_0 + \mathbf{r}'\mathbf{x})dP(\mathbf{r}), \quad k = 1, 2\ldots \end{aligned}$$

$$
\begin{aligned}
\liminf_l \int_{R^n} -u(x_0+\mathbf{r}'\mathbf{x}_k)dP(\mathbf{r}) &\geq \int_{R^n} \liminf_l -u(x_0+\mathbf{r}'\mathbf{x}_l)dP(\mathbf{r}) \\
&\geq \int_{R^n} -u(x_0+\mathbf{r}'\mathbf{x})dP(\mathbf{r})
\end{aligned}
$$

for every sequence $\mathbf{x}_l \longrightarrow \mathbf{x}$ which completes the proof. □

Since x_0 is a given parameter, $\mathbf{r}$ has a bounded support and the feasible set of investment strategies is compact, assumption of boundedness of interval I represents no addition restriction.

Comparing the smooth approach (see Theorem 1) with the non-smooth approach (see Corollary 5), when assuming convergence of ARA measures, the full information about utility functions of the decision maker is not needed. This advantage can be used in the situation when we have full information about ARA measure of decision maker, but the portfolio selection problem can not be solved, because it is impossible to express analytically the exact form of utility function. In this case we can use approximation by another suitable utility function. The stability result in Theorem 1 can be useful for examination of quality of the approximation. Assuming hypoconvergence of expected utility functions, we can obtain a stability result for larger class of utility functions than the class given by assumption (b). On the other hand, to verify this assumption, the full information about utility functions is needed what can be unreachable. Typically, a verification of assumptions (b) and (c) is less demanding than a verification of the assumption of hypoconvergence. However, from the theoretical point of view, variational analysis approach offers the more general stability results.

Acknowledgements

This work was partially supported by the Grant Agency of the Czech Republic (grants 201/05/H007, 402/05/0115 and 201/05/2340). The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s.

References

[1] J. G. Kallberg, W. T. Ziemba: Comparison of alternative utility functions in portfolio selection problems, *Management Science* **29** (1983), 1257–1276.

[2] M. Kopa: Portfolio selection problem and stability of optimal portfolio, *WDS'03 Proceedings*, Part I 2003, 13 – 18.

[3] M. Kopa: Stability of optimal solution in portfolio selection problem, *Mathematical Methods in Economics 2003 Proceedings*, pp. 166 – 170.

[4] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer Verlag, Berlin, 2004.

[5] J. W. Pratt: Risk aversion in the small and in the large, *Econometrica* **32** (1964), 122–136.

[6] W. Römisch: Stability of Stochastic Programming Problems, Stochastic Programming (A. Ruszczyński and A. Shapiro eds.), Chapter 8 in Handbooks in Operations Research and Management Science, Elsevier, Amsterdam 2003.

Electricity market game

Jan Kubát
University of Economics, Návazná 1161, 165 00 Praha 6
kubatj@vse.cz

Abstract

The regional oligopolistic model of the electricity market in the Czech Republic and its neighbor states is developed. Electricity producers are the players in this non-cooperative dynamic game. Their strategies are their productions, sales, bids on the cross-border auctions and corresponding cross-border flows of electricity and their investments into building new power plants in discrete time intervals. Each producer is maximizing its discounted profit, the excess between incomes from electricity sales and costs, where costs include electricity generation costs, bids on the cross-border auctions and investments costs, in each time interval. The main aim of the model is to describe and solve the competition in the cross-border auctions together with the classic oligopoly game of the electricity producers in each time interval and to show the time development on the supply side. The equilibrium solution is supposed to reach in every time interval and so in general too. For finding solution it is designed the iterative method according to the fictive game principle. This problem should have result in the unique pure Nash equilibrium. The demand of electricity is supposed to rise in all time intervals. For the model calibration there were used real numbers but also many approximations, that were determined with respect to obtained model solutions. The number of iteration depends on the initial values of all variables and so finding good initial values is one of the problems to be solved.

Keywords

Electricity, model, market game, oligopoly, Nash equilibrium
JEL: C72

1 Introduction

Political and economical development influence changes in the electricity sector in the last years. The integration process of electricity distribution systems speeds up and the market within European Union changes into one common market without borders between member states functioned on the economic competition. Also the continuing unbundling process brings changes into behavior of the electricity market „players". According to EU Directives, full liberalization must be accomplished by 2007 in all member states. Households will be free to choose their electricity provider, in the Czech Republic it has been since 2006. There are many models describing electricity markets over the world based on game theory. Some models describe the competition between providers and producers [7] , some depict the competition mainly on supply side [4], some focuses on the cross-border trade and depicts also the regulators side as [2] and some are dynamic [6]. This model describes the market model with electricity as the recursive dynamic oligopolistic game of regional electricity producers. This model is based on the model developed in [3], the main change is in the adding the time dimension. The demand side is defined by regional demand functions. So the model does not take care about regional structure of electricity traders and customers. It is assumed that electricity producers generate a constant amount of electricity in discrete time interval and that customers, defined here as regional demands, can consume everything what is supplied. This model describes the competition of electricity producers in selling the electricity. How play this producer's game means which strategies select by each producer, where strategies are volume of electricity to produce in each region, volume of electricity to sell in each region, sum of money to bid in cross-border auctions and so the market regional transmission scheme, all these strategies in discrete time intervals. Finally each producer can decide how to invest into building new power plants. The producers choose their strategies following only one wish, to maximize their total profits. The main aim of this article is to make the formal definition of the model.

2 The model

Endogenous variables (producer's strategies)

$Q_r^v(t)$ volume of electricity that generates the producer v in region r in the time interval t,
$S_r^v(t)$ volume of electricity that offers (and sells) the producer v in the region r in the time interval t,
$T_{r^*,r}^v(t)$ volume of electricity that „transports" the producer v from the region r^* into the region r in t,
$a_{r^*,r}^v(t)$ money amount that pays v on the cross-border auction to „transport" no more than $T_{r^*,r}^v$ from the region r^* into the region r in t,
$I_r^v(t)$ investments of v in r into the new producing capacity (without change in unit costs) in t,
$J_r^v(t)$ investments of v in r into the new producing capacity (with change in unit costs) in t.

Parameters

$K_{r^*,r}^v$ maximal volume of electricity that can be „transported" from the region r^* into r,
B_r „reference price" in the region r,
b_r propensity to consumption in the region r,
R_r regional cost index of renewing power plants,
S_r regional cost index of building new power plants.

Computed variables

P_r regional market price,
$n_r^v(t)$ unit cost of production,
$Kap_r^v(t)$ production capacity.

Profit function

Z^v total profit of the producer v.

Sets

v	set of producers,	$v = 1, 2, ..., V$,
r	set of regions,	$r = 1, 2, ..., R$.
t	set of discrete time intervals,	$t = 1, 2, ..., N$.

Constants

d the discount factor,
l the time lag between time of investment I and time that this new capacity is added,
k the time lag between time of investment J and time that this new capacity is added,
g index of reducing unit costs,
Ik maximum capacity that can be added each time interval by investment type I,
Jk maximum capacity that can be added by investment type J.

Producers maximize their profits given by profit function (1).

$$Z^v = \sum_t d^t \left\{ \sum_r S_r^v(t) P_r(t) - \sum_r n_r^v(t) Q_r^v(t) - \sum_{r^*,r} a_{r^*,r}^v(t) - \sum_r R_r\, I_r^v(t) + \sum_r S_r\, J_r^v(t) \right\} \quad \text{for all } v, r \text{ and } t \qquad (1)$$

The first addend in (1) shows the producers incomes, the second one shows the production costs, the third one means the sum of all bids on the cross-border auctions and the fourth and fifth mean amounts of money on investments going to growing up producing capacities set up in the time interval t. Producers are limited by their maximal producing regional capacities that they cannot exceed (2).

$$Q_r^v \le Kap_r^v \quad \text{for all } v, r \text{ and } t \qquad (2)$$

If any producer want to „transport" some electricity between regions r and r^* he must buy the allowance to do it. He must engage the cross-border auctions between these regions to „transport" the amount of electricity which is given by (3). The „transport" means trade flows because physical electricity flows according to the physical laws.

$$T^v_{r^*,r}(t) \le \frac{a^v_{r^*,r}(t)}{\sum_v a^v_{r^*,r}(t)} K_{r^*,r} \qquad \text{for all } v,\ r^* \ne r \text{ and } t \qquad (3)$$

This definition of cross-border auction does not correspond with the reality exactly, because in the real auction the capacity is allocated to auction's competitors due the sequence of theirs offers. The price of auction is defined by the lowest price from the successful bids, so the format of this auction can be marked as the Uniform-price auction. However (3) simulates the competition on the cross-border auction and the price for all producers is the same for each unit as in Uniform-price auction format.
All producers must satisfy the balance equation (4) that equalizes the supply of producer v into region r with his corresponding regional production and regional flows of electricity realized by this producer in each region and time interval.

$$S^v_r(t) = Q^v_r(t) + \sum_{r^*} T^v_{r^*,r}(t) - \sum_{r^*} T^v_{r,r^*}(t) \qquad \text{for all } v,\ r^* \ne r \text{ and } t \qquad (4)$$

Than it is needed to equalize the total supplies (sells) and total production for each producer in each time interval, the equation (5).

$$\sum_r S^v_r(t) = \sum_r Q^v_r(t) \qquad \text{for all } v \text{ and } t \qquad (5)$$

The demand function is postulated in the simplest way as a linear function (6), which defines the regional price. This price is regional unique.

$$P_r(t) = B_r(t) - b_r \sum_v S^v_r(t) \qquad \text{for all } r \text{ and } t \qquad (6)$$

Now let us have a look on the dynamics. The demand function (6) changes from time to time only by different parameter $B_r(t)$. The key dynamic is given by capacity function (7).

$$Kap^v_r(t) = Kap^v_r(t-1) + I^v_r(t-l) + J^v_r(t-k) \qquad \text{for all } v,\ r \text{ and } t. \qquad (7)$$

Producers can choose whether kind of investment they realized if any. The investment type I mean the growing up the old power plants capacities by renewing old reactors and making them more powerful. The investment type J represents building of new power plant. If producer makes investment I, his unit costs stays unchanged, but if producers build new power plant, the investment type J, his unit costs will change (reduce) according to (8).

$$vn^v_r(t) = vn^v_r(t-1)\left(1 - g\frac{J^v_r(t-k)}{Kap^v_r(t-1)}\right) \qquad \text{for all } v,\ r \text{ and } t \qquad (8)$$

The investments cannot exceed the exogenous given maximum investments level (9).

$$I^v_r(t) \le Ik \quad J^v_r(t) \le Jk \qquad \text{for all } v,\ r \text{ and } t \qquad (9)$$

Finally it is necessary to fill in the non zero condition (10) into the model.

$$Q^v_r(t) \ge 0 \quad T^v_{r^*,r}(t) \ge 0 \quad S^v_r(t) \ge 0 \quad a^v_{r^*,r}(t) \ge 0 \quad I^v_r(t) \ge 0 \quad J^v_r(t) \ge 0$$
$$\text{for all } v,\ r^* \text{ a } r,\ r^* \ne r \text{ and } t. \qquad (10)$$

According to [1] this model can be formulated as the normal form game. The pay off functions are (1), strategies are the endogenous model variables restricted by the constraints (2) to (10). Nash equilibrium of this game (in the pure strategies) must satisfy all conditions. In case of profit maximization of each producer with other producer fixed strategies the equilibrium strategies have to stay unchanged.
On the basis of the computed solutions can be formed the regional utility of producing, trading and consuming electricity as the function O_r (11).

$$O_r = \sum_t d^t \left\{ \tfrac{1}{2}\sum_{r^*}\sum_v a^v_{r,r^*}(t) + \tfrac{1}{2}\sum_{r^*}\sum_v a^v_{r^*,r}(t) + \tfrac{1}{2} b_r \left(\sum_v S^v_r(t)\right)^2 \right\} + \sum_{v|v\in r} z^v \qquad (11)$$

O_r is defined as discounted sum of money paid by producers on the cross-border auctions from and to the region r, the halves sum as in the nowadays reality, plus regional consumers benefit and plus total profit of domestic producers.

3 The iteration method

Iteration method consists in sequence unique producer profit maximization. The other producer strategies are fixed. The iteration method ends after between two sequenced iterations steps stay the profits of all producers unchanged, respectively differ by the given small constant. One iteration step means onetime maximization of each producer. The problem to be solved in each iteration step is relatively similar. The problem of iteration method is convergence. The iteration method needn't lead into the equilibrium in spite of the equilibrium, in pure strategies, exists. With the growing up number of producer grows up the number of iteration steps required to reach equilibrium results. The speed of reaching equilibrium depends on the initial solution also. The proof that the obtained equilibrium result is really equilibrium comes directly from the theorem about Nash equilibrium [5]. The obtained equilibrium of this model does not have to be unique generally. The unverified answer on the uniqueness question can be in the different initial value solutions. This procedure evaluates the uniqueness of equilibrium. The serious mathematic proof was not done, but it is expected that the reached equilibrium is unique. Alternative way to find the equilibrium could be the Kuhn-Tucker's conditions applied on the normal form of the model.

4 Model calibration

The current model calibration is only the pilot version and it is still in processing. The production capacities come from the estimation of maximal yearly production (GWh/year), that mainly take care about regional relations. These estimations were proportionally decreased due to simplify computing. Unit costs were forecasted and given into the model in the units CZK/MWh. The cross-border profiles (capacities) come from official numbers from [8] that were multiplied by the same coefficient as yearly productions. Other parameters were more or less absolutely unknown and were set up so that the model results or the model results relations, corresponds to the reality. The pilot calibration consists from 5 regions (Czech Republic and neighbor states), 4 years and 14 producers (3 per state except Slovakia). In this pilot calibration is expected that the parameter $B_r(t)$ raise by 3 % each year. Producers can not invest outside theirs home region. The capacity lags (l and k) are set to be both 1 for the reason to show reactions in only 4 time intervals. The discount factor d is set to 0.95. The parameters for the first time period (year) are the same as in [3] so do first year equilibrium strategies. Some parameters are shown in the appendix and are depicted in italics. In the initial solution were productions set to maximal capacities of producers, all to offer in home region and for all years.

5 Model results

The model was implemented and computed in the software GAMS by the CONOPT 3 nonlinear solver. As it is seen from the table 1 in the appendix, the regional prices and offers raise in the time, due to increasing demand and also due to increasing producer capacities and lowering theirs unit costs by the investments. The money necessary to be paid on the cross-border auctions per unit raise too. The investments are summed in the table 2. Most producers invest to building new power plants mainly because of lowering the unit costs.

6 Discussion and conclusion

The model was successfully defined and solved. The reasonable equilibrium strategies were calculated by the model. This model benefit is in the setting up the cross-border competition and so in the equilibrium prices on these auctions and also in the equilibrium interregional „market" flows of electricity, all this in the discrete time. The model proofed that the low cross-border profiles disadvantage mainly the customers and producers with higher unit costs can gain better profits than some producers with lower costs. This one turns over with the raising the cross-border profiles. One of the most disputable facts is the calibration because most of parameters are still more or less good estimations and predictions. There are two main directions in the future development of this model. First, defined the demand side competition and second, solve the problem as a cooperative game. The first one finds answer on the questions about vertical market competition and the second one about possible fusions and merges between producers. It will be also suitable to make better predictions about development of the regional demand and define scenarios.

7 Appendix – result tables

Table 1: Equilibrium regional offers and prices

region	*B* (t = 1)	*b*	Total regional offers t = 1	t = 2	t = 3	t = 4	Regional prices t = 1	t = 2	t = 3	t = 4	Total regional surplus (O)
CZE	*800*	*0.08*	1 000	1 140	1 347	1 487	720	733	741	755	1 731 380
SVK	*830*	*0.12*	780	953	1 097	1 197	736	741	749	763	554 213
PL	*790*	*0.04*	2 143	2 710	3 265	3 697	704	705	707	715	2 414 828
A	*1000*	*0.14*	1 126	1 315	1 522	1 620	842	846	848	866	1 391 086
GE	*1200*	*0.03*	9 200	10 190	11 182	11 643	924	930	938	962	9 271 277

Table 2: Equilibrium investments into increasing capacity (*J* with the change of unit costs)

	Maximal capacity t = 1	Production unit costs t = 1	Investments *I* t = 1	t = 2	t = 3	Investments *J* t = 1	t = 2	t = 3	total profit	total profit per first year capacity
CZE1	*1 100*	*650*	0	0	0	10 000	10 000	10 000	385 935	351
CZE2	*400*	*690*	0	0	0	10 000	10 000	10 000	117 943	295
CZE3	*200*	*710*	0	10 000	0	10 000	10 000	10 000	75 998	380
SVK1	*800*	*700*	0	0	0	10 000	10 000	10 000	147 857	185
SVK2	*150*	*750*	0	10 000	0	10 000	10 000	10 000	49 552	330
PL1	*2 000*	*640*	0	0	0	10 000	10 000	0	555 574	278
PL2	*1 500*	*690*	0	0	0	10 000	10 000	0	85 731	57
PL3	*1 200*	*700*	0	0	0	10 000	10 000	0	42 915	36
A1	*600*	*780*	0	0	0	10 000	10 000	10 000	206 857	345
A2	*550*	*790*	0	0	0	10 000	10 000	10 000	172 320	313
A3	*300*	*800*	0	1 995	0	10 000	10 000	10 000	109 978	367
GE1	*3 500*	*840*	0	0	0	10 000	10 000	10 000	1 214 431	347
GE2	*3 000*	*850*	0	0	0	10 000	10 000	10 000	973 345	324
GE3	*3 000*	*860*	0	0	0	10 000	10 000	10 000	805 171	268

References

[1] CHOBOT M., TURNOVCOVÁ A.: Modely rozhodovania v konfliktných situáciách a za neurčitosti. Bratislava, ALFA, 1980.

[2] DAXHELET O., SMEERS Y.: Cross-Border Trade: A Two-Stage Equilibrium Model of the Florence Regulatory Forum Proposals. CMI Working Paper, 2002.

[3] KUBÁT J.,: Hra na trhu s elektřinou. Den doktorandů FIS, Praha, 2006.

[4] LISE, KEMFERT, TOL S. J.: Strategic Action in the Liberalized German Electricity Market. Fondazione Eni Enrico Mattei, Working Paper, 2003.

[5] MAŇAS M.: Teorie her a její aplikace. Praha, SNTL 1991.

[6] PINEAU P. O., MURTO P.: An Oligopolistic Investment Model of the Finnish Electricity Market. Annals of Operations Research Volume 121, 2003.

[7] SPEAR S. E.: The Electricity Market Game. GSIA Working Paper, 2001.

[8] ČEPS (online): www.ceps.cz.

Methods for the Multipletours Traveling Salesman Problem Making the Final Solution in One Time

Petr Kučera, Milan Houška, Martina Beránková
Czech University of Agriculture in Prague
Faculty of Economics and Management
Department of Operational and Systems Analysis
kucera@pef.czu.cz, houska@pef.czu.cz, berankova@pef.czu.cz

Abstract

We define the multipletours traveling salesman problem (MTTSP) as a special type of the vehicle routing problem. A capacity for each city (except the central one) as well as a capacity of vehicles (the same for all) is given. For each vehicle the sum of capacities on its route must not exceed its capacity.

This problem is NP-hard. Heuristics (approximation methods), solving it, usually only separate cities into groups, each group being determined for a particular route. Then the cities on these groups are to be ordered on the routes using some of the methods for the classical traveling salesman problem (TSP).

The aim of this contribution is to propose methods making groups of cities and determining their order on the routes in the same time, i.e. solving the whole task at once. One of these approaches is modification of the saving method (originally designed for the TSP), which chooses edges (straight routes) between two cities into the solution according to their comparison with routes between the same cities via another fixed city (the central one in the modification for the MTTSP). Another possibility is to exploit so called Habr frequencies, which are values assigned to single edges, specifying a given edge in comparison with all edges non-incident with this edge. Some modifications of these approaches are tested on several instances.

Keywords

Multiple-tours traveling salesman problem, savings method, Habr frequency.
JEL: C61

1 Introduction

We are given n cities (places, points) and a distance (cost for transportation) for each pair of them. The traveling salesman problem (TSP) is a task to find a cyclic route (Hamiltonian cycle) of minimum possible length passing exactly once through each of these cities. It belongs to the NP-hard problems, for which there is no efficient algorithm finding their theoretical optimum. So the only way how to obtain efficiently or in a reasonably short time some solution is to use some of heuristics (approximation methods), which give only "good" or "close to optimal" solution, not exactly optimum.

In practice we often meet with a task how to distribute certain material from (or to) the central point to (or from) finite number of places using a circular tour. When the capacity of one vehicle is not sufficient we have to design more than one circular tour, use more than one vehicle or the vehicle must make more tours. Similar situation occurs in passenger transport where we have to take in mind not only the vehicle capacity but also a reasonable duration of one tour. Such a problem we call multiple-tours traveling salesman problem (MTTSP).

The amount and choice of different heuristics for the MTTSP are smaller than for the traveling salesman problem. In addition, almost all do not solve this task completely, but only separate cities into groups so that each group contains cities on a route for one vehicle. Then some method for the TSP must be used to determine the order of the cities on the single cycles. The aim of this contribution is to propose some methods making these two steps – splitting cities into groups and ordering them in

each group to create the tour – in one time. Furthermore, differently from ordinary methods, these methods will work for tasks with non-symmetric cost matrix (directed edges), too.
Let us introduce some notation for the MTTSP. The central city will get index 0 and the other cities numbers from 1 to n. The cost matrix will be denoted by **C** (and so single costs c_{ij}, $i, j = 0, \ldots, n$).

2 Savings method

The savings method, first published in [1], is a heuristics for the TSP. It is based on comparing lengths of the straight route between any two cities and the route via another chosen city. The algorithm is here:

Savings Method:

1) Choose arbitrarily one city. Let us denote it 0.
2) For all pairs of other cities (i, j) compute the savings $s_{ij} = c_{i0}+c_{0j}-c_{ij}$.
3) Process edges (straight routes between pairs of cities) according to the descending order of the savings s_{ij}. If after adding this edge all the edges so far added to the route form the set of vertex disjoint paths, then add it to the solution. Repeat this until the Hamiltonian path containing all the cities except the central one has been created.
4) In the end add the city 0 to close the cyclic route.

It is recommended to try to choose all cities as the city 0 and from the obtained results to choose the best one.
For solving the MTTSP the edges from/to the central city are important, because they are more frequently used than other "ordinary" edges. So modifying this method for the MTTSP, the choice the central city 0 as the city 0 from the algorithm for the TSP will suffice. The main idea is clear: to add edges according to their savings so that they could form a feasible solution in a similar sense as in the previous algorithm. But a question what to do arises, when, adding an edge, the sum of capacities of the cities on some connected part of the solution exceeds the capacity of the vehicle. The procedure proposed here is based on experience with testing modifications of the savings method for the time bounded transportation problem in [4]:

Modified Savings Method for the MTTSP:

1) For all pairs of other cities (i, j) compute the savings $s_{ij} = c_{i0}+c_{0j}-c_{ij}$.
2) Process edges (straight routes between pairs of cities) according to the descending order of the savings s_{ij}. If after adding this edge all the edges so far added to the route form the set of vertex disjoint paths and for each path the sum of the capacities of the cities lying on it does not exceed the capacity of the vehicle, then add it to the solution. Repeat this until each city lies on some of the paths and joining arbitrary two paths the allowed capacity is exceeded.
3) In the end add the city 0 to all the paths to create cyclic routes.

3 Habr Frequencies

The disadvantage of savings is that they compare a given edge with only one route via only one city chosen for all the computation. Habr (author of e.g. [2] and [3], however, in Czech only) introduced so called frequencies, which compare the edge with all the 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 $F_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n}(c_{ij} + c_{kl} - c_{il} - c_{kj})$. This form obviously shows its sense. There exists another form, called 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 edges with the same importance. But in the case of MTTSP the edges from/to the central city are more important (more frequently and often used) than the others. Now we show how big this difference is: Let us suppose that p vehicles (cycles) will be used 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 capacity of the vehicle). Let us consider a randomly chosen (with an uniform probability distribution, without respect to the costs) solution with p cycles. The probability of the choice of an edge non-incident to the central city is $\frac{n-p}{n(n-1)}$ (we consider directed edges) while for the edges from/to the central city this probability is equal to p/n. So the edges incident to the central city are $\frac{pn-p}{n-p}$-times more important than the others (they occur in the solution with $\frac{pn-p}{n-p}$-times greater probability). Thus the frequencies for the MTTSP will be computed by the formula

$$F_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n}(c_{ij} + c_{kl} - c_{il} - c_{kj}) + \frac{pn-p}{n-p}\sum_{m=1}^{n}(2c_{ij} + c_{m0} - c_{i0} - c_{mj} + c_{0m} - c_{im} - c_{0i}) \qquad (1)$$

or modified frequencies $F`_{ij}$ can be computed by a formula derived from (1) by an analogous linear transformation as in the general case above, which we do not mention here.

Now a heuristics based on the Habr frequencies similar to the one derived from the savings method in the previous chapter can be proposed:

Habr Frequencies Approach to the MTTSP:

1) For all pairs of other cities (i, j) compute the frequencies according to (1).
2) Process edges (straight routes between pairs of cities) according to the ascending order of the frequencies. If after adding this edge all the edges so far added to the route form the set of vertex disjoint paths and for each path the sum of the capacities of the cities lying on it does not exceed the capacity of the vehicle, then add it to the solution. Repeat this until each city lies on some of the paths and joining arbitrary two paths the allowed capacity is exceeded.
3) In the end add the city 0 to all the paths to create cyclic routes.

4 Mayer method and FVL method

In the chapter 5 the results of the savings method and the Habr frequencies approach are compared with the methods used in [5]. Let us briefly describe them:

The Mayer method chooses the remotest city from the central one as the first city of the group for the first route (vehicle). Then the closest cities to some of already chosen ones are consecutively added until capacities of these cities and of the vehicle do not enable to continue. Note that this procedure is similar to the Prim (Jarník) algorithm for the minimum spanning tree. Then the remotest city from the central one of the remaining cities is taken and the same procedure is applied on the rest of the cities to create groups for the other routes. In [5] two versions of this method were used. The first one finished making the group for a route as soon as an added city causes the vehicle capacity excess for the first time, while the second version in this case did not consider such a city and tried to find another one which would have sufficiently small capacity to be joined to this group.

The Fernandez de la Vega - Lueker (FVL) method is originally designed for the bin packing problem which is NP-hard, too. We are given n elements, for each of them the weight being determined, and the task is to place them into the minimum number of classes (bins) not exceeding a given capacity. In the application for the MTTSP the elements with their weights are the cities with their capacities and the bins are the groups for the single routes. Using the FVL method, the elements are separated according their weights into two groups: the large (heavy) elements and the small (light) elements (the border between these two groups is to be set depending on the required accuracy of the solution). The large elements are placed into the bins by the optimum way. Then, in the original FVL method, the small elements are added by the first-fit way. In [5] two versions for the MTTSP for the small cities are used again. The first one considers the distances (the cities remote from the central one are processed first), the second one the capacities (the biggest of the small cities are processed first).

5 Test Computations and Their Results

For testing the same randomly generated cases, which were used in [5], were taken. Let us remind that the central city was located to the middle of the attended region where there were 12 other cities. The vehicles with the same capacity as in [5] were used, too.
First five of these cases were solved by the modified savings method from chapter 2 and the Habr frequencies approach from chapter 3. The results showed, that it is not necessary to compute the remaining tasks. The savings method gave in four and the Habr frequencies approach in three of these five cases better result than the Mayer method, which it one of the mostly used, and in the cases, when they were worse, they loose at most 0.5 p.c. These methods were successful even in the case 4, where the savings method modified for the time bounded transportation problem with the same cost matrix (cf. [4]) was not.
The results are summarized in the percentage form, where 100 p.c. is the result of the Mayer method, in the table 1, where they are compared also with the results obtained by other methods in [5]. Bold characters mean that the solution consists of four cycles. Otherwise the solution has five cycles.

	Mayer v.1	Mayer v.2	FVL v.1	FVL v.2	Savings	Habr Freq.
Case 1	100.0%	**97.4%**	**97.4%**	**97.4%**	90.9%	90.9%
Case 2	100.0%	109.6%	**132.7%**	**130.5%**	98.8%	100.7%
Case 3	100.0%	106.9%	98.8%	**118.9%**	100.5%	100.5%
Case 4	100.0%	101.1%	**105.2%**	**117.5%**	97.9%	**93.5%**
Case 5	100.0%	116.9%	**98.7%**	**103.1%**	91.6%	93.1%

Table 1: Test Cases Results

6 Conclusion

The methods for the MTTSP, which only separate cities into the groups for single routes, e.g. the Mayer method, are popular, perhaps because they are simpler than the methods proposed here. The solution of the instances of the TSP, which has to follow then, does not seem to be important complication. Especially in the cases, when short cycles with only a few cities are expected.
Nevertheless, the methods proposed here have several advantages. They solve the MTTSP completely and it is not necessary to solve any other, though small, additional problems. But it is neither the only nor the greatest their advantage. In the most of the test cases solved here they found better solution than the Mayer method. Of course, for the solid analysis of their results they are to be tested on different types of test instances. Another interesting property is that they give different solutions from the Mayer method (and from the FVL method, too, cf. [5]). And the last great advantage, which we mention here, is that these methods work for the tasks with the non-symmetric cost matrix (directed edges), too, while e.g. all methods from [5] do not.

References

[1] Clarke, G., Wright, J.W.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points, Operations Research, 12, p. 568-581, 1964

[2] Habr, J.: Jednoduché optimalizační metody pro ekonomickou praxi, SNTL Prague, 1964

[3] Habr, J.: Prognostické modelování v hospodářské praxi, SNTL Prague, 1976

[4] Kučera, P.: Tree Approach to the Time Bounded Transportation Problem, Mathematical Methods in Economics, Hradec Králové, 2005

[5] Kučera, P., Dömeová, L.: Solution of the Multipletours Traveling Salesman Problem with Vehicle Capacity Preference, Mathematical Methods in Economics, Prague, 2003

Empirical applications of threshold autoregressive models

Alexandr Kuchynka
Dept. of Statistics and Operations Research
Faculty of Economics
University of West Bohemia in Pilsen
Email: alexk@kso.zcu.cz

Abstract

In this paper, we show how nonlinear models of the threshold type which are able to capture state-dependent behaviour can be successfully applied to several economic phenomena, even those considered challenging for a long time. Special attention is devoted to modelling the relation between identical or similar goods traded in parallel markets or different locations, including relationship between stock-index futures price and spot price (based on the cost-of-carry model).

Keywords
threshold autoregressive models, nonlinear cointegration, cost of carry model, law of one price

JEL: C32, G12

Acknowledgment

Financial support from Czech Science Foundation GACR under Grant Nr. 402/05/2394 is gratefully acknowledged.

1 A general model

In this paper, we present a unifying view which is behind so called threshold autoregressive models (hereafter TAR) and their multivariate extensions leading to the notion of threshold cointegration. We concentrate on three cases where the use of a threshold model is motivated by the presence of transaction costs:
a) identical asset traded on two institutionally different markets with transaction costs represented by the market fees
b) identical (or similar) good traded on two spatially separated markets where only costs are those of transportation
c) underlying asset and its derivative (spot and futures in particular)

First case includes situations where so called "parallel markets" exist (for instance, think of countries where official and parallel exchange rates coexist and where arbitrage is costly, but still possible, or situation where identical stocks are traded both on the stock exchange and the over-the-counter system, like in the Czech Republic at the Prague Stock Exchange and RMS). The latter case is studied in the work of Hanousek and Němeček (1998), nevertheless, their methodology leads to a univariate linear model (where transaction costs are included to an additional explanatory variable) rather than a genuine multivariate threshold one. Although the logic behind all above mentioned cases is similar, the methodology for analyzing them can differ slightly depending on the data at hand.

The benchmark model for parallel markets employed in this paper builds on Garbade and Silber (1983) (augmented by transaction costs) who developed a simple framework to study price discovery across futures and spot markets. The model describes a situation when an identical asset (a stock, say) is traded on two markets (denoted by A and B). It assumes that costs of buying one stock on market A and selling it on market B (or vice versa) are given by

$$C(P_t^A, P_t^B) = t_A P_t^A + t_B P_t^B \tag{1}$$

where P_t^A, P_t^B are prices of the stock on markets A and B, respectively, and t_A, t_B are transaction costs on the markets expressed as a proportion of the trade volume.

No other costs are considered. We suppose that arbitrageurs react to a price inbalance between the markets (with one period delay) only if the arbitrage covers at least the transaction costs, that is, when the price difference exceeds the transaction costs:

$$\left|P_t^A - P_t^B\right| > C$$

which implies, after substituting

$$P_t^A > P_t^B \frac{1+t_B}{1-t_A} \text{ or } P_t^B > P_t^A \frac{1+t_A}{1-t_B}$$

This condition can be approximately restated in logs as $\left|p_t^A - p_t^B\right| > c$, where $p_t^i = \ln P_t^i, i = A, B$.
Similarly, arbitrage between spatially separated markets will take place when the log price difference exceeds transportation costs between the two locations (cf. Obstfeld and Taylor, 1997, or Lo and Zivot, 2001).

If the price on market B in the previous period was high enough relative to the price observed on market A that even after accounting for transaction costs it is profitable to buy a stock on market A and sell it on market B and vice versa.

The dynamics of the system can be described by the following set of equations:

$$\begin{aligned}
&\Delta p_t^A = \beta_4\left(p_{t-1}^B - p_{t-1}^A - c\right) + \varepsilon_t^A \\
&\Delta p_t^B = -\beta_3\left(p_{t-1}^B - p_{t-1}^A - c\right) + \varepsilon_t^B
\end{aligned} \qquad \text{if } p_t^B > p_t^A + c$$

$$\begin{aligned}
&\Delta p_t^A = -\beta_1\left(p_{t-1}^A - p_{t-1}^B - c\right) + \varepsilon_t^A \\
&\Delta p_t^B = \beta_2\left(p_{t-1}^A - p_{t-1}^B - c\right) + \varepsilon_t^B
\end{aligned} \qquad \text{if } p_t^A > p_t^B + c \tag{2}$$

$$\begin{aligned}
&\Delta p_t^A = \varepsilon_t^A \\
&\Delta p_t^B = \varepsilon_t^B
\end{aligned} \qquad \text{otherwise}$$

where $0 \le \beta_i \le 1, i = 1,2,3,4$ are reaction coefficients and $\varepsilon_t^A, \varepsilon_t^B$ zero-mean random shocks.

If we take the (lagged) pricing error $z_{t-1} = p_{t-1}^A - p_{t-1}^B$ as the transition variable, it is easy to show that the resulting model can be rewritten as a two-dimensional threshold vector error correction model.

2 Cost of carry model

The theoretical relationship between index and futures prices is given by the so called cost-of-carry model:

$$F_{t,T} = S_t e^{(r_t - d_t)(T-t)} \tag{3}$$

where S_t is the value of the underlying index at time t, $F_{t,T}$ is the "fair" index futures price at time t with maturity T, r_t is the risk free interest rate, d_t the dividend yield on the security index, and (T-t) is time to maturity of the futures contract. This is a pricing model which establishes that investors would be essentially indifferent to buying or selling underlying assets or the futures contract. Nevertheless, a number of factors (transaction costs, dividend and interest rate risk) prevents this relation from holding exactly. Thus, we can define the pricing error (or the basis) as the difference between the true (log) futures price $f_{t,T}$ and the "fair" (log) futures price:

$$z_t = f_{t,T} - s_t - (r_t - d_t)(T-t) \tag{4}$$

where $s_t = \ln S_t$. The mispricing series is expected to be stationary. If we assume that the (log) cash index and the (log) futures index series adjusted for the costs of carry are I(1), this implies that cash and adjusted futures index series should be cointegrated with a cointegrating vector (1,-1).

As in the benchmark model, the linkage in prices between a stock index and index futures is usually supposed to be maintained by arbitrage. If the futures price is high relative to the theoretical price, arbitrageurs will buy stocks underlying the index and sell the futures contracts; in the opposite case they will do the reverse. Again, in the presence of transaction costs it should be acknowledged that the widely employed linear cointegration analysis is not suitable to deal with this case and a nonlinear error correction approach would be more appropriate.

Applying the concept of threshold cointegration (Balke and Fomby, 1997) to the present case, we can formulate a bivariate threshold error correction model (TECM) as follows:

$$\begin{aligned} \Delta f_t &= \alpha_i^f z_{t-1} + \varepsilon_t^f \\ \Delta s_t &= \alpha_i^s z_{t-1} + \varepsilon_t^s \end{aligned} \tag{5}$$

where f_t is the *demeaned* log futures price series, s_t is the *demeaned* log spot price series, ε_t^f and ε_t^s are possibly heteroskedastic and cross-correlated white noises, and i = 1,2,3.

Using intraday data, subtracting daily means from log prices ensures that any constant due to dividends and interest rates for that day is removed (for more details, see Dwyer et al., 1995). For this reason, equations (5) do not contain constant terms. Moreover, lag dependence is ignored in (5) for simplicity. The mispricing term can be constructed as the difference of demeaned spot and futures log prices.

The regimes are given by the following relations:

$$\begin{aligned} &i = 1 && \text{if } z_{t-1} \le r_L \\ &i = 2 && \text{if } r_L < z_{t-1} < r_U \\ &i = 3 && \text{if } z_{t-1} \ge r_U \end{aligned} \tag{6}$$

where r_L, r_U are lower and upper thresholds, respectively, which can be restricted to be symmetric.

Subtracting equations in (5) it becomes obvious that the model can be formulated in a univariate setting as a threshold autoregressive (TAR) model for the pricing error z_t.

It is reasonable to suppose that in the inner regime the pricing error can contain a unit root, however, in both outer regime does exhibit a mean reversion. In the inner regime arbitrage is not profitable and therefore no significant adjustment occurs, on the other hand, if the pricing error crosses the boundaries given by r_L, r_U, the arbitrage begins to be profitable inducing a mean reversion in the basis.

Alternatively, if a gradual rather than discrete switch between regimes is assumed, a smooth transition autoregressive (STAR) model can be employed.

3 Law of one price in spatially separated markets: a brief review

One of the open questions in econometric research concerns price linkages (i.e. market integration) in spatially separated markets. In a frictionless world, arbitrage forces should ensure that law of one price (LOP) applies. Nevertheless, as argued by Heckscher (1916), several factors like transaction costs or tariff and nontariff barriers in the international trade drive a wedge between prices of similar goods in spatially separated markets. In the following, we sketch the most important contributions of three generations of market integration modelling depending on the methodology that was applied.

First-generation models are linear in essence. For instance, Parsley and Wei (1996) used monthly data of 51 goods and services for a group of 48 cities with a sample period from 1975 to 1992. Based on panel unit root tests, they were able to reject unit roots in the data and, in addition, they find some evidence of nonlinearities implying that convergence occurs faster for larger price differences.

Second generation authors explicitly considered threshold models beside standard panel unit root tests. O´Connell and Wei (1997) used the same data as Parsley and Wei (1996) and their results are highly supportive of the nonlinear adjustment view. Obstfeld and Taylor (1997) reintroduced Heckscher´s commodity points into the analysis of LOP and purchaising power parity (PPP) and suggested a methodology based on a threshold autoregression as a parsimonious specification for nonlinear adjustment. They used the dataset which comprises (at the disaggregated level) consumer price indices for four cities in the US and Canada sampled at monthly frequency in the years 1980 to 1995. When performing LOP tests, the authors focus on three quasi-tradable baskets (clothing, food and fuel) although they admit that even these categories may contain substantial nontradable components (related to the distribution and marketing services). The main finding was that threshold models are able to generate much quicker adjustment (which is in accordance with economic intuition) that linear AR models.

Lo and Zivot (2001) marked the beginning of the third generation which can be characterized by the shift from univariate modelling to a multivariate approach. As they stressed, the advantage of multivariate models over univariate ones is that the former allow to detect potential nonlinearities and asymmetries in the adjustment of individual prices and provides more information regarding the dynamics of the data. Moreover, when testing for threshold cointegration, procedures based on multivariate models are supposed to be more powerful. Authors applied their methodology to a set of U.S. disaggregated price data for 41 goods and service categories and 29 cities from 1986 to 1996. Differently from previous papers, no panels are formed and 1148 (41 x 28, excluding the benchmark city) bivariate systems of log prices are investigated. The results are mixed: although substiantial evidence for nonlinearity was found, however, the estimated thresholds does not appear to be in accordance with commodity arbitrage theory.

4 Spot and futures: empirical results

Early attempts to examine the price relationship between index futures and spot employed the idea of linear cointegration (see Ghosh, 1993, Wahab and Lashgari, 1993, Koutmos and Tucker,1996, and Tse,1999, among others).

Yadav et al. (1994) examined hourly data on the FTSE 100 index futures traded on the LIFFE. In their paper, they test a linear autoregression model for the pricing error against a threshold model and find some evidence in favor of the nonlinear model.

Dwyer et al. (1995) studied minute-by-minute data on the S&P 500 futures and the index itself in a threshold error correction model, finding again that a TECM offers a more adequate description of the reality than a linear ECM. They discovered no signs of asymmetry in the pricing error, and, in addition, their estimated thresholds well correspond to other sources of data from the market.

Martens et al. (1998) examine nonlinear adjustment between S&P 500 index and index futures traded at the Chicago Mercantile Exchange markets for the period of May and November 1993. The authors mention a leading effect of futures prices; that is, shocks causing deviations from no-arbitrage relation appear at the futures market first and later are followed by the underlying index. As mentioned by the authors, besides arbitrage there exists another factor explaining this effect: infrequent trading. As noted by Miller et al. (1994), the futures price reacts immediately when a new information appears in the market, however, this is not true for the stock index because not all stocks are traded within every short time period.

Applying nonlinear models to other data than S&P 500 is less common; Lin et al. (2003) successfully applied threshold framework to intraday data of Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and TAIEX futures. Robles-Fernandez et al. (2004) investigated intraday prices in the Eurostoxx50 cash and futures index and find nonlinear dynamic relationships between both sets of prices, however without imposing a particular model for the adjustment process. Bruzda (2005) focused on data from Warsaw Stock Exchange (FW20 futures contract with WIG20 as underlying index).

Conclusion

In the previous research, analyses of market integration and spot-future relationship have often been biased because the role of transaction costs has been neglegted. This omission is well known to lead to misleading results. However, with the recent theoretical developments in the area of nonlinear models (and threshold ones in particular), it has become possible to model this phenomena adequately.

References

[1] Balke, N., and F. Fomby (1997); Threshold cointegration; International Economic Review, 38:3, 627-645

[2] Bruzda, J. (2005); The Cost-of-Carry Model for the FW20 Futures Contracts Threshold Cointegration Framework; Working Paper.

[3] Dwyer, G.P., P. Locke, and W.Yu (1995); Index Arbitrage and Nonlinear Dynamic between the S&P 500 Futures and Cash; Working Paper Series, Federal Reserve Bank of Atlanta.

[4] Garbade, K., and W. Silber (1983); Price Movements and Price Discovery in Futures and Cash Markets; Review of Economics and Statistics, 289-297.

[5] Ghosh, A. (1993); Cointegration and error correction models: intertemporal causality between index and futures prices; Journal of Futures Markets, 13, 193-198.

[6] Hanousek, J., and L. Němeček (1998); Mispricing and lasting arbitrage between parallel markets in the Czech Republic; CERGE-EI working paper series.

[7] Heckscher, E. F. (1916); Vaxelkursens Grundval vid Pappersmyntfot; Ekonomisk Tidskrift 18, 309 – 312.

[8] Koutmos, G., and M. Tucker (1996); Temporal Relationship and Dynamic Interactions between Spot and Futures Stock Markets; The Journal of Futures Markets 16, 55-69.

[9] Lin, Ching-Chung, Shen-Yuan Chen, and Dar-Yeh Hwang (2003); An Application of Threshold Cointegration to Taiwan Stock Index Futures and Spot Markets; Review of Pacific Basin Financial Markets and Policies, 16(3), 291-304.

[10] Lo, M.C., and E. Zivot (2001); Threshold Cointegration and Nonlinear Adjustment to the Law of One Price; Macroeconomic Dynamics, 5(4), 533 – 576.

[11] Martens, M., Kofman, P.,and T.C.F. Vorst (1998); A Threshold Error-Correction Model for Intraday Futures and Index Returns; Journal of Applied Econometrics 13, 245-263.

[12] Miller, M. H., J. Muthuswamy, and R. E. Whaley (1994); Mean reversion of Standard & Poor's 500 index basis changes: Arbitrage-induced or statistical illusion?; Journal of Finance, 49, 479-513.

[13] Obstfeld, M., and A.M. Taylor (1997); Non-Linear Aspects of Goods-Market Arbitrage and Adjustment: Heckscher´s Commodity Points Revisited; The Journal of the Japanese and International Economics 11, 441 – 479.

[14] O´Connell, P.G.J., and S.-J.Wei (1997); ´The bigger they are, the harder they fall´ How price differences across U.S. cities are arbitraged; Working Paper, National Bureau of Economic Research, No. 6089.

[15] Parsley, D.C., and S.-J.Wei (1996); Convergence to the law of one price without trade barriers or currency fluctuations; Quarterly Journal of Economics 111, 1211-1236.

[16] Robles-Fernandez, M.D., Nieto L., and M.A. Fernandez (2004); Nonlinear Intraday Dynamics in Eurostoxx 50 Index Markets; Studies in Nonlinear Dynamics and Econometrics 8(4).

[17] Tse, Y. (1999); Price Discovery and Volatility Spillovers in the DJIA Index and Futures Markets; The Journal of Futures Markets 19, 911-930.

[18] Wahab, M., and M. Lashgari (1993); Price dynamics and error correction in stock index and stock index futures markets: A cointegration approach; Journal of Futures Markets, 13, 711-742.

[19] Yadav, P. K., P. F. Pope, and K. Paudyal (1994); Threshold autoregressive modeling in finance: The price differences of equivalent assets; Mathematical Finance, 4, 205-221.

Practical Application of Monte Carlo Simulation in MS Excel and its Add-ins – The Optimal Mobile Phone Tariffs for Various Types of Consumers in the Czech Republic

Martina Kuncova
University of Economics, W.Churchill Sq. 4, 13067 Prague 3

kuncovam@vse.cz

Abstract

Simulation methods belong to the suitable instruments that can be used in the real world situations to better understand the reality. The economists usually try to find the best way how to optimize profit, minimize costs or improve their behavior. Simulation models can help in this situation especially when there is no possibility to use analytical tools. Computer simulation means using computer models to imitate real life or make predictions. Monte Carlo simulation is a method for iteratively evaluating a deterministic model using sets of random numbers as inputs. This method is often used when the model is complex, nonlinear, or involves more than just a couple uncertain parameters. The Monte Carlo method is just one of many methods for analyzing uncertainty propagation, where the goal is to determine how random variation, lack of knowledge, or error affects the sensitivity, performance, or reliability of the system that is being modeled. Monte Carlo simulation is categorized as a sampling method because the inputs are randomly generated from probability distributions to simulate the process of sampling from an actual population. It is necessary to choose a distribution for the inputs that most closely matches data we already have, or best represents our current state of knowledge. The data generated from the simulation can be represented as probability distributions (or histograms) or converted to error bars, reliability predictions, tolerance zones, and confidence intervals. This contribution will guide you through the process of performing a Monte Carlo simulation using Microsoft Excel or its add-ins for consumers to decide which mobile phone tariff would be optimal. Although Excel will not always be the best place to run a scientific simulation, the basics are easily explained with just a few simple examples.

Keywords

Monte Carlo Method, Simulation, MS Excel, Crystal Ball, mobile phone tariff

JEL: C15

1 Introduction to Simulation

Simulation nowadays means usually a technique for imitation of some real situations, processes or activities that already exist in reality or that are in preparation. It is an attempt to model a real-life situation on a computer. The reasons for this are various: to study the system and see how it works, to find where the problems come from, to compare more model variants and select the most suitable one, etc. Simulation is used in many contexts, including the modeling of natural or human systems in order to gain insight into their functioning, simulation of technology for performance optimization, safety engineering, testing, training and education. Simulation can be used to show the eventual real effects of alternative conditions and courses of action. Key issues in simulation include data acquisition of valid source of information, selection of key characteristics and behaviors, the use of simplifying approximations and assumptions within the simulation, and also fidelity and validity of the simulation outcomes.

It is impossible to create a simulation model without computer. Computer simulation has become a useful part of modeling not only the natural systems in biology, physics and chemistry, but also the human systems in economy. Traditionally, all systems are modeled by mathematical models, which attempt to find analytical solutions to problems that enables the prediction of the behavior of the system from a set of parameters and initial conditions. Computer simulation is often used as an adjunction to modeling systems for which other analyses are too mathematically complex or too difficult to be solved analytically.

2 Monte Carlo Simulation

Monte Carlo simulation was named for Monte Carlo in Monaco, which is full of casinos containing games of chance. These games (such as dice, slot machines and roulette wheels) exhibit random behavior which is similar to how Monte Carlo simulation selects variable values at random to simulate a model. For each uncertain variable defined in model, it is necessary to specify the type of probability distribution that should be used for random variates generation to obtain the values that correspond with reality or with our expectations. In short Monte Carlo methods are stochastic techniques based on the use of (pseudo)random numbers and probability statistics to investigate problems.

It is also possible to see the Monte Carlo method from the mathematical point of view. It can be seen as a general technique of numerical integration and in this sense it is possible to construct a definite integral for the application of the Monte Carlo method. In fact this might be a problem for some managers, so in this article we do not use this kind of specification (see www.riskglossary.com).

3 Monte Carlo Simulation in Economy

The mathematical point of view might raise a presumption that the Monte Carlo method can be used only for complicated and mathematically well defined problems. In this article I would like to show how easily it can be used for analyzing or solving usual economic problems without any specific mathematical theory.

In economy we must face a lot of decisions that have to be made, and pay a lot of money afterwards often without knowing whether we have done right or wrong. A lot of these decisions are dependent upon how much money we earn, how much money we spend, how much time we may dedicate to something, etc. Some of these things are given (each day has “only” 24 hours, so it is not possible to dedicate to something more time) but some are uncertain (we cannot say exactly how much money we will spend next month). Usually people are able to describe the expenses as “something between 8 and 12 thousand crowns” or “15 thousand crowns at a medium”. Although it seems to be vague, inaccurate and insufficient, with some knowledge of statistical distributions we are able to use given information and even make a decision or recommendation.

We may think of nearly every uncertainty, that can be described as “how much...” , as an uncertain variable. On the basis of obtainable information we should select the type of probability distribution (that corresponds with our expectations about the values of the variable, and we are able to define all the parameters for). The most typical and frequent distribution types are normal, triangular, uniform (discrete uniform), Poisson, lognormal and exponential ones. Mathematical specification of these variables and the calculations derived from them might be complicated (especially when non trivial distribution is chosen). But via the simulation Monte Carlo and via MS Excel and its add-ins it is possible to analyze the problem and find a solution or a recommendation for each specified situation.

4 Monte Carlo Simulation using MS Excel and its add-ins

As one of the well known, often used and wide spread software, MS Excel spreadsheet serves as a good environment where to start simulation, since almost nearly all people working with computer know how to work with it and create a simple simulation model is easy, although Excel is not the best place to run a scientific simulation. Excel contains a pseudo random number generator that was tested for sufficiency in 1991 by Law and Kelton. The function is invoked using the Excel function =RAND()[1]. When this function is entered in a cell in an Excel spreadsheet, it generates a uniformly distributed pseudo random number between 0 and 1. Its values can be easily updated by pressing the Calculation Key F9 (every press means new simulation experiment). Via this generator it is possible to generate random variates having any other distribution – see Table 1.

Distribution	Parameters	Excel Expression
Uniform	a,b	=a+RAND()*(b-a)
Normal	μ,σ	=NORMINV(RAND(), μ, σ)
Lognormal	μ,σ	=LOGINV(RAND(), μ, σ)
Exponential	1/λ	=(-1/ λ)*LN(RAND())

Table 1 - The Excel expressions for generation of random variates from given distribution

If the distributions described in the Table 1 are sufficient to describe all uncertain variables that we have, it is possible to use an Excel sheet to solve the problem (see part 5) without anything else – only define the interconnections between the variables and specify the decision function.

There exists a lot of other distribution that might be necessary to obtain right values. One of the most suitable for various cases is triangular distribution. It is possible to derive the Excel expression for random variates generated from this distribution (Kuncova, 2005), but it is a little bit of mathematics for managers to do it. The best way for this case and for other specific distributions is to take advantage of the Excel add-ins that were made to help in this situations. They are: Crystal Ball, @Risk, Lumenaut, Simtools, Formlist, MonteCarlito, Simulacion 4.0, SimulAr, Risk Analyzer, etc. One of the best known is Crystal Ball (www.decisioneering.com). Everything might be the same as in the usual Excel sheet. The first step to using Crystal Ball is to determine which model inputs are uncertain. Then use all the information to create proper probability distribution (or assumption) for the cells – just choosing among the displayed probability distribution functions of various distributions and then change the parameters for the chosen distribution. The next step is to identify a decision function called a forecast. A forecast is a formula cell (or cells) that we want to measure and analyze. After that we may run a simulation for 1000 (or more) trials (so 1000 simulation experiments or forecasts). Simulation results are displayed in interactive histograms (or frequency charts), and there are also the most common statistics which describe each of the decision function. Part 5 of this contribution shows how to use Excel and Crystal Ball to find the optimal mobile phone tariff via Monte Carlo simulation.

5 Practical Application – The Optimal Mobile Phone Tariff

To see how Monte Carlo simulation model might look like in Excel we consider a simple but usual problem – what mobile phone tariff is the best. Let us describe a problem: a company (or a single person or subject) would like to know which of the standard tariffs of mobile phone operators in the Czech Republic would be the best (it means the total price will be minimal). The subject uses the tariff called Eurotel Gold and wants to know if there exist a better one with nearly the same monthly fee (about 1000 crowns), or if it is better to pay more or less fee. The subject does not know much about the previous calls (or the sum of the calls changes rapidly) but it is possible to set minimum and

[1] The function RAND() is called NÁHČÍSLO() in the Czech version of MS Excel.

maximum number of minutes probably called per month to each mobile phone network in Czech Republic and also minimum and maximum number of sms sent (see Table 2) – so we suppose the minutes to be generated via the uniform distribution.

	Number of minutes called per month			Generated minutes
Calls	distribution	parameters (min, max)		
T-mobile	uniform	60	120	62,0681868
Eurotel	uniform	30	90	47,3693326
Vodafone	uniform	30	100	37,0638813
Local calls	uniform	15	30	17,0133686
sms	uniform	50	80	65
			Total calls	163,514769

Table 2 – Expected intervals for minutes called per month and generated values

According to the official web sites of all the mobile phone operators (www.t-mobile.cz, www.eurotel.cz, www.vodafone.cz) we create the table (Table 3) of all the standard tariffs that are offered. In the first three rows there are the tariffs with the monthly fee around 1000 crowns.

Tariffs		**Monthly**	**Free minutes**	**Own Calls**	**Other Calls**	**sms**
T-mobile	***T250 HIT***	**1178,1**	250	3,33	3,33	1,19
Eurotel	***Gold Max***	**1059,1**	200	2,98	4,76	1,19
Vodafone	***300 Naplno***	**1190**	300	4,17	4,17	1,19
Eurotel	***Bronz Max***	214	30	3,69	6,19	1,19
	Silver Max	660	100	3,21	5,24	1,19
	Gold Max	1059	200	2,98	4,76	1,19
	Platinum Max	1892	400	3,45	4,52	1,19
	Diamant Max	3689	1000	1,19	4,05	1,19
T-mobile	***T30 HIT***	226	30	3,81	5,71	1,19
	T80 HIT	536	80	3,33	4,28	1,19
	T250 HIT	1175	250	3,33	3,33	1,19
	T500 HIT	2130	500	2,86	2,86	1,19
Vodafone	***Rozjezd Naplno***	60	0	5,95	5,95	1,19
	50 Naplno	298	50	5,36	5,36	1,19
	150 Naplno	714	150	4,76	4,76	1,19
	300 Naplno	1190	300	4,17	4,17	1,19
	500 Naplno	1904	500	3,96	3,96	1,19

Table 3 – Standard mobile phone tariffs and the prices paid for minutes called, sms send and number of free (non paid) minutes

5.1 Creating simple model in Excel

The calculations that have to be made are simple. We need to know the total costs per month. Total costs (*TC*) consist of the monthly fee (*MF*), price per sms sent (*S*), and price per calls (*PC*) to each mobile net that exceed the number of free minutes. This is the problematical part, but if we simplify it, it is possible to divide out the free minutes among the four nets. So the total costs are:

$$TC=MF+S+PC$$

$$S = ps * sms$$

ps...price per sms, sms...number of sms

$$PC = p(T)*c(T) + p(E)*c(E) + p(V)*c(V) + p(L)*c(L)$$

p(T) ... price per minute call to the T-mobile net
c(T) ... number of minutes that are supposed (generated) to be called to T-mobile net, that exceed ¼ of the free minutes
p(E) ... price per minute call to the Eurotel net
c(E) ... number of minutes that are supposed (generated) to be called to Eurotel net, that exceed ¼ of the free minutes
p(V) ... price per minute call to the Vodafone net
c(V) ... number of minutes that are supposed (generated) to be called to Vodafone net, that exceed ¼ of the free minutes
p(L) ... price per minute call to the Local net
c(T) ... number of minutes that are supposed (generated) to be called to Local net, that exceed ¼ of the free minutes

The calculations are in Table 4 in the last column. Here we may see that for this simulation experiment, the minimum total calls are achieved with the tariff "T-mobile T80 HIT, the next is "Vodafone 300 Naplno".

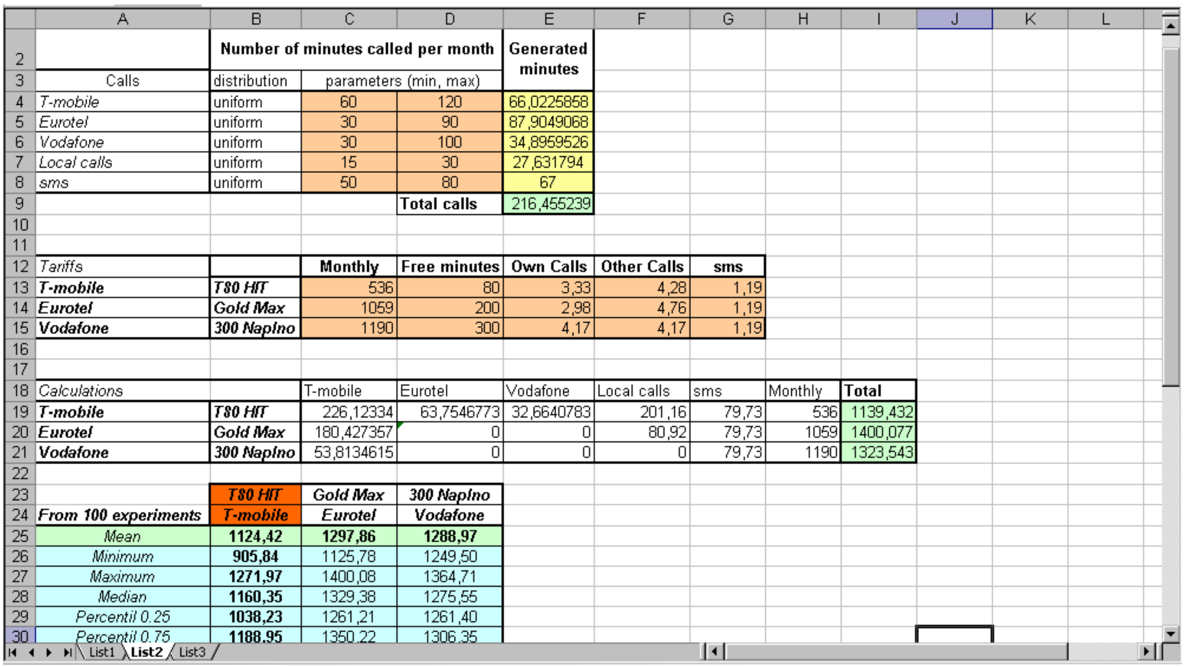
Calculations		T-mobile	Eurotel	Vodafone	Local calls	sms	Monthly	**Total**
T-mobile	***T250 HIT***	87,622	67,972	0	0	80,92	1178,1	1414,615
Eurotel	***Gold***	156,661	156,661	0	0	80,92	1059,1	1453,343
Vodafone	***300 Naplno***	0	32,993	0	0	80,92	1190	1303,913
Eurotel	***Bronz Max***	503,328	278,27	237,148	78,953	80,92	214	1392,621
	Silver Max	334,38	185,897	109,052	0	80,92	660	1370,251
	Gold Max	184,75	98,077	0	0	80,92	1059	1422,748
	Platinum Max	0	0	0	0	80,92	1892	1972,92
	Diamant Max	0	0	0	0	80,92	3689	3769,92
T-mobile	***T30 HIT***	309,803	430,602	218,759	72,83	80,92	226	1338,916
	T80 HIT	229,147	269,263	110,473	1,0911	80,92	536	**1226,896**
	T250 HIT	87,622	67,972	0	0	80,92	1175	1411,515
	T500 HIT	0	0	0	0	80,92	2130	2210,92
Vodafone	***Rozjezd Naplno***	528,438	493,326	272,579	120,516	80,92	60	1555,781
	50 Naplno	409,038	377,408	178,55	41,566	80,92	298	1385,484
	150 Naplno	244,250	216,161	39,563	0	80,92	714	1294,895
	300 Naplno	57,6	32,993	0	0	80,92	1190	1361,514
	500 Naplno	0	0	0	0	80,92	1904	1984,92

Table 4 – The results of the calculations

So, we may choose these two and add the tariff that is actually used (Eurotel Gold Max) and create the data table for 100 (or 1000 or more) experiments – it means that Excel according to the given formulas generates minutes called and calculates the total costs, everything 100 times. Than we calculate the basic statistical characteristics (mean, minimum, maximum, median, percentiles, etc.). The functions in Excel are called nearly the same: MEAN[2], MIN, MAX, MEDIAN, PERCENTIL. So the short form of Excel table that solves this problem is on Figure 1. We may see that for the given situation it is

[2] The function MEAN is called PRŮMĚR in the Czech version

better to use tariff “T-mobile T80 HIT”, because all the statistical characteristics are better in comparisons with other two tariffs.



	A	B	C	D	E	F	G	H	I
2		Number of minutes called per month			Generated minutes				
3	Calls	distribution	parameters (min, max)						
4	T-mobile	uniform	60	120	66,0225858				
5	Eurotel	uniform	30	90	87,9049068				
6	Vodafone	uniform	30	100	34,8959526				
7	Local calls	uniform	15	30	27,631794				
8	sms	uniform	50	80	67				
9				Total calls	216,455239				
10									
11									
12	Tariffs		Monthly	Free minutes	Own Calls	Other Calls	sms		
13	T-mobile	T80 HIT	536	80	3,33	4,28	1,19		
14	Eurotel	Gold Max	1059	200	2,98	4,76	1,19		
15	Vodafone	300 Naplno	1190	300	4,17	4,17	1,19		
16									
17									
18	Calculations		T-mobile	Eurotel	Vodafone	Local calls	sms	Monthly	Total
19	T-mobile	T80 HIT	226,12334	63,7546773	32,6640783	201,16	79,73	536	1139,432
20	Eurotel	Gold Max	180,427357	0	0	80,92	79,73	1059	1400,077
21	Vodafone	300 Naplno	53,8134615	0	0	0	79,73	1190	1323,543
22									
23		T80 HIT	Gold Max	300 Naplno					
24	From 100 experiments	T-mobile	Eurotel	Vodafone					
25	Mean	1124,42	1297,86	1288,97					
26	Minimum	905,84	1125,78	1249,50					
27	Maximum	1271,97	1400,08	1364,71					
28	Median	1160,35	1329,38	1275,55					
29	Percentil 0.25	1038,23	1261,21	1261,40					
30	Percentil 0.75	1188,95	1350,22	1306,35					

List1 List2 List3

Figure 1 – Simple simulation model in MS Excel

5.2 Using Crystal Ball for more complicated functions

The key to Monte Carlo simulation is generating the set of random inputs. If we have more information about the calls made or if we are sure what distribution should be used to generate number of minutes or number of phone calls, it is better to use some Excel add-in – Crystal Ball is one of the well known.

Now, let us suppose to solve a problem for a client that does not call much (only 10-40 calls per month), but sometimes the call lasts 40 minutes. The presumptions are that the distribution for the length of the call should be triangular with the minimum, likeliest and maximum values set to 0.5, 5 and 40 minutes, respectively. We may have the same model (table) as before, but define which values are uncertain (which cells) and what distribution should be used for each uncertain variable. For example we suppose that the length of each call has the triangular distribution with minimum 0,5 minute, maximum 40 minutes, and likeliest 5 minutes – see Figure 2 (we have to click in the cell and then choose Define-Define Assumptions. This is the Crystal Ball extra menu). Than we identify which forecast should be analyzed (the cell for counting the total costs for the given tariff) and than the simulation (usually for 1000 experiments but we may try more) may start.

The example was changed so that we may set the probability of calling to each mobile net (30% to each mobile net, the rest to the Local), set the expected number of calls per month (and set the distribution for this type of uncertain value – her discrete uniform between 10 and 40) and set the distribution of the length of the call (here triangular distribution – Figure 2). Then we may see each phone call, its length and the total price (costs) for the number of calls in one month. Crystal Ball can show only one single step (one experiment – Figure 3) or it may run a simulation (usually 1000 experiments) and after that we may see the histogram of the forecasted cell (total costs) and all the statistics necessary. If we have each list for each tariff, we obtain everything separately for all the tariffs – see Figure 4.

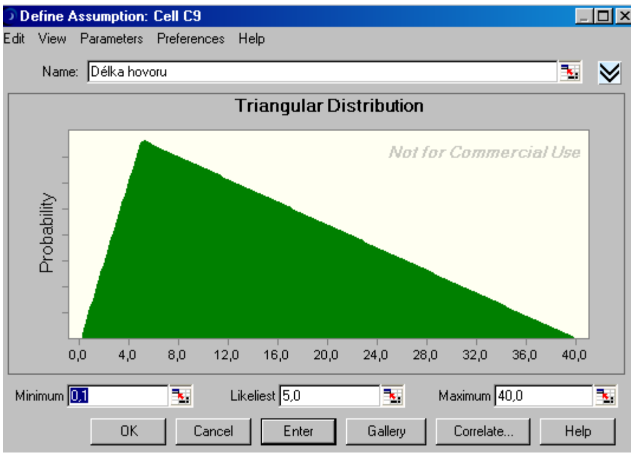
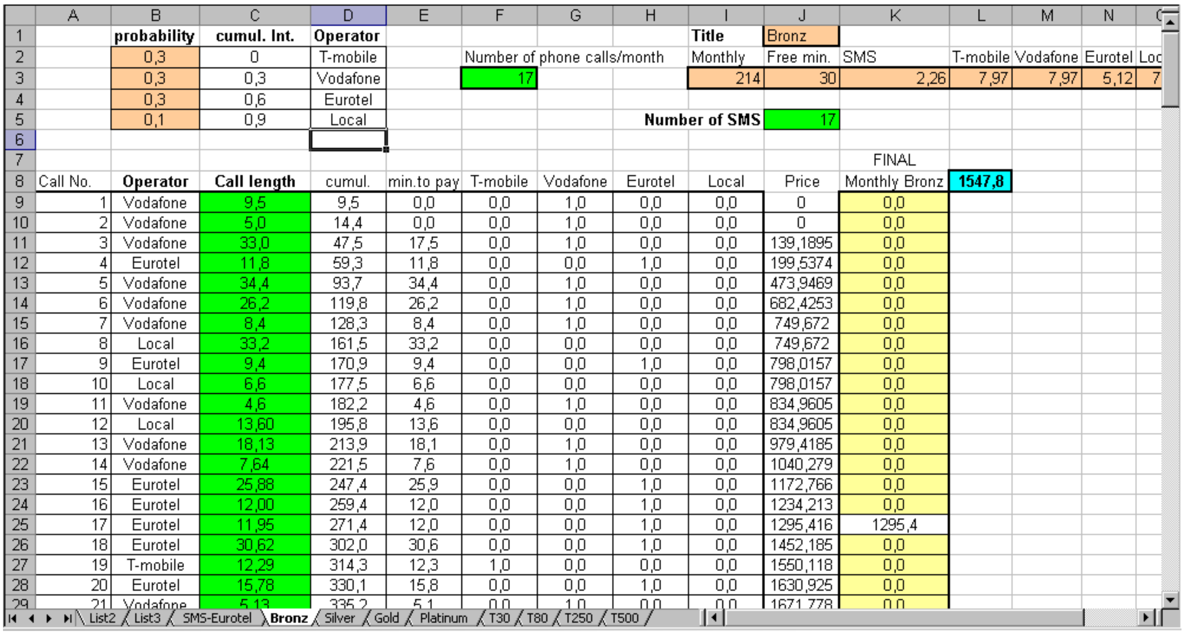


Figure 2 – Defining Assumption Cell in Crystal Ball – the length of the call



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		probability	cumul. Int.	Operator					Title	Bronz					
2		0,3	0	T-mobile		Number of phone calls/month			Monthly	Free min.	SMS	T-mobile	Vodafone	Eurotel	Loc
3		0,3	0,3	Vodafone		17			214	30	2,26	7,97	7,97	5,12	7
4		0,3	0,6	Eurotel											
5		0,1	0,9	Local				Number of SMS		17					
6															
7											FINAL				
8	Call No.	Operator	Call length	cumul.	min.to pay	T-mobile	Vodafone	Eurotel	Local	Price	Monthly Bronz	1547,8			
9	1	Vodafone	9,5	9,5	0,0	0,0	1,0	0,0	0,0	0	0,0				
10	2	Vodafone	5,0	14,4	0,0	0,0	1,0	0,0	0,0	0	0,0				
11	3	Vodafone	33,0	47,5	17,5	0,0	1,0	0,0	0,0	139,1895	0,0				
12	4	Eurotel	11,8	59,3	11,8	0,0	0,0	1,0	0,0	199,5374	0,0				
13	5	Vodafone	34,4	93,7	34,4	0,0	1,0	0,0	0,0	473,9469	0,0				
14	6	Vodafone	26,2	119,8	26,2	0,0	1,0	0,0	0,0	682,4253	0,0				
15	7	Vodafone	8,4	128,3	8,4	0,0	1,0	0,0	0,0	749,672	0,0				
16	8	Local	33,2	161,5	33,2	0,0	0,0	0,0	0,0	749,672	0,0				
17	9	Eurotel	9,4	170,9	9,4	0,0	0,0	1,0	0,0	798,0157	0,0				
18	10	Local	6,6	177,5	6,6	0,0	0,0	0,0	0,0	798,0157	0,0				
19	11	Vodafone	4,6	182,2	4,6	0,0	1,0	0,0	0,0	834,9605	0,0				
20	12	Local	13,60	195,8	13,6	0,0	0,0	0,0	0,0	834,9605	0,0				
21	13	Vodafone	18,13	213,9	18,1	0,0	1,0	0,0	0,0	979,4185	0,0				
22	14	Vodafone	7,64	221,5	7,6	0,0	1,0	0,0	0,0	1040,279	0,0				
23	15	Eurotel	25,88	247,4	25,9	0,0	0,0	1,0	0,0	1172,766	0,0				
24	16	Eurotel	12,00	259,4	12,0	0,0	0,0	1,0	0,0	1234,213	0,0				
25	17	Eurotel	11,95	271,4	12,0	0,0	0,0	1,0	0,0	1295,416	1295,4				
26	18	Eurotel	30,62	302,0	30,6	0,0	0,0	1,0	0,0	1452,185	0,0				
27	19	T-mobile	12,29	314,3	12,3	1,0	0,0	0,0	0,0	1550,118	0,0				
28	20	Eurotel	15,78	330,1	15,8	0,0	0,0	1,0	0,0	1630,925	0,0				
29	21	Vodafone	5,13	335,2	5,1	0,0	1,0	0,0	0,0	1671,778	0,0				

Figure 3 – The Excel table and Crystal Ball – one step of simulation for the tariff Eurotel Bronz

When we compare the results (Table 5), we may suggest two tariffs – "T-mobile T250" and "Vodafone 150 Naplno" because all the others have worse statistical characteristics for the given situation (histograms on Figure 5).

Once we have created the tables and defined the forecasts, it is possible to change distributions for uncertain variables – for example for different clients – and see and compare the results. In this way it is easy to create models in a form of Excel table for everyday situations when some calculations may depend upon uncertain variables.

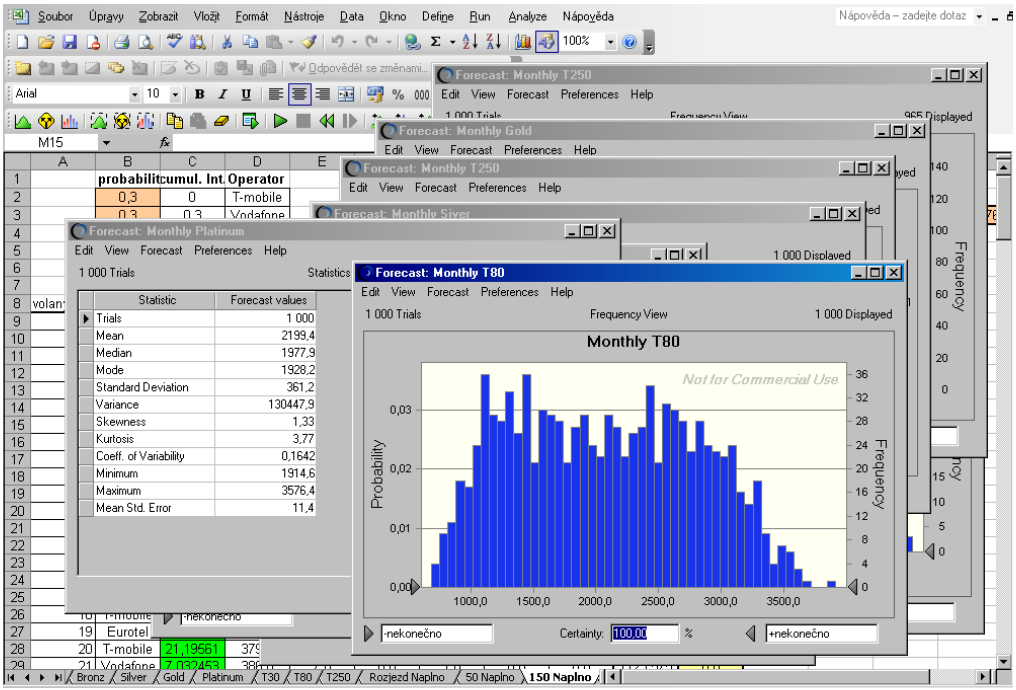


Figure 4 – The results by Crystal Ball – each window for different tariff

Tariff / Statistic	***Mean***	***Median***	***Minimum***	***Maximum***
T-mobile T30	2511	2496	708	5000
T-mobile T80	2073	2090	630	3767
T-mobile T250	**1818**	**1763**	**1195**	**3315**
Eurotel Bronz	2721	2705	737	5205
Eurotel Silver	2402	2383	713	4954
Eurotel Gold	2055	2040	1082	4200
Eurotel Platinum	2204	1975	1914	3603
Vodafone Rozjezd Naplno	2401	2427	604	4439
Vodafone 50 Naplno	2073	1999	624	4078
Vodafone 150 Naplno	**1841**	**1811**	**725**	**3813**

Table 5 – The main statistics from 1000 simulation experiments for total costs (in crowns) for chosen mobile phone tariffs

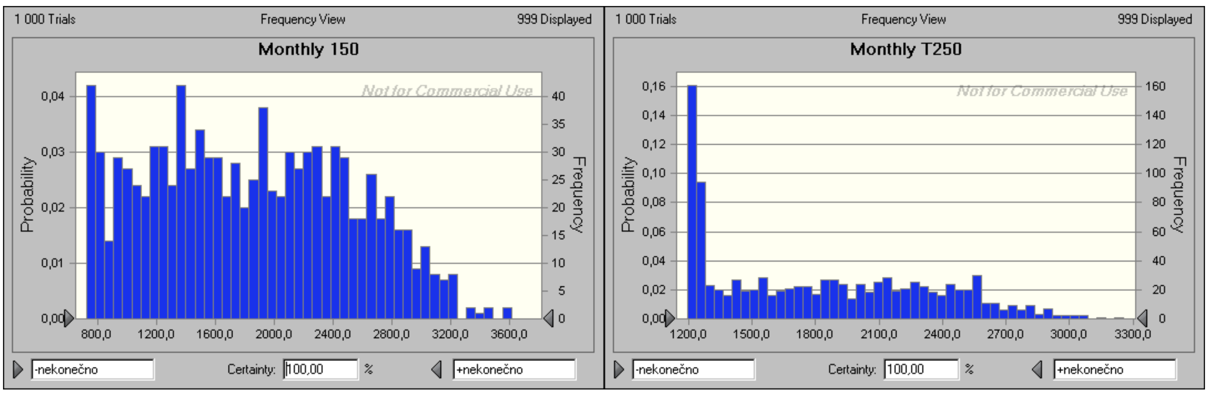


Figure 5 – The comparison of the histograms of the two best tariffs

6 Conclusion

The Monte Carlo method is another standard numerical method that is widely used not only in mathematics and physics, but also in finance to quantifying risk. In a situation where a decision is influenced by uncertainty that can be estimated by some statistical distribution, it is possible to create a simple model in MS Excel and using the Excel generator of random numbers with repeatedly counted the key function, then analyze and compare the results. The Excel is a good tool for it because we might see what may happen on one page. The Excel add-in Crystal Ball is a good solution in case we have more precisely described data or we have an imagination of what statistical distribution use to generate the corresponding numbers. It is easier to make a complicated model and try more than 100 simulation experiments in Crystal Ball than in Excel itself (because every using of the random generator in Excel in a cell slows the recomputation). Monte Carlo simulation in Excel might be an easy way how to analyze and solve complicated economic problems.

References

[1] Banks, J.: Handbook of Simulation. John Willey & Sons, USA 1998
[2] Dlouhý, M., Fábry, J., Kuncová, M.: Simulace pro ekonomy, Oeconomica, Praha 2005
[3] http://en.wikipedia.org/wiki/Monte_Carlo_method
[4] http://www.decisioneering.com/monte-carlo-simulation.html
[5] http://www.eurotel.cz
[6] http://www.t-mobile.cz
[7] http://www.vodafone.cz
[8] http://www.riskglossary.com/link/monte_carlo_method.htm
[9] http://www.vertex42.com/ExcelArticles/mc/
[10] http://www.visionengineer.com/mech/monte_carlo_example.shtml
[11] Kuncova, M.: Inventory Optimisation in a Short Supply Chain in MS Excel Using Triangular Distribution for Demand Generation. Ostrava 06.09.2005 – 08.09.2005. In: ŠTEFAN, Jan (ed.). *ASIS 2005.* Ostrava : MARQ, 2005
[12] Law, A.: Simulation Modelling and Analysis. MC-Graw Hill, Boston 2000
[13] Wittwer, J.W., "Monte Carlo Simulation in Excel: A Practical Guide" From Vertex42.com, June 1, 2004, http://vertex42.com/ExcelArticles/mc/

Computer support of courses of linear optimization models

Milada Lagová, Jana Kalčevová

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Prague 3

lagova@vse.cz, kalcevov@vse.cz

Abstract

Currently there is a number of program systems those are able to solve linear models of large dimensions at professional level. However there are not appropriate for pedagogical purposes. Therefore the system LP for teaching of linear programming algorithms is developed at Department of Econometrics, University of Economics in Prague. The system LP is built in Microsoft Access environment by reason of appropriate format of input and output screens and user's interactive contact with LP system. The system LP has two basic modules those could be used by all operational research and optimization methods courses.

Module SIMPLEXOVÁ METODA (simplex method) solves linear programming problem by simplex algorithm or dual simplex algorithm. Student can enter own data using screen, load data from database or generate unlimited number of new data using method "from result to input". He can watch solution and also steps on screen or can print solution and steps. Solutions of primary and dual problems with sensitivity analysis are presented together. In addition student can use manually controlled computational procedure without numerically difficult computation.

Module DOPRAVNÍ PROBLEM (transportation problem) searches initial solution of classical transportation problem by three methods (northwest corner method, minimum cost method and Vogel's approximation method). This module also solves given transportation problem by modified distribution method. The method for initial solution is selected from the menu and solution computation can be also selected. Inputs are given by using screen, loaded from database or generate, as above. Outputs shows optimum test, formation of closed cycles and change of solution.

Both introduced models are programmed and used in selected courses of Department of Econometrics at present. During trial run was detected that it is possible and useful to expand these modules by other linear programming methods from these courses. So four new modules are developed now – module with upper and lower bounds, module for parametric programming, module for integer programming and multicriterial programming module. All these modules are connected with SIMPLEXOVÁ METODA and it is possible to export data from this module in the others (with appropriate modification).

Keywords

Linear programming, Software, Models, Methods, Education, Teaching, Simplex Method, Transportation problem.

1 Introduction

The department of econometrics of Economic University (VŠE) in Prague makes use of computers to modernization and teaching extension already from the installation of the first computer NCR ELLIOTT 4100 in 1967. Possibilities were very limited at that time because solving systems for operation research problems were not available as well as pedagogical programs for teaching. Teachers and students together created own softwares according to their possibilities and requirements. The situation changed since that time and now the set of professional softwares exists, e.g. Lindo and Lingo, MPL, Xa or optimization modules of spreadsheets those solve especially linear programming problems. However these softwares are not pedagogically visual and mostly not user-friendly. In addition they have not implemented, except simplex method and integer linear programming, the other special methods and linear programming problems which are introduce to students of Faculty of Informatics and Statistics and other our faculties. The department of econometric used for example STORM, MORE and DS Win sooner but these softwares have limited dimension and archaic complicated control. Also the licenses ware limited and now they are invalid. Consequently we come back to our starts and try to create own pedagogical program system LinP focused on solving of linear programming (LP) problems by the simplex method and its modifications. The second focus was solving of transportation problem that is a part of most courses at our department. The LinP system comes from pedagogical and programming experiences of VŠE teachers and it is used in his beta-version in several special courses now.

2 Object and possibilities of system

The system objective is teaching support of LP and making easy to master its algorithms actively for all students. The system is constructed for calculation of optimal solution with all consecutive analyses (e.g. for simplex method they are values of structural variables, slack and surplus, reduced costs and dual prices, sensitivity analyses for right hand sides and prices, and objective value of course) and also for monitoring of iterations. Detailed statement of all iterations serves for method illustration and for eventual verification of student's calculation. In addition this system makes possible manual solution when student intervenes in simplex method and the system only transforms the simplex table. In this case the student is bound to know whole method algorithm but he is relieved of long numerical calculations. Students can solve their own examples, use examples from system database (a set of typical problems appropriate to practice of different algorithms) or generate unlimited number of examples for manual solving.

This system is not only for students but also for teachers. It can make easy to prepare examples for lessons, exercises and tests. The teacher can prepare typical examples in database (one-phase or two-phase simplex method) or on the other hand atypical examples (repeated input and output of variable in or from base, alternative optimal solution, unbounded solution of primary or dual problem, cycling of base in degenerated problem etc.). The teacher can also generate examples for lessons, homeworks and tests with given properties including numerical difficulty.

The examples those can be solved by this system correspond with authors' object, i.e. to support of teaching and not to solve practical LP problems. Therefore the problem solved by simplex method can have maximal fifteen constraints and fifteen variables (after test semester the authors extend problem dimension for examples solved for economic interpretation, problem dimension for algorithm illustration is unchanged, i.e. eight variables and eight constraints). Transportation problem has maximal eight sources and eight destinations.

3 Problem structure

Program system LinP is realized in Microsoft Access environment that is very appropriate for input and output screen's structure and interactive contact of user with LinP system.

LinP system has two basic modules those are defined for using in all courses of operation research and optimization methods. The module SIMPLEXOVÁ METODA solves LP problem by simplex method or dual simplex method according to user selection. The module DOPRAVNÍ PROBLÉM makes possible to solve an initial solution of the classical transportation problem by three methods (Northwest corner method, minimum cost method and Vogel's approximation method – VAM) and then to solve the optimal solution by modified distribution method.

Based on experiences from test run and needs of education several other special modules are proposed:

- Modified simplex method for bounded variables,
- Solving LP problems with parametric right hand sides or prices (including possibility of their postoptimal change),

- Solving LP problems with integer variables,
- Multiobjective LP problem.

All these new modules will be connected with the basic module SIMPLEXOVÁ METODA so, that the problem engaged in this module will be transformed in other selected module by extension about other necessary input data. By this way one LP problem can be solved not only by simplex method but also by others.

4 System input data

The data unit of LinP system is "the problem". The problem is defined by integer number given by system. User can enter input data by three ways. The way for input is selected by menu "Výběr úlohy" (the problem selection).

Výběr úlohy:
z obrazovky
ze souboru
generování

The first choice: After selection "z obrazovky" (from the screen) the user can see the input screen of given module. The data are sectioned into two parts:

Identification data,

Input data

Identification data are basic data characterizing LP problem, such as number, name, dimension, date etc. These dates are predefined but the user can change them (except problem number). There is a place for notes in this part also.

Input data are entered in the table for given module and it is filled by usual way. Generally all numbers can be in interval from -999 999 to 999 999 unless stated otherwise. The LinP system verified number correctness.

The data can be arbitrarily changed and solved until time of saving. At that time the system saves actual problem with given identification number. The user so can change the problem as long as he is comfortable.

The second choice: After selection „ze souboru" (from the file) the problem from database is loaded. The list of saved problems is in "Seznam úloh" (the list of problems) which is display after this choice. The list of problems contains the identification data of each problem in database. The problems are sectioned by module, so the list of problems for simplex method displays only problems solved by simplex method or dual simplex method. The list can be actualized, existing problems can be deleted, new problems can be created by copying and so can be created modifications of a problem.

Every user of LinP system has his database. In it user's problems are included. The standard typical problems LP are here and teacher can use them for lessons and student for exercises.

The third choice: The third choice for input data is their generation. The user selects properties of generated problem (dimension, input data values, the type of solution finish – i.e. unique optimal solution, alternative solution, unbounded solution or no feasible solution). The dimension is limited by three constraints and four variables (dimension for manual solving) for simplex method and eight sources and eight destinations for transportation problem. The generation runs in series with maximal 25 examples. The series is saved and the user can print the solutions of generated problems and the input of them (for tests, homeworks etc.). In the case of simplex method the way "from solution to input" is used on the base of general matrix expression of simplex table (see [1] for example). The complications grow up for choice of appropriate intervals for input data. For arbitrary numbers the generated problems are out of place for manual solving. Therefore input screen offers menu with specified values for input generated data. The system automatically solves the generated problem and drops the problem with inappropriate base determinant in whatever iteration. The user can select this limited value as he wants.

The generation of transportation problem is easy because how it is known the balanced transportation problem has the optimal solution always (see for example [2]). The user selects the balanced transportation problem or the unbalanced one with fictional source or destination and selects dimension of the problem. He can also select the method for initial solution. The system again generates the series of problems with solution and input data.

5 Calculation

In module SIMPLEXOVÁ METODA the user has two possibilities:

- Calculation by system
- Manual controlled calculation

Calculation by system is dependent on user selection of the method – simplex or dual simplex. The system automatically verifies that the problem is solvable by selected method – that for simplex method the right hand sides are nonnegative and for dual simplex method the signs of prices. If the input is correct the system adds slack and surplus (with respect to type of inequation) and artificial variables if needed and run iterations. The calculation finishes in the time of end identification – if optimal solution is found (the system notifies an alternative solution) or it is detected that the problem has unbounded or unfeasible solution. The LinP system has no protection for degeneration clearing but the number of steps is watched and in the case of cycling the process is stopped. For illustration of this problem the famous example of Beal [3] is saved in database.

Manual controlled calculation enables active interference in its run. Student selects pivot column and row and he is not checked and restricted by system. He can start in initial solution or in any iteration of solved problem. After pivot selection the user presses button "Transformace" (transformation) and the table will be recalculated. The student can continue or, in the case of mistake, go back and select new pivot. Of course the end of calculation the student selects. This operation is possible to use for presentation of different "bad approaches" – for example the bad selection of pivot row, to illustration of modified optimum test, for calculation of alternative solutions or for practice of simplex or dual simplex method without boring eliminations. It is possible to solve parametric programming problems or postoptimal analyses too.

In module DOPRAVNÍ PROBLÉM the LinP system adds the fictional source or destination and informs user about this fact on the screen. The user selects one from three selected methods for initial solution (northwest corner method, minimum cost method and Vogel's approximation method). The algorithm of each method is responded to interpretation in Department of Econometric courses. The students can so verify their own calculations. The user can choose if he wants the solution of this problem also.

6 System output data

The results can be in three types: in screen type, in print type or saved in XLS-file. The user can select outputs on screen or print outputs from the menu in bottom part of window. There are three types of result tables:

1. Výsledky (results)
2. Krokování (steps)
3. Zadání (input)

The first choice: The results in the module SIMPLEXOVÁ METODA includes all data for problem, i.e. identification and input data, optimal values of variables, values of slacks and surpluses, reduced costs and dual prices, objective value and sensitivity analyses. It is not only for optimal solution display but also for its analyses. In module DOPRAVNÍ PROBLÉM is presented balanced transportation table, input, initial and optimal solution of this problem.

The second choice: The steps show whole iteration process. For simplex method it is the series of iterations displayed in simplex tables, for transportation problem they are all found solutions with optimum test, selected input variable and closed cycle.

The third choice: The problem input is set for student's homework or test. It includes identification data of problem and student, input data in table and notes from teacher.
The same structure is used for output in MS Excel file. The students can use it for their seminar works for example. The LinP system can save the outputs in MS Word format also.
The structure is similar for all other modules of this system and student can use them for his education in special algorithms of LP. The basic module SIMPLEXOVÁ METODA is connected with other modules and the given problem can be extend to integer linear programming problem, parametric programming or multiobjective linear programming problem.

7 Conclusion

The LinP system is teaching program system for teaching support of LP in the range of Department of Econometric VŠE lessons. It solves the small problems and so it is impossible to use it commercially. LinP is an interactive system with well-arranged input data, calculation control and results outputs. The users have not problems with control of this program (it is very user-friendly) and so the beginners can use it. On the other side the students of special courses (mathematical methods in economics) use it for their practice of special algorithms and methods those are not in other softwares. The student can follow all iterations and in special cases he can influence the calculation. The teachers can use this system also for preparation of typical examples, the generation of problems with given properties for exercises, lessons and tests. The test working in this semester shows that the LinP system is quite popular for its intuitive control and user-friend environment.

References

[1] Beneš, K.: Generování úloh lineárního programování řešených simplexovou metodou, Diplomová práce, 1986.

[2] Roušarová, Š.: Generování úloh lineárního programování řešených distribuční simplexovou metodou, Diplomová práce, 1987.

[3] Lagová M., Jablonský J.: Lineární modely, VŠE, 2004.

[4] Lagová M.: Lineární modely v příkladech, VŠE, 2002.

Derivation of exchange rate computer-agent models using dynamic clearing conditions

Ladislav Lukáš
University of West Bohemia in Pilsen, Faculty of Economics,
Department of Statistics and Operations Research,
Husova 11, 30614 Plzeň
lukasl@kso.zcu.cz

Abstract

Paper presents various dynamic nonlinear models based upon market clearing conditions applied to FX rate. Building of models follow classical concept of computer-agent techniques. Basically, there are two kinds of traders being represented by computer agents applying different trading rules based upon either technical analyses or fundamental ones. The technical analysis concerns with identifications of both trends and trend reverses using more or less sofisticated procedures to predict future price movements from those of the recent past. The fundamental analysis searches and looks for various reasons and/or events thus explaining market actions in order to set up future expectations. In each trading period agents select proper trading rules to determine their speculative positions on the foreign exchange money market, thus making instant interactions between chartists and fundamentalists. The volume of international trade is neglected comparing to the volume of speculative exchange money trade transactions made by the traders. The crucial role in the models play market shares of chartists and fundamentalists expressing their instant weight on the FX money market during trading period, or more properly their ratio. This ratio is expressed as a compound function using relative distances between estimated and observed spot FX rate values as arguments. The presented models incorporate also various forms of central bank and/or money market makers influences, whose actions are introduced by computer-agents, too. Constitutive equations describing trading strategies of all interacting agents including central bank either reversing or targeting strategies are discussed in detail. These expressions form the basic terms in dynamic clearing conditions, which provide the computer-agents models considered.

Keywords

Computer-agent models, FX money market, market clearing conditions, nonlinear dynamic models.

JEL: C63, E44, E58, F31.

1 Introduction

The foreign exchange (FX) money markets attract lasting interest of both academic and applied finance families. Standard analysis of FX rate examines economic fundamentals to explain FX rate movements, but, in many cases, the fundamentals-based models fail to explain the past adequately, or predict the future reliably. On the contrary, a lot of FX money market practitioners apply various forms of technical analyses either in short or longer period quite successfully. There is well-known that technical trend signals can affect traders and price behaviour, generating excess market reactions without any fundamental reason. Hence, the instant processing of market messages plays specific role, thus leading to permanent interactions among traders. Since an empirical evidence of trading volumes in spot FX money markets shows that the overwhelming part of the turnover is merely due to short-term speculative trading transactions, one usually accepts neglecting of volumes due to an international trade transactions.

Following the classical ideas, there are typical two kinds of traders applying different trading rules based upon either technical analyses or fundamental ones. The technical analysis concerns with identifications of both trends and trend reverses using more or less sofisticated procedures to predict future price movements from those of the recent past. The fundamental analysis searches and looks for various reasons and/or events thus explaining market actions in order to set up future expectations.

Basically, these two different concepts are used for building various problem-oriented computer-based agent interaction models. The basic idea of such models is rather simple - in mathematical form to allow a self-development of FX money market under driving forces caused by interactions among computer agents which are designed to model traders', i.e. chartists' and fundamentalists', behaviour. Since these models are formulated in abstract forms, their main issues are usually, rather than prediction of actual foreign exchange market movements, studying complex dynamic effects on the base of non-linear simulation models, and investigation of possible quantitative and/or qualitative responses upon various modes of control.

For more details, the interested readers are referred to the book [2] as regards the financial market theory, to the papers [3],[4],[6],[9]-[12] as to both various aspects of technical and fundamental analyses and influences of central bank, as well, to the papers [1],[5] as for some theoretical aspects of building mathematical models in economics and finance and computer-based agent models in particular, and finally, to the papers [7],[8] to see author's contributions.

2 Clearing conditions

Based upon general concept of market equilibrium conditions we are able to formulate FX money market clearing conditions, too. We shall accept the following assumptions to hold:

i) set of trading periods is defined by $\{[t_i, t_{i+1}[\}$, where $t_0 < t_1 < t_2 < \ldots$ forms a discrete set of equidistant time values,

ii) the volume of international trade is neglected comparing to the volume of speculative exchange money trade transactions,

iii) formulation of market clearing conditions may depend upon the number of various computer agent types being involved. In order to simplify notation we introduce $t = t_i$ to denote the current trading period, while the past ones will be denoted (t-1), (t-2), etc. We shall focused upon three basic abstract forms of FX money market clearing conditions at the period t, as follows

$$m(t)\, d^C(t) + (1-m(t))\, d^F(t) = 0, \tag{1}$$

with two types of agents acting, only – i.e. chartists and fundamentalists,

$$m(t)\, d^C(t) + (1-m(t))\, d^F(t) + d^B(t) = 0, \tag{2}$$

with three agents – chartists, fundamentalists and a central bank,

$$m(t)\, d^C(t) + (1-m(t))\, d^F(t) + d^B(t) = q^M(t), \tag{3}$$

with four agents – chartists, fundamentalists, the central bank, and market makers.

Within these equations, let $d^C(t)$, $d^F(t)$, $d^B(t)$, $q^M(t)$ denote demand of chartists, fundamentalists and the central bank, and market makers quotation, respectively, at the trading period t. The market shares of chartists and fundamentalists are denoted $m(t)$ and $(1-m(t))$, respectively. These quantities also express the weight of chartists and fundamentalists on the FX money market during that period. In order to proceed further we need analytic descriptions of all used quantities.

Let the FX rate $S(t)$ represents a temporal equilibrium between two currencies available during the trading period t, as usual $m_I(t) = S(t)\, m_{II}(t)$, or in an additive form $log(m_I(t)) = log(S(t)) + log(m_{II}(t))$. We assume that $t \to S(t)$ is a real positive discrete function defined on $\{..., t-2, t-1, t, t+1, ...\}$, which actually represents a trajectory of a discrete stochastic process with continuous state space.

In general, the *chartists* deliver a buy/sell signal, if the exchange rate is rising/declining. Technical analysis is based upon evaluation of some information set available to the chartist. Since that information set is usually represented by a truncated history of past rates $\{S(t-i),\ i = 1,...,n_C+1\}$, with given integer $n_C \geq 1$, the corresponding rules are built in correspondence with an idea of moving averages or ratios. Hence, we may adopt an excess demand of chartists $d^C(t)$ to be a mix of both systematic and unsystematic components in the following form

$$d^C(t) = a^{C,1}\varphi(.) + a^{C,2}\omega(.), \tag{4}$$

where $a^{C,1}$, $a^{C,2} > 0$ represent the corresponding reaction coefficients levering both modes together.

The systematic component $\varphi(.)$ is composed in the following form

$$\varphi(.) = \sum_{(i)} \alpha_i \, log(S(t-i)/S(t-i-1)) = \sum_{(i)} \alpha_i \, (log(S(t-i) - log(S(t-i-1)), \; i = 1,..., n_C \,, \tag{5}$$

where $\alpha_i \geq 0$ denotes a relative attention paid to the lagged ratio $S(t-i)/S(t-i-1)$. There is quite natural to assume that a norm condition holds, i.e. $\sum_{(i)} \alpha_i = 1$, and further, since the most attention is paid to the recent information, and such attention usually decays in time, so that $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_{nc}$ is to hold in real conditions.

The unsystematic component $\omega(.)$ takes the form

$$\omega(.) = \delta(t-1), \tag{6}$$

where $\delta(t-1) \sim N(0,\sigma_C^2)$, i.e. a typical noise with zero mean and time invariant finite variance σ_C^2.

Finally, the formula (4) yields

$$d^C(t) = a^{C,1}(\sum_{(i)} \alpha_i \, log(S(t-i)/S(t-i-1))) + a^{C,2}\delta(t-1) \,. \tag{7}$$

In particular, for $n_C = 3$, we get the expression which has already appeared in [7].

$$d^C(t) = a^{C,1} (\alpha_1 \, log(S(t-1)/S(t-2)) + \alpha_2 \, log(S(t-2)/S(t-3)) + \alpha_3 \, log(S(t-3)/S(t-4))) + a^{C,2}\delta(t-1). \tag{8}$$

In principle, the *fundamentalists* deliver a buy/sell signal, if the expected future FX rate is above/bellow the spot rate, and any rate-related news is found to be a relatively important factor in formation of FX rate forecast. Usually, modeling of fundamentalist expectation process relies upon the fact that fundamentalists believe the exchange rate is to return to its fundamental, i.e. equilibrium value $S^F(.)$, whenever having been declined.
Hence, we may adopt an excess demand of fundamantalists $d^F(t)$ within the period t, as follows

$$d^F(t) = a^{F,1}(\mathrm{E}^F_t [S(t+1)] - S(t)) / S(t), \tag{9}$$

where $\mathrm{E}^F_t [S(t+1)]$ expresses the expected future FX rate made by fundamentalist agent at the period t, and $a^{F,1} > 0$ stands for reaction coefficient coping with relative distance between expected future FX rate and the spot rate measured by this rate.

Now, two crucial questions occur: i) how to define the quantity $\mathrm{E}^F_t \,[S(t+1)]$, and ii) how to describe the development of the fundamental value $S^F(.)$.

Obviously, the simplest way how to express such expectation formation is based upon a strict convex combination, as follows

$$\mathrm{E}^F_t [S(t+1)] = \gamma S^F(t-1) + (1-\gamma) \, S(t-1), \tag{10}$$

where γ expresses the expected adjustment speed of the exchange rate towards its fundamental value, with $\gamma \in \,]0,1[$. We should stress that at the trading period t the past value sets of merely both $S(t-i)$ and $S^F(t-i)$, $i = 1,...$, are known so that the expectation value can be calculated.
Hence, we may adopt more complicated expression of forming that expectation, which is sometimes called an anchoring heuristics

$$\mathrm{E}^F_t [S(t+1)] = \gamma S^F(t-1) + (1-\gamma) \, (\sum_{(j)} \beta_j \, S(t-j)), \tag{11}$$

where $\beta_j \geq 0$ denote norm weights of $S(t\text{-}j)$, $j = 1,...,n_F$, so that $\sum_{(j)} \beta_j = 1$ holds, with an integer n_F given. These weighted spot values may be applied by fundamentalists in order to make an individual decision making of the future FX rate.
In particular, for $n_F = 1$, $\beta_1 = 1$ we get (10), and taking $n_F = 2$, β_1 , $\beta_2 = 0.5$ it yields

$$\mathrm{E}^{\mathrm{F}}{}_{t}\,[S(t+1)] = \gamma S^{\mathrm{F}}(t\text{-}1) + (1\text{-}\gamma)\,(S(t\text{-}1) + S(t\text{-}2))\,/\,2. \tag{12}$$

As regards the development of the fundamental value $S^F(.)$ there is quite natural to assume that it is due to the rate-related news and behaves like a jump process. The logarithm of $S^F(.)$ is given by

$$logS^{\mathrm{F}}(t) = logS^{\mathrm{F}}(t\text{-}1) + a^{\mathrm{F},2}\varepsilon(t), \tag{13}$$

where $a^{F,2}$ is a coefficient representing the probability of news shock which may hit FX money market, i.e. $0 < a^{F,2} < 1$, sometimes called the jump arrival time intensity, too. For example, $a^{F,2} = 0.25$ means that a shock hits the market each 4-th trading period on average. The quantity $\varepsilon(t) \sim N(0,\sigma_F^2)$, i.e. a typical noise with zero mean and time invariant finite variance σ_F^2, again.

Further, we need both definition of time-dependent *market shares*, i.e. quantities $m(t)$ and $1\text{-}m(t)$, of chartists and fundamentalists at the trading period t, respectively. In general, these quantities should include a mode of trading rules selection, and their ranges must be]0,1] for any feasible combination of their arguments.

The market evidence shows that such rule selections depend mainly upon expected future performance possibilities, which are derived prospectively from set of past observations $\{S(t\text{-}i),\ S^F(t\text{-}j),\ i, j = 1,...\}$ being known. Hence, we may adopt the expression of $m(t)$ in the following form

$$m(t) = (1 + w(t))^{-1}\ \text{, or equivalently } w(t) = (1\text{-}m(t))/m(t) \tag{14}$$

where the values of the introduced function $w(t)$ can change within the interval $[0,+\infty[$. Following [8], the $w(t)$ is defined in form of a compound function, as follows

$$w(r(t)) = \chi_0 + (\textstyle\sum_{(k)} \chi_k\, r(t)^{2k})^{1/p}\ , \tag{15}$$

where $\chi_0 \geq 0$, and $\chi_k > 0$, $k = 1,...,n_S$ are real constants given, and an integer $n_S \geq 1$, and $p \geq 1$, too. The function $r(t)$ introduced below expresses a relative distance between $S^F(t\text{-}1)$ and $S(t\text{-}1)$, in a way similar to (9) in some sense

$$r(t) = (S^{\mathrm{F}}(t\text{-}1) - S(t\text{-}1))/S(t\text{-}1). \tag{16}$$

For sake of consistency, one should explain a seeming discrepancy between the arguments, i.e. periods t and $(t\text{-}1)$, on the left and right side, respectively. It stems from an idea to construct $r(t)$ in a way involving $\mathrm{E}^{F}{}_{t}\,[S(t+1)]$, i.e. expectation of future FX rate made at the period t, and the last spot rate, i.e. $S(t\text{-}1)$, too.
Both are supposed to have the most significant influence upon a proper selection of trading rules. Actually, substituting (10) into $(\mathrm{E}^{F}{}_{t}[S(t+1)]\text{-}S(t\text{-}1))/S(t\text{-}1)$ it yields (16). Thus, the $r(t)$ simply carries the fact that the expectation formation procedure is related to the period t.

Now, we are ready to handle the clearing condition (1). However, the clearing conditions (2) and (3) contain $d^B(t)$ and $d^M(t)$ more, respectively. As regards an empirical evidence of *central bank* influence, there is a lot of literature at disposal.
Basically, two strategies are most commonly reported:

i) reversing strategy, which tries to reverse the past FX rate trends, known also as the leaning against the wind strategy,
ii) targeting strategy, which goals to support a convergence of $S(t)$ to a target FX rate selected a priori.

Using the first strategy, the central bank always trades against the observed trends and thereby counters the action of chartists. While using the second one, the central bank buys/sells an undervalued/overvalued currency hoping to push the FX rate towards its target value, which can be for simplicity identified as the fundamental value. We also assume that the interventions take place every trading period, and that the traders, simulated by computer/based agents, are not able to identify whether a change in FX rate is triggered by the central bank or by any other rate-related factor.

The excess demand of central bank when implementing the reversing strategy may be expressed by

$$d^{B,1}(t) = a^{B,1}\,(\log S(t\text{-}2) - \log S(t\text{-}1)), \tag{17}$$

in order to stay in compatibility with (5). Implementing the targeting strategy, the $d^{B}(t)$ may be given, using (16), in the following form

$$d^{B,2}(t) = a^{B,2}\,(S^{F}(t\text{-}1) - S(t\text{-}1))/S(t\text{-}1) = a^{B,2}\,r(t), \tag{18}$$

where $a^{B,1}$, $a^{B,2}$ > 0 are given reaction coefficients of the corresponding strategies. In principle, it is possible to introduce a convex combination of both strategies

$$d^{B}(t) = \eta\, d^{B,1}(t) + (1\text{-}\,\eta)\, d^{B,1}(t), \tag{19}$$

where η expresses a leverage between both intervention strategies, with $\eta \in [0,1]$. Of course, there is evident that both expressions (17) and (18) can be generalized by introducing more entries from the set of observed past rates.

At last, we shall define the quantity $d^{M}(t)$ expressing quotation of $S(t)$ done by the *market makers* at the beginning of trading period t. Naturally, the simplest form of quotation adjustment is based upon the last known value $S(t\text{-}1)$, as follows

$$q^{M}(t) = a^{M}\,(S(t) - S(t\text{-}1))/S(t), \tag{20}$$

where $a^{M} > 0$ is a quotation adjustment coefficient, the preferable value of which is $a^{M} = 1$.

Substituting (20) into (3) yields

$$S(t) = S(t\text{-}1) + D(t)(S(t)/a^{M}), \tag{21}$$

where $D(t) = m(t)\, d^{C}(t) + (1\text{-}m(t))\, d^{F}(t) + d^{B}(t)$ denotes total excess demand. That expression renders a sense of quotation made by market makers at the period t, simply due to prospective iterations.

Finally, combining all equations in proper way we get the desired dynamic models for the FX rate $S(t)$ in correspondence with the abovementioned three various market clearing conditions.

<u>At first</u>, substituting (9) into the equation (1) yields directly

$$S(t)\, m(t)\, d^{C}(t) + (1\text{-}m(t))\, a^{F,1}(\mathrm{E}^{F}_{t}\,[S(t\text{+}1)] - S(t)) = 0, \tag{22}$$

which represents an equivalent form of the market clearing condition (1), too. Now, solving (22) for the unknown quantity $S(t)$, the resulting difference equation has the following form

$$S(t) = \mathrm{E}^{F}_{t}\,[S(t\text{+}1)]/\{1\text{-}\, d^{C}(t)/(a^{F,1}\, w(t))\}, \tag{23}$$

where the expressions (11) and (15) with all subordinated ones have to be inserted in, thus rendering selected values from the set of past observations $\{S(t\text{-}i), S^{F}(t\text{-}j), i, j = 1,...\}$ into the right hand side. Further, we need $(1\text{-}m(t)) \neq 0$ to hold, which can be simply achieved by $\chi_0 > 0$.

<u>At second</u>, substituting both (9) and (19) into the equation (2) yields

$$S(t)\, m(t)\, d^{C}(t) + (1\text{-}m(t))\, a^{F,1}(\mathrm{E}^{F}_{t}\,[S(t\text{+}1)] - S(t)) + S(t)\, d^{B}(t) = 0, \tag{24}$$

which is an equivalent form of the market clearing condition (2). Solving (24) for $S(t)$ yields

$$S(t) = \mathrm{E}^{F}_{t}\,[S(t\text{+}1)]/\{1\text{-}(m(t)\, d^{C}(t) + d^{B}(t))/(a^{F,1}(1 - m(t)))\}, \tag{25}$$

where the expressions (11), (15) and (19) with all subordinated ones have to be inserted in under the usual assumption that the denominator is not equal to zero.

At third, substituting (9), (19) and (20) into the equation (3) yields

$$S(t)\, m(t)\, d^{C}(t) + (1\text{-}m(t))\, a^{F,1}(E^{F}_{t}[S(t+1)] - S(t)) + S(t)\, d^{B}(t) - a^{M}(S(t) - S(t\text{-}1)) = 0, \qquad (26)$$

which is an equivalent form of the clearing condition (3). Now, solving (26) for $S(t)$ yields

$$S(t) = \{\, (1\text{-}m(t))E^{F}_{t}[S(t+1)] + a^{M} S(t\text{-}1) \,\}/\{\, a^{F,1}(1 - m(t)) + a^{M} - (m(t)\, d^{C}(t) + d^{B}(t))\}, \qquad (27)$$

where the expressions (11), (15) and (19) with all subordinated ones have to be inserted in, and with non-zero denominator assumed, again.

3 Conclusions

The given FX money market clearing conditions serve as a suitable platform for computer simulations to run. Some models have been already realized in OOP Java and numerical experiments reported. The others are currently under construction. However, still a lot of numerical simulations and studies are to be done. They should be focused mainly upon

i) identification and calibration of model constitutive parameters,

ii) detection and explanation of various endogenous nonlinear effects already observed and reported in the field of time-dependent FX rate development.

Acknowledgement: This research work was supported both by the grant No. 402/05/2392 of the Grant Agency of Czech Republic and by funds of the project LC06075 of the Ministry of Education of Czech Republic.

References

[1] Barnett, W. – Serletis, A.: *Martingales, Nonlinearity, and Chaos.* J.Econ. Dynamics and Control, 24 (2000), 703-724

[2] Blake, D.: *Financial Market Analysis.* (*Analýza finančních trhů*), Grada Publishing, Praha, 1995

[3] Fiess, N.M. – MacDonald, R.: *Towards the fundamentals of technical analysis*: *analysing the information content of High, Low and Close prices*. Economic Modelling, 19 (2002), 353-374

[4] LeBaron B.: *Technical trading rule profitability and foreign exchange intervention*. J.Int. Economics, 49 (1999), 125-143

[5] LeBaron B.: *Agent-based computational finance*: *Suggested readings and early research*. J.Econ. Dynamics and Control, 24 (2000), 679-7002

[6] Lui, Y.H. – Mole, D.: *The use of fundamental and technical analyses by foreign exchange dealers*: *Hong Kong evidence*. J.Int. Money and Finance, 17 (1998), 535-545

[7] Lukáš, L.: *Market clearing conditions used in FX rate simulation models*. Proceedings 22-th Int.Conf. Mathematical Methods in Economics, Masaryk University Brno, 2004, ISBN 80-210-3496-3, 184-188

[8] Lukáš, L.: *Variety of Agent-based Models for Computer Simulation of FX Rate*. 9-th Annual Conf. Computing in Economics and Finance, July 11-13, 2003, Seattle, Washington, USA

[9] Saacke, P.: *Technical analysis and the effectiveness of central bank intervention*. J.Int. Money and Finance, 21 (2002), 459-479

[10] Skouras, S.: *Financial returns and efficiency as seen by an artificial technical analyst*. J.Econ. Dynamics and Control, 25 (2001), 213-244

[11] Westerhoff, F.H.: *Speculative behavior, exchange rate volatility, and central bank intervention.* CEJOR (2001) 9: 31-50, Physica-Verlag, A Springer-Verlag Co., 2001

[12] Westerhoff, F.H.: *Central bank intervention and feedback traders.* Int. Fin. Markets, Inst. and Money, 13 (2003), 419-427

Application of Dynamic Models and an SV Machine to Inflation Modelling

Dušan Marček
The Silesian University, Faculty of Philosophy and Science, Opava, Czech Republic
and
The Faculty of Management Science and Informatics University of Žilina, Žilina, Slovak Republic

Abstract

Based on work [1] we investigate the quantifying of statistical structural model parameters of inflation in the Slovak economics. Dynamic and SVM´s (Supper Vector Machine) modelling approaches are used for automated specification of a functional form of the model in data mining systems. Based on dynamic modelling, we provide the fit of the inflation models over the period 1993-2003 in the Slovak Republic, and use them as a tool to compare their forecasting abilities with those obtained using SVM´s method. Some methodological contributions are made to dynamic and SVM´s modelling approaches in economics and to their use in data mining systems. The study discusses, analytically and numerically demonstrates the quality and interpretability of the obtained results. The SVM´s methodology is extended to predict the time series models.

Keywords

Support vector machines, data mining, learning machines, time series analysis and forecasting, dynamic modelling.
JEL: C22, C51

1 Introduction

The paper is about learning from a database. In contemporary statistical data mining systems, potential inputs are mainly chosen based on traditional statistical analysis. These include descriptive statistics, data transformations and testing. Input selection relies mostly on correlation and partial autocorrelation, cross-correlation analysis, cluster analysis, classification techniques, statistical tests and other statistical tools [3], [4], [5]. Although all these tools are in reality linear, they are deemed to provide a useful tool for the determination of the input lag structure and the selection of inputs.

In economic and finance, where our understanding of real phenomena is very poor and incomplete, it seems to be more realistic and more useful, instead of making unrealistic mathematical assumptions about functional dependency, to take the data as they are and try to represent the relationships among them in such a way that as much information as possible would be preserved. Very frequently, in such cases more sophisticated approaches are considered. Knowledge discovery in databases is a non trivial process for identifying valid, novel, potentially useful and ultimately understandable patterns in data. The usual steps in this process are data selection, data preprocessing, data transformation, data mining, evaluation and interpretation of found knowledge. Our purpose is to present quantitative procedures for use in data mining systems that routinely predict values of variables important in decision processes and to evaluate their fit to the data and forecasting abilities.

This contribution discusses determining an appropriate causal model for inflation modelling and forecasting in the Slovak Republic. We use these models as a tool to compare their fit to the data and forecasting abilities with those obtained using SVM´s method. The paper is organised as follows. In the next section of this article, we briefly describe the analysed data. Section 3 discusses building a structural model by modelling strategy described as being a "specific to general" methodology and provides a fit of the SV regression model, discusses the circumstances under which SV regression outputs are conditioned and corresponding interpretation of SV regression results is also considered. Section 6 extends the SVM´s methodology for economic time series forecasting. A section of conclusions will close the paper.

2 Analysed Data

The character of econometric models is mainly determined by theoretical economic approaches, which are more or less target used in constructing them. The period from 1993 to 1996 of the Slovak economy, in the field of econometric modelling, was a period of enormous growth of data quality and quantity. In this period at the Institute of Economics of the Slovak Academy of Science was constructed a database with approximately 160 time series of relevant macroeconomic indicators. This database contains variables relating to the following block of the economy.

Examples of some variables (for details see [7]):

- Monetary block:
  - demand deposits of enterprises,
  - insurance companies,
  - households - disposable incomes, final consumption, incomes/volumes, interest rates of time and savings deposits, interest rates of time and savings deposits, volume of time and savings deposits,
  - liquid liabilities,
  - volume of credit enterprises and households,
  - net domestic asset,
  - aggregate credit rate.
- State Budget block:
  - state budget deficit/revenues/expenditures,
  - tax (income tax of physical/legal persons), consumption taxes, customs duties,
  - expenditures (current, capital, other).
- Population block:
  - wages incomes of population (real, social, nominal).
- Block Prices and Labour productivity:
  - labour supply economically active population,
  - labour demand (employment number, increment of fixed capital),
  - unemployment (economically active population, number of unemployment rate),
  - consumer price index.
- GDP block. Its main components are:
  - private/public consumption,
  - public gross capital formation,
  - net export.
- Foreign Trade block. Its main components are:
  - exchange rates and prices of partner imports,
  - domestic production,
  - prices of production (imports and exports).

This text focuses primarily on quarterly data series. Occasionally other frequencies may be used. A part of quarter year time series of the period 1993 – 2004 was gathered together and included in a data matrix form into our information system. The data required by the forecasting system are tested for reliability and analysed to detect obvious or likely mistakes. Observations can be excluded or corrected through an appropriate management person, who may then decide whether or not to include the observation in the forecasting process. The some approach is used in analysing historical data in order to select the model form and develop the initial values for estimating the model parameters or improving forecasting performance.

3 Automated Modelling Strategy

The strategy for selecting an appropriate model is based on so called a "specific to general" methodology [2]. This strategy is well known under the common name Dynamic Modelling in Economics (DME). The DME methodology leads to two stage modelling procedure. In the first stage the researcher use simple economic theory or can incorporate some prior knowledge which might be used to formulate and estimate a model and, if found to be unsatisfactory, in the second phase is generalised until it is acceptable.

Next, we will demonstrate these phases for modelling economic time series, say inflation which may be explained by the behaviour of another variables. According to the inflation theory [8], the variable inflation is explained by the unemployment rates and wages. In this section we will present the DME approach in the modelling and investigating of the relationship between the dependent variable of inflation measured by *CPI* (Consumption Price Index) and two independent variables the unemployment rate (*U*), and aggregate wages (*W*) in the Slovak Republic. In the next section the SV regression (SVM´s method) is applied. Finally, the results are compared between a dynamic model based on statistical modelling and the SV regression model.

To study the modelling problem of the inflation quantitatively, the quarterly data from 1993Q1 to 2003Q4 was collected concerning the consumption price index *CPI*, aggregate wages *W* and unemployment *U*. These variables are measured in logarithm, among others for the reason that the original data exhibit considerable inequalities of the variance over time, and the log transformation stabilises this behaviour. Fig. 1a illustrates the time plot of the *CPI* time series. This time series shows a slight decreasing trend without apparent periodic structure. Using simple economic inflation theory the model formulation may be

$$CPI_t = \beta_0 + \beta_1 W_t + \beta_2 U_t + u_t \tag{1}$$

where u_t is a white noise disturbance term, $\beta_0, \beta_1, \beta_2$ are the model parameters (regression coefficients). Using time series data the model is estimated as

$$C\hat{P}I_t = 11.3302 - 1.355\, W_t + 1.168 U_t \qquad R^2 = 0.374, \quad \mathrm{DW} = 0.511 \tag{2}$$

Model (2) is not satisfactory. The estimated coefficients do not have the correct signs, there is evidence of first order positive autocorrelation. The model does not well fit the data inside the estimation period. R^2 is often refered loosely as the amount of variability in the data explained or accounted for by the regression model (only 37 percent of the variance in CPI_t is explained by the model).

There are various methods and criteria for automated selecting the lag structure of dynamic models from a database. As we have mentioned above, autoregression, partial autoregression and cross-correlation functions can provide powerful tools to determine the relevant structure of a dynamic model. So after the data are transformed, the first thing to do is to perform a suitable differencing of the input and output series, and to analyse autocorrelation, partial autocorrelation and cross-correlation function of the series to produce an appropriate model of the input-output series (transfer function models). In this procedure the orders of AR processes are usually determined within the procedure itself using an information criterion, e.g. Bayesian or Akaike information criterion.

Experimenting with these methods [1], the following reasonable model formulation was found

$$\hat{CPI}_t = 0.5941 - 0.0295\, W_{t-1} - 0.00359 U_{t-1} + 0.84524\, CPI_{t-1},$$
$$(0.229) \qquad (0.3387) \qquad (0.1035) \qquad R^2 = 0.7762 \qquad (3)$$

where the standard deviations of the model parameters are presented in parentheses.

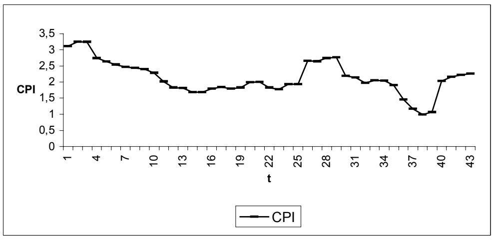


Fig. 1a Natural logarithm of quarterly inflation from January 1993 to December 2003

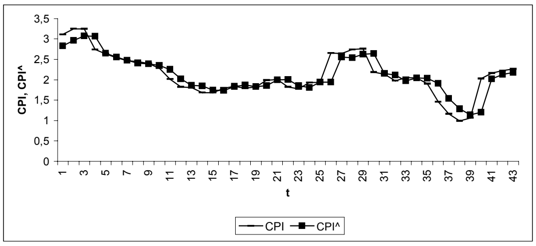


Fig. 2b Natural logarithm of actual and fitted inflation values (model (4))

Finally another attempt was made supposing a more sophisticated dependence of current inflation on the previous observation performed with the help of SV regression. As it is well known, we can not have a model in which the coefficients are statistically insignificant. We made an arbitrary decision which deleted “insignificant” explanatory variables. Then the equation (3) becomes the first-order autoregressive process, i.e.

$$\hat{CPI}_t = \beta_0 + \beta_1\, CPI_{t-1} = 0.292 + 0.856\, CPI_{t-1}. \qquad (4)$$
$$(0.158) \quad (0.072) \qquad R^2 = 0.776$$

A graph of the historical and the fitted values for inflation is presented in Fig. 1a. The model follows the pattern of the actual very closely.Statistical modelling approach based on dynamic models have found extensive practical application. These models naturally arise in areas where either a correlative or causal structure exists between variables that are temporally or spatially related. These models are also useful in many types of process and quality-control problems and everywhere, where the value of the dependent variable at time t is related to the adjustment to the controllable process variables at previous time periods t *-1*.

4 Causal Models, Experimenting with Non-linear SV Regression

In this section we will discuss the problem of selecting the appropriate functional form of the SV regression model [6]. We demonstrate here the use of the SV regression framework for estimating the model given by Eq. (4). If CPI_t exhibits a curvilinear trend, one important approach for generating an appropriate functional non-linear form of the model is to use the SV regression in which the CPI_t is regressed either against CPI_{t-1} or the time by the form

$$\hat{CPI}_t = \sum_{i=1}^{n} w_i \varphi_i(\mathbf{x}_t) + b \qquad (5)$$

where $\mathbf{x}_t = (CPI_{t-1}, CPI_{t-2}, ...)$ is the vector of time sequence of the regressor variable (regressor variable CPI_{t-1} - causal model) or $\mathbf{x}_t = (1,2,...,43)$ is the vector of time sequence (time series model), b is bias, $\varphi(.)$ is a non-linear function (kernel) which maps the input space into a high dimensional feature space, w_i are the that are subject of learning. Our next step is the evaluation of the goodness of last three regression equations to the data insite the estimation period expressed by the coefficient of determination (R^2), and the forecast summary statistics the Root Mean Square Error (RMSE) for each of the models out of the estimation period.

One crucial design choice in constructing an SV machine is to decide on a kernel. The choosing of good kernels often requires lateral thinking: many measures of similarity between inputs have been developed in different contexts, and understanding which of them can provide good kernels depends on the insight into the application´s domains. Tab.1shows SVM´s learning of the historical period illustrating the actual and the fitted values by using various kernels and presents the results for finding the proper model by using the quantity R^2. As shown in Tab. 1, the model that generate the "best" $R^2 = 0.9999$ is the time series model with the RBF kernel and quadratic loss functions. In the cases of causal models, the best R^2 is 0.9711 with the exponential RBF kernel and ε-insensitive loss function (standard deviation σ = 0.52). The choice of σ was made in response to the data. In our case, the CPI_t, CPI_{t-1} time series have σ = 0.52. The radial basis function defines a spherical receptive field in $\Re$ and the variance σ^2 localises it.

Tab. 1 SV regression results of three different choice of the kernels and the results of the dynamic model on the training set (1993Q1 to 2003Q4). In two last columns the fit to the data and forecasting performance respectively are analysed. See text for details.

Fig.	MODEL	KERNEL	σ	DEGREE-d	LOSS FUNCTION	R^2	$RMSE$
5	causal	Exp. RBF	1		ε - insensitive	0.9711	0.0915
5	causal	RBF	1		ε - insensitive	0.8525	0.0179
5	causal	RBF	0.52		ε - insensitive	0.9011	0.0995
5	causal	Polynomial		2	ε - insensitive	0.7806	0.0382
5	causal	Polynomial		3	ε - insensitive	0.7860	0.0359
5	time series	RBF	0.52		quadratic	0.9999	1.1132
4	dynamic					0.7760	0.0187

5 Conclusion

In Support Vector Machines (SVM´s), a non-linear model is estimated based on solving a Quadratic Programming (QP) problem. The use of an SV machine is a powerful tool to the solution many economic problems. It can provide extremely accurate approximating functions for time series models, the solution to the problem is global and unique.

In this paper, we have examined the SVM´s approach to study linear and non-linear models on the time series of inflation in the Slovak Republic. The benchmarking of that model was performed between traditional statistical techniques and SVM´s method in regression tasks. The SVM´s approach was illustrated on the conventional regression function. As it visually is clear from Tab. 1, this problem was readily solved by a SV regression with excellent fit of the SV regression models to the data. Tab. 1 present also forecast statistics (RMSEs) for the ex post time periods. From the Tab. 1 is shown that too many model parameters results in overfitting, i.e. a curve fitted with too many parameters follows all the small fluctuations, but is poor for generalisation. Our experience shows that SV regression models deserve to be integrated in the range of methodologies used by data mining techniques, particularly for control applications or short-term forecasting where they can advantageously replace traditional techniques.

Acknowledgment
This work was supported by the grants VEGA 1/2628/05 and GAČR 402/05/2768.

References

[1] Adamča, J., Marček, M., Pančíková, L. (2004) Some results in econometric modelling and forecasting of the inflation in Slovak economics. Journal of Economics, 52, No. 9: 1080 – 1093.

[2] Holden, K. (1997) "Developments in Dynamic Modelling in Economics", In J. Močkoř et al. editors, Proceedings of the Mathematical Methods in Economics´, Faculty of Economics, VŠB – Technical University of Ostrava (the Czech Republic): 56-61.

[3] Jensen, D. (2002) Knowledge Evaluation, In: Handbook of Data Mining and Knowledge Discovery. Oxford University Press, Oxford: 485-489.

[4] Jensen, D. D., Cohen, P. R. (2000) Multiple Comparisons in Induction Algorithms, In: Machine Learning, 38: 309-338.

[5] Kass, G. (1975) Significance Testing in Automatic Interaction Detection (A.I.D), In: Applied Statistics 24: 178-189.

[6] Kecman, V. (2001) Learning and Soft Computing – Support Vector Machines, Neural Networks and Fuzzy Logic Models. Massachusetts Institute of Technology.

[7] Páleník, V., Bors, L., Kvetan,, V., Vokoun, J. (1998) "Construction and Verification of Macroeconomic Model ISWE97q3" Journal of Economics, Vol. 46, No. 3: 428-466.

[8] Phillips, A. W. (1957) The Relation between Unemployment and the Rate of Change of Money Wages in the United Kingdom. Economica: 1861-1957.

Extended IS-LM model - construction and analysis of behavior

David Martinčík
Department of Economics and Finance, Faculty of Economics,
University of West Bohemia, Husova 11, 306 14 Plzeň
martinci@kef.zcu.cz

Blanka Šedivá
Department of Mathematics, Faculty of Applied Sciences,
University of West Bohemia, Univerzitní 22, 306 14 Plzeň
sediva@kma.zcu.cz

Abstract

The article presents an extended IS-LM model, its construction and modification, an analysis of its behavior and a statistical verification on data of the Czech economy. The extended IS-LM model is used for analyses of efficiency of macroeconomics stabilization policy in Czech Republic under period 1998-2005.

IS-LM model in various modifications always presents the standard approach of the mainstream macroeconomics to the efficiency of the stabilization policy. The fundamental diagram of the model was developed by John Richard Hicks in 1937 and it included only 3 market – the aggregate market of products and services (real sector of the economy) and the aggregate money market and the aggregate market of others financial assets (monetary sector).

The platform of extended model is connection between the real and the monetary sector by way of interest rate (keynesian transmission mechanism). The model describes an internal balance of the economy that is determined by the fiscal policy, the monetary policy and the exchange rate policy; further the policy of inflationary expectation and the characteristics of yield curve. These five determinants of aggregate product are statistically tested by the help of linear regression model including different time lags.

Keywords

IS-LM model, yield curve, expected inflation, macroeconomic stabilization policy
JEL: E12

1. IS-LM and yield curve – basic model

The IS-LM model includes at the same time two balances, the balance of aggregate market of goods and services (product) and the balance of aggregate market of money and other financial assets.
We use the linear and static model.

The IS curve describes the balance of aggregate market of product:

$$Y = \alpha\,(A - b\, i_{RL} + v\, E_{D/F} {*} P_F / P_D) \qquad (1)$$

where

Y	real gross domestic product (it is income)
α	static multiplier of autonomous expenditure (i.e. influence the autonomous expenditure on real gross domestic product) $\alpha = 1 / (1 - mpc{*}(1 - mrit))$; where *mpc* is marginal propensity to consumption and *mrit* is marginal rate of income taxation

A autonomous expenditure (private consumption and investment expenditure, public expenditure, foreign demand a import expenditure, all are independent on income, interest rate and real exchange rate)
b autonomous expenditure dependence on long term real interest rate
i_{RL} real long term interest rate
v autonomous expenditure dependence on real exchange rate
$E_{D/F}*P_F/P_D$ real exchange rate (nominal exchange rate multiple by the foreign and domestic price level ratio)

The curve LM describes the balance of aggregate market of money and other financial assets:

$$i_{NS} = 1/h\,(k\,Y - M/P_D) \tag{2}$$

where
i_{NS} nominal short term interest rate
h demand for real money dependence on short term nominal interest rate (so-called speculative demand for real money)
k demand for real money dependence on real gross domestic product (so-called transaction/income demand for real money)
M/P_D real money (nominal quantity of money divide by domestic price level)

It is needful to put in the yield curve and expected inflation between the curve IS and LM, for extension of the model. This way we transfer the short term nominal interest rate (which is determinate on aggregate market of money) to long term real interest rate (which influences the private consumption and investment expenditure).
The characteristic of the yield curve is following:

$$i_{NL} = i_{NS} + \varepsilon + \lambda + \sigma \tag{3a}$$

where
i_{NL} nominal long term interest rate
ε expectation of future nominal short term interest rate development
λ liquidity premium
σ risk premium

Through the connection of yield curve and expected inflation we obtain the long term real interest rate:

$$i_{RL} = i_{NS} + \varepsilon + \lambda + \sigma + \pi^e \tag{3b}$$

where is
π^e expected inflation

The LM curve (2) extended of yield curve and expected inflation (3b) is following:

$$i_{RL} = 1/h\,(k\,Y - M/P_D) + \varepsilon + \lambda + \sigma + \pi^e \tag{4}$$

We solve equation IS (1) and extended LM (4) and get the determination of gross domestic product (income):

$$Y = \beta\,A + \beta\,b/h\,M/P_D + \beta\,v\,E_{D/F}*P_F/P_D + \beta\,b\,\pi^e - \beta\,b\,(\varepsilon + \lambda + \sigma) \tag{5a}$$

where $\beta = \alpha h/(h + \alpha bk)$

we replace

$\beta_1 = \beta$	autonomous expenditure multiplier, it is fiscal policy multiplier
$\beta_2 = \beta\ b/h$	real money multiplier, it is monetary policy multiplier
$\beta_3 = \beta\ v$	real exchange rate multiplier
$\beta_4 = \beta\ b$	multiplier of expected inflation
$\beta_5 = \beta\ b\ (\varepsilon + \lambda + \sigma)$	equation term containing the yield curve
$R = E_{D/F} * P_F/P_D$	real exchange rate

so we get the equation:

$$Y = \beta_1 A + \beta_2 M/P_D + \beta_3 R + \beta_4 \pi^e + \beta_5 \qquad (5b)$$

This equation we first multiple by domestic price level (we drop the index, so $P_D = P$) and then date for period t and t-1. These two equations (for date t and t-1) we deduct and obtain the "increase version"[1]:

$$\Delta(YP) = \beta_1 \Delta(AP) + \beta_2 \Delta M + \beta_3 \Delta(RP) + \beta_4 \Delta(\pi^e P) + \beta_5 \Delta P \qquad (5c)$$

where

Δ	increase
YP	nominal gross domestic product (nominal income), i.e. $YP = Y * P$
AP	nominal autonomous expenditure, i.e. $AP = A * P$
M	nominal quantity of money
RP	real exchange rate multiplied by domestic price level, i.e. $RP = R * P$
$\pi^e P$	expected inflation multiplied by domestic price level, i.e. $\pi^e P = \pi^e * P$

The equation (5c) we could read this way:
Increase of nominal gross domestic product = fiscal policy + monetary policy + exchange rate policy multiple by price level + expected inflation multiple by price level + yield curve parameters multiple by price level (all terms are multiplied by relevant multipliers)

The equation (5c) indicates the determinants of nominal gross domestic product and is the equation of aggregate demand.
The different dating of separate variables - determinants of nominal product, is useful for time lag analysis.
The date basis was quarterly time series of Czech economy from 1998 to 2005 (period of relatively stable economics development after currency turbulences in 1997). There were 32 768 linear regress equations analyzed in MATLAB. These equations were developed from 8 possibilities by each variable. It was allowed the time lag from 0 to 6 quarters (7 possibilities), and the missing variable (+1). The total number of equations is $8^5 = 32\ 768$ then.
The equation (5c) was transformed to econometric form (accidental variable μ) and there was also added the absolute equation term β_0.

The estimated linear regress equation is following:

$$\Delta(YP_t) = \beta_0 + \beta_1 \Delta(AP_{t-n}) + \beta_2 \Delta M_{t-n} + \beta_3 \Delta(RP_{t-n}) + \beta_4 \Delta(\pi^e P_{t-n}) + \beta_5 \Delta P_{t-n} + \mu \qquad (6)$$

where are the indexes:

t	current period
$t-n$	last period, whereas n = 0, …, 6

[1] The increase version is useful in the situation, when we don't know the total autonomous expenditure in the economy. Further we use this simple abstraction: the private and foreign components of autonomous expenditure are constant and then the change of total autonomous expenditure must be caused only by public expenditure. This way the total autonomous expenditure could be replaced by the fiscal policy.

The 25 best solutions according to the determination R^2 are in the table:

Variable time lag (n)					Multiplier							Interval estimation
$\Delta(AP)$	ΔM	$\Delta(RP)$	$\Delta(\pi^e P)$	ΔP	β_0	β_1	β_2	β_3	β_4	β_5	R^2	
5	4	5	2	4	17912	0,66	-0,10	-76145	-1281	-408790	0,86	no
5	1	5	0	4	13251	0,84	0,04	-70722	1782	-359330	0,85	no
5	4	5	4	4	17114	0,60	-0,09	-62195	914	-362170	0,85	no
4	2	4	4	3	11607	1,29	0,15	-125320	-1462	-481010	0,85	no
5	4	5	0	4	15894	0,73	-0,05	-70697	1048	-351880	0,84	no
2	4	1	5	4	18777	-0,54	-0,11	-29605	3179	-332460	0,84	yes
2	3	6	4	4	19231	0,55	-0,19	-76768	2850	-275370	0,84	yes
5	2	4	6	3	15464	-0,62	0,14	-92270	-3160	-472970	0,84	yes
0	3	6	0	2	17640	-0,68	-0,25	-50684	1786	188260	0,83	yes
5	4	5	1	4	16048	0,56	-0,09	-62107	780	-278090	0,83	no
5	4	5	5	4	16892	0,62	-0,09	-58465	930	-372070	0,83	no
4	2	4	1	3	12057	1,07	0,11	-96070	898	-438850	0,83	no
4	6	4	5	3	18855	0,84	-0,16	-53918	2336	-473070	0,83	yes
2	4	5	1	4	17022	-0,42	-0,10	-41533	1517	-184900	0,83	no
4	2	4	6	3	12827	0,93	0,13	-102200	-827	-507680	0,82	no
4	2	4	2	3	12595	1,03	0,13	-101200	596	-500330	0,82	no
4	2	4	3	3	12210	1,17	0,14	-111020	-634	-508360	0,82	no
2	4	5	5	4	19051	-0,51	-0,11	-29419	1949	-363580	0,82	no
5	4	5	6	4	16724	0,68	-0,09	-69758	234	-341230	0,82	no
5	6	5	0	4	14943	0,82	-0,02	-70673	1914	-365770	0,82	no
5	3	5	0	4	14939	0,84	-0,02	-72676	1680	-375160	0,82	no
5	4	5	3	4	16685	0,66	-0,09	-68893	-77	-336320	0,82	no
5	4	5	NaN	4	16709	0,65	-0,09	-68476	NaN	-336220	0,82	no
4	4	5	6	4	19206	-0,79	-0,11	-59733	-2052	-259250	0,82	yes
5	5	5	0	4	14654	0,88	0,01	-76813	1773	-404320	0,82	no

(NaN – not a number, i.e. a variable having non-numeric value
Interval estimation „no" – one or more variables interval estimation contains zero)

The different time lags and different multipliers β_1 till β_5 (there are also negative values often) are the effect of unstability of transmissions mechanisms in the Czech economy.
The low sizes of multipliers (especially β_1 and β_2) detect the low efficiency of stabilization economics policy generally.
Next the absolute equation term β_0 is statistically significant and it means, that the growth of Czech economy is the effect of the other factors (long run growth factors), not effect of the terms of the equation of aggregate demand (short run economics policy). The increase of Czech nominal gross domestic product per quarter is between 12 and 19 billions CZK, so the annual long run trend is between 48 and 76 billions CZK and it is out of the control of stabilization policy.

2. First modification – dynamic model

The first modification of the basic model consists in its transformation in dynamic form. We replace the variables increases by theirs growth rate. The equation (5c) is transformed to dynamic form in the way that both sides of equation we divide by YP_{t-1}, further we multiple each term of right side by "suitable one" (e.g. the first term we multiple by AP_{t-1} / AP_{t-1}). This way the multipliers are changed to elasticity and the increases to percentual growth rate.

The dynamic equation is following:

$yp = \gamma_1\, ap + \gamma_2\, m + \gamma_3\, rp + \gamma_4\, \pi^e p + \gamma_5\, p$ (7a)

where

yp	nominal gross domestic product (nominal income) growth rate
ap	nominal autonomous expenditure growth rate
m	nominal quantity of money growth rate
rp	growth rate of real exchange rate multiplied by domestic price level
$\pi^e p$	growth rate of expected inflation multiplied by domestic price level
p	domestic price level growth rate = inflation
γ_1	nominal gross domestic product (yp) elasticity to autonomous expenditure
γ_2	yp elasticity to nominal quantity of money
γ_3	yp elasticity to real exchange rate multiplied by domestic price level
γ_4	yp elasticity to expected inflation multiplied by domestic price level
γ_5	yp elasticity to inflation

The estimated linear regress equation is following:

$yp_t = \gamma_0 + \gamma_1\, ap_{t-n} + \gamma_2\, m_{t-n} + \gamma_3\, rp_{t-n} + \gamma_4\, \pi^e p_{t-n} + \gamma_5\, p_{t-n} + \mu$ (7b)

where

γ_0	absolute equation term
μ	accidental variable

the indexes:

t	current period
$t-n$	last period, whereas $n = 0, \ldots, 6$

The 25 best solutions according to the determination R^2 are in following table:

Variable time lag (n)					Multiplier							Interval estimation
ap	m	rp	$\pi^e p$	p	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	R^2	
6	3	4	3	1	0,02	-0,05	-0,41	-0,16	-0,01	0,73	0,90	no
0	3	6	0	0	0,02	-0,10	-0,29	-0,08	0,01	0,57	0,90	no
3	4	4	1	0	0,02	-0,06	-0,21	-0,07	0,01	0,45	0,90	no
3	3	4	1	1	0,02	-0,06	-0,35	-0,13	0,01	0,60	0,89	no
6	3	4	0	1	0,02	-0,05	-0,38	-0,13	0,01	0,66	0,89	no
0	4	4	1	0	0,02	0,05	-0,25	-0,08	0,01	0,42	0,89	no
1	4	4	1	0	0,02	-0,05	-0,24	-0,07	0,01	0,49	0,89	no
5	4	1	4	0	0,02	-0,04	-0,25	-0,10	0,01	0,51	0,89	no
3	4	1	1	0	0,02	-0,07	-0,21	-0,07	0,01	0,52	0,89	no
3	4	1	4	0	0,02	-0,04	-0,23	-0,10	0,01	0,55	0,89	no
0	3	4	0	1	0,02	-0,04	-0,40	-0,11	0,01	0,62	0,89	no
6	3	4	5	1	0,02	-0,05	-0,43	-0,14	0,01	0,67	0,89	no
1	4	1	4	0	0,02	-0,04	-0,26	-0,09	0,01	0,55	0,89	no
3	3	4	3	1	0,02	-0,03	-0,41	-0,15	-0,01	0,71	0,88	no
4	4	1	4	0	0,02	-0,03	-0,24	-0,09	0,01	0,54	0,88	no
4	4	4	1	0	0,02	0,04	-0,23	-0,08	0,01	0,42	0,88	no
5	4	4	1	0	0,02	-0,03	-0,23	-0,07	0,01	0,44	0,88	no
0	4	1	4	0	0,02	-0,03	-0,23	-0,10	0,01	0,54	0,88	no
6	3	4	1	1	0,02	-0,03	-0,37	-0,13	0,01	0,62	0,88	no
6	4	1	4	0	0,02	0,02	-0,24	-0,10	0,01	0,53	0,88	no
0	3	6	3	0	0,02	-0,11	-0,33	-0,09	0,00	0,58	0,88	no
3	3	4	0	1	0,02	-0,03	-0,39	-0,13	0,01	0,66	0,88	no
4	3	4	0	1	0,02	-0,03	-0,40	-0,11	0,01	0,68	0,88	no
0	3	4	5	1	0,02	-0,05	-0,44	-0,12	0,01	0,62	0,88	no
2	4	1	4	0	0,02	0,01	-0,25	-0,10	0,01	0,52	0,88	no

(Interval estimation „no” – one or more variables interval estimation contains zero)

This model shows better results (determination R^2) than the basic model. The sizes of elasticities are to near to zero and the interval estimations are unreliable then. There are often negative values and it shows the unstability of transmissions mechanisms as well as the basic (statical) model. The stable estimation value shows only the absolute term – the long run growth of economy.

3. Second modification – time lag in product (convergent geometrical series of multipliers efficiency)

The efficiency of the economics policy is spread out the time and is going through the private consumption and investment expenditures.
Therefore, the stabilization policy is exogenous in the model and its efficiency describes the sum of geometrical series of the private expenditure increase.
The impact of stabilization policy on gross domestic product proceeds in exponential form:

$$\omega_k = v^{k+1} \tag{8}$$

where $v < 1$

We transform the equation (5c) to the form:

$$\sum_{k=0}^{4} \omega_k \, \Delta(YP_{t+k}) = \beta_1 \, \Delta(AP_{t-n}) + \beta_2 \, \Delta M_{t-n} + \beta_3 \, \Delta(RP_{t-n}) + \beta_4 \, \Delta(\pi^e P_{t-n}) + \beta_5 \, \Delta P_{t-n} \tag{9}$$

where the indexes:
t current period
t-n last period, whereas $n = 0, \ldots, 2$
t+k future period[2], whereas $k = 0, \ldots, 4$

Estimated econometric equation is following:

$$\sum_{k=0}^{4} v^{k+1} \, \Delta(YP_{t+k}) = \beta_0 + \beta_1 \, \Delta(AP_{t-n}) + \beta_2 \, \Delta M_{t-n} + \beta_3 \, \Delta(RP_{t-n}) + \beta_4 \, \Delta(\pi^e P_{t-n}) + \beta_5 \, \Delta P_{t-n} + \mu \tag{10}$$

The 25 best solutions according to the determination R^2 are in following table:

Variable time lag (n)					Multiplier							Interval estimation
$\Delta(AP)$	ΔM	$\Delta(RP)$	$\Delta(\pi^e P)$	ΔP	β_0	β_1	β_2	β_3	β_4	β_5	R^2	
0	2	1	0	0	9758	-0,71	-0,05	-36359	571	393990	0,81	no
0	2	1	1	0	9627	-0,58	-0,07	-24332	876	404800	0,81	no
1	2	0	1	0	8298	-0,23	-0,05	-31611	1872	410210	0,80	no
0	1	1	0	0	9177	-0,71	0,03	-42967	1191	289730	0,80	no
0	2	0	1	0	8832	-0,27	-0,07	-15927	1924	392340	0,80	no
0	2	1	2	0	9533	-0,65	-0,06	-35707	228	429010	0,80	no
0	2	1	NaN	0	9674	-0,70	-0,06	-36676	NaN	429200	0,80	no
1	2	0	1	0	10040	-0,25	-0,07	-34751	2016	436290	0,80	no
0	2	2	1	0	9260	-0,35	-0,09	-6361	1858	384870	0,79	no
0	0	1	0	0	9113	-0,69	0,01	-40573	1053	325430	0,79	no
0	2	NaN	1	0	9249	-0,41	-0,09	NaN	1915	391200	0,79	no
1	1	0	1	0	7319	-0,23	0,03	-41333	1911	351750	0,79	no

[2] If the total size of multiplier is 2, then in five “round” (this is period t+4) is the efficiency 1,9375. So it is acceptable to ignore the period t+5 and further.

0	2	0	1	0	10527	-0,26	-0,09	-19161	2097	415250	0,79	no
2	2	0	1	0	8356	-0,08	-0,05	-27931	2145	375940	0,79	no
0	2	1	1	0	11389	-0,59	-0,10	-23161	1097	426820	0,79	no
0	2	1	0	0	11549	-0,75	-0,07	-38237	690	414840	0,79	no
1	2	0	1	0	12320	-0,26	-0,10	-37909	2224	457290	0,79	no
1	0	0	1	0	7369	-0,24	0,01	-40056	1758	382390	0,79	no
2	2	0	1	0	10068	-0,07	-0,07	-30492	2305	399810	0,78	no
0	2	1	2	0	11341	-0,69	-0,09	-37886	173	457240	0,78	no
0	2	1	NaN	0	11447	-0,73	-0,09	-38620	NaN	457380	0,78	no
0	2	2	1	0	11040	-0,37	-0,11	-6079	2032	407820	0,78	no
2	1	0	1	0	7361	-0,10	0,03	-38665	2218	308890	0,78	no
0	2	NaN	1	0	11030	-0,42	-0,11	NaN	2086	413880	0,78	no
1	2	2	1	0	8871	-0,17	-0,07	-19359	1844	373050	0,78	no

(NaN – not a number, i.e. a variable having non-numeric value
Interval estimation „no" – one or more variables interval estimation contains zero)

In this model is the number of equations lower the in the previous. It was allowed the time lag from 0 to 2 quarters and it gives only four possibilities (incl. missing of variable), the number of equation is then 4^5=1024. This impacts the decreasing of attained determination R^2. It is fully according to the theory on the average higher time lag by monetary policy (ΔM) then by the fiscal policy (ΔAP). They are often contained the unstable or negative sizes of multipliers and this validate the unstability of transmissions mechanisms and low efficiency of stabilization policy over again. Only the absolute term, that includes the long run growth factors, shows stable estimates again.

References

[1] BERNANKE, B. S., WOODFORD, M., *Inflation Forecasts and Monetary Policy*. NBER Working Paper, No. 6157, 1997.

[2] DORNBUSCH, R., FISCHER, S., *Makroekonomie.* Praha, SPN, 1994.

[3] HENDRY, D. F., *Dynamic Econometrics.* Oxford, Oxford University Press, 2001.

[4] HU, Z., *The Yield Curve and Real Activity.* IMF Staff Paper, Vol. 40, No.4, International Monetary Fund, December 1993.

[5] PESARAN, M. H., WICKENS, M. R., *Handbook of Applied Econometrics, volume I: Macroeconomics.* Oxford, Blackwell Publisher Inc., 1999.

Dynamic production inventory model

Nora Mikušová [1], Veronika Miťková [2]

[1] Department of Operations Research and Econometrics, Faculty of Business Informatics,
University of Economics in Bratislava, Slovak Republic
nora.mikusova@fhi.euba.sk

[2] Department of Economics, Faculty of Social and Economic Sciences,
Comenius University in Bratislava, Slovak Republic
veronika.mitkova@fses.uniba.sk

Abstract

A lot of companies have own production inventory. In this paper we assume a dynamic model of production inventory. For solving a model of production inventory system we use a necessary condition of optimal control theory in this paper. In the first part of the paper we introduce the theory of inventory in short, and about the expenditures, which belongs to the inventory. In the next one we give some theoretical background for theory of optimal control and its necessary conditions. For dynamic model we use Hamiltonian function and the necessary condition for its extremes. With this function we can solve the dynamic problem as a static one in every time period. In the last part we present a model of production inventory. This model is a bounded control and we are minimizing the cost of production and at the same time keep the production as close as possible to its goal level. We assume one control and one state variable with one difference equation for changing state variable in time. The state variable is production inventory of the company and the control variable is production level. In the end of this paper we show what happens if some parameters of the problem are changing.

Keywords

Theory of optimal control, Production inventory, Dynamic optimization, Bounded control, Hamiltonian

JEL: C61 - Optimization Techniques; Programming Models; Dynamic Analysis

1 Introduction

Many companies use a production inventory system to manage fluctuations in consumer demand for the product. Such a system consists of manufacturing plant and finished good warehouse to store those products which are manufactured but not immediately sold. Once a product is made and put into inventory it incurs inventory holding costs of two kinds. *The cost of physically storing* the products, insuring it and likewise and *the carrying charges* of having the firm's money invested in the unsold inventory.

Having products inventory has advantages and disadvantages at the same time. Advantage could be that the company is immediately available to meet demand. Another example of products inventory is using the warehouse to store excess production during low demand periods in order to be available during high demand periods, which allows smoothing of production schedules of the manufacturing plant. Company has to think about the level of production inventory. If the level is too high, it could mount up the cost of firm. [3]

We can make a conclusion from above mentioned. High level of inventories (in production inventory, but also in inventory) freeze finances of company, it needs new manufacturing plant and warehouse. Low level of inventories could stop the production or in the case of production inventory it could make "troubles" during high demand period. Therefore companies would like to manage their level of inventory in the way of continuous production and satisfy their demand, with minimum of inventory cost. [2]

2 Theory of optimal control

In optimal control problems variables are divided into two classes: *state variables* $x(t)$ and *control variables* $u(t)$. One can see that both of this variables are functions. The movement of state variables is governed by first order differential equations.

Control problem is to find control function $u(t)$ and state functions $x(t)$ on the time interval $t_0 \leq t \leq t_1$, which satisfies the constraints and maximizes (minimizes) the objective.
The simplest control problem is to select a piecewise continuous control function $u(t)$, $t_0 \leq t \leq t_1$, which satisfyies:

$$\max \int_{t_0}^{t_1} f(t, x(t), u(t)) dt \tag{1}$$

$$x'(t) = g(t, x(t), u(t)) \tag{2}$$

$$x(t_0) = x_0, \qquad x(t_1) \textit{ free} \tag{3}$$

where $x'(t)$ is the differentiation of function $x(t)$ with respect to time t, $dx(t)/dt$. Functions f and g are assumed to be continuously differentiable functions of three independent arguments, none of which is a derivate. The *control variable* $u(t)$ must be a piecewise continuous function of time. The *state variable* $x(t)$ changes over time according to the differential equation (2) governing its movement.

The control u influences the objective (1), booth directly, through its own value, and indirectly through its impact on the evolution of the state variable x. An optimal control problem may have several state variables and several control variables. Each state variable evolves according to a differential equation. The number of state variable may be greater or smaller than the number of control variables. [3]

3 Model

The optimization problem is to balance the benefits of production smoothing versus the cost of holding inventory. We consider a factory producing a single homogenous good and having a finished goods warehouse. This company has received an order for B units of product to be delivered by time T. To state the model we define the following quantities:

- $I(t)$ inventory level at time t (state variable),
- $P(t)$ production rate at time t (control variable),
- $S(t)$ sales rate at time t represents the demand at time t (exogenous variable),
- T length of planning period,
- $I(0)$ initial inventory level at time 0,
- $I(T)$ inventory level at the end of planning period T,
- $P^{\wedge}$ production goal level,
- h inventory holding cost,
- c production cost.

Production goal level, $P^{\wedge}$, can be interpreted as the desired or most efficient level at witch we want to run the factory. With this notation we can state the conditions of the model.

The first condition is the differential equation for changing the inventory level (state variable) at time $I'(t) = P(t) - S(t)$. It says, that the inventory at time t is increased by the production rate and decreased by the sales rate.

For inventory we have initial and ending condition $I(0) = A$, $I(T) = B$, it means, that at the beginning of planning period, $t = 0$, company has A units of production inventory and at the end of this period, $t = T$, must achieve B units of production inventory.

The objective function of the model is

$$\min \int_0^T \left[hI(t) + \frac{c}{2}(P(t) - \hat{P})^2 \right] dt .$$

The interpretation of the objective is that we want to minimize the inventory cost and keep the production as close as possible to its goal level. The quadratic term $(P(t) – P^{\wedge})^2$ impose „penalties“ for having P not being close to its corresponding goal level.

The model which we solve has a following notation

$$\max \int_0^T -\left[hI(t) + \frac{c}{2}(P(t) - \hat{P})^2 \right] dt \tag{4}$$

$$I'(t) = P(t) - S(t) \tag{5}$$

$$I(0) = A \qquad I(T) = B \tag{6}$$

Hamiltonian function with its derivations (first order conditions for extreme of optimal control theory)

$$H = -hI - \frac{c}{2}(P - \hat{P})^2 + \lambda(P - S)$$

$$H_P = -c(P - \hat{P}) + \lambda \tag{7}$$

$$-H_I = h = \lambda'$$

$$H_\lambda = I' = P - S$$

For finding optimal P, I, λ, at first we integrate λ', for solving λ and this we substitute into H_P. Then we eliminate P and put it in the condition $I' = P – S$, which we again integrate for solving I.

$$\lambda' = h \qquad -c(P - \hat{P}) + \lambda = 0 \qquad I' = P - S$$

$$\lambda = ht + k_1 \qquad P = \hat{P} + \frac{\lambda}{c} \qquad I' = \hat{P} + \frac{ht + k_1}{c} - S$$

$$P = \hat{P} + \frac{ht + k_1}{c} \qquad I = \hat{P}t + \frac{ht^2}{2c} + \frac{k_1 t}{c} - St + k_2$$

k_1 and k_2 are the constants of integrating, which we find throw the boundary condition $I(0) = A$ and $I(T) = B$.

$$I(0) = 0 \rightarrow k_2 = A$$

$$I(T) = B \rightarrow \hat{P}T + \frac{hT^2}{2c} + \frac{k_1 T}{c} - ST + A = B \rightarrow k_1 = \frac{Bc}{T} - \hat{P}c - \frac{hT}{2} + Sc - \frac{Ac}{T}$$

From this we can solve $P(t)$, $I(t)$, $\lambda(t)$ which maximizes the objective function in every time period t and satisfy conditions of the model

$$I(t) = \frac{ht(t - T)}{2c} + \frac{Bt}{T} + \frac{A(T - t)}{T} \qquad P(t) = \frac{h(t - \frac{T}{2})}{c} + \frac{B - A}{T} + S$$

$$\lambda(t) = c\left(\frac{B - A}{T} - \hat{P} + S \right) + h\left(t - \frac{T}{2} \right) .$$

For next analyze consider exogenous demand of production is $S(t) = t^3 – 12t^2 + 32t + 30$, company has received an order for 50 units of his production (B = 50) delivered by 8 month (T = 8). At the beginning of planning period this company has in the warehouse 10 units of production (A = 10). Goal level of production for the company is 30 unit or production ($P^{\wedge} = 30$), cost of production is one unit ($c = 1$) and inventory holding cost is 3 unit ($h = 3$). We can solve the model with these constants, make graphic analysis of inventory, production and demand level in functionality of time. We can study movement of the optimal solution at time and in varying of one or more constant of the problem.

The optimal solution of the model for before mentioned exogenous variables is

$$I(t)=\frac{3t^2-14t+20}{2} \qquad P(t)=t^3-12t^2+35t+23 \qquad S(t)=t^3-12t^2+32t+30$$

The movement of inventory, production and demand we show graphically on figure 1:

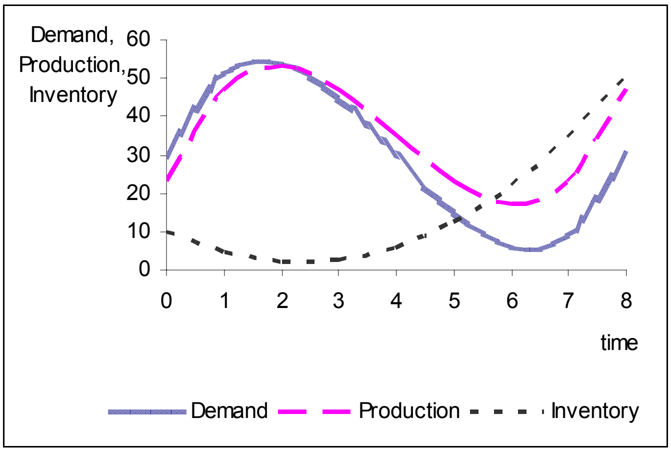


Figure 1: Demand, inventory and production level of firm in the case of order 50 units of production and low beginning inventory level

Value of the objective function for the model in general case

$$\frac{c}{2}\left[\frac{T^7}{7}-4T^6+\frac{214T^5}{5}-\frac{T^4}{2}(397+\hat{P})\right]+\frac{T^3}{6}\left[c(673+24\hat{P})+3h\right]+\frac{7T^2}{2}\left[5c(23-\hat{P})+h\right]+\frac{Tc}{2}\left[529-\hat{P}(46-\hat{P})\right]$$

The value of the objective function is 840, 96 units for the considered case. That means if the company follows this model in every period, its costs for this project are minimal.

Now we can consider change in some parameters of the problem and study the changes in the function of production, inventory, demand and in the objective function. Consider, that the subscriber needs his order in four months (not in eight like we consider in the beginning). This situation for the company we can see on the figure 2.

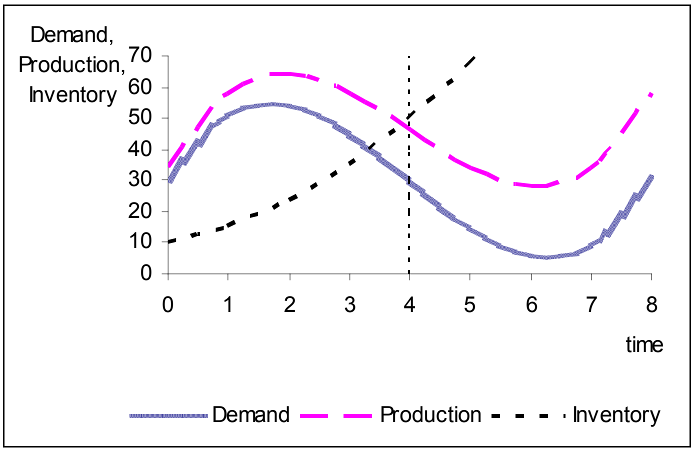


Figure 2: Demand, inventory and production level of firm in the case of order 50 units of production and shortest order time

Demand function did not change in this case, because it was imported in the model as exogenous variable. The optimal value of inventory and production function changed for this case. The inventory function is more sloping up, because we need more inventory at shorter time. The production curve and its beginning level have also changed. Company starts its production in higher level, because it needs to produce more production in shorter time – the company must be „in hurry“ with its production. Inventory costs for this case are 605, 55 units.

4 Conclusion

We showed a small dynamic production inventory model and we solved it according the first order condition of optimal control theory. In the end we changed the time parameter and showed how the optimal solution is changed. Of course, it is possible to change more or all the parameters of the model. We can say that this model is a basic one, but we can extend it in order to better cope with the real situation. We can consider more state and control variables and we also can make more first order differential equation for governing the state variables at the time and we also can add more condition for inventory and production, which does not depend on time.

References

[1] Fendek, M. *Kvantitatívna mikroekonómia*, Bratislava, IURA Edition, 1999, ISBN 80-88715-54-7

[2] Ivaničová, Z., Brezina, I. *Kvantitatívne metódy pre manažérov*, Bratislava, IURA edition, 1997, ISBN 80-88715-32-6

[3] Kamien I. M., Schawartz N. L. *Dynamic optimization: the calculus of variations and optimal control in economics and management*, Amsterdam: ELSEVIER SCIENCE B. V., 1995, ISBN 0-444-01609-0

[4] Laščiak A. a kolektív, *Optimálne programovanie,* Bratislava, ALFA, 1990, ISBN 80-05-00663-2

[5] Lincoln B. *Dynamic Programming and Time-Varying Delay Systems,* Lund, BLOMS I LUND TRZCHKERI AB, 2003, ISSN 0280-5316

Gender Gap: A Case of Some European Countries

Veronika Miťková, Nora Mikušová

Ing. Veronika Miťková: Department of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia, e-mail: veronika.mitkova@fses.uniba.sk

Ing. Nora Mikušová: Department of Operations Research and Econometrics, Faculty of Business Informatics, University of Economics in Bratislava, Slovakia, e-mail: nora.mikusova@fhi.euba.sk

Abstract

The aim of this paper is explaining a gender gap for the case of some European countries using the model of Elul, Silva-Reus and Volij (2002). Model is based on the fact that difference in earnings between men and women is a consequence of demographic fact: men marry younger women rather than a consequence of explicit discrimination in which women receive lower salary for doing precisely the same work. We calculate an optimal savings of a representative man of each country at the parameters of an economy to see whether woman marries him or not.

Keywords

Gender gap, average wage, Nash equilibrium model
JEL: C78

1 Introduction

We base the explanation of the gender gap on a demographic fact that men are allowed to marry only younger women (one at most). This is the only one difference between man and woman. The consequence of this fact is that man has an advantage in planning his professional life because he decides about his occupation and location before his spouse does so. Woman has less flexibility in choosing her location since if she decides to marry she must accept the location chosen by man.

2 Life cycle

Let us assume two regions or cities, city I and city II. For simplicity we can consider city I as one single city and city II as the rest of the world. Cities will represent a specialization in the labor market. The labor endowment is one of the two types: labor I and labor II. Set up the parameter β ($0 \le \beta \le 1$) which represents the degree of city-specificity of labor in this model. The labor I is effective in city I but only β effective in city II. Analogously, labor II is effective in city II and only β effective in city I. For example, as Elul et al. (2002) states, each city may use different language and only worker with knowledge of this language is effective in this city. Without this knowledge he is less productive and only β efficient out of the city of his type. This example holds true for Slovakia, where 10,8% of population is Hungarian nationality. For part of Hungarians without speaking Slovak language it is difficult find a job out of their city.

There exist only three periods in the life of the couple. Each agent, be the man or woman, lives for two periods and is endowed with unit of labor only when young. All agents choose their city of residence in their first period of life[1]. Let us consider the representative man. Man maximizes his income by moving to the city of his type. He is working in the period of his youth, consumes a part of his earnings and saves the rest for his old age. On the beginning of the next period a man proposes a woman to marry him. There is probability of 50% that a woman and a man are the same labor types

[1] Another way is to consider that agents are born in the city of their type.

and there is a 50% chance that they are the opposite types. In the next period a woman is born. She is immediately randomly matched to a man from the previous generation. A woman is obliged to decide whether to marry a man or not. Her decision depends on whether a man is her own type or not and on his savings. In this period of life a man retires. In the third period a woman retires and spends her savings and a man is dead.

If a woman is matched to a man of her own type, then the benefits of marriage are clear and she will accept his proposal. If she is matched to a man of different labor type, living in a city of a type different from her, then she faces a dilemma. The only benefit from a marriage in this model is that the single consumption is substituted by joint consumption. If she decides marry the man, she receives a salary corresponding to her lower productivity denoted as β. She can reject his proposal, moves to the city of her type and gets a higher salary. Her choice depends on terms of man's offer, that is his savings, taking in mind the joint consumption and her own consumption without marriage. Her decision also depends on factors of economy such as wages and interest rates. The only difference between man and woman is that a man has no information about his potential wife, whereas a woman has this information when she is young.

3 The model

Let us denote interest rate as *r*, wage as *w*, savings as *s* and consumption as *c*. The lower index denotes the period and the upper index means *M* for man, *D* for different labor types and *S* for similar labor types.

If a woman rejects a proposal of a man, they live their lives separately in the second period. Man consumes only his savings from the first period of his life with accumulated interest $\left(s^M r_2\right)$, so his utility function he wishes to maximize is $\ln\left(s^M r_2\right)$. Similarly a woman moves to the city of her type, earns wage w_2, split her income equally between consumption and savings. Her utility function is $\ln\left(w_2/2\right)+\ln\left[\left(w_2/2\right)\cdot r_3\right]$ and she wishes to maximize it.

If a woman accepts a proposal, she must decide how to divide resources between joint consumption and savings. She will accept a proposal only if the utility accruing to marriage exceeds that of remaining single. Woman's subgame is defined according to Elul et al. (2002) as

$$\begin{aligned} &\max \ln\left(c^i\right)+\ln\left(s^i r_3\right)\\ &s.t.\quad c^i \geq s^M r_2\\ &\qquad\;\; c^i + s^i = I^i + s^M r_2 \end{aligned} \tag{1}$$

and the solution of this game is:

$$\begin{aligned} c^i &= \begin{cases} \dfrac{I^i + s^M r_2}{2}, & if\ s^M r_2 < I^i \\ s^M r_2, & if\ s^M r_2 \geq I^i \end{cases}\\ s^i &= \begin{cases} \dfrac{I^i + s^M r_2}{2}, & if\ s^M r_2 < I^i \\ I^i, & if\ s^M r_2 \geq I^i \end{cases} \end{aligned} \tag{2}$$

where *I* denotes joint consumption.

The decision of a woman depends on whether she is a same labor type as man or not.

If a couple is the same labor type, a woman always accepts a proposal, since savings of a man is higher than zero. The consumption and savings of the couple are the following:

$$c^S\left(s^M, w_2, r_2\right)=\begin{cases} \dfrac{w_2 + s^M r_2}{2}, & if\ s^M r_2 < w_2 \\ s^M r_2, & if\ s^M r_2 \geq w_2 \end{cases} \tag{3}$$

$$s^S\left(s^M, w_2, r_2\right)=\begin{cases}\dfrac{w_2+s^M r_2}{2}, & \text{if } s^M r_2 < w_2 \\ w_2, & \text{if } s^M r_2 \geq w_2\end{cases} \tag{4}$$

If a couple is the different labor type, woman faces a dilemma. If the man's contribution is large enough to compensate for drop in woman's salary $s^M r_2 \geq (1-\beta)w_2$, woman accepts his proposal. Their joint consumption and saving are the following:

$$c^D\left(s^M, w_2, r_2\right)=\begin{cases} s^M r_2, & \text{if } s^M r_2 < (1-\beta)w_2 \\ \dfrac{\beta w_2 + s^M r_2}{2}, & \text{if } (1-\beta)w_2 \leq s^M r_2 \leq \beta w_2 \\ s^M r_2, & \text{if } \beta w_2 < s^M r_2 \end{cases} \tag{5}$$

$$s^D\left(s^M, w_2, r_2\right)=\begin{cases} \dfrac{w_2}{2}, & \text{if } s^M r_2 < (1-\beta)w_2 \\ \dfrac{\beta w_2 + s^M r_2}{2}, & \text{if } (1-\beta)w_2 \leq s^M r_2 \leq \beta w_2 \\ \beta w_2, & \text{if } \beta w_2 < s^M r_2 \end{cases} \tag{6}$$

A man's subgame is to maximize his expected utility $\ln\left(w_1 - s^M\right)+\frac{1}{2}\ln\left(c^S\right)+\frac{1}{2}\ln\left(c^D\right)$. The solution of this game is the optimal rate of man's savings depending on his wage, woman's wage and the interest rate. One important note is that man decides about his savings while he doesn't know what type of woman he will be matched to.

The solution is one of the four following cases:

I. If man's savings hold the condition $s^M > w_2/r_2$, a couple always marries and a woman contribute nothing to their joint consumption, it means that a woman is a free-rider.

II. If man's savings satisfy the condition $\left(\beta w_2/r_2\right) \leq s^M \leq \left(w_2/r_2\right)$, a couple always marries and when they are of the same type, woman will contribute to their joint consumption.

III. If man's savings are from the following interval $\left[(1-\beta)w_2/r_2\right] \leq s^M \leq \left(\beta w_2/r_2\right)$, a couple still marries and a woman will always contribute to their joint consumption.

IV. If man's savings are low such that $s^M \leq (1-\beta)w_2/r_2$, a couple marries only if they are the same labor type.

4 Demographic and economic facts

We are interested to demographic facts as median of age of first marriage for man and woman and the difference between the medians. The difference in medians is an underlying fact of this model as we mentioned above. To use this model there must be an age gap. Economic indicators used in this model are the average wages of both genders and an interest rate. As a wage we used average gross annual earnings in industry and services. An interest rate we considered EMU convergence criterion series - annual data.[2]

We will consider the case of six European countries: Cyprus, Germany, Hungary, Portugal, Slovakia and Sweden in year 2004. Cyprus has the highest wage gap: average wage of women is only 67,3% of average wage of men. In opposite, Sweden has the less wage gap, women earn 87,2% of wage of men. Germany has the highest average wages for both genders and also the highest age gap. Hungary has the highest interest rate. We may consider Portugal to be in somewhere in the middle odf the range according to every factor. We included Slovakia to this research because of compare it to other countries.[3]

[2] Both data sets from The Statistical Office of the European Communities (EUROSTAT).

[3] Facts on the basis of available data.

Let us set one hypothetic country. On the example of this country we will se how the model works and how the optimal solution depends on the parameters of the economy.

4.1 Slovakia

Slovakia has the lowest average wages among the countries included to this research. An average wage of men was 532 Euro and of women 385.6 Euro in 2004. A woman earns in average 72.5% wage of a man. The interest rate was 5.2% in 2004. The median age of first marriage is 27.85 years for man and 25,37 years for woman. The difference is 2.48 years, so Slovakia fulfills the condition of the model: be a man older than a woman at the first marriage.

4.2 Cyprus

The average wage in Cyprus was 1 864.5 Euro for man and 1 254.2 Euro for woman in year 2004. A women earn on average only 67.3% wage of a men. This country reaches the highest wage gap. The interest rate was 5.8% in 2004. The median age of the first marriage of men was 27.46 years and of women 25.47 years. The difference of medians was 1.99 years. It means that we can consider Cyprus for our purposes and apply the model to this country.

4.3 Germany

The average wage in Germany was 3 598.3 Euro for a man and 2 824.8 Euro for a woman in year 2004. The average wages are the highest of considered countries. Women earn on average 78.5% wage of men. The interest rate was 4.04%. Germany reaches the highest difference in medians, which is 4.06 years. One can obtain it as difference between the median age of first marriage of man (31.49 years) and woman (27.43 years).

4.4 Hungary

The average wage in Hungary was 633.8 Euro for man and 526.8 Euro for woman in year 2004. Women earn on average 83.1% wage of men. The interest rate was 8.19% in 2004, which is the highest among the considered countries. The median age of men was 27.41 years and of women 25.42 years. These demographic data are very similar to the case of Germany. The difference of medians was 1.99 years.

4.5 Portugal

We covered Portugal to our research as testing sample, because demographic and economic indicators of this country are in the middle of the range of other countries. The average wage was 1 382.9 Euro for man and 1 078.6 Euro for woman in 2004. The interest rate was 4.14%. The median age of men was 27.47 years and of women 25.42 years. The difference is 2.05 years, what underlies our model.

4.6 Sweden

Sweden is a country with the lowest wage gap. In average women earn 87.2% of earnings of men. The average wage of men was 2 898.5 Euro and of women 2 527.5 Euro. The interest rate was 4.42%. The age gap is 2.03 years, since median of the first marriage for men is 32.03 years and for women 30 years. Swedes are the oldest among other inhabitants in time of their first marriage.

The demographic data correspond with Elul et al. (2002), who states this statistics as 1.8 years in United States and 2.5 years in Europe. What matter is that there exists an age gap though very small and man has a head start at planning his professional life. This is a case for example of Slovakia: marriage occurs typically at a critical time when a man is beginning his carrier life and a woman is still studying, finishing her studies or she has just finished them.

Costa and Kahn (2000) states that university educated couples are concentrated in metropolitan areas. We suppose it is because of increasing opportunities for matching jobs for man and woman simultaneously.

4.7 Hypothetic country

Let us consider a hypothetic country with the following parameters: Average wage of men is 1 000 Euro and average wage of women is 300 Euro. The wage gap is very large, woman earns in average only 30% of average wage of man. The interest rate is given 4%. The age gap let be 2 years.

5 Applied model

For every one country we are able to construct the table with the intervals for a man within he allocates his savings. The consumptions change on the intervals. Note, that for the second and the fourth intervals is the consumption (for the same labor types and for the different labor types) the same. According to these consumptions we are able construct a man's utility function in the next step. Maximizing his utility function we find man's optimal savings. We get an optimal amount of man's savings, which differs according to the intervals mentioned earlier in the chapter 3.

We look for a perfect Nash equilibrium of the marriage game. That is, the solution must induce Nash equilibrium on each subgame (subgame of a man and subgame of a woman). There exist intervals with no solution of the marriage game for parameter of efficiency β. Such a case arises if a subgame has no Nash equilibrium (the case of Cyprus for example).

Let us consider the case that parameter β = 0.75. We choose that value according to example by Elul et at. (2002). For each country we find out a perfect Nash equilibrium defined by man's savings at given parameters of economy (wages and interest rate).

5.1 Slovakia

	intervals of man's savings	$\mathbf{c^S}$	$\mathbf{c^D}$
I.	$s^M \geq 341.24$	$1.13s^M$	$1.13s^M$
II.	$341.24\beta < s^M < 341.24$	$192.8 + 0{,}565s^M$	$1.13s^M$
III.	$341.24(1-\beta) \leq s^M \leq 341.24\beta$	$192.8 + 0{,}565s^M$	$192.8\beta + 0.565s^M$
IV.	$s^M < 341.24(1-\beta)$	$192.8 + 0{,}565s^M$	$1.13s^M$

	intervals of man's savings	**man's optimal savings**	
I.	$s^M \geq 341.24$	$s^M = 226$	doesn't hold true
II.	$341.24\beta < s^M < 341.24$	$s^M = 218.13$	hold's true for $0.37 \leq \beta \leq 0.63$
III.	$341.24(1-\beta) \leq s^M \leq 341.24\beta$	$s^M = 5.04 - 127.97 \pm \sqrt{D}/2$ [4]	hold's true for $\beta \geq 0.64$
IV.	$s^M < 341.24(1-\beta)$	$s^M = 218.13$	hold's true for $\beta \leq 0.36$

The first interval does not hold true for average Slovak man; he has never savings higher than 341.24 Euro. It means that for an average Slovak couple is never true that woman contribute nothing to the joint consumption of the household (the first interval). A couple always marries and when they are the same type woman will contribute to the joint consumption if her parameter of efficiency β is less than 0.63 and higher than 0.37 (the second interval). The third case arises if the parameter β is higher than 0.64. A couple marries if a man and a woman are the same type if parameter of efficiency is less than 0.36.

[4] D is a discriminant of quadratic equation.

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 341.24$	
II.	$255.93 < s^M < 341.24$	
III.	$85.31 \leq s^M \leq 255.93$	214.14 Euro
IV.	$s^M < 85.31$	

For the parameter β = 0.75 the Nash equilibrium is realized within the third interval at man's savings 438.17 Euro. For an average Slovak couple holds true for this case that couple always marries and a woman always contributes to the joint consumption.

5.2 Cyprus

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 1\ 119.8$	$1.12s^M$	$1.12s^M$
II.	$1.119{,}8\beta < s^M < 1.119{,}8$	$627.1 + 0.56s^M$	$1{,}12s^M$
III.	$1.119{,}8(1-\beta) \leq s^M \leq 1.119{,}8\beta$	$627.1 + 0.56s^M$	$627.1\beta + 0.56s^M$
IV.	$s^M < 1.119{,}8(1-\beta)$	$627.1 + 0.56s^M$	$1.12s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 1119.8$	$s^M = 932.25$	doesn't hold true
II.	$1119.8\beta < s^M < 1119.8$	$s^M = 770.14$	hold's true for $\beta \leq 0.69$
III.	$1119.8(1-\beta) \leq s^M \leq 1116.8\beta$	$s^M = 46.2 - 419.9\beta \pm \sqrt{D}/2$	
IV.	$s^M < 1119.8(1-\beta)$	$s^M = 770.14$	$\beta \leq 0.31$

The first interval does not hold true for average Cypriot man; he has never savings higher than 932.25 Euro. It means that for an average Cypriot couple is never true that woman contribute nothing to the joint consumption of the household. Couple always marries and when they are the same type woman will contribute to the joint consumption if her parameter of efficiency β is less than 0.69. The third case depends on the parameter β. A couple marries if a man and a woman are the same type if parameter of efficiency is less than 0.31.

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 1119.8$	
II.	$839.85\beta < s^M < 1119.8$	
III.	$279.95 \leq s^M \leq 839.85$	438.17 Euro
IV.	$s^M < 279.95$	

For the parameter β = 0.75 the Nash equilibrium is realized within the third interval at man's savings 438.17 Euro. For an average Cypriot couple holds true that couple always marries and a woman always contributes to the joint consumption.

5.3 Germany

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 2\,414.36$	$1.17s^M$	$1.17s^M$
II.	$2\,414.36\beta < s^M < 2\,414.36$	$1\,412.4 + 0.58s^M$	$1.17s^M$
III.	$2\,414.36(1-\beta) \leq s^M \leq 2\,414.36\beta$	$1\,412.4 + 0.58s^M$	$1\,412.4\beta + 0.58s^M$
IV.	$s^M < 2\,414.36(1-\beta)$	$1\,412.4 + 0.58s^M$	$1.17s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 2\,414.36$	$s^M = 1\,799.15$	doesn't hold true
II.	$2\,414.36\beta < s^M < 2\,414.36$	$s^M = 1\,467.91$	hold's true for $0.39 \leq \beta \leq 0.6$
III.	$2\,414.36(1-\beta) \leq s^M \leq 2\,414.36\beta$	$s^M = -5.81 - 92.89 \pm \sqrt{D}/2$	hold's true for $0.66 \leq \beta \leq 0.99$
IV.	$s^M < 2\,414.36(1-\beta)$	$s^M = 1\,467.91$	hold's true for $\beta \leq 0.38$

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 2\,414.36$	
II.	$1\,810.77 < s^M < 2\,414.36$	
III.	$603.59 \leq s^M \leq 1\,810.77$	758.73 Euro
IV.	$s^M < 603.59$	

5.4 Hungary

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 450.26$	$1.17s^M$	$1.17s^M$
II.	$450.26\beta < s^M < 450.26$	$263.4 + 0.585s^M$	$1.17s^M$
III.	$450.26(1-\beta) \leq s^M \leq 450.26\beta$	$263.4 + 0.585s^M$	$263.4\beta + 0.585s^M$
IV.	$s^M < 450.26(1-\beta)$	$263.4 + 0.585s^M$	$1.17s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 450.26$	$s^M = 316.9$	doesn't hold true
II.	$450.26\beta < s^M < 450.26$	$s^M = 256.91$	holds true for $\beta \leq 0.57$
III.	$450.26(1-\beta) \leq s^M \leq 450.26\beta$	$s^M = -10.4 - 168.85\beta \pm \sqrt{D}/2$	holds true for $\beta \leq 0.72$
IV.	$s^M < 450.26(1-\beta)$	$s^M = 256.91$	holds true for $\beta \leq 0.43$

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 450.26$	
II.	$337.69 < s^M < 450.26$	
III.	$112.56 \leq s^M \leq 337.69$	122.95 Euro
IV.	$s^M < 112.56$	

5.5 Portugal

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 989.54$	$1.09s^M$	$1.09s^M$
II.	$989.54\beta < s^M < 989.54$	$539.3 + 0.545s^M$	$1.09s^M$
III.	$989.54(1-\beta) \leq s^M \leq 989.54\beta$	$539.3 + 0.545s^M$	$539.3\beta + 0.545s^M$
IV.	$s^M < 989.54(1-\beta)$	$539.3 + 0.545s^M$	$1.09s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 989.54$	$s^M = 691.45$	doesn't hold true
II.	$989.54\beta < s^M < 989.54$	$s^M = 560.1$	holds true for $0.44 \leq \beta \leq 0.57$
III.	$989.54(1-\beta) \leq s^M \leq 989.54\beta$	$s^M = -25.35 - 371.08\beta \pm \sqrt{D}/2$	holds true for $0.73 \leq \beta \leq 0.99$
IV.	$S^M < 989.54(1-\beta)$	$s^M = 560.1$	holds true for $\beta \leq 0.43$

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 989.54$	
II.	$742.16 < s^M < 989.54$	
III.	$247.39 \leq s^M \leq 742.16$	256.29 Euro
IV.	$s^M < 247.39$	

5.6 Sweden

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 2\,318.8$	$1.09s^M$	$1.09s^M$
II.	$2\,318.8\beta < s^M < 2\,318.8$	$1\,263.75 + 0.545s^M$	$1.09s^M$
III.	$2\,318.8(1-\beta) \leq s^M \leq 2\,318.8\beta$	$1\,263.75 + 0.545s^M$	$1\,263.75\beta + 0.545s^M$
IV.	$s^M < 2\,318.8(1-\beta)$	$1\,263.75 + 0.545s^M$	$1.09s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 2\,318.8$	$s^M = 1\,449.25$	doesn't hold true
II.	$2\,318.8\beta < s^M < 2\,318.8$	$s^M = 1159.4$	holds true for $0.5 \leq \beta$
III.	$2\,318.8(1-\beta) \leq s^M \leq 2\,318.8\beta$	$s^M = -289.85\beta - 144.93 \pm \sqrt{D}/2$	
IV.	$s^M < 2\,318.8(1-\beta)$	$s^M = 1159.4$	holds true for $\beta \leq 0.5$

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 2\,318.8$	
II.	$1\,739.1 < s^M < 2\,318.8$	
III.	$579.7 \leq s^M \leq 1\,739.1$	986.6 Euro
IV.	$s^M < 579.7$	

The German, Hungarian, Portugal and Sweden couples situations are the same as for the case of Slovakia or Cyprus. A couple always marries and woman will always contribute to their joint consumption whatever labor type she is. The parameters of an economy don't allow a woman to be single, it is always profitable for her to accept a man's proposal.

5.7 Hypothetic country

	intervals of man's savings	c^S	c^D
I.	$s^M \geq 277.78$	$1.08s^M$	$1.08s^M$
II.	$277.28\beta < s^M < 277.78$	$150 + 0.54s^M$	$1.08s^M$
III.	$277.28(1-\beta) \leq s^M \leq 277.78\beta$	$150 + 0.54s^M$	$150\beta + 0.54s^M$
IV.	$s^M \leq 277.28(1-\beta)$	$150 + 0.54s^M$	$1.08s^M$

	intervals of man's savings	man's optimal savings	
I.	$s^M \geq 277.78$	$s^M = 500$	always holds true
II.	$277.28\beta < s^M < 277.78$	$s^M = 74.18$	doesn't hold true
III.	$277.28(1-\beta) \leq s^M \leq 277.78\beta$	$s^M = -34.72 - 104.174\beta \pm \sqrt{D}/2$	holds true for $0.5 \leq \beta \leq 0.73$
IV.	$s^M \leq 277.28(1-\beta)$	$s^M = 74.18$	holds true for $\beta \geq 0.79$

	intervals of man's savings	Nash equilibrium
I.	$s^M \geq 277.78$	500 Euro
II.	$208.33 < s^M < 277.78$	
III.	$69.44 \leq s^M \leq 208.33$	
IV.	$s^M < 69.44$	

We set the parameters of hypothetic country such that always holds true that woman marries a man and she contributes nothing to the joint consumption. This situation is caused by very large age gap where a woman earns only 30% of man's wage. For her it is profitable to marry him and he has enough savings to be a woman a free-rider. The Nash equilibrium occurs in the first interval and optimal man's savings are 500 Euro.

6 Conclusion

We based our research on the model of Elul et al. (2002). The original model is a general equilibrium model of a competitive firm and a household. We have applied a part of this model concerned to the household the case of some European counties. The results are that in every country an average man has never enough savings to be his wife a free-rider. The optimal solution for efficiency parameter 0.75 in every country is indicates that couple always marries and a woman always contributes to the joint consumption. For each country holds true that a man with an average wage has not enough savings to sustain his wife also with average wage. It means that parameters of selected European countries are similar (at least from the point of view of this model).

References

[1] Costa, D.L., Kahn, M.E. (2000) Power couples: change in the locational choice of the college educated, 1940-1990. *Quarterly Journal of Economics* 115. 1287-1315.

[2] Elul. R, Silva-Reus, J., Volij, O. (2002) Will you marry me? A perspective on the gender gap. *Journal of Economic Behavior ad Organizaton* 49. 549-572.

[3] Siow, A. (1998) Differential fecundity, markets and gender roles. *Journal of Political Economy* 106, 334-354.

Decision Support System for Portfolio Selection

Vladimír Mlynarovič
Faculty of Social and Economic Sciences, Comenius University in Bratislava Slovak Republic
vladimir.mlynarovic@fses.uniba.sk

Abstract

The developed decision support system allows in Excel environment to describe the set of efficient portfolios and analyze alternatives of investment strategy applicable to individual types of pension funds (conservative, balanced, growth) It consists of the following stages: definition of fund's investment style parameters, optimization of investment strategy, presentation of investment strategy alternatives and their description and analysis of sensitivity of investment strategy to changes in expectations. Within the definition of investment style parameters the following is specified:

a) limits for individual asset classes, namely money market instruments, bonds and equity,
b) limits also for individual assets, contemplated within asset classes,
c) currency (local, USD, EUR, SKK) of individual asset classes, to which yields of individual assets are recomputed,
d) expected return of individual assets (historical yields for the whole period, historical yields for the specific sub-period, expert values or long run equilibrium values based on Black –Litterman approach)

Optimization of investment strategy consists in effective approximation of efficient portfolios frontier in the space of expected return and risk (measured by standard deviation, lower semi – standard deviation, lower semi–absolute deviation, below target risk or conditional value at risk) and identification of four alternatives of investment strategy, portfolios, having the following attributes:

a) portfolio with global minimum risk,
b) portfolio with maximum expected return,
c) portfolio with the same value of profitability index and safety index,
d) portfolio with specified relation of profitability index and safety index.

The computed alternatives of investment strategy are presented graphically and described through a system of attributes, specifying their yield and risk-related characteristics, both from general point of view and from the point of view of individual assets selected for the portfolio.

Keywords

Decision support system, portfolio selection, pisk measures, efficient frontiers
JEL: G11

1 Portfolio selection models

Allocation of assets is a key decision in terms of achieving investment goals, and its basic responsibility is to decide on classes and proportion of assets to be included in the investment portfolios. Whereas the active shifts of asset weights as well as selection of specific securities within each asset class impact on portfolio's returns, overall performance is driven by or depends on allocation of portfolio assets, i.e. allocation of portfolio within different asset classes. It is public knowledge and it is accepted that asset allocation exercises greater influence on aggregate portfolio's returns vis-à-vis other individual decision. Experience of investors indicates that global allocation of assets or allocation in different classes of international assets is a source of the greatest differences among individual portfolios in terms of their performance. (Some pension fund-related researches document that the investment policy, i.e. strategic allocation of assets, explains 94% of total variance of actual yields, while source of the remaining 6% consists in market timing and selection of securities.)

Global asset markets provide significant opportunities for enhancement of investment returns. Development of consistent and rigorous approach to allocation of assets is a presumption for utilization of advantages related to global market investments, while consisting of the following steps:

- Selection and justification of what asset classes to be contemplated in the asset allocation task. The currency structure of a mix of assets should be reflected either in currency hedging or in currency component contemplated as a separate asset class in the process of asset allocation.
- Estimation/forecasting of risk and returns parameters of selected asset classes used in optimization models should be constructed under quantitative or qualitative models or combination of thereof.
- Building optimal portfolios at the given parameters using selected types of optimization model.
- Confirmation of candidates for optimal portfolios via testing of *sample* performance, historical simulation and *what if* analysis.

Taking a decision on strategic asset allocation (SAA) is a presumption of solving tasks related to *active asset allocation* or of such occasional shifts in asset weights, which make use of advantages ensuing from favorable markets and favorable economic conditions for specific sub-categories of the given category of assets. Historical and empiric data indicate that *tactical asset allocation can* provide an opportunity for quantifiable increase of portfolio's returns.

The process of tactical allocation may give good opportunities to increase long-term returns of the portfolio. Based on statistical experience it ensues that active changes in asset weights result, in average terms, in higher returns compared to statistic allocation. However, tactical allocation may also bring about substantial losses. If a drop of the asset's market value resulting from weak bases of the company is misinterpreted as a temporary correction, the investor may record significant capital losses, provided that the asset continues impairing. Consequently, it is inevitable to manage such a risk, which arises from deviations from strategic portfolio, namely through special limits. Therefore the key issue of tactical allocation is represented by:

5. Estimate or setting of explicit limits for admissible deviations from the set structure for asset allocation – benchmark portfolio.

1.1 Modeling efficient frontier

From technical point of view it is necessary to solve series of problems related to quadratic programming known as *Markowitz* problems of portfolio selection. The historical discussion pointed out that the efficient frontier in view of expected return and risk, measured by variance, is a suitable and elegant method of getting description of effective portfolios in modern portfolio-related theory. Hence, the presented approach is based on solution of quadratic programming tasks searching for portfolios, with which the expected portfolio return is achieved at the lowest possible risk measured at portfolio's return variance. Formally the problem can be written in the form

$$\min \quad V_P = \mathbf{w}^T \mathbf{C} \mathbf{w}$$

subject to

$$\mathbf{E}^T \mathbf{w} = E_P$$
$$\mathbf{e}^T \mathbf{w} = 1$$
$$\mathbf{w}^l \leq \mathbf{w} \leq \mathbf{w}^u$$

where

$\mathbf{C}$ - the covariance matrix, $n \times n$, where n is the number of assets,

w - the vector of portfolio weights,

E - the vector of expected returns,

E_P – expected portfolio return,

$\mathbf{w}^l$ - the vector of lower bounds on portfolio weights,

$\mathbf{w}^u$ - the vector of upper bounds on portfolio weights,

e - the vector which elements equal 1.

To provide for an effective execution of the solution process in Excel environment, specially developed *VBA procedures* were utilized, which make the formulation and solution of problem sequence by solver automatic and provide approximation of efficient frontier.

1. 2 Benchmark – strategic portfolio in profitability and safety space

The approximation of efficient portfolios frontier generally contains an indefinite number of portfolios, while one of them is selected as benchmark or strategic portfolio. It must reflect preferences of the investor and be sufficient information for decision-making in the context of the already identified effective portfolios. One of the effective possibilities is to transform the set efficient frontier in the space of "*expected return – variance"* into the space of *"safety* – profitability" and subsequently to select such portfolio from the given frontier, as having certain required characteristic, e.g. its profitability index corresponds to safety index. The problem of selecting the portfolio is examined in this part of strategic portfolio structuring as a problem of multiple-criteria optimization and in order to find the portfolio with such a special characteristic, compromise programming methods with efficient frontier in the "*profitability – safety"* space are used. The given portfolio stands for so-called *well balanced portfolio*. The approach may be modified in a way that the investor will not use the anti–ideal portfolio as a reference portfolio but instead, it will use the portfolio with specified proportion between the safety and profitability indices that can be interpreted as expression of its relation to risk. The given proportion should be different for individual types of pension funds, while the conservative fund shall have the highest value. The final portfolio is then identified as the effective portfolio with the following characteristic: its distance from the ideal portfolio towards the reference portfolio is minimal.

Formally, two small modifications of the above problem provide so called global minimum variance portfolio, with characteristics $(V_P, E_P) = (V_*, E_*)$, and portfolio with maximum expected return with characteristics $(V_P, E_P) = (V^*, E^*)$. These two portfolios restrict the efficient frontier and its approximation is achieved by solving a series of the above problems for $E_* \le E_P \le E^*$. *Ballestero and Romero* [8] define the index of profitability θ *and the* index of safety ψ, where

$$\theta = \frac{E_P - E_*}{E^* - E_*}, \quad \psi = \frac{V^* - V_P}{V^* - V_*}$$

It is clear that $0 \le \theta \le 1$, $0 \le \psi \le 1$ and (V_*, E^*), respectively (1, 1), is so called ideal portfolio in the mean - variance space, respectively in the profitability – safety space. Analogically (V^*, E_*), respectively (0, 0), is the corresponding anti-ideal portfolio. Finally, these definitions result in the approximation of a normalized efficient frontier in the space profitability - safety.

The portfolio selection problem is very often examined as a multiple criteria optimization problem and techniques of compromise programming [6, 8] are being used to find the efficient portfolio with a specified property. The set of efficient portfolios in the profitability – safety space can be formally written as the set

$$F = \{(\theta, \psi) \mid T(\theta, \psi) = 0,\ 0 \le \theta \le 1,\ 0 \le \psi \le 1\}$$

where $T(\theta,\psi) = 0$ is the efficient frontier in the profitability - safety space.

The efficient portfolio that minimise maximum deviation α from the ideal portfolio with anti-ideal portfolio as a reference portfolio can be find as the optimal solution of the problem in the form

$$\min \alpha$$

subject to

$$\alpha \geq 1-\theta$$

$$\alpha \geq 1-\psi$$

$$(\theta,\psi) \in F$$

or, equivalently, can be identify as the intercept of the line $\theta = \psi$ with the efficient frontier in the profitability - safety space. It is so called well balanced portfolio in this space.

1.3 Alternative models of portfolio selection

The model of portfolio selection in the mean – variance space, as presented and applied at strategic portfolio structuring for individual types of funds, is broadly used in fund management. It is used for allocation of assets for the purpose of setting fundamental fund management policy and also for the management of individual assets that form the portfolio, for risk management as well as for performance measurement, etc.

It is further used for specifying proportions of fund allocated to passive (index) management and for different types of active management. Its utility is determined by the following facts:

- if the rate of return has a normal distribution of probability, which was usually considered presumption fulfilled for common stock, then the model is consistent with "expected utility maximization" principle,
- quadratic programming problems, representing technical execution of the model, are solvable considering the existing knowledge of mathematical programming methodology.

Nevertheless, in recent years one can observe radical changes in investment environment. There are different financial instruments with asymmetric distribution of yield, such as options and bonds. Besides, recent statistical studies have shown that normal distribution of return is not recorded with all common stock. As a result, one can never rely on a standard model of portfolio selection.

In the past there were several risk measures proposed, different from the variance, including s*emi-absolute deviation* and below target *risk*. There are also models explicitly examining *skewness* of return distribution. Relatively new measure of lower partial *risk* comprise also *Value at Risk*, which is widely used for market risk measurement. This risk measure is very popular in conservative environment as probability of huge loss, larger than let's say $VaR_{0.99}$, is very low, provided that the portfolio's returns have normal distribution. However, considering the existing methodologies of non-linear programming it is impossible to find out a portfolio with the lowest *VaR*. For this reason the *CVaR* (*conditional value at risk* or *expected loss*) becomes more and more attractive risk measure, namely with regard to its theoretical and computing features. That is to say, it is possible to find a portfolio maximizing $CVaR_{0.99}$, which is a good approximation of portfolio combined with minimum *VaR.*

Based on the above facts the concept of strategic portfolio structuring is being gradually complemented in terms of system so that it is possible, during optimization of strategic portfolios, to apply models based on the following measures of lower partial risk:

***Lower Semi-Variance* (*Lower Semi-Standard Deviation*)**

The most well known lower partial risk the lower semi-variance of Markowitz [1] then can be written for portfolio **w** in the form

$$V_{-}(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - E(\mathbf{w})\right|_{-}^{2}\right]$$

where $R(\mathbf{w})$ is the rate of return of portfolio **w** and

$$|u|_{-} = \max\{0, -u\}$$

Lower semi-standard deviation $\sigma_{-}(\mathbf{w})$ is the square root of $V_{-}(\mathbf{w})$. These two measures are more appealing to the practitioners feelings against risk, but they were largely ignored since they are essentially the same as variance when portfolio return follows normal distribution.

Lower Semi-Absolute Deviation

Lower semi-absolute deviation of $R(\mathbf{w})$is defined as follows:

$$W_{-}(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - E(\mathbf{w})\right|_{-}\right]$$

As it is shown in Ogryczak and Ruszczynski [3] this measure is a convex function of **w** like lower semi/standard deviation of above, and most portfolios on the efficient frontier generated by mean – lower semi-absolute deviation model is consistent with the principle of *maximization of expected utility*.

Below –Target Risk

Let τ be the target rate of return. Then the below target risk of order □ is defined in the form

$$T_{B}^{\alpha}(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - \tau\right|_{-}^{\alpha}\right]^{1/\alpha}$$

One can show that this measure is again consistent with the principle of *maximization of expected utility*. The measure is also a convex function of **w** for all □.

Value at Risk and Conditional Value at Risk

VaR (value at risk) is a relatively new lower partial risk measure and it is widely used for the measurement of market risk. Let $L(\mathbf{w})$ (= $-R(\mathbf{w})$) be the loss associated with portfolio **w**. Then $VaR_{\beta}(\mathbf{w})$, $0 < \beta < 1$, is defined as the smallest number ω_{β} such that

$$\mathrm{P}\{L(\mathbf{w}) \geq \omega_{\beta}\} = 1 - \beta$$

When $R(\mathbf{w})$ follows normal distribution we seldom experience a loss exceeding $VaR_{\beta}(\mathbf{w})$, where β is over 0.95. Unfortunately it is not valid when $R(\mathbf{w})$ exhibits a fat tail distribution. Also $VaR_{\beta}(\mathbf{w})$ is not a convex function associated with portfolio **w**. Therefore $VaR_{\beta}(\mathbf{w})$ is not an adequate measure of risk for portfolio optimization.

Conditional value at risk (*CVaR*) is an alternative measure of risk which maintains advantages of *VaR*, yet free from computational disadvantages of *VaR*. The measure is defined as follows:

$$CVaR_{\beta}(\mathbf{w}) = \frac{1}{1-\beta}\mathrm{E}\left[L(\mathbf{w}) \middle| L(\mathbf{w}) \geq VaR_{\beta}(\mathbf{w})\right]$$

From the above definition one can see that $CVaR_\beta(\mathbf{w})$ is always located to the left of $VaR_\beta(\mathbf{w})$. Rockafellar and Uryasev [4] proved that $CVaR_\beta(\mathbf{w})$ can be minimized over usual set of feasible portfolios using convex minimization algorithms.

2 Developed decision support system for portfolio selection

To solve problems related to strategic, tactical and operating management of portfolio of individual funds and the whole pension funds management company an integrated system is being gradually developed in excel environment. The system is supposed to support decision-making process on the basis of systems SAP, Mathematica, Industrial Optimization and Neural Network, which make use of information sources of Bloomberg and Reuters systems. Currently the developed system allows solving the following problems.

The system allows to structure and to analyze alternatives of investment strategy applicable to individual types of funds and it consists of the following stages:

- Definition of fund's investment style parameters (Figure 1 and Figure 2)
- Optimization of investment strategy
- Presentation of investment strategy alternatives and their description
- Analysis of sensitivity of investment strategy to changes in expectations

2. 1 Definition of fund's investment style parameters

Within the definition of investment style parameters the following is specified:

1. limits for individual asset classes, namely money market instruments, bonds and equity, in form of lower and upper weight limits of assets in the portfolio,
2. limits also for individual assets, contemplated within asset classes,
3. currency (local, USD, EUR, SKK) of individual asset classes, to which yields of individual assets are recomputed,
4. expected return of individual assets (historical yields for the whole period, historical yields for the specific sub-period, expert values, long run equilibrium returns), the period of optimization and the type of returns,
5. the type of volatility (the same weighs for all period, or exponential weights),
6. the existence of risk free asset,
7. the measure of risk.

2. 2 Optimization and presentation of investment strategies

Optimization of investment strategy consists in effective approximation of efficient portfolios frontier in the space of expected return and risk (Figure 3) and identification of four alternatives of investment strategy, portfolios, having the following attributes (Figure 4):

- portfolio with global minimum risk,
- portfolio with maximum expected return,
- portfolio with the same value of profitability index and safety index,
- portfolio with specified relation of profitability index and safety index.

The calculated alternatives of investment strategy are presented graphically and described through a system of attributes, specifying their yield and risk-related characteristics, both from general point of view and from the point of view of individual assets selected for the portfolio.

Expected efinitions

Value	Fund	Profitability	Safety	Currency	Returns			
1	Conservative	0.1	0.8	SKK	Historical	Monthly	0.94	Relative
2	Balanced	0.45	0.55	LOCAL	Expected	Quarterly	0.92	Logarithmic
3	Growth	0.75	0.25	EUR	Expert	Semianually	0.9	
4				USD	Scenario	Annualy	0.88	

Volatility	Decay factor	R .F. Asset	Risk	Run		
Standard		NO	StDeviation	Short	RFR, p.a.	2.80%
Exponential	0.94	YES	CondVaR	Long	VaR CL	0.95
			Lower_SAD		Target_ret, p.a.	8.00%
			Below Target		Target_ret, p.p.	0.64%

Figure 1: Definition of investment style parameters

Lower and Upper Bounds on Asset Classes

	Money market		Bond Market		Money and Bonds		Corporations		Equity Market	
Fund	lower	upper	lower	upper	Lower	upper	Lower	upper	lower	upper
Conservative	0.1	1	0	0.9	0	1	0	0.25	0	0
Balanced	0.08	1	0	0.75	0.5	1	0	0.35	0	0.5
Growth	0.05	1	0	0.7	0.2	1	0	0.4	0	0.8

Figure 2: Limits on the portfolio weights

2. 3 What if analysis

Within the sensitivity analysis of expected return and risk of investment strategy with respect to changes in expectations, four scenarios of possible changes in expected return are examined:

- optimistic (trend of the existing positive changes in expected return in all asset classes)
- pessimistic (opposite of the optimistic scenario)
- combined, while it is presumed that the existing situation in money market will be preserved and positive changes in bond market will be accompanied by negative changes in equity market,
- combined, while it is presumed that the existing situation in money market will be preserved and negative changes in bond market will be accompanied by positive changes in equity market,

The result of analysis is the final expected change in the strategic portfolio's return and identification of portfolio, which would be optimal in the new situation. In addition, such changes in expected returns in individual asset classes have been identified, with which the expected return of strategic portfolio would drop to zero.

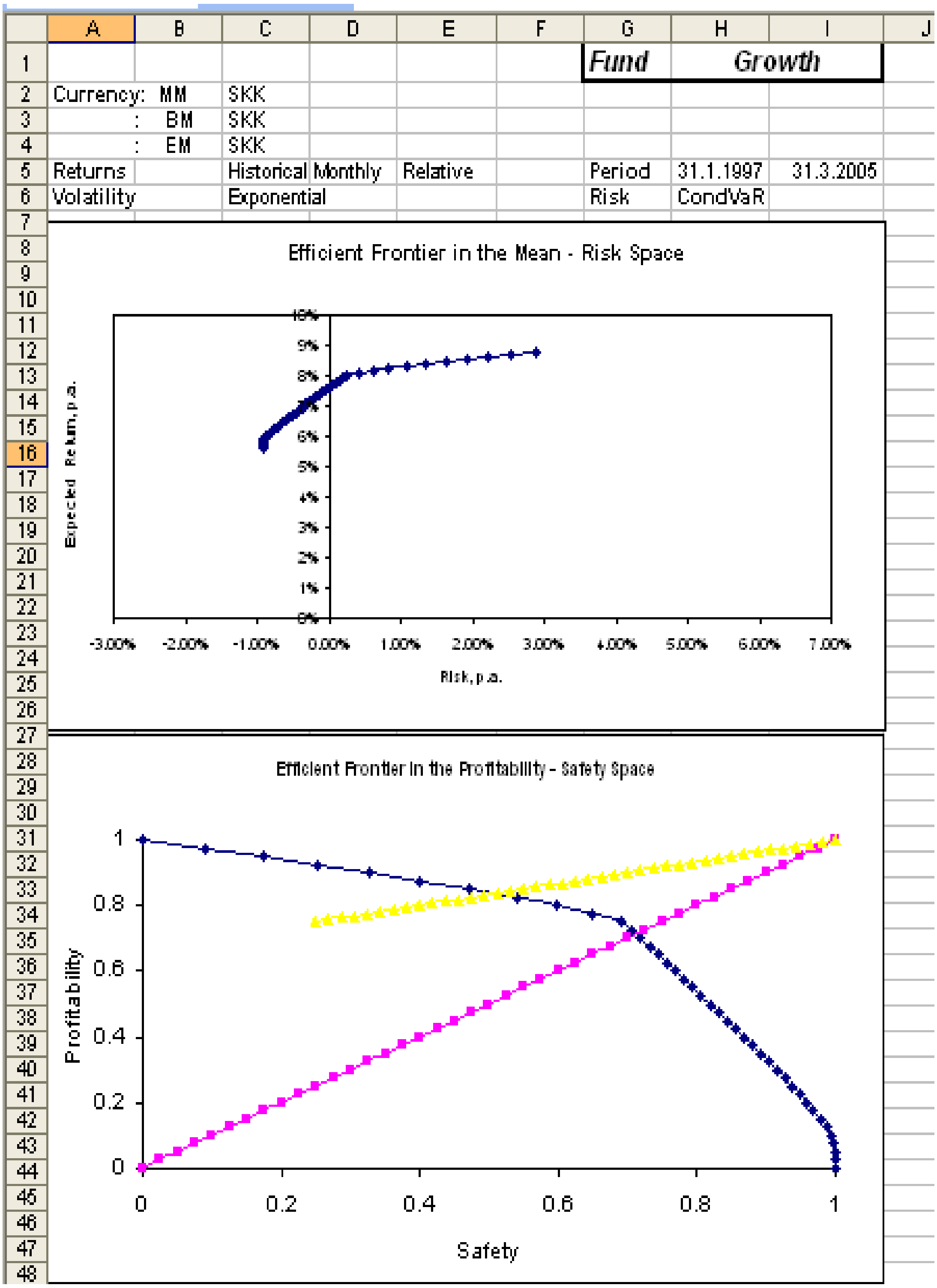


Figure 3: Approximations of the efficient frontiers

	A	B	C	D	E	F	G	H	I	J
1	Monthly	Historical	Relative					Fund		Growth
2	Exponential			Limits		Portfolio				
3	Period	31.1.1997	31.3.2005			GMR	P-S balanced	P-S structure	Maximum return	Portfolio manager
4	Expected return, p.a					5.62%	7.88%	8.28%	8.80%	5.46%
5	Risk, p.a. - CondVaR					-0.92%	0.17%	0.94%	2.88%	
6	Standard deviation,p.a.					0.67%	1.38%	1.66%	2.36%	3.08%
7	Minimum return (5% sign level) An					4.52%	5.61%	5.55%	4.92%	0.40%
8	Probability of nonpositive return An					0.00%	0.00%	0.00%	0.01%	3.79%
9	Probability for return less than 10% An					100.00%	93.80%	85.04%	69.34%	92.98%
10	Probability for return more than 20% An					0.00%	0.00%	0.00%	0.00%	0.00%
11	VaR p.a. - absolute loss		95%			-4.51%	-5.60%	-5.54%	-4.90%	-0.09%
12	Return Risk Ratio An					-1.24	-1.41	-1.49	-1.80	-58.22
13	Skewness per period					1.10	1.40	1.02	0.36	-0.65
14	Kurtosis per period					3.87	6.85	5.01	3.32	3.90
15	Dratio per period					0.00	0.00	0.00	0.02	0.43
16	Semivariance per period					0.00%	0.08%	0.13%	0.35%	0.66%
17	Potential return per period					169.01	8.23	5.26	2.22	1.06
18	Risf Free Rate per period					0.23%	0.23%	0.23%	0.23%	0.23%
19	Sharpe ratio per period					4.80	2.36	1.95	1.32	0.73
20	Index of profitability (P)					0.00%	71.22%	83.66%	100.00%	-4.90%
21	Index of safety (S)					100.00%	71.22%	50.97%	0.00%	75.05%
22	ALL MARKETS			Lower	Upper	100.00%	100.00%	100.00%	100.00%	100.00%
23	Money Market			5%	100%	84.01%	46.71%	43.23%	35.00%	45.15%
24	Currency:		BBOR3M	0%	20%	20.00%	20.00%	20.00%	20.00%	0.00%
25	SKK		BBRO6M	0%	20%	20.00%	20.00%	20.00%	15.00%	0.00%
26			BBRO9M	0%	20%	20.00%	6.71%	3.23%	0.00%	20.00%
27			BBRO1Y	0%	20%	20.00%	0.00%	0.00%	0.00%	20.00%
28			CZK3M	0%	0%	0.00%	0.00%	0.00%	0.00%	0.00%
29			CZK6M	0%	0%	0.00%	0.00%	0.00%	0.00%	0.00%
30			CZK9M	0%	0%	0.00%	0.00%	0.00%	0.00%	0.00%
31			CZK1Y	0%	0%	0.00%	0.00%	0.00%	0.00%	0.00%
32			USD3M	0%	5%	0.00%	0.00%	0.00%	0.00%	0.00%
33			USD6M	0%	5%	0.00%	0.00%	0.00%	0.00%	0.00%
34			USD9M	0%	5%	0.00%	0.00%	0.00%	0.00%	0.00%
35			USD1Y	0%	5%	0.00%	0.00%	0.00%	0.00%	0.00%
36			EUR3M	0%	5%	0.00%	0.00%	0.00%	0.00%	1.22%
37			EUR6M	0%	5%	0.00%	0.00%	0.00%	0.00%	3.93%
38			EUR9M	0%	5%	0.00%	0.00%	0.00%	0.00%	0.00%
39			EUR1Y	0%	5%	4.01%	0.00%	0.00%	0.00%	0.00%

Fund_setup / Optimize / Style report / WhatIf / Graph / VCV / Statistics / R

Figure 4: Style report

References

[1] Markowitz, H. (1959) Portfolio Selection: Efficient Diversification of Investments. John Wiley and Sons.

[2] Zeleny, M. (1982) Multiple Criteria Decision Making. McGraw Hill, New York.

[3] Ogryczak, W and A. Ruszczynski (1999) From stochastic dominance to mean – risk models: semi-deviation as risk measures. *Euro. J. Operat. Res.* **116**, 35-50.

[4] Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value at risk. J. Risk **2**, 21-41.

[5] Konno, H. - Waki , H.–A. Yuuki (2002): Portfolio Optimization under Lower Partial Risk Measures. *Asia/Pacific Financial Markets* **9**, 127-140

[6] Mlynarovič, V. (2001): Finančné investovanie. Teória a aplikácie. IURA Edition, Bratislava, ISBN 80-89047-16-5.

[7] Martellini, L. – Simsek, K. – Goltz, F. (2005):Structured Forms of Investment Strategies in Institutional Investor's Portfolios. An Edhec Risk and Asset Management Research Centre Publication. April 2005. EDHEC Bussines School, Lile – Nice.

[8] Ballestero, E. – C. Romero: Portfolio Selection: A Compromise Programming Solution. Journal of Operational Research Society 47, 1996, 1377 – 1386.

Peer group games in economics of communication networking

Alexandru Murgu
BT Networks Research Centre, Adastral Park
Ipswich, IP5 3RE, United Kingdom
alexandru.murgu@bt.com

Abstract

In communication networks there are scenarios of interaction configurations where the hierarchical information structure influences the economic outcomes of the agent groups. Economic situations such as network resource auctioning, communication channel signaling, traffic sequencing decisions and flow control actions are related to peer group games. The economic outcomes of an agent are determined by its position in the hierarchy. The important relational group for an agent in such hierarchy consists of the leader, the agent itself and the intermediate agents in the given hierarchy between the agent and the leader. The hierarchy is defined by a rooted directed tree where the leader is located in the root and each agent is a distinct node in the tree. This tree structure characterizes the peer group interactions, that is, each agent peer group corresponds to the agents in the unique path connecting the agent node with the root in the tree. The tree connected peer group occurrences are modelled as sets of agents involving the peer group structure describing the organization's social configuration and a vector that gives the individual economic possibilities of the agents. A game theory view for analyzing these constraints on the economic cooperative behavior of Internet communication traffic classes is considered in this paper.

Keywords

Communication networks, agent peer group, hierarchical organization, trees, game theory, cooperative behavior.

JEL: C44, C73, D52, D74.

1 Introduction

The Internet infrastructure is so pervasive that it is common that people are connected from a range of different locations: office, home, conference places, or meeting rooms. This is sometimes called mobile computing and it forces designers of applications to take into account two new requirements: (1) users may connect to the network from arbitrary locations (usually with different network addresses), and (2) they are not permanently connected. Thus, connectivity is intrinsically transient, and machine disconnection is the normal way of operating. Collaborative work over the Internet is often pursued by manipulating electronic artifacts that are shared with others by exploiting the mediation of some server. The client/server approach is possible and common, in all cases where a reliable and permanent network infrastructure is available in order to connect the participating nodes. However, in many cases, people would like to collaborate while they are supported through a much looser architecture, since they cannot or do not want to afford the cost of setting up a central server. The resulting reference architecture is a network of peers, each of which contributes to the overall logical structure in an equivalent way. Moreover, peers cannot be assumed to be always on-line, that is, a peer is not always reachable by others. Since the network connection is intrinsically intermittent, peers may dynamically join and leave an ad-hoc community. For example, PeerVerSy [2] is a management tool allowing the users for freely accomplishing their computations and collaborative actions as check-in and check-out also when they are not able to communicate with the machine that holds the reference copy of an electronic artifact. Concurrency control is basically optimistic and an automatic reconciliation step is performed when connection is established again, possibly arising conflicts. The client/server approach assumes the availability of the network infrastructure even in the frequent case that no concurrent work is done on a particular item. It is perfectly reasonable and desirable that one could check-in a file which is under his/her control if no other developers want to manipulate it. Similarly, check-out operations can be performed also when the latest version of an artifact is available somewhere (not necessarily on repository servers), but for example on the local file system or on the file system of any connected node. In the PeerVerSy approach there are no well known servers. Instead, it is based on the notion of the authority for a set of items (it owns them), and the copy of an artifact owned by the authority is the master copy. In addition to the master copy, other peers can keep a local copy (replica) of the documents they do not own in order to allow users to work on them even when the authority is not reachable or when the peer is disconnected from the network. In fact, a user can perform both check-in and check-out operations also from the local copies of a document and the only difference between the master copy and a replica is that a check-in of a new version becomes definitive and available for all users only when the authority accepts the changes and updates the master copy. Intuitively, the peer-to-peer protocol may affect the number of conflicts generated during the collaboration [8].

The modelling described in this paper aims at describing the performances of a server mediated optimistic peer-to-peer protocol with a particular emphasis on the economics of network interactions. Specifically, there are network configuration situations where the social interactions of the peers influence the potential economic possibilities of the cooperating groups formed by entities which we call agents. A helpful approach in the modeling of the group social interactions is the game theoretical framework allowing the analysis of this kind of cooperative behavior. Commonly, the potential individual economic opportunities/outcomes are described by a cooperative game with transferable utility (TU-game) based on the structure of the agent set induced by the social configuration of the interacting agents [6]. This game theory approach naturally include the economic situations where the set of agents is hierarchically structured and where the potential individual economic outcomes are determined by the behavior rules induced by the organization structure. Sequencing situations, flow control situations and communication signaling situations are typical cases for this setting, where every agent in a strict hierarchy has a relationship with the leader either directly or indirectly with the help of one or more other agents. The economic gains of an agent are restricted by his position in the hierarchy. For an agent in such hierarchy, the important elements are the leader, the agent himself and all the intermediate agents that exist in the given hierarchy between the agent and the leader and we call such a group of agents a peer group. The game theoretical approach is based on peer groups of agents and the integrative view of the economic possibilities and the organization structure. The agents hierarchy is described by a rooted directed tree with the leader located in the root and each other agent in a distinct node. This tree uniquely determines the peer group structure: each agent's peer group corresponds to the agents in the unique path connecting the agent's node with the root in the tree. Tree-connected peer group situations are introduced as triplets consisting of the set of agents involved, the peer group structure describing the organization social configuration and a real-valued vector that gives the potential individual economic possibilities of the agents. To each tree-connected peer group situation we associate a TU-game, which we call a peer group game, having the agents regarded as players and the characteristic function being defined for each coalition by pooling the individual economic possibilities of those members with the corresponding peer groups inside of the coalition [4]. The peer groups are essentially the only coalitions that can generate a non-zero payoff within a peer group game.

A *TU*-game is a pair $\langle N, v\rangle$, where $N = \{1,2,...,n\}$ is a finite set of players and $v : R^N \to R$ is a characteristic function that assigns to each coalition $S \subset N$ the reward $v(S)$ of the coalition, with $v(\varnothing) = 0$. The game $\langle N, v\rangle$ is said to be superadditive if $v(S \cup T) \ge v(S) + v(T)$, for all $S, T \subset N$ with $S \cap T = \varnothing$. If $v(\{i\}) = 0$ for each $i \in N$, then $\langle N, v\rangle$ is called zero-normalized game. A game $\langle N, v\rangle$ is said to be monotonic if $v(S) \le v(T)$, for all $S \subset T \subset N$. We say that the game $\langle N, v\rangle$ is convex if it satisfies $v(S \cup T) + v(S \cap T) \ge v(S) + v(T)$, for all $S, T \subset N$.

2 Peer Group Games

Let $N = \{1,...,n\}$ be a finite set of agents with social as well as individual economic characteristics. The social features are given by a strict hierarchy defining the agents' relationships. Such a hierarchy can be described by a rooted directed tree *T* with $N = \{1,...,n\}$ as node set, agent 1 (the leader) as root, and each other agents located in a node. By a rooted directed tree we mean a directed graph with one distinguished node as root and for each node there is a unique directed path from the root to that node. Chain-like hierarchies will be represented by line-graphs or chains. By a line-graph or chain we mean a tree whose nodes are located on a single directed path. The individual attributes are the agents potential economic possibilities, described by the vector $a \in R^N$, where a_i is the gain which can be obtained by agent *i* if all his superiors in the hierarchy cooperate with him. The social constraints in the economic behavior are described by the means of *T*- connected peer groups of agents. For each agent $i \in N$, the subset of *N* consisting of all he agents in the path of *T* connecting 1 to i is called the *T*-connected peer group of agent i and denoted by $[1,i]$. The peer group of agent i imposes the social constraints on the agent i economic behavior. Agent i can only become 'effective' if it is in cooperation with all the other members of his peer group. A peer group structure on *N* induced by *T* is a mapping *P* which associates to each agent i in *N* the peer group of agent , $P : N \to 2^N$, where $P(i) = [1,i]$. Hence $P(i) = \{j \in N | j \prec i\}$, where $j \prec i$ means that *j* lies on the path from the root 1 to i. A *T*-connected peer group situation is a triplet $\langle N, P, a \rangle$, where *N* is the set of agents involved, *P* is the peer group structure on *N* induced by *T* and $a \in R^n_+$ is the vector describing the individual potential economic possibilities. To each *T*-connected peer group situation, we associate a TU cooperative game called the peer group game. Given a *T*-connected peer group $\langle N, P, a \rangle$, we call the corresponding peer group game the *TU*-game $\langle N, v_{P,a} \rangle$, or shortly $\langle N, v \rangle$, with $N = \{1,...,n\}$ and

$$v(S) = \sum_{i:P(i) \subset S} a_i \text{ , } \forall S \subset N \text{, } v(\varnothing) = 0 \tag{1}$$

where if $1 \notin S$, then we set $v(s) = 0$. The peer groups are essentially the only payoff generating coalitions within a peer group game. Each peer group game can be expressed as a nonnegative combination of the unanimity games corresponding to the peer groups. Recall that for $T \subset N$, the unanimity game corresponding to *T* is defined by

$$u_T(S) = \begin{cases} 1, \text{if } S \supset T \\ 0, \text{otherwise} \end{cases} \tag{2}$$

Let $u_{[1,i]}$ be the unanimity game corresponding to the peer group $[1,i]$ of agent i. Then

$$v([1,i]) = \sum_{i=1}^{n} a_i u_{[1,i]} \tag{3}$$

where a_i is called the Harsanyi dividend [4] of the peer group $[1,i]$. Let $\alpha, \beta \geq 0$, $a, b \in R^n_+$ and let $\langle N, P, a \rangle$, $\langle N, P, b \rangle$, $\langle N, P, \alpha a + \beta b \rangle$ be peer groups corresponding to the tree hierarchy *T* described by *P*. Then for the corresponding peer group games $v_{P,a}$, $v_{P,b}$, $v_{P,\alpha a + \beta b}$ we have

$$v_{p,\alpha a + \beta b} = \alpha v_{P,a} + \beta v_{P,b} \tag{4}$$

Consequently, peer groups form a cone $\{\langle N, v_{P,a}\rangle | a \in R^n_+\}$ generated by the independent subset $\{u_{[1,i]} | i \in N\}$ of unanimity games corresponding to peer groups. Denote by $PGG(N,P)$, the set of all peer group games with player set *N* and peer group structure *P*. From (3) it follows that peer group games are monotonic and that the leader is a veto player because $v(s) = 0$ for each $S \subset N$ with $1 \notin S$. So, the cone of peer group games is also a subcone in the cone of monotonic games with 1 as veto player [10]. From the convexity property it follows that peer group games are also superadditive and the game $w = v - a_1 u_{[1,1]}$ is zero-normalized and superadditive [14, 15]. A game $\langle N, v\rangle$ is called *T*-component additive game if $\langle N, v\rangle$ is a superadditive zero-normalized game with $R_T(v) = v$, where *T* is a tree and

$$R_T(v)(S) = \sum_{U \in S/T} v(U) \tag{5}$$

where S/T is the set of connected components of *S* in *T*. For each $i \in N \setminus \{1\}$, $u_{[1,i]}$ is a *T*-component additive game because

$$R_T(u_{[1,i]}(S)) = \sum_{S/T} u_{[1,i]}(U) = \begin{cases} 1, \text{if } [1,i \subset S] \\ 0, \text{otherwise} \end{cases} \tag{6}$$

So, $R_T(u_{[1,i]}) = u_{[1,i]}$ and

$$R_T(w) = \sum_{i \in N \setminus \{1\}} a_i R_T(u_{[1,i]}) = \sum_{i \in N \setminus \{1\}} a_i u_{[1,i]} = w \tag{7}$$

The fact that peer group games are convex games and veto rich games implies good properties described below [1, 4, 13].

Proposition 1. *For peer group games the following properties of solution concepts hold:*

i) *The bargaining set* $M(v)$ *coincides with the core* $C(v)$*;*

ii) *The kernel* $K(v)$ *coincides with the pre-kernel* $K^*(v)$ *and the pre-kernel consists of a unique point which is the nucleolus of the game;*

iii) *The nucleolus* $Nu(v)$ *occupies a central position in the core and is the unique point satisfying*

$$Nu(v) = \{x \in C(v) | s_{ij}(x) = s_{ji}(x), \forall i, j\} \tag{8}$$

where

$$s_{ij}(x) = \max\{v(S) - x(S) | i \in S \subset N \setminus \{j\}\} \tag{9}$$

iv) *The core* $C(v)$ *coincides with the Weber set* $W(v)$*, that is*

$$\text{conv}\{m^\sigma(v) | \sigma \text{ is a permutation of the players}\} \tag{10}$$

and $m^\sigma(v)$ *is the marginal vector w.r. to* σ.

v) *Shapley value* $\Phi(v)$ *is the center of gravity of the extreme points of the core and is given by*

$$\Phi_i(v) = \sum_{j: i \in P(j)} \frac{a_j}{|P(j)|}, \quad i \in N \tag{11}$$

where $P(j)$ *is the peer group of player j and* $|P(j)|$ *means the number of elements in* $P(j)$*;*

vi) *The* τ*-value is given by*

$$\tau(v) = \alpha(a_1,0,...,0) + (1-\alpha)(M_1(v), M_2(v),..., M_n(v)) \tag{12}$$

where

$$M_i(v) = \sum_{j:i\in P(j)} a_j \tag{13}$$

and $\alpha \in [0,1]$ *is such that*

$$\sum_{i=1}^{n} \tau_i(v) = v(N) \tag{14}$$

■

Line-graph peer group games arise from studying cooperative behavior in network transactions (auctions, resource allocation). We prove that the nucleolus of a pg-game corresponding to a line-graph is the unique solution of *n* equations. In the following, let $\langle N, v\rangle$ be the pg-game corresponding to $\langle N, P, a\rangle$, where $P = \{\{1\}, \{1,2\}, \{1,2,3\},..., \{1,2,...,n\}\}$, $a \in R_+^n$ and *z* is the nucleolus of $\langle N, v\rangle$.

Proposition 2. *For each* $i \in \{1,..., n-1\}$*, the nucleolus satisfies the equation*

$$z_{i+1} = \min\{z_i, Z_i - A_i\} \tag{15}$$

where

$$Z_i = \sum_{k=1}^{i} z_k \ , \ \ A_i = \sum_{k=1}^{i} a_k \tag{16}$$

Proof. Since *z* is a core element we have

$$z_1 \ge a_1 \ge 0 \text{ and } z_k \ge 0 \text{ for } k \in \{2,..., n\}$$

$$z(T) \ge v(T) \text{ for all } T \in 2^N$$

Since *z* is a pre-kernel element we have

$$s_{i,i+1}(z) = s_{i+1,i}(z)$$

for each $i \in \{1, 2,..., n-1\}$. In view of (4) it is sufficient to prove that for $i \in \{1, 2,..., n-1\}$

$$s_{i+1,i}(z) = -z_{i+1}$$

$$s_{i,i+1}(z) = \max\{-z_i, A_i - z_i\}$$

Take a coalition *S* with $i+1 \in S$ and $i \notin S$. Let *U* be the largest peer group in *S* if there is one, otherwise, let $U = \varnothing$. Then

$$e(S,z) = v(S - z(S) = v(U) - z(S) \le v(U) - z(U) - z_{i+1} \le -z_{i+1} = e(\{i+1\}, z)$$

So,

$$s_{i+1,i}(z) = \max\{e(S,z) | i \notin S, i+1 \in S\} = e(\{i+1\}, z) = -z_{i+1}$$

Take a coalition *S* with $i \in S$, $i+1 \notin S$. Then $S = S_1 \cup \{i\} \cup S_2$, where $S_1 = S \cap [1, i-1]$ and $S_2 = S \cap \{i+2\},..., n$. For $i = 1$ we interpret $[1, i-1]$ as the empty set $\varnothing$ and $[i+2, n] = \varnothing$ if $i = n-1$ or $i = n$. Let T_1 be the largest peer group in S_1, if there are peer groups in S_1, otherwise, let $T_1 = \varnothing$. We consider two cases.

Case 1. Let $T_1 = [1, i-1]$. Then

$$e(S,z) = v(S) - z(S) = v([1,i]) - z(S) \le v([1,i]) - \sum_{k=1}^{i} z_k = e([1,i], z)$$

For $i = 1$ one obtains

$$s_{1,2}(z) = \max\{e(S,z) | 1 \in S, 2 \notin S\} = e([1,1], z) = v(\{1\}) - z_1 = a_1 - z_1 = \max\{-z_1, a_1 - z_1\}$$

So, then result holds for $i = 1$.

Case 2. Let T_1 be a proper subset of $[1, i-1]$, where $i > 1$. Then

$$e(S,z) = v(S) - z(S) = v(T_1) - z(S) \le v(T_1) - z(T_1) - z_i \le -z_i = e(\{i\}, z)$$

Hence, for $i \ge 2$ we have

$$s_{i,i+1}(z) = \max\{v(S) - z(S) | i \in S, i+1 \notin S\} = \max\{e([1,i], z), e(\{i\}, z)\} = \max\{A_i - Z_i, -z_i\}$$

■

Proposition 3. *The nucleolus x of* $\langle N, v \rangle$ *is the unique solution of the n equations*

$$Z_n = A_n \tag{17}$$

$$z_i = \min\{z_{i-1}, Z_{i-1} - A_{i-1}\}, \ \ i = 2,...,n \tag{18}$$

Proof. Let $M = \{x \in R^n | x_{i+1} = \min\{x_i, X_i - A_i\}, \text{for each } i \in \{1,2,...,n-1\}\}$. For $x \in M$, the first coordinate uniquely determinates the other coordinates $x_2, x_3,...,x_n$, i.e., x_1 determines x_2, then x_3 is uniquely determined by x_1 and x_2, and so on. Note also that for $x, x' \in M$, if $x_1 > x_1'$, then $x_k > x_k'$ for each $k \in \{2,...,n\}$, so

$$\sum_{i=1}^{n} x_i > \sum_{i=1}^{n} x_i'$$

This implies that there is at most one element in *M* with the sum of the coordinates equal to A_n. On the other hand, in view of Proposition 2 the nucleolus *z* is an element in *M*, and $Z_n = A_n$ because *z* is a core element, so the efficiency condition holds. Hence $\{z\} = \{x \in M | X_n = Z_n\}$.

■

Let $\langle N, P, a \rangle$ be as before a *T*-connected peer group situation. Consider the case where gains are made via binary interactions of a central agent 1 with each of the other agents $i \in N$, resulting in a gain a_i, with the communication restrictions by the tree *T*. The graph restricted game $v = w_{|T}$ is

$$v = \sum_{i=1}^{n} a_i u_{[1,i]} \tag{19}$$

and

$$w_{|T}(S) = \sum_{U/T} w(U) \tag{20}$$

where S/T is the set of connected components of *S* in *T*.

3 Coordination and Information Sharing

Internet IP traffic was dominated up to a few years ago by Web related applications as reported by many traffic measurement studies. The traffic characteristics are changing due to many factors such as the massive commercial deployment of the Internet through ADSL access points, unprecedented increase of link transmission capacities and the emergence of new applications. Peer-to-peer (P2P) services have rapidly spread over the Internet and have a great impact on the networks in terms of traffic as well as of usage patterns. P2P applications are changing the way the customers utilize the network, that is, as long as the web like applications were dominant, the client/server paradigm was realistic. For P2P applications, any user terminal can become a server and any peer can download information from this server and consequently, driven by P2P protocols, the usage of the network becomes symmetric. When observing a link connecting different ADSL areas to IP backbone networks, a significant proportion of data flows in the upstream direction are generated by P2P protocols installed on ADSL terminals. These terminals play a server role for peers and transmit a significant amount of data into the network. P2P applications generate a significant part of traffic in current commercial IP networks. Beyond predominance in terms of volume, P2P applications have features which greatly impact the characteristics of traffic. For instance, files are segmented into chunks of limited size and these chunks may be separately and asynchronously downloaded by users. Because any end user terminal can become a server, the bandwidth of connections corresponding to P2P data transfers are naturally of rather small bit rates [9]. From traffic engineering viewpoint (buffer dimensioning, assessing the link transmission capacities, bandwidth provisioning), Internet traffic is extremely irregular and exhibits a fractal nature, making any traffic control attempt difficult because of the safety margins to account for traffic spikes. Another key characteristic of Internet traffic which is related in some sense to self-similarity is the long range dependence. Predominance of P2P traffic tends to remove the long range dependence and the self-similarity in IP networks carrying traffic to residential customers. The connections carrying P2P traffic have usually small bit rates and their superposition process can be well approximated by a smooth Gaussian process. The batch scheduling of data flows and the asynchronous transaction schedules for P2P traffic smoothing as data aggregates motivates a careful look into the coordination mechanisms regarding the interacting traffic processes in the network [3, 5, 7].

Generically, a group coordination situation is a triplet (σ_0, p, α) where σ_0 is the initial order, $p = (p_i)_{i \in N}$ with $p_i > 0$ is the processing time of agent i, and $\alpha = (\alpha_i)_{i \in N}$, where α_i is the cost per unit of time for i [11]. The urgency index of the agent i is given by $u_i = \alpha_i / p_i$. It is known that it is optimal to serve the agents according to their urgency, that is, the most urgent first. This order can be obtained by neighbor switches. The corresponding cost savings game is a nonnegative combination of unanimity games on neighbors that switch,

$$v = \sum_{(k,l), k<l} g_{k,l} u_{[k,l]} \tag{21}$$

where

$$g_{k,l} = \left[p_k \alpha_l - p_l \alpha_k \right]^+ \tag{22}$$

We think of group coordination situations in which the initial order $\sigma_0 = (1, 2, ..., n)$ of n interacting agents is such that the following relation between urgency indices holds

$$u_2 > u_3 > \cdots > u_n \tag{23}$$

The optimal order is obtained only by neighbor switches between 1 and some other agents, which means that all $g_{k,l}$ with $k \neq 1$ are zero. These coordination situations lead to peer group games of the form

$$v = \sum_{i=2}^{n} g_{1,i} u_{[1,i]} \tag{24}$$

where

$$g_{1,i} = [p_1\alpha_i - p_i\alpha_1]^+ \tag{25}$$

Such a peer group game corresponds to a *T*-connected peer group situation $\langle N, P, a\rangle$, where *T* s the line-graph with arcs $(1,2),(2,3),...,(n-1,n)$, corresponding to the initial order in the coordination situation, and where $a_i = g_{1,i}$ for each $i \in N \setminus \{1\}$ and $a_1 = 0$.

In the spirit of the notion of authority needed to control the contention for a set of items which are the subject of agents interactions (that is, the copy of an artifact owned by the authority being the master copy and other peers having to keep a local copy (replica) of the artifacts they do not own), let us extend the sequencing games to the case of an information collecting situation. In this case, a single agent called the decision maker, has to choose an action *a* from some action set and the reward resulting from this choice depends upon the true state of the world, which is not known by the decision maker. The decision maker has some information about the true state of the environment described by a partition of the finite set Ω of all possible states. An element of such a partition is called an event and if the true state is obtained, then the decision maker knows which event happens, that is, he knows which element of his partition contains the true state. Next to the decision maker, there are other agents that have information about the uncertainty and these agents can be consulted by the decision maker. An information sharing situation with the agent *k* as the decision maker is denoted by the tuple $\langle N, k, (\Omega,\mu), \{F_i\}_{i\in N}, A_k, r_k\rangle$, where $N = \{1,2,...,n\}$ is the set of all agents including the decision maker *k*, Ω is the finite set of all states of the world and μ is a positive measure on Ω which means that $\mu(\omega) > 0$ for all $\omega \in \Omega$ and $\sum_{\omega\in\Omega}\mu(\omega) = 1$. The information of agent $i \in N$ is represented by the partition F_i of Ω. The set A_k is the finite set of actions available for the decision maker and $r_k(\omega, a_k)$ is his reward if ω is the true state and if action a_k is chosen. Assume that

$$\sum_{I\in F_k}\max_{a_k\in A_k}\sum_{\omega\in I} r_k(\omega, a_k)\mu(\omega) > 0 \tag{26}$$

which means that is the decision maker works on its own, then it can achieve a nonnegative expected reward. Sharing the information from the set of agents $S \setminus \{k\} \subset N$ results in the expected reward

$$v_k = \sum_{I\in F_S}\max_{a_l\in A_k}\sum_{\omega\in I} r_k(\omega, a_k)\mu(\omega) \tag{27}$$

where $F_S = \left\{\bigcap_{i\in S} I_i \middle| I_i \in F_i, \bigcap_{i\in S} I_i \neq \varnothing\right\}$ is the partition of Ω that describes the total information of all the agents in *S*. The cooperative game $\langle N, v_k\rangle$ related to such information sharing situation with decision maker *k* is defined by the player set *N* and the function $v_k(S)$ that assigns $v_k(S) = 0$ to coalition *S* if $k \notin S$ and $v_k(S)$ is given by (1) if $k \in S$. An information sharing situation is denoted by a tuple $\langle N, (\Omega,\mu), \{F_i, A_i, r_i\}_{i\in N}\rangle$, where for each agent $i \in N$, we define a nonempty finite action set A_i and a reward function $r_i : \Omega \times A_i \to R$. The agent i has to choose an action $a_i \in A_i$ and it obtains the reward $r_i(\omega, a_i)$ depending on the chosen action and the true state ω. Assume that

$$\sum_{I\in F_i}\max_{a_i\in A_i}\sum_{\omega\in I} r_i(\omega, a_i)\mu(\omega) \geq 0 \tag{28}$$

for all $i \in N$, that is, on its own, any agent can obtain a nonnegative expected reward. If a group *S* of agents decides to cooperate and share their information, then the total expected reward for this group is given by

$$v(S) = \sum_{i\in S}\sum_{I\in F_S}\max_{a_i\in A_i}\sum_{\omega\in I} r_i(\omega, a_i)\mu(\omega) \tag{29}$$

with the convention that $v(\varnothing) = 0$. Suppose that we have an information sharing situation $\left\langle N, k, (\Omega,\mu), \{F_i\}_{i\in N}, A_k^*, r_k^*\right\rangle$ leading to the cooperative game $\left\langle N, v_k\right\rangle$. This game is also characterizing the situation $\left\langle N, (\Omega,\mu), \{F_i, A_i, r_i\}_{i\in N}\right\rangle$, where $A_k = A_k^*$, with A_i is an arbitrary nonempty finite set for $i \neq k$, $r_k = r_k^*$ and $r_i = 0$ for $i \in N\setminus\{k\}$. So, a group coordination situation with *k* as action taker can be transformed into an information sharing situation where all the agents except *k* have a trivial reward function, the zero function. Conversely, if we have an information sharing situation $\left\langle N, (\Omega,\mu), \{F_i, A_i, r_i\}_{i\in N}\right\rangle$, consider the group coordination situations $\left\langle N, k, (\Omega,\mu), \{F_i\}_{i\in N}, A_k, r_k\right\rangle$ for all $k \in N$ with the corresponding cooperative games $\left\langle N, v_k\right\rangle$, $k \in N$. Having the agent *k* a veto player in $\left\langle N, v_k\right\rangle$, we define

$$v = \sum_{i\in N} v_i \tag{30}$$

that is, an information sharing situation can be decomposed into $|N|$ group coordination situations. Moreover, an information sharing game is a game with a population monotonic value allocation scheme $\{a_{iS}\}_{i\in S,\, S\in 2^N\setminus\{\varnothing\}}$ defined by $a_{iS} = 0$ if $i \notin S$ and $a_{iS} = v_i(S)$, with

$$v(S) = \sum_{i\in S} v_i(S) = \sum_{i\in S} a_{iS} \tag{31}$$

and

$$a_{iS} \leq a_{iT} \text{ for all } i \in S \subset T \tag{32}$$

We introduce the so-called local information coordination games $\left\langle N, k, (\Omega,\mu), \{F_i\}_{i\in N}, A_k, r_k\right\rangle$ where given that ω is the true state, a group *S* of agents knows that the event $I_S(\omega)$ happens, which is that element of F_S containing ω. Consequently, each $\omega' \in I_S(\omega)$ occurs with the conditional probability

$$\mu(\omega')/\mu(I_S(\omega)) \tag{33}$$

The local information coordination game $\left\langle N, v_{k,\omega}\right\rangle$ assigns to each coalition S, $k \in S$, the expected reward

$$v_{k,\omega} = \max_{a_k\in A_k}\sum_{\omega'\in I_S(\omega)} r_k(a_k,\omega')\,\mu(\omega')/\mu(I_S(\omega)) \tag{34}$$

if one knows that ω is the true state and $v_{k,\omega}(S) = 0$, otherwise. We have $v_{k,\omega}(S) = v_{k,\omega'}(S)$ for all $\omega' \in I_S(\omega)$ and for coalitions *S* with $k \in S$,

$$v_k(S) = \sum_{I\in F_S}\max_{a_k\in A_k}\sum_{\omega'\in I} r_k(a_k,\omega')\mu(\omega') = \sum_{I\in F_S}\mu(I)v_{k,\overline{\omega}}(S) = \sum_{I\in F_S}\sum_{\omega\in I}\mu(\omega)v_{k,\omega}(S) \tag{35}$$

for some $\overline{\omega} \in I$. Also,

$$v_k(S) = \sum_{\omega\in\Omega}\mu(\omega)v_{k,\omega}(S) = 0 \tag{36}$$

if $k \notin S$.

4 Cost Allocation of Internet Traffic

In this section we address the problem of communication costs associated to the interacting traffic classes in the Internet [9]. We start by presenting a result describing the composition rules of the local information coordination games.

Proposition 4. *If* $\langle N, v_k \rangle$ *is an information coordination game and* $\{\langle N, v_{k,\omega} \rangle\}_{\omega\in\Omega}$ *are the corresponding local coordination games, then*

$$v_k = \sum_{\omega\in\Omega} \mu(\omega) v_{k,\omega} \tag{37}$$

■

Define

$$M_i(T, v_k) = v_k(T) - v_k(T \setminus \{i\}) \tag{38}$$

as the marginal contribution of agent i (interpreted in this section as being an Internet traffic class) to coalition $T \setminus \{i\}$. An information coordination game $\langle N, v_k \rangle$ is called *k*-concave if

$$M_i(S, v_k) \geq M_i(T, v_k) \tag{39}$$

for all $i \in N \setminus \{k\}$ and $S, T \subset N$ with $\{i, k\} \subset S \subset T$. If

$$M_i(S, v_k) \leq M_i(T, v_k) \tag{40}$$

for $\{i, k\} \subset S \subset T$, $i \neq k$, then the game is called *k*-convex. Then, the following statement holds:

Proposition 5. *If all the local information coordination games* $\{\langle N, v_{k,\omega} \rangle\}_{\omega\in\Omega}$ *are k-concave* (*convex*), *then the corresponding information coordination game* $\langle N, v_k \rangle$ *is k-concave* (*convex*).

Proof. We only check that $\langle N, v_k \rangle$ is *k*-concave, given that all local games are *k*-concave. For all $i \neq k$, *S* and *T* with $\{i, k\} \subset S \subset T$

$$M_i(S, v_k) - M_i(T, v_k) = \sum_{\omega\in\omega} \left[M_i(S, v_{k,\omega}) - M_i(T, v_{k,\omega}) \right] \mu(\omega) \geq 0$$

■

We call a monotonic game $\langle N, v \rangle$ with veto player *k* a total big boss game with *k* as the big boss if the game itself and all sub-games $\langle T, v \rangle$, $k \in T$, are big boss games, That is, a monotonic game $\langle N, v \rangle$ with veto player *k* is a total big boss game with big boss *k* if and only if

$$v(T) - v(S) \geq \sum_{i\in T\setminus S} M_i(T, v) \tag{41}$$

for all *S*, *T* with $k \in S \subset T$.

Proposition 6. *Let* $\langle N, v_k \rangle$ *be an information coordination game. Then* $\langle N, v_k \rangle$ *is a total big boss game with big boss k if and only if it is k-concave.*

Proof. Let $\langle N, v_k \rangle$ be an information coordination game. Assume first that it is k-concave and let $k \in S \subset T$. Assuming $T \setminus S = \{i_1, i_2, ..., i_h\}$, then

$$v_k(T) - v_k(S) = \sum_{r=1}^{h} [v_k(S \cup \{i_1, i_2, ..., i_r\}) - v_k(S \cup \{i_1, i_2, ..., i_{r-1}\})] =$$
$$= \sum_{r=1}^{h} M_{i_r}(S \cup \{i_1, i_2, ..., i_r\}, v_k) \geq \sum_{r=1}^{h} M_{i_r}(T, v_k) = \sum_{i \in T \setminus S} M_i(T, v_k)$$

So, k-concavity implies that $\langle N, v_k \rangle$ is a total big boss game with k as big boss. Suppose now that $\langle N, v_k \rangle$ is a total big boss game with k as big boss; we prove that

$$M_i(U, v_k) \geq M_i(U \cup \{j\}, v_k)$$

for all $U \subset N$ and $i, j \in N \setminus \{k\}$ such that $\{i, k\} \subset U \subset N \setminus \{j\}$. Also,

$$v_k(U \cup \{j\}) - v_k(U \setminus \{i\}) \geq M_j(U \cup \{j\}, v_k) + M_i(U \cup \{j\}, v_k)$$

On the other hand

$$v_k(U \cup \{j\}) - v_k(U \setminus \{i\}) = v_k(U \cup \{j\}) - v_k(U) + v_k(U) - v_k(U \setminus \{i\}) =$$
$$= M_j(U \cup \{j\}, v_k) + M_i(U, v_k)$$

To prove that $\langle N, v_k \rangle$ is k-concave, we take $S, T \subset N$ with $\{i, k\} \subset S \subset T$ and assuming that $T \setminus S = \{i_1, i_2, ..., i_h\}$, we have

$$M_i(S, v_k) \geq M_i(S \cup \{i_1\}, v_k) \geq M_i(S \cup \{i_1, i_2\}, v_k) \geq \cdots \geq M_i(T, v_k)$$

Thus, $M_i(S, v_k) \geq M_i(T, v_k)$.

■

In describing the interaction costs among the Internet traffic classes, we consider a total big boss game with k as big boss and where for all T with $k \in T$, the core $C(T, v)$ of the sub-game (T, v) is defined to be

$$C(T, v) = \left\{x \in R^T \middle| 0 \leq x_i \leq M_i(T, v), i \in T \setminus \{k\}, \sum_{i \in T} x_i = v(T)\right\} \tag{42}$$

The τ-value, $\tau(T, v)$ and the nucleolus $Nu(T, v)$ [12] coincide and are equal to the center z of the core $C(T, v)$ defined by

$$z_j = \begin{cases} M_j(T, v)/2, & \text{if } j \in T \setminus \{k\} \\ v(T) - \sum_{i \in T \setminus \{k\}} M_i(T, v)/2, & \text{if } j = k \end{cases} \tag{43}$$

Taking an IC game $\langle N, v_k \rangle$ and denoting by P_k the set $\{S \subset N | k \in S\}$ of all coalitions containing the decision maker, we call a scheme $\lfloor b_{i,S} \rfloor_{i \in S, S \in P_k}$ an allocation scheme if $\lfloor b_{i,S} \rfloor_{i \in S}$ corresponds to a core element of the sub-game $\langle S, v_k \rangle$. Such an allocation scheme $\lfloor b_{i,S} \rfloor_{i \in S, S \in P_k}$ is called a bi-monotonic allocation scheme if for all $S, T \in P_k$ with $S \subset T$, we have $b_{i,S} \geq b_{i,T}$ for all $i \in S \setminus \{k\}$ and $b_{k,S} \leq b_{k,T}$. In a bi-monotonic allocation, the big boss is better off in larger coalitions and the other players are worse off. Let the scheme $\lfloor b_{i,S} \rfloor_{i\in S, S\in P_k}$ be defined by $b_{k,S} = v_k(S)$ and $b_{i,S} = 0$ if $i \in S \setminus \{k\}$. If $\langle N, v \rangle$ is a total big boss game with *k* as the big boss and $x \in C(N,v)$, then *x* can be extended to a bi-monotonic allocation scheme. Since $x \in C(N,v)$, we can find for each $i \in N \setminus \{k\}$ and $\alpha_i \in [0,1]$, such that $x_i = \alpha_i M_i(N,v)$, then

$$x_k = v(N) - \sum_{i\in N\setminus\{k\}} \alpha_i M_i(N,v) \tag{44}$$

Then, $\lfloor b_{i,S} \rfloor_{i\in S, S\in P_k}$ defined by $b_{i,S} = \alpha_i M_i(S,v)$ for all $i \in S \setminus \{k\}$ and $b_{k,S} = v(S) - \sum_{i\in S\setminus\{k\}} \alpha_i M_i(S,v)$ is a bi-monotonic allocation game.

In synchronous sequencing only data batches having equal processing time (that is, the same time class) can be switched [8]. If $(N,\ \sigma_0,\ \alpha,\ p)$ is a sequencing step for an aggregate $S \subset N$, a reordering σ is synchronous admissible if it leaves the aggregate $N - S$ fixed and the switching takes place only between streams with equal batch processing time, then

$$|P(\sigma_0, j)| = |P(\sigma, j)| \text{ for all } j \in N - S \tag{45}$$

and

$$p_i = p_j \text{ for all } i, j \in N \text{ with } i = \sigma(j) \tag{46}$$

The optimal processing order for each aggregate $S \subset N$ can be obtained: a sequencing order is optimal if streams with equal data processing time are arranged in decreasing order of scheduling potentials. If $RIG(S)$ denote the set of synchronous admissible rearrangements of aggregate $S \subset N$, then a synchronous sequencing game (N, r) arising from a sequencing $(N,\ \sigma_0,\ \alpha,\ p)$ is defined as

$$r(S) = \max_{\sigma\in RIG(S)} \left\{ \sum_{i\in S} \alpha_i C_{\sigma_0, i} - \sum_{i\in S} \alpha_i C_{\sigma, i} \right\} \tag{47}$$

Synchronous sequencing games have a nonempty core as synchronous sequencing games are permutation games. A cooperative game $\langle N, v \rangle$ is called a permutation game if there is an $n \times n$ matrix A such that

$$v(S) = \max_{\sigma\in\pi_S} \left\{ \sum_{i\in S} a_{ij} - \sum_{i\in S} a_{i\sigma(j)} \right\} \qquad 48)$$

for all $S \subset N$, where $S \neq \varnothing$, $v(\varnothing) = 0$ and π_S is the set of all permutations of S. Let $(N,\ \sigma_0,\ \alpha,\ p)$ be a sequencing step and $\langle N, r \rangle$ be the corresponding synchronous sequencing game. Let *a* be such that

$$a_{ij} = \begin{cases} \alpha_i C_{\sigma j}, & \text{if } p_i = p_j \\ \infty, & \text{otherwise} \end{cases} \tag{49}$$

For an aggregate $S \subset N$ a reordering is synchronous admissible if and only if it is a permutation of players of *S* with equal processing time,

$$\begin{aligned} r(S) &= \max_{\sigma\in RIG(S)} \left\{ \sum_{i\in S} \alpha_i C_{\sigma, i} - \sum_{i\in S} \alpha_i C_{\sigma, j} \right\} = \\ &= \max_{\sigma\in\pi_S} \left\{ \sum_{i\in S} a_{ij} - \sum_{i\in S} a_{i\sigma(j)} \right\} \end{aligned} \tag{50}$$

The last entry follows from $a_{ij} = \infty$ if $p_i \neq p_j$.

5 Concluding Remarks

In this paper we have described some results of the modelling of economic interactions occurring in communication networks concerning the peer-to-peer architecture. The P2P paradigm has recently captured scientific and academic researchers attention since the explosive growth of the number of .le sharing application users that generate a large fraction of the Internet traffic. P2Pbased applications pose challenging research problems related to reliability, scalability, resource organization, indexing, dimensioning, discovery and coordination in decentralized architectures. Various models such as stochastic fluid models are analytic models that have recently drawn the attention of many researchers for the performance evaluation of complex communication systems. Fluid variables may be used to approximate discrete variables to tackle the state space explosion problem that typically occurs when analyzing discrete-state based models. The contribution of the modelling effort activity is on two levels: on one hand, models are used to describe the integration of the information sharing and group coordination activities and on the other hand, to quantify the impact of the peer-to-peer protocol on the reference cooperative work scenario (coalitions of users cooperating on electronic artifacts). For peer-to-peer applications, each agent can become a server and any peer can download information from this server and consequently, driven by P2P protocols, the usage of the network becomes symmetric. Files are segmented into chunks of limited size and these chunks may be separately and asynchronously downloaded by users. Because any end user terminal can become a server, the bandwidth of connections corresponding to P2P data transfers are naturally of rather small bit rates. The connections carrying P2P traffic have usually small bit rates and their superposition process can be well approximated by a smooth Gaussian process. The batch scheduling of data flows and the asynchronous transaction schedules for P2P traffic as data aggregates has been a serious motivation for the analysis of the coordination mechanisms regarding the interacting traffic processes in the network.

References

[1] Arin J, Feltkamp V (1997) The nucleolus and kernel of veto rich transferable utility games. Int. J. of Game Theory 26:61–73.

[2] Balzarotti, D. et al. (2002), Freeing cooperation from servers tyranny, in: E. Gregori, L. Cherkasova, G. Cugola, F. Panzieri and G. P. Picco, editors, Web Engineering and Peer-to-Peer Computing, Lecture Notes in Computer Science **2376**, Springer-Verlag, pp. 235–246.

[3] Bar-Noy, A., Ladner, A., 2003. Windows Scheduling Problems for Broadcast Systems, SIAM Journal on Computation, Vol. 32, No. 4, pp. 1091-1113.

[4] Curiel I (1997) Cooperative game theory and applications. Kluwer Academic Publishers, Boston

[5] Curiel I., Potters J., Rajendra Prasad V., Tijs S., Veltman B., 1994. Sequencing and cooperation, Operations Research, Vol. 42, pp. 566–568.

[6] Driessen TSH (1988) Cooperative games, solutions and applications. Kluwer Academic Publishers, Dordrecht.

[7] Kalai E, Zemel E (1982) Totally balanced games and games of flow. Math. Op. Res. 7:476–478.

[8] Keidar, I., Khazan, R., 2002. A Virtually Synchronous Group Multicast Algorithm for WANs: Formal Approach, SIAM Journal on Computation, Vol. 32, No. 1, pp. 78-130.

[9] Murgu, A., 2004. Dynamic Behavior of Traffic Network Equilibria in IP Service Aggregation, IIAS-Transactions on Systems Research and Cybernetics, Vol. 4, No. 2, pp. 17-28.

[10] Potters J, Reijnierse H (1995) G-component additive games. Int. J. Game Theory 24:49–56.

[11] Rasmusen E (1989) Games and information, an introduction to game theory. Blackwell, Oxford UK & Cambridge USA.

[12] Schmeidler D (1969) The nucleolus of a characteristic function game. SIAM J. App. Math. 17:1163–1170.

[13] Shapley LS (1953) A value for n-person games. Ann. Math. Stud. 28:307–317.

[14] Shapley LS (1971) Cores of convex games. Int. J. Game Theory 1:11–26.

[15] Topkis DM (1998) Supermodularity and complementarity. Princeton Univesity Press, Princeton, NJ.

New Keynesian Model of the Small Open Czech Economy*

Karel Musil and Osvald Vašíček
Faculty of Economics and Administration of Masaryk University
Department of Applied Mathematics and Computer Science
Lipová 41a, 602 00 Brno
e–mail: musilk@mail.muni.cz, osvald@econ.muni.cz

Abstract

The paper introduces a New Keynesian Dynamic Stochastic General Equilibrium (NK DSGE) model. This model is related to the New Open Economy Macroeconomics (NOEM). It is strictly based on microeconomic foundations and consists of representative agents. They are representative households and firms, a central monetary authority and a foreign economy. The economic model is log–linearized and transformed to a rational expectations (LRE) model, which is solved. Parameters of the solved model are estimated by Bayesian simulation techniques using a priory set information. The estimated model together with the impulse responses seems to give a reasonable approximation of behavior of the Czech economy.

Keywords

New Keynesian DSGE model, New Open Economy Macroeconomics, monetary policy, Taylor rule, inflation targeting, rational expectations, Bayesian simulations
JEL: C15, C51, E12, E17, E52

1 Introduction

Most of the central banks in the developed economies (including Czech National Bank) implement monetary policy through a money market to influence a short run interest rate with respect to the inflation target. For more information see [1]. Every monetary regime is very close to an economic paradigm. If a central monetary authority influences a level of short run interest rate to alter a long run interest rate and agents' decisions about their consumption and investment subsequently, the result will be a change in a dynamics of behavior of the economy (and in inflation as well). This transmission mechanism is a base stone of the New Keynesian Economy.

The inflation targeting regime is a regime stemming from short run nominal interest rate toward inflation rate. The fundamental question is how these quantities are connected and what influences them. Some answers could be offered by a model approach. We proceed from a model of P. Liu [2] and articles of J. Gali and T. Monacelli [3], T. Lubik and F. Schorfheide [4].

2 Model

The model is a New Keynesian Dynamic Stochastic General Equilibrium (NK DSGE) model strictly based on microeconomic foundations. It is a small open economy model of the Czech economy and its structure is closely related to the New Open Economy Macroeconomic (NOEM) approach. The model consists of representative agents, who optimize their behavior. They are households, firms, a central monetary authority and foreign sector (agents in a foreign economy, or more exactly agents in the rest of the world).

The model is in a gap form. Variables marked by small letters are logarithms of the origin variables (originally marked by capital letters).

*This paper has been worked as a part of the project MŠMT research center identification code 1M0524 and with the grant support of GAČR No. 402/05/2172.

2.1 Households

A representative household maximizes its utility function subject to its budget constraint:

$$E_t \sum_{t=0}^{\infty} \beta^t \left(\frac{(C_t - hC_{t-1})^{1-\sigma}}{1-\sigma} - \frac{N^{1+\varphi}}{1+\varphi} \right) \qquad s.t. \qquad P_t C_t + E_t \left(\frac{D_{t+1}}{1+r_t} \right) \leq D_t + W_t N_t,$$

where β is a discount factor, C_t is a consumption which contains a habit formation factor h, σ is an inverse elasticity of intertemporal substitution, N_t are hours of labor, φ is an inverse elasticity of labor supply, P_t is overall Consumer Price Index, D_t is nominal pay–off on a portfolio held at $t-1$, r_t is a nominal interest rate and W_t denotes a nominal wage.

The consumption C_t is divided between consumption of domestically ($C_{H,t}$) and foreign ($C_{F,t}$) produced goods:

$$C_t \equiv \left((1-\alpha)^{\frac{1}{\eta}} C_{H,t}^{\frac{\eta-1}{\eta}} + \alpha^{\frac{1}{\eta}} C_{F,t}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}},$$

where η is an elasticity of substitution between home and foreign goods and α is a degree of openness. The optimal allocation between goods are following demand function for the i-th good:

$$C_{H,t}(i) = \left(\frac{P_{H,t}(i)}{P_{H,t}} \right)^{-\epsilon} C_{H,t} \qquad\qquad C_{F,t}(i) = \left(\frac{P_{F,t}(i)}{P_{F,t}} \right)^{-\epsilon} C_{F,t},$$

where $P_{H,t}$ is a price index of home produced goods (goods for domestic and foreign consumption), $P_{F,t}$ is an import price index and ϵ is an elasticity of substitution between goods produced across two countries. The overall Consumer Price index P_t is $P_t \equiv \left\{ (1-\alpha) P_{H,t}^{1-\eta} + \alpha P_{F,t}^{1-\eta} \right\}^{\frac{\eta}{\eta-1}}$.

Log–linear approximation of the first order condition is ($\pi_t = p_t - p_{t-1}$):

$$w_t - p_t = \varphi n_t + \frac{\sigma}{1-h} (c_t - hc_{t-1})$$

$$c_t - hc_{t-1} = E_t (c_{t+1} - hc_t) - \frac{1-h}{\sigma} (r_t - E_t \pi_{t+1}) \tag{1}$$

The optimizing behavior is influenced by terms of trade, uncovered interest parity and purchase power parity that includes some rigidities.

The connection of representative households abroad are following relations (variables with a superscript * hold for the foreign economy):

- terms of trade:

$$\Delta s_t = \pi_{F,t} - \pi_{H,t} \tag{2}$$

- uncovered interest rate parity:

$$\Delta E_t q_{t+1} = (r_t^* - E_t \pi_{t+1}^*) - (r_t - E_t \pi_{t+1}) \tag{3}$$

- condition for a parallel optimizing of domestic and foreign households:

$$c_t - hc_{t-1} = y_t^* - hy_{t-1}^* - \frac{1-h}{\sigma} q_t \tag{4}$$

- law of one price gap (we use it in the form of $\Psi_t = \frac{P_t^*}{Z_t P_{F,t}}$ and the law of one price gap implies $\Psi \neq 1$) combined together with terms of trade and an expression for the real exchange rate, see [5]:

$$\psi_t = -[q_t + (1-\alpha) s_t] \tag{5}$$

Used variables are log–linearized and express: S_t terms of trade, Z_t nominal exchange rate, Q_t real exchange rate, Ψ_t law of one price gap ($\psi_t = log \Psi_t$), $\pi_{H,t}$ domestic inflation and $\pi_{F,t}$ imported inflation.

2.2 Firms

A representative monopolistically competitive firm produces a differential good Y_j according to the production function described by $Y_t = A_t N_t(j)$, where $a_t = logA_t$ is describing a technological progress following an AR(1) process:

$$a_t = \rho_a a_{t-1} + \epsilon_t^a. \tag{6}$$

The log of real marginal cost is derived from the production function and is following:

$$mc_t = w_t - p_{H,t} - a_t = \frac{\sigma}{1-h}(c_t - hc_{t-1}) + \varphi y_t + \alpha s_t - (1+\varphi)a_t. \tag{7}$$

The representative firm tries to set a new price $\bar{P}_{H,t}$ every period to maximize its profit (measured by the current value of households' dividends subject to a demand constraint):

$$\max \sum_{k=0}^{\infty} \theta^k \frac{1}{R_{t+k}} E_t(Y_{t+k}(\bar{P}_{H,t} - MC_{t+k}^n)) \qquad s.t. \qquad Y_{t+k} \leq \left(\frac{\bar{P}_{H,t}}{P_{H,t+k}}\right)(C_{H,t+k} + C_{H,t+k}^*),$$

where MC_t^n are nominal marginal costs.

The firm sets its new price in a Calvo price effect style. Every period only part of firms $(1-\theta_H)$ resets their price optimally. The rest of the firms (θ_H) adjusts their prices by indexing in accordance with the last period inflation. In this case the present domestic inflation depends on the new prices of optimizing firms and the prices indexed with respect to the last period inflation:

$$\pi_{H,t} = (1-\theta_H)(\bar{p}_{H,t} - p_{H,t-1}) + \theta_H^2 \, \pi_{H,t-1}.$$

Using the first order condition of the representative firm we get the New Keynesian Phillips Curve (NKPC) after some amendment:

$$\pi_{H,t} = \beta(1-\theta_H)E_t\pi_{H,t+1} + \theta_H\pi_{H,t-1} + \lambda_H mc_t, \tag{8}$$

where $\lambda_H = \frac{(1-\beta\theta_H)(1-\theta_H)}{\theta_H}$.

Similar behavior can be applied for domestic importer firms. According to the Calvo price effect only θ_F as a fraction of these firms can not reoptimize their prices. The Phillips Curve expressing the imported inflation is following:

$$\pi_{F,t} = \beta(1-\theta_F)E_t\pi_{H,t+1} + \theta_F\pi_{F,t-1} + \lambda_F\psi_t, \tag{9}$$

where $\lambda_F = \frac{(1-\beta\theta_F)(1-\theta_F)}{\theta_F}$.

The overall inflation is influenced by the degree of openness, the domestic and imported inflation:

$$\pi_t = (1-\alpha)\pi_{H,t} + \alpha\pi_{F,t}. \tag{10}$$

2.3 Central Bank

The central bank monetary policy development can be approximated by a causal relation of the modified Taylor rule (in a gap form), see [6]. The inflation targeting is obtained in the relation implicitly. It is in a development of the inflation gap (π_t). π_t is a deviation of the consumer price inflation from its target:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\phi_1\pi_t + \phi_2 y_t), \tag{11}$$

where ρ_r is a degree of interest rate smoothing (backward–looking parameter of the interest rate gap), ϕ_1 and ϕ_2 are relative weights on inflation gap and output gap of the domestic economy.

2.4 Foreign Economy

The domestic economy is a small open economy. The foreign sector is exogenous in this model. It is described by an equation for output and interest rate as following AR(1) processes:

$$y_t^* = \lambda_1 y_{t-1}^* + \epsilon_t^{y^*} \tag{12}$$

$$r_t^* - E_t\pi_{t+1}^* = \rho_{r^*}(r_{t-1}^* - \pi_t^*) + \epsilon_t^{r^*}, \tag{13}$$

where $E_t\pi_{t+1}^*$ are rational expectations put exogenously from the foreign data.

A condition for the equilibrium of the domestic goods market respects the influence of the foreign output. The domestic output is divided between domestic consumption and foreign consumption (exports) of home made goods ($Y_t = C_{H,t} + C_{H,t}^*$). Together with an optimal expenditure functions of representative domestic and foreign households (for more details see [5]) we get:

$$y_t = (2-\alpha)\alpha\eta s_t + (1-\alpha)c_t + \alpha\eta\psi_t + \alpha y_t^*. \tag{14}$$

2.5 Linearized System

The log–linearized model consists of equations (1) – (14). These 14 equations are rearranged and completed by exogenous domestic and foreign shocks. The system is following:

$$\psi_t = -[q_t + (1-\alpha)s_t] \tag{15}$$
$$\Delta s_t = \pi_{F,t} - \pi_{H,t} + \epsilon_t^s \tag{16}$$
$$\Delta E_t q_{t+1} = -\{(r_t - E_t\pi_{t+1}) - (r_t^* - E_t\pi_{t+1}^*)\} + \epsilon_t^q \tag{17}$$
$$\pi_t = (1-\alpha)\pi_{H,t} + \alpha\pi_{F,t} \tag{18}$$
$$\pi_{F,t} = \beta(1-\theta_F)E_t\pi_{F,t+1} + \theta_F\pi_{F,t-1} + \lambda_F\psi_t + \epsilon_t^{\pi_F} \tag{19}$$
$$\pi_{H,t} = \beta(1-\theta_H)E_t\pi_{H,t+1} + \theta_H\pi_{H,t-1} + \lambda_H mc_t + \epsilon_t^{\pi_H} \tag{20}$$
$$mc_t = \frac{\sigma}{1-h}(c_t - hc_{t-1}) + \varphi y_t + \alpha s_t - (1+\varphi)a_t \tag{21}$$
$$a_t = \rho_a a_{t-1} + \epsilon_t^a \tag{22}$$
$$c_t - hc_{t-1} = E_t(c_{t+1} - hc_t) - \frac{1-h}{\sigma}(r_t - E_t\pi_{t+1}) \tag{23}$$
$$c_t - hc_{t-1} = y_t^* - hy_{t-1}^* - \frac{1-h}{\sigma}q_t \tag{24}$$
$$y_t = (2-\alpha)\alpha\eta s_t + (1-\alpha)c_t + \alpha\eta\psi_t + \alpha y_t^* \tag{25}$$
$$r_t = \rho_r r_{t-1} + (1-\rho_r)(\phi_1\pi_t + \phi_2 y_t) + \epsilon_t^r \tag{26}$$
$$y_t^* = \lambda_1 y_{t-1}^* + \epsilon_t^{y^*} \tag{27}$$
$$r_t^* - E_t\pi_{t+1}^* = \rho_{r^*}(r_{t-1}^* - \pi_t^*) + \epsilon_t^{r^*} \tag{28}$$

3 Solving the Model and Results

The model contains 11 equations for endogenous variable and 3 equations for exogenous processes (see (22), (27) a (28)). The log–linearized model is possible to rewrite into the rational expectations (LRE) model:

$$0 = Ax_t + Bx_{t-1} + Cy_t + Dz_t \tag{29}$$
$$0 = FE_t(x_{t+1}) + Bx_t + Hx_{t-1} + JE_t(y_{t+1}) + Ky_t + LE_t(z_{t+1}) + Mz_t \tag{30}$$
$$E_t(z_{t+1}) = Nz_t + E_t(\epsilon_{t+1}) \tag{31}$$
$$E_t(\epsilon_{t+1}) = 0, \tag{32}$$

where x_t is the endogenous state vector, y_t is the endogenous vector of unobservable variables and z_t is the exogenous stochastic process where $x_t = \{y_t, q_t, r_t, \pi_t, \pi_{F,t}, r_t^*, y_t^*\}, y_t = \{\psi_t, s_t, c_t, mc_t, \pi_{H,t}\}$, $z_t = \{a_t, \epsilon_t^s, \epsilon_t^q, \epsilon_t^{\pi_H},\ \epsilon_t^{\pi_F}, \epsilon_t^r, \epsilon_t^{y^*}, \epsilon_t^{r^*}\}$. The matrices of system are: $A_{3\times7}$, $B_{3\times7}$, $C_{3\times5}$, $D_{3\times8}$, $F_{10\times7}$, $G_{10\times7}$, $H_{3\times7}$, $J_{10\times5}$, $K_{10\times5}$, $L_{10\times8}$ and $N_{10\times8}$.

The general equilibrium (GE) rule is expressed by solving the model (29) – (32) in the following form:

$$S_{t+1} = \Gamma_1 S_t + \Gamma_2 w_{t+1} \tag{33}$$
$$Y_t = \Lambda S_t + v_t, \tag{34}$$

where the vector $S_t = \{x_t, y_t\}$ is a state vector, Y_t is a vector of observed variables, Γ_1 and Γ_2 are matrices of the model deep parameters, Λ is a matrix defining a relationship between observed and state variables, w_t is a vector of a innovations and v_t a vector of measurement error.

The initial state conditions $S_0 \sim N(\bar{S}_0, \Sigma_0)$ is evaluated by the Kalman filter algorithm. For the estimation of the 12 parameters of the model $(h, \sigma, \eta, \varphi, \theta_H, \theta_F, \phi_1, \phi_2, \rho_r, \rho_r^*, \rho_a, \lambda_1)$ and 8 parameters representing the standard deviations of shocks $(\sigma_a, \sigma_s, \sigma_q, \sigma_{\pi_H}, \sigma_{\pi_F}, \sigma_r, \sigma_{y^*}, \sigma_{r^*})$ is used the Bayesian method. The Bayesian estimation needs data and any ordered information about a prior density of the parameter vector, which is obtained in the simulated Markov Chain for every parameter. The estimation uses the Metropolis–Hasting Monte Carlo Markov Chain (MCMC) method to simulate a posterior density of the parameter vector. All inferences about parameters are included in the estimated posterior density of the parameter vector.

Quarterly data from I.Q 1995 to IV.Q 2005 was used for the estimation. The model is a gap model. Data (including foreign variables) are entered as a deviation of variables from their long run equilibrium development except of the terms of trade (s_t). The inflation gap is a deviation of the overall inflation from the inflation target of the central bank.

Table 1 contains the estimation of parameters. Figure 1 shows the posterior and prior marginal density plot of parameters. It was used 100 000 Markov Chain (MC) draws for every parameter. Parameters α and β were fixed at 0.4 and 0.99 respectively.

Parameter	Prior Mean	Posterior Median	95% Posterior Interval
h	0.50	0.8934	$\langle 0.8301;\ 0.9566\rangle$
σ	1.00	0.8793	$\langle 0.2581;\ 1.5007\rangle$
η	1.00	0.3822	$\langle 0.2860;\ 0.4784\rangle$
φ	1.00	1.3804	$\langle 0.8654;\ 1.8953\rangle$
θ_H	0.50	0.6532	$\langle 0.5929;\ 0.7135\rangle$
θ_F	0.50	0.4284	$\langle 0.3014;\ 0.5554\rangle$
ϕ_1	1.50	1.3070	$\langle 1.0863;\ 1.5276\rangle$
ϕ_2	0.25	0.5188	$\langle 0.2552;\ 0.7825\rangle$
ρ_r	0.50	0.6496	$\langle 0.5537;\ 0.7454\rangle$
ρ_r^*	0.70	0.6820	$\langle 0.4900;\ 0.8740\rangle$
ρ_a	0.70	0.9694	$\langle 0.9297;\ 1.0090\rangle$
λ_1	0.70	0.7917	$\langle 0.7057;\ 0.8778\rangle$
σ_a	$\langle 0;\infty\rangle$	0.8257	$\langle 0.3715;\ 1.2800\rangle$
σ_s	$\langle 0;\infty\rangle$	14.4810	$\langle 12.3251;\ 16.638\rangle$
σ_q	$\langle 0;\infty\rangle$	4.7321	$\langle 3.0487;\ 6.4155\rangle$
σ_{π_H}	$\langle 0;\infty\rangle$	2.9937	$\langle 2.1483;\ 3.8391\rangle$
σ_{π_F}	$\langle 0;\infty\rangle$	7.2143	$\langle 4.1161;\ 10.3124\rangle$
σ_r	$\langle 0;\infty\rangle$	1.8503	$\langle 1.3461;\ 2.3545\rangle$
σ_{y^*}	$\langle 0;\infty\rangle$	0.3518	$\langle 0.2618;\ 0.4418\rangle$
σ_{r^*}	$\langle 0;\infty\rangle$	0.4281	$\langle 0.3257;\ 0.5306\rangle$

Table 1: Posterior Estimates of the Parameters and Innovations

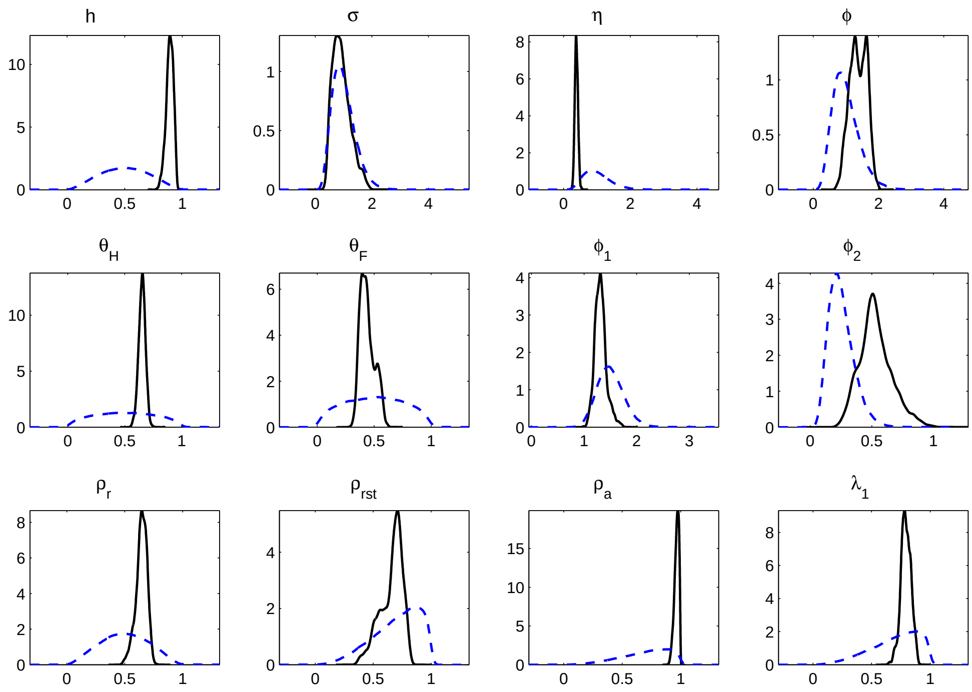


Figure 1: Posterior and Prior Marginal Density Plot of Parameters

All parameters are statistically significant and the model with estimated parameters has following basic characteristics. There is relatively high degree of habit formation in consumption ($h = 0.89$), low rate of substitution between domestic and foreign goods ($\eta = 0.38$), relatively low labor supply elasticity ($\varphi = 1.38$), duration of an average price contact for domestic firms is 3 quarters ($\theta_H = 0.65$) and for import firms is 2 quarters ($\theta_F = 0.43$), relatively high inertia for using technologies ($\rho_a = 0.97$). The monetary policy is described by following modified Taylor rule: $r_t = 0.6496\, r_{t-1} + (1 - 0.6496)\,(1.3070\,\pi_t + 0.5188\, y_t) + \epsilon_t^r$.

The posterior marginal densities are sharper than priors. It indicates a reduction in entropy of information.

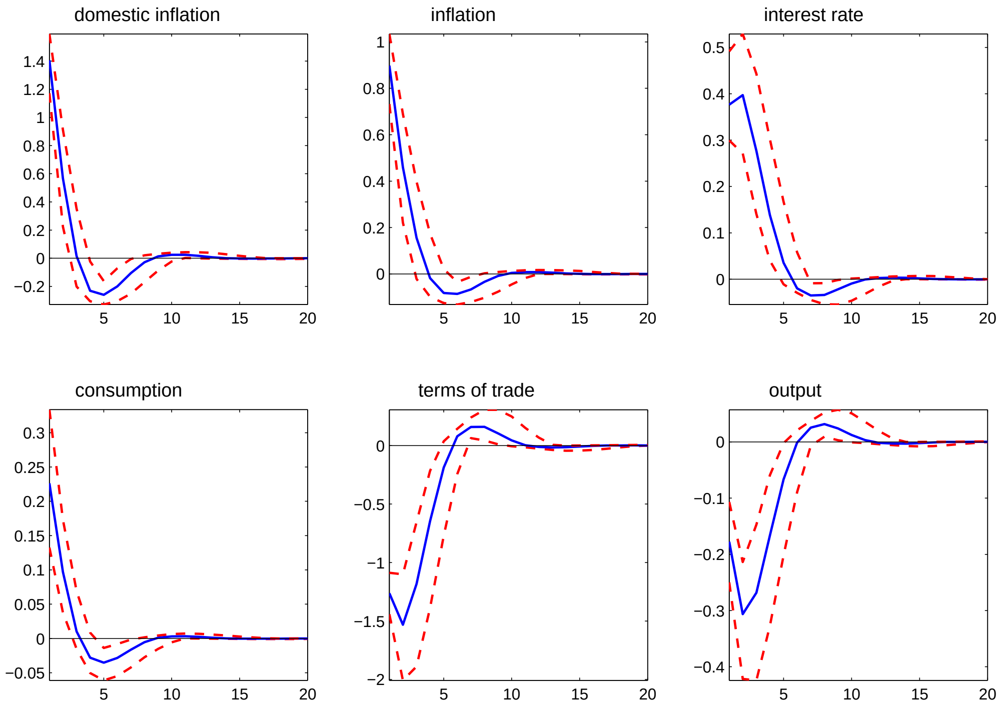


Figure 2: Impulse Response Functions from One Unit of Inflation Innovation

For detailed behavior analysis we introduced 10 sets of impulse response functions. Figure 2 shows the impulse responses from one unit of inflation innovation which influences a conducting of monetary policy. This impulse response seems to describe the behavior of the economy quite well. It increases domestic inflation and the central bank must higher interest rate according to the modified Taylor rule. The impact is a fall of output and subsequently a fall of consumption (first there is the intertemporal substitution of consumption due to a change of the interest rate). Terms of trade worsen as a reaction to the higher domestic price level. The economy comes back to its steady state level quickly (about 3 years) due to a strong reaction of the central bank.

4 Conclusions

The estimated parameters reflect the Czech economy characteristics and the model seems to give quite suitable approximation of behavior with respect to the results. The model can be used to simulate monetary policy and help to analyze the policy conclusions to model uncertainty. Using the transmission mechanism of monetary policy, the monetary rule describes the inflation targeting behavior of the central bank. It is confirmed by the analysis of the impulse response function.

References

[1] Hurník, J. (2003): Nová keynesovská ekonomie aneb měnová politika bez peněz. In *Bankovnictví*, 18. 9. 2003, page 21, Section: Měna a regulace.

[2] Liu, P. (2005): A Small New Keynesian Model of the New Zealand Economy. Working Paper, Reserve Bank of New Zealand and The Australian National University.

[3] Gali, J., Monacelli, T. (2005): Monetary Policy and Exchange Rate Volatility in a Small Open Economy. Review of Economic Studies.

[4] Lubik, T., Schorfheide, F. (2005): A Bayesian look at new open economy macroeconomics. In *NBER Macroeconomics Annual 2005*, vol 20, NBER, MIT Press.

[5] Vašíček, O., Musil, K. (2006): Behavior of the Czech Economy: New Open Economy Macroeconomics DSGE Model. Working Paper 23/2006, Research Centre for Competitiveness of Czech Economy. ISSN 1801–4496.

[6] Musil, K. (2006): Konceptuální model malé uzavřené ekonomiky. In *MEKON 2006*, Ostrava, VŠB-TU Ostrava, ISBN 80-248-1013-1.

Testing of Hysteresis in Unemployment[1]

Daniel Němec Dalibor Moravanský

Department of Applied Mathematic and Computer Science
Faculty of Economics and Administration, Masaryk University
Lipová 61a, 602 00 Brno, Czech Republic
nemecd@econ.muni.cz, dalibor@econ.muni.cz

Abstract
Traditional approaches for testing and describing of hysteresis hypothesis are based on unit root tests and on various specifications of the Phillips curve. From this point of view, suggestions about roots of hysteresis phenomenon are related mainly to the process of a wage setting within the frame of the Insider-Outsider model. In this contribution, a little more general wage bargaining model is presented. From the theoretical point of view, we are able to decide whether hysteresis effect is the rule rather than the exception. Using macroeconomic data of the Czech Republic, parameters of this model are estimated and compared with empirical results within the framework of traditional approaches. Conclusions about the patterns of Czech unemployment are made.

Keywords
hysteresis, unemployment, Phillips curve, wage bargaining model
JEL: C2, E24, J6

1 Introduction

Hysteresis hypothesis is related to the idea that inertia in unemployment might be explained by the tendency for actual unemployment to cause an upward movement in equilibrium unemployment. Hysteresis is a property of dynamic systems. These systems are path-dependent systems because the long-run solution does not only depend on the long-run values of the exogenous variables but on initial conditions of each state variable.

Consider the one-dimensional linear case where a state variable X_t follows the following law of motion:

$$X_t = aX_{t-1} - Z_t, \tag{1}$$

where Z_t denotes the exogenous variable. There is a unique steady-state value $\bar{X}$ dependent solely on the steady-state level $\bar{Z}$, if a differs from unity (non-hysteric case):

$$\bar{X} = \frac{\bar{Z}}{a-1}, \tag{2}$$

If $a = 1$, there is not a unique steady-state value of X anymore. The long-run solution of equation (1) is given by:

$$X_t = X_0 - \sum_{i=1}^{t} Z_i, \tag{3}$$

where we consider time dependent values of Z rather than its steady-state level. In this case of hysteresis, any temporary disturbance in Z will have a permanent effect on X.

Applying general principles of hysteresis phenomenon on unemployment, we get some implications of hysteresis hypothesis in economy. These implications concern, in particular, the question of equilibrium unemployment and price stability. Thus, the main implications are as follows [2][3]:

- The growth rate of nominal wages depends solely on the variation in the unemployment rate. There are real rigidities in the economy.
- The economy does not converge to a stationary equilibrium independently of the initial conditions and a transitory shock has permanent effects on the economy.

[1]This contribution was supported by the Grant Agency of the Czech Republic, grant number 402/05/2172

- Non-accelerating inflation rate of unemployment (NAIRU) automatically follows in the path of the actual unemployment rate, and any rate of unemployment is consistent with steady inflation.
- Equilibrium rate of unemployment can be reduced if actual unemployment declines as the result, for example, of expansionary demand policy.

The items mentioned above are related to macroeconomical behaviour of the economy, and are important for both theoretical and practical aspects of the efficiency of fiscal and monetary policies. But what is in the background? Asking this question we want to know what are causes of hysteresis effect or, more precisely, the microfoundations of hysteresis hypothesis. The literature offers three principal explanations of this phenomenon:

- Insider–outsider hypothesis, where employed insiders do not take care about unemployed outsiders. Thus, wage setting does not take into account the level of unemployed persons. Moreover, unemployed workers cannot find jobs at lower wages because firms do not accept underbidding [5].
- Hypothesis of detoriation of human capital, where variations in nominal wages depend solely on the short-term unemployment, and not on the total stock of unemployed people.
- The role of capital stock means that adverse demand shock leads to a reduction in the capital stock if firms close plants and scarp the capital. This decrease of the capital stock leads to rising unemployment, which cannot be lowered immediately after the economy has recovered.

In our contribution we deal with the first explanation of them. Insider–outsider approach (developed by Blanchard and Summers [1]) answers the question how hysteresis may be generated. But this approach is restricted with a unique insiders behaviour. The insiders target employment equals the employment in previous period and wages are fixed according to this assumption. This behaviour may be seen as a specific case of a more general process of wage bargaining. For a better plausibility of hysteresis hypothesis, one may examine whether or not a more general wage setting model offers a possibility of hysteresis effect as well.

In this paper, a little more general wage bargaining model is presented and discussed. From the theoretical point of view, we will be able to decide whether a hysteresis effect is the rule rather than the exception. Using macroeconomic data of the Czech Republic, parameters of our models will be estimated.

2 Traditional Approaches

There are many approaches to examining the hysteresis hypothesis. The simplest of them is to check whether the unemployment rate contains a unit root. This approach is exactly in accordance with examining of the equation (1). But it is clear that the fact that unemployment data exhibit this feature, says nothing about the source of this phenomenon. The most common approach to investigating hysteresis effects is based on various specifications of the Phillips Curve. As an example, we can present the specification taken from Gordon [3].

Consider a simple version of the natural rate hypothesis relating inflation π_t and unemployment U_t:

$$\pi_t = \alpha\pi_{t-1} + \beta(U_t - U_t^*). \tag{4}$$

Hysteresis can arise when U_t^* depends on the lagged unemployment rate U_{t-1}, in addition to its microeconomic determinants represented by Z_t.

$$U_t^* = \eta U_{t-1} + Z_t. \tag{5}$$

After substituting equation (4) into (5) and transforming results, we get

$$\pi_t = \alpha\pi_{t-1} + \beta(1-\eta)U_t + \beta\eta(U_t - U_{t_1}) - \beta Z_t. \tag{6}$$

Full hysteresis occurs when $\eta > 0$ implying that there is no longer a unique U_t^*. Another variants of the Phillips curve operate with wage inflation. Of course, there is a lack of microeconomical foundations for hysteresis too. To deal with this problem, a wage bargaining model must be developed and analyzed.

3 A Wage Bargaining Model

The model presented in this section is based on the methodology of Stiassny [8]. This model is able to examine hysteresis hypothesisy and beyond this it can be helpful at leat (from the empirical point of view) to discuss properties of the labor market.

In our model, the bargainers are firms and unions. But it must be noted that bargaining process concerns all employees and it is not necessary for bargaining to be organized in unions. Interests groups for bargaining may be formed by all insiders at an informal basis, too. The reasons for this may be found in [5] because insider-outsider relations may be applied on the insider-insider relations as well.

The wage setting process is divided into two stages. In the first stage there is bargaining between unions and firms for the next period's wage w_{t+1}. In the second stage, the wage is given and the firms decide about the employment unilaterally. We suppose that unions want to maximize a quadratic utility function. The role of this form of utility function is to provide weights, which are given, respectively, to the unemployment and wage target. In following equations, these weights are given by the parameter α.

$$U_{t+1} = -(\Delta w_{t+1} - \Delta w^*_{t+1})^2 - \alpha(u^e_{t+1} - u^*_{t+1})^2, \tag{7}$$

with

$$\begin{aligned}\Delta w^*_{t+1} = \quad & \Delta p^e_{t+1} + \Delta prod^e_{t+1} - \eta\Delta(1 - t^e_{t+1}) \\ & -\zeta(w_t - p_t - prod_t + \eta(1 - t_t) - s).\end{aligned} \tag{8}$$

Except for the unemployment rate u, lower case letters represent logs. In our case, the next period's utility depends on the deviations of wages and unemployment from their target values marked with asterisks. The target for gross wages is formulated as an error correction mechanism, where Δp^e_{t+1} denotes expected inflation rate in the next period, $\Delta prod^e_{t+1}$ expected growth in productivity, $\Delta(1 - t^e_{t+1})$ refers to expected change of employee's tax ($(1 - t)$ is defined as $ln(1 + T)$), η corresponds to tax elasticity of target wage ($0 \leq \eta \leq 1$). Scale parameter s depends on the desired income distribution between wages and profits. If ζ is zero, the unions formulate their target wage only in differences. If ζ is unity, the target wage is formulated only in levels. Depending on the value of ζ, the wage target can deviate from the long-run target $(p + prod - \eta(1 - t) + s)$ in the short run. The scale parameter will be ignored because we can assume that income is distributed between wages and profits uniformly over time.

Firms maximize their profits

$$\Pi_{t+1} = \Pi(w_{t+1}, \ldots). \tag{9}$$

The negotiated wage is the solution of the following Nash bargaining problem [6] [7]:

$$\begin{aligned}\max_{w_{t+1}} : [\quad - \quad & (\Delta w_{t+1} - \Delta p^e_{t+1} - \Delta prod^e_{t+1} + \eta\Delta(1 - t^e_{t+1}) \\ & +\zeta(w_t - p_t - prod_t + \eta(1 - t_t)))^2 \\ & -\alpha(u^e_{t+1} - u^*_{t+1})^2 - d_U]^\lambda \times [\Pi(w_{t+1}, \ldots) - d_F]^{1-\lambda},\end{aligned} \tag{10}$$

subject to labor demand:

$$\begin{cases} n_t = & \delta n_{t-1} - (1 - \delta)\delta_0(w_t + (1 + \tilde{t}_t) - p_t - prod_t) \\ & +(1 - \delta)e_t, \\ u_t \equiv & l_t - n_t, \end{cases} \tag{11}$$

where the value of (d_U, d_F) represents disagreement point (the utility and the profit achieved by unions and firms in the case when bargaining has failed) and λ defines the relative bargaining power of the insiders. In the equation of labor demand, n denotes employment and l labor force. The relevant wage for firms is W times $(1 + \tilde{T})$, where $\tilde{T}$ represents the employer's tax rate (employer's contribution to social insurance). All other shocks are represented by e_t. The parameter δ corresponds to an inertia in employment ($0 \leq \delta < 1$). The parameter $(1 - \delta)$ can be interpreted as a rate of voluntary quitting [4]. "True wage"' elasticity of labor demand may be expressed by the parameter δ_0.

We are able to derive an expression for expected unemployment by taking first difference of equation (11) and by applying the expectation operator.

$$\begin{aligned}u^e_{t+1} = \quad & u_t + \delta\Delta u_t + (1 - \delta)\delta_0(\Delta w_{t+1} + \Delta(1 + \tilde{t}^e_{t+1}) - \Delta p^e_{t+1} - \Delta prod^e_{t+1}) \\ & +\Delta l^e_{t+1} - \delta\Delta l_t - (1 - \delta)\Delta e^e_{t+1}\end{aligned} \tag{12}$$

Now, we can solve the Nash bargaining problem (10). The first order condition for w_{t+1} is:

$$\begin{aligned}\Delta w_{t+1} = \quad & \Delta p^e_{t+1} + \Delta prod^e_{t+1} - \eta\Delta(1 - t^e_{t+1}) - \alpha(1 - \delta)\delta_0(u^e_{t+1} - u^*_{t+1}) \\ & -\zeta(w_t - p_t - prod_t + \eta(1 - t_t)) + \frac{1}{2}K,\end{aligned} \tag{13}$$

with K equal to

$$\frac{1 - \lambda}{\lambda}\frac{d\Pi_{t+1}}{dw_{t+1}}\frac{(U_{t+1} - d_U)}{(\Pi_{t+1} - d_F)} \tag{14}$$

We assume that unions do not take into account that their wage setting may affect the general price level. Expression (14) depends on relative unions power λ and on their relative position in the case of disagreement and on the elasticity of profits on wages $\epsilon_{\Pi,W}$, if $d_F = 0$. The value of K is highly dependent on an institutional framework inside the economy. Thus this value will be changing only very slowly. We assume K to be approximately constant. In the case of monopoly union ($\lambda = 1$), K vanishes. Combining equations (12) and (13) we get:

$$\begin{aligned}\Delta w_{t+1} = \quad & \Delta p^e_{t+1} + \Delta prod^e_{t+1} - \frac{\eta\Delta(1-t^e_{t+1}) + \alpha((1-\delta)\delta_0)^2\Delta(1+\tilde{t}^e_{t+1})}{1+\alpha((1-\delta)\delta_0)^2} \\ & -\frac{\alpha(1-\delta)\delta_0}{1+\alpha((1-\delta)\delta_0)^2}(u_t + \delta\Delta u_t - u^*_{t+1}) \\ & +\frac{1}{2(1+\alpha((1-\delta)\delta_0)^2)}K - \left(\frac{\zeta}{1+\alpha((1-\delta)\delta_0)^2}\right) \\ & \times(w_t - p_t - prod_t + \eta(1-t_t)) + k\end{aligned} \tag{15}$$

with

$$k = \frac{\alpha(1-\delta)\delta_0}{1+\alpha((1-\delta)\delta_0)^2}((1-\delta)\Delta e^e_{t+1} - \Delta l^e_{t+1} + \delta\Delta l_t) \tag{16}$$

We assume that insiders target unemployment u^*_{t+1} is formulated as follows:

$$u^*_{t+1} = \mu u^{nat}_t - (1-\mu)\sum_{i=0}^{m}\gamma_i u_{t-i}. \tag{17}$$

In this expression, the sum of coefficients γ_i is equal to one and $0 \leq \mu \leq 1$. We suppose that unions pay attention to the labor market equilibrium unemployment rate u^{nat}_t and to a weighted sum of previous unemployment rates. If unemployment remains high for some time, the unions accustom to higher rates of unemployment. The parameter μ determines an influence of the natural rate of unemployment on the unemployment target. Inserting the target unemployment rate (17) into (15), we obtain:

$$\begin{aligned}\Delta w_{t+1} = \quad & \Delta p^e_{t+1} + \Delta prod^e_{t+1} \\ & -\frac{\eta\Delta(1-t^e_{t+1}) + \alpha((1-\delta)\Delta_0)^2\Delta(1+\tilde{t}^e_{t+1})}{1+\alpha((1-\delta)\delta_0)^2} \\ & -\frac{\alpha(1-\delta)\delta_0}{1+\alpha((1-\delta)\delta_0)^2}\left(\textstyle\sum_{i=0}^m g_i u_{t-i} - \mu u^{nat}_t\right) \\ & +\frac{1}{2(1+\alpha((1-\delta)\delta_0)^2)}K - \left(\frac{\zeta}{1+\alpha((1-\delta)\delta_0)^2}\right) \\ & \times(w_t - p_t - prod_t + \eta(1-t_t)) + k\end{aligned} \tag{18}$$

Now, the sum of coefficients g_i equals to μ because the expression (15) implies: $u_t - \delta\Delta u_t - u^*_{t+1} = u_t + \delta u_t - \delta u_{t-1} - \sum_{i=0}^m \gamma_i u_{t-i} + \mu\sum_{i=0}^m \gamma_i u_{t-i} \Rightarrow \sum_{i=0}^m g_i = 1 + \delta - \delta - \sum_{i=0}^m \gamma_i + \mu\sum_{i=0}^m \gamma_i = \mu$.

Not all variables are observable. Therefore the basis for our empirical estimations is the following equation:

$$\begin{aligned}\Delta w_{t+1} = \quad & a\Delta p^e_{t+1} + b\Delta prod^e_{t+1} + \frac{\eta\Delta(1-t^e_{t+1})}{1+\alpha((1-\delta)\delta_0)^2} \\ & -\frac{\alpha((1-\delta)\delta_0)^2\Delta(1+\tilde{t}^e_{t+1})}{1+\alpha((1-\delta)\delta_0)^2} - \frac{\alpha(1-\delta)\delta_0}{1+\alpha((1-\delta)\delta_0)^2}(1-\mu)\Delta u_t \\ & -\frac{\alpha(1-\delta)\delta_0}{1+\alpha((1-\delta)\delta_0)^2}\mu u_t - \left(\frac{\zeta}{1+\alpha((1-\delta)\delta_0)^2}\right) \\ & \times(w_t - p_t - prod_t + \eta(1-t_t)) + v_t(\mu u^{nat}_t, constant, \ldots) + \epsilon_t\end{aligned} \tag{19}$$

Sum of coefficients of the u_t's equals to μ, v_t is an intercept term. The parameters a and b should be near to one. The coefficient at $prod^e_{t+1}$ might be less than one if unions smooth productivity fluctuations. The values of the parameters μ and ζ are of key concern for the validity of hysteresis hypothesis.

4 Model Properties

In this section, we will answer the question whether this bargaining model implies hysteresis. The "particular"' solution of the dynamic system defined by the wage setting equation (18) and the labor demand (11) is

$$\begin{pmatrix} 1 & 1 & \frac{\alpha(1-\delta)\delta_0\mu}{\Omega} & \frac{-\alpha(1-\delta)\delta_0\mu}{\Omega} & \frac{-\eta\zeta}{\Omega} & \frac{-\alpha(1-\delta)\delta_0^2\mu}{\Omega} & \frac{\alpha(1-\delta)\delta_0\mu}{\Omega} \\ 0 & 0 & \frac{-\zeta}{\Omega} & \frac{\zeta}{\Omega} & \frac{-\eta\zeta\delta_0}{\Omega} & \frac{\zeta\delta_0}{\Omega} & \frac{\alpha(1-\delta)\delta_0^2\mu}{\Omega} \end{pmatrix} \tag{20}$$

$$\times \begin{pmatrix} p & prod & e & l & (1-t) & (1+\tilde{t}) & u^{nat} \end{pmatrix}^T,$$

where $\Omega = \zeta + \alpha\delta_0^2\mu - \alpha\delta_0^2\delta\mu$. The first row in equation (20) represents the long run solution for the real wage, and the second row the long run solution for the unemployment, which equals to the NAIRU. The solution for u indicates that shocks in the growth rates in e, l or t lead to permanent effects onto u (and thus onto the NAIRU) whenever $\zeta > 0$. Thus we can say that this bargaining model leads to hysteresis effects regardless what is the value of the unemployment-target parameter μ. If $\zeta = 0$ then the long run solution will be:

$$\begin{pmatrix} 1 & 1 & \frac{1}{\delta_0} & \frac{-1}{\delta_0} & 0 & -1 & \frac{1}{\delta_0} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{21}$$

If union's wage target is solely formulated in differences ($\zeta = 0$), there are no hysteresis effects. The NAIRU equals to natural rate of unemployment. Moreover, setting $\mu = 0$, we get the long-run solution for unemployment independently on its natural rate. In this case, it is clear from equations (18) or (19) that wage setting does not depend on the unemployment level. This is the most typical property of hysteresis in unemployment.

The main conlusion of this section may be such that hysteresis effect is the rule rather than the exception for our bargaining model. The steady state unemployment is compatible with any level of unemployment and inflation. Only when the level of real wages plays no role in wage bargaining and $\mu > 0$, there is no hysteresis effect. And the NAIRU is restricted by the natural rate of unemployment.

It should be noted that this variants emphasize two sources of the hysteresis phenomenon. If $\mu = 0$ then we interpret this in such a way that unions and the public accomodate to higher unemployment rates when unemployment remains high for some time. A "natural rate of unemployment" is thus variable in determining of unemployment-target. If the parameter ζ is significant, then the source of hysteresis is due to the fact that unions strive for a long-run real wage target. If unemployment rises, unions are willing to accept a deviation from their wage target (depending on the parameter α in their utility function, equation (7)), but as long as $\alpha < \infty$, not to a sufficient extent. Therefore, unemployment does not return to its previous natural level.

5 Empirical Results

In this section, we estimate equations (6) and (19) along with some other variants of them. As a first step, we test the stationarity properties of unemployment rate. We test for unit roots using augmented Dickey-Fuller test. Table 1 reports the results. The null of a unit root can not be rejected for the unemployment rate even at the 10 percent significance level.

Table 1: Augmented Dickey-Fuller test for unit root

ADF t-statistic	**Number of lags**	**AR(1) estimate**
-0.8702	1	0.1873
	Critical values	
1%	5%	10%
-3.640	-2.949	-2.616

As a next step, we estimate equation (6) using quarterly data for net inflation and for the level of unemployment from the first quarter 1994 to the last quarter 2005. Many specifications have been made. Two best estimations are as follows. The first estimation was made using π_t as the quarterly inflation (version 1), the second estimation used π_t as the yearly inflation (version 2). Quarterly differences in unemployment were used in both cases. Our results are shown in Table 2.

We can identify hysteresis in both specifications. In the version 2, we can see the presence of full hysteresis because the estimated coefficient for level effects is zero ($\eta = 1$). Estimated parameters η in both versions of the model imply the high path dependence of U^* on history of actual unemployment.

Table 2: Regression results of Phillips Curve equation

	Version 1		**Version 2**	
Parameter	*Estimation*	*t-stat*	*Estimation*	*t-stat*
$-\beta Z_t$	$\mathbf{1.9600}$	3.9526	$\mathbf{3.4021}$	2.4993
α	$\mathbf{0.3048}$	2.0859	$\mathbf{0.6959}$	6.4965
$\beta(1-\eta)$	$\mathbf{-0.1694}$	-3.4107	-0.2964	-2.2996
$\beta\eta$	$\mathbf{-0.8530}$	-2.4186	$\mathbf{-1.6463}$	-2.6953
R^2	0.6354		0.8971	
Durbin-Watson	1.6879		1.021	
Re-computation of the parameters				
β	-1.0224		-1.9427	
η	0.8343		0.8474	
Z_t	1.9170		1.7512	

Statistically significant coefficients at 5 percent level are bold.

To estimate the parameters in wage bargaining model, we have used the quartely data from the first quarter 1995 to the last quarter 2005 (all variables, except unemployment rates, are in logs). Wage bargaining is made every year (not necessary at the beginning of the year) and the achieved agreement serves for this period. Thus we will use year to year differences. This procedure solves the problem of seasonal adjustment of our data. Another problem occurs because of the same development of employee's and employer's tax rates (rates of social contributions). We use only one of them to avoid this problem. Adaptive expectations are used for changes in the price level.

We estimate equation (19) and its various specifications using the hypothesis of constant coefficients (by ordinary least squares). This estimation method is sufficient enough because our estimation period is quite short and we do not consider important changes in the parameters. We are especially interested in coefficients at Δu_t and u_t which are important for evaluting the value and significance of μ (the "unemployment target" parameter). The coefficient of the error correction term is of a great importance, too. According to previous theoretical discussion, both are necessary for identification of hysteresis.

The main results are presented in Table 3. This table contains parameters estimation at various specifications of equation (19) without the error correction term, thus in this case we assume that ζ equals zero. Re-estimation with the error correction term did not affect values of previously estimated parameters and the parameter of this term has been insignificant (and relativ small). The coefficient of u_t was not statistically significant in any of the eight versions (including the versions with the error correction term). This result leads to the conclusion that hysteresis in the Czech Republic matters. The value of the parameter ζ may be considered as zero. It may be surprising that in the Czech republic the bargaining process is not related to the level of wages. This might be explained by considering the income distribution between wages and profits as passable and hence there is no tendency to change it.

Table 3: Regression results without error correction term

Version	1a	2a	3a	4a
Dep. variable		Δw_{t+1}		
Intercept v	0.0522	$\mathbf{0.1002}$	$\mathbf{0.1051}$	0.0551
Δp_t^e	$\mathbf{0.4337}^*$	0.2595	0.2632	$\mathbf{0.4444}^*$
$\Delta prod_t^e$	$\mathbf{0.8251}$	$\mathbf{0.5496}^*$	0.3634	$\mathbf{0.6574}$
Δu_t	-0.0047	$\mathbf{-0.0066}^*$	$\mathbf{-0.0123}$	$\mathbf{-0.0100}$
u_t	-0.0016	-0.0054	-0.0045	-0.0005
Δt	—	$\mathbf{-2.5034}^*$	$\mathbf{-2.6093}^*$	—
Δl_t	—	—	$\mathbf{1.8389}^*$	$\mathbf{1.7685}^*$
R^2	0.71	0.74	0.76	0.74
Durbin-Watson	1.51	1.53	1.69	1.64

Statistically significant coefficients at 5 percent level are bold, at 10 percent level bold with asterisk.

Although it seems that changes in taxes play important role in wage bargaining, it must be noted that changes in taxes were very small during the whole investigated period (only one little change occured at the beginning of 1996). Thus the version (4a) in Table 3 is the most relevant for our interpretations.

The period from 1995 to 2005 is characterized by the decreasing rate of inflation, especially from the end of

1990s. From this point of view, it is obvious that unions with adaptive expectations have taken into account this fact. The value of Δp_t^e 0.44 reflects this reality. The value of the productivity parameter $\Delta prod_t^e$ is statistically significant and it is quite high on the whole but it is not unity as might be expected. The reason for this may be that productivity is hard to measure exactly. Unions may thus take only 65% of expected productivity growth into wage bargaining because the true productivity growth is smaller than the expected one. Changes in employment are significant to explain the wage growth, too. One percent change in previous labor force leads to two percent change in nominal wage rate of growth. This may be interesting because one migth expect the opposite sign here. Insiders should moderate their wage pressures regarding rising stock of labor suppliers. But the positive sign is a mark indicating the power of insiders and the plausibility of insider-outsider hypothesis. The intercept term contains especially unobservable variables. Its interpretation in respect to wage bargaining may be the base for this process. The basic claim of unions is to have the annual wage growth of 5.5%, which is further reduced or sharped in considerations of expected economic development of other relevant variables.

The necessary condition for a plausibility of hysteresis hypothesis is $\mu = 0$ because ζ equals zero. This condition is satisfied. Wage growth does not depend on the number of the unemployed people. It depends on the variation in the unemployment rate only. Long run unemployment is not related to the natural rate of unemployment, so any rate of unemployment is consistent with steady inflation. Therefore, we can accept hysteresis hypothesis for the Czech Republic.

6 Conclusions

In our paper, we examined various tests of hysteresis in unemployment, including a wage bargaining model. This model is very useful for testing hysteresis hypothesis. The results based on this model are more plausible because of microeconomic foundations standing in the background. From the theoretical point of view, this model shows that hysteresis is rather the rule than the exception.

Using the macroeconomic data of the Czech Republic, our estimations shows that there are strong hysteresis effects. Unit roots approach and Phillips curve approach confirm the validity of hysteresis hypothesis in the Czech Republic. We have investigated two specifications of wage bargaining model, too. Our empirical results imply insignificance of the error correction term. Thus, it is clear that insiders in the Czech Republic do not concern with level of wages but they are interested only in the wage growth. The source of hysteresis is therefore caused by the unemployment-target parameter μ. This parameter was insignificant and not clearly different from zero. This result may be interpreted in such a way that unions and the public become accustomed to higher unemployment rates when unemployment level remains high for some time. It is therefore incorrect the opposite opinion that Czech unions follow a policy of preserving ”full” employment (represented approximately by the natural rate of unemployment).

Because the hysteresis is present, we can accept that unemployment could be reduced by using expansionary policies without negative inflation consequences. The NAIRU is thus compatible with any level of unemployment as the hysteresis hypothesis suggests.

References

[1] Blanchard, O. J., Summers, L. H. (1986) Hysteresis in Unemployment. NBER Working Paper Series, Working Paper No. 2035, October

[2] Cahuc, P., Zylbeberg, A. (2004) Labor Economics. The MIT Press, Cambridge, Massachusetts, ISBN 0-262-03316-X

[3] Gordon, R.J. (1989) Hysteresis in History: Was There Ever a Phillips Curve? *The American Economic Review*, Vol. 79, No. 2, May, 220–225

[4] Layard, R., Nickell, S., Jackman R (2005) Unemployment: Macroeconomic Performance and the Labour Market. Oxford University Press, New York, second edition, ISBN 0199-27917-9

[5] Lindbeck, A., Snower, D. J. (1986) Wage Setting, Unemployment, and Insider-Outsider Relations. *The American Economic Review*, Vol. 76, No. 2, May, 235–239

[6] Nash, J. F. (1950) The Bargaining problem, *Econometrica*, Vol. 18, No. 2, 155–162

[7] Nash, J. F. (1953) Two-Person Cooperative Games, *Econometrica*, Vol. 21, No. 1, 128–140

[8] Stiassny, A. (1998) Wage Setting and Hysteresis in Unemployment. *Empirica*, Vol. 25, 79–107

Investment under Monetary Uncertainties

Václava Pánková[1]
VŠE Praha, W.Churchilla 4, 130 67 Praha 3
pankova@vse.cz

Abstract

Optimizing of an investment decision involves uncertainty about future rewards from the investment as its implicit constraint. As a consequence, there is an evidence that investment is sensitive to volatility and uncertainty over the economic environment. A small open economy with transition characteristics aspires to attract big investments from abroad, that is why inflation uncertainty and / or exchange rate uncertainty can play an important role. Though both this variables usually are controlled by a National Bank, their future values are not known.

It is generally assumed that such uncertainties have a negative effect on investment. Nevertheless, there is also an influence of facts as irreversibility of investment decisions and opportunity cost of possibility to wait rather then to invest. So, apart from transaction motives, a speculative motive can also take place here. That is why impacts of monetary uncertainties can differ according to the type of industry.

Monetary uncertainties are rising from a volatility of relevant variable, a value of which, though not observable, can be anticipated as its permanent part. The permanency is supposed to subject an adaptive expectation process. After a formalization of permanent inflation, respective exchange rate, their influence on investment in CR is estimated. The fifteen branches of the Czech industry exhibit different responses to common macroeconomic determinants.

Keywords

investment, monetary uncertainties, panel data
JEL: C23, D21, E58

1 Investment and economic uncertainties

The investment behavior of a firm is supposed to be formalized as an optimizing problem

$$V_t \rightarrow MAX$$

subject to

$$I_t - \delta K_t = K_t^*,$$

the solution of which gives an optimal investment.

In the formulation, $V_t = \sum_{\tau=0}^{\infty} \beta_{t+\tau} [\pi(K_{t+\tau}) - P_{t+\tau} h(I_{t+\tau})]$ is a firm´s value, when firm with capital K_t has a profit $\pi(K_t)$ from which it is necessary to subtract investment $P_t\, h(I_t)$, P_t being a unit production price, $\beta_{t+\tau} = \prod_{\theta=1}^{\tau} (1 + r_{t+\theta})^{-\tau}$, r is an interest rate.

Optimizing of an investment decision also involves uncertainty about future rewards from the investment as its implicit constraint. As a consequence, there is an evidence that investment is sensitive to volatility and uncertainty over the economic environment. Usually, uncertainties are rising in monetary characteristics as inflation, interest rate or exchange rate. As the uncertainties in the economic environment are important determinants of investment, their nature and impact are in focus of recent studies.

Capital as one of the most important productive inputs can be characterised by a certain capital mobility, a degree of which is influencing an economic growth. A more open capital account shows

[1] Financial support of GACR 402/04/0756 is gratefully acknowledged by the author

out a higher productive performance than economies with restricted capital mobility. A small open economy with transition characteristics tends to be an acceptor of capital and aspires to attract big investments from abroad, that is why inflation uncertainty and / or exchange rate uncertainty can play an important role. Though both this variables are controlled by a National Bank, their future values are not known. It is generally assumed that such uncertainties have a negative effect on investment. Nevertheless, there is also an influence of facts as irreversibility of investment decisions and opportunity cost of possibility to wait rather then to invest. So, apart from transaction motives, a speculative motive can also take place here. That is why impacts of monetary uncertainties can differ according to the type of industry.

Theoretically, an uncertainty can be understood as a temporary component of relevant variable, the other component being its permanent part. An evidence of different effects from permanent and temporary changes is referred e.g. in Byrne and Davis (2004). An alternative approach introducing an uncertainty as a discount factor of future prices is given e.g. in [3].

2 Permanency as a part of a model

Monetary uncertainties are rising from a volatility of relevant variable, a value of which, though not observable, can be anticipated as its permanent part. The permanency is supposed to subject an adaptive expectation process, details e.g. in [2].

A variable X is supposed to split in two unobservable parts: a permanent one and a temporary one

$$X_t = X_t^P + X_t^T .$$

The permanent value is anticipated to subject an adaptive expectation process as

$$\Delta X_t^P = X_t^P - X_{t-1}^P = \lambda(X_t - X_{t-1}^P) \quad \text{with} \quad 0 \le \lambda \le 1 .$$

It means

$$X_t^P = \lambda X_t + (1-\lambda) X_{t-1}^P$$

with the following interpretation. In year t a permanent value is a weighted average of an actual one and a previous permanent value. The previous permanent value follows the same schema, so

$$X_{t-1}^P = \lambda X_{t-1} + (1-\lambda) X_{t-2}^P \qquad \text{a. s. o.}$$

By a substitution we then have

$$X_t^P = \lambda X_t + \lambda(1-\lambda) X_{t-1} + \lambda(1-\lambda)^2 X_{t-2} + \lambda(1-\lambda)^3 X_{t-3} + ... \qquad (1)$$

what means that a current value has the greatest weight and the weights decline steadily by going back in the past.

Then, we can estimate a model

$$Y_t = \beta_0 + \beta_1 X_t^P + u_t \qquad (2)$$

in variants. Constructing (1) under different choice of λ between zero and one ($\lambda = 0.1, 0.2, \ldots, 0.9$), we compute (2). We than choose such a λ which produces a best fit of (2) according to the R-squared.

3 Application to the Czech industry

The fifteen branches of the Czech industry are studied. After a formalization of permanent inflation, respective exchange rate, their influence on investment in CR is estimated. As a common scheme

$$I = \beta_0 + \beta_1 X_1^P + \ldots + \beta_j X_j^P + \beta_{j+1} W_{j+1} + \ldots + \beta_k W_k + u \qquad (3)$$

can be written with j permanent values of monetary variables with uncertainties in question and k-j other relevant exogenous variables as e.g.. level of wages, GDP per capita, a.s.o. In (3), I as an investment is an endogenous variable, β_0 is a constant and β_i´s are parameters of an econometric model. To demonstrate an influence of the common economic environment on different industrial branches the seemingly unrelated regression will be appropriate here to get individual sets of parameters under an assumption of correlated disturbances. Thus, an eventual diversity of monetary uncertainties impacts could be proved. Unfortunately, only four years of data (1999 – 2003) observations were available in the sources of ČSÚ (Czech Statistical Bureau), that is why the model was dramatically restricted and an other estimation method used. So, existing in the same economic environment .the investment in an industrial branch is exposed by the same but only one permanent value of an X^P variable

$$I = \beta_0 + \beta_1 X^P + u$$

and W´s are dropped. A technique of panel data (with 60 observations) pooled regression was used which allows at least for distinguishing in a constant β_0.

For a quick survey, directions of deviations from a mean (in parentheses) are given in Table 1. Constructing permanent exchange rate CZK/EUR according to (1) with five lags, λ = 0.5 was found as giving optimal results (highest R- squared by valid t – tests). Repeating the same principal by using permanent inflation as an exogenous variable, λ =0.7

industry	*inflation rate (3433,26)*	*exchange rate (2719,13)*
Mining and quarrying of energy producing materials	+	-
Mining and quarrying except energy producing	-	-
Food products, beverages, tobacco	+	+
Textile and textile products	-	-
Wood and wood products	-	-
Pulp. paper and paper products, printing	-	-
Chemicals and chemical products	-	-
Rubber and plastic products,.	+	+
Other nonmetallic and mineral products	-	+
Basic metals and fabricated metal products	-	+
Machinery and equipment	+	-
Electrical and optical equipment	+	+
Transport equipment	+	+
Other industry (n.e.c.)	+	+
Electricity, gas and water supply	-	-

Table 1.

Conclusions

Investments in the Czech industry, especially foreign investments, are not coming equally to all industrial branches. It can be taken for granted, that the investors use to study the economic conditions and make their expectations about economic environment. Also their timing is well-considered. A hypothesis of different impacts of monetary uncertainties into industrial branches, which theoretically should be a consequence of such a behaving, seems to be validated for the Czech industry when followed in the beginning of this decade. Nevertheless, the small time dimension of the data sample does not allow to state it as a really confirmative result.

References

[1] Byrne, J. P., Davis, E.P.: Permanent and temporary inflation uncertainty and investment in United States, Elsevier, Economic Letters 85, pp 271 – 277, 2004

[2] Dougherty, Ch.: Introduction to Econometrics, Oxford Univ. Press, 1992

[3] Hallert, A. H., Peersman, G., Piscitelli, L.: Investment under Monetary Uncertainty: A Panel Data Investigation, Bank of England Working Paper, 2003

[4] ČSÚ, internet

Appendix

Using panel data and pooled OLS regression with a group distinction, we get by PcGive:

(i) For z5 = permanent exchange rate

DPD - Modelling invest by OLS

sigma	2051.568	sigma^2	4208932
R^2	0.9426599		
no. of observations	60	no. of parameters	16
Using robust standard errors			
constant:	yes	time dummies:	0
group dummies:	14	time*group:	0
number of individuals	15 (derived from year)		
longest time series	4 [1 - 4]		
shortest time series	4 (balanced panel)		

(ii) w7 = permanent inflation

DPD - Modelling invest by OLS

sigma	2035.807	sigma^2	4144510
R^2	0.9435376		
no. of observations	60	no. of parameters	16
Using robust standard errors			
constant:	yes	time dummies:	0
group dummies:	14	time*group:	0
number of individuals	15 (derived from year)		
longest time series	4 [1 - 4]		
shortest time series	4 (balanced panel)		

The Fuzzy Weighted Average Operation in Decision Making Models

Ondřej Pavlačka, Jana Talašová
Department of Mathematical Analysis and Applied Mathematics,
Faculty of Science, Palacký University Olomouc,
tř. Svobody 26, 771 46 Olomouc, Czech Republic,
onpav@centrum.cz, talasova@inf.upol.cz

Abstract

The paper deals with the operation of fuzzy weighted average of fuzzy numbers. The operation can be applied to the aggregation of partial fuzzy evaluations in the fuzzy models of multiple-criteria decision making and to the computation of the expected fuzzy evaluations of alternatives in the discrete fuzzy-stochastic models of decision making under risk. Normalized fuzzy weights figuring in the operation have to form a special structure of fuzzy numbers; its properties will be studied. The practical procedures for setting the normalized fuzzy weights will be shown. The operation of a fuzzy weighted average of fuzzy numbers will be defined and an effective algorithm of its calculation will be described.

Keywords: Multiple-criteria decision making, decision making under risk, normalized fuzzy weights, fuzzy weighted average, fuzzy weights of criteria, fuzzy probabilities

JEL: C44

1 Introduction

The standard operation of a weighted average is used for aggregating the partial evaluations of alternatives in the most of multiple-criteria decision making models and for computing the expected evaluations of alternatives in the discrete models of decision making under risk. The applied normalized weights, i.e. non-negative real numbers whose sum is equal to one, express different importance of considered criteria in the first case and probabilities of states of the world in the second case.

Whereas the probabilities of states of the world have a unique mathematical meaning, weights of criteria are understand differently in different models. The most general definition says that the weights of criteria are non-negative real numbers whose ordering expresses the importance of criteria. According to the definition, the weights mean the measurements of criteria importance that are defined on an ordinal scale. But in most of the multiple-criteria decision making models, the weights of criteria represent some kind of cardinal information concerning the importance of criteria. In the model of multiple-criteria decision making that is described in [7] the weights of criteria represent shares of the corresponding partial objectives of evaluation in the overall one. This concept of criteria weights will be considered in this paper.

The weights of criteria as well as the probabilities of states of the world are usually set subjectively, i.e. they are more or less uncertain. In this paper, it will be shown how this kind of uncertain information can be expressed by means of the tools of the fuzzy sets theory. A special structure of fuzzy numbers, so called normalized fuzzy weights, will be introduced for that purpose. The operation of a fuzzy weighted average of fuzzy numbers, where the normalized fuzzy weights are used, will be defined. An effective computing algorithm of the fuzzy weighted average will be described.

2 Applied Notions of the Fuzzy Sets Theory

A fuzzy set A on a universal set X, $X \neq \emptyset$, is given by its membership function $A : X \to [0,1]$. For any element $x \in X$, $A(x)$ is called the membership degree of the element x to the fuzzy set A. A set $Supp\ A = \{x \in X \mid A(x) > 0\}$ is called a support of A. Sets $A_\alpha = \{x \in X \mid A(x) \geq \alpha\}$, $\alpha \in (0,1]$, are α-cuts of A. A set $Ker\ A = \{x \in X \mid A(x) = 1\}$ is called a kernel of A. A fuzzy set A is called normal if $Ker\ A \neq \emptyset$.

A fuzzy number is defined as a fuzzy set C on the set of all real numbers $\Re$ that fulfils the following conditions: a) C is a normal fuzzy set, b) C_α, $\alpha \in (0,1]$, are closed intervals, c) $Supp\ C$ is a bounded set. It can be proved (see [4]) that the membership function $C(x)$ is upper semicontinuous and that there exist real numbers $c^1 \leq c^2 \leq c^3 \leq c^4$ (so called significant values of C) such that $[c^1, c^4] = Cl(Supp\ C)$, where $Cl(Supp\ C)$ means the closure of $Supp\ C$, $[c^2, c^3] = Ker\ C$, $C(x)$ is non-decreasing for $x \in [c^1, c^2]$ and non-increasing for $x \in [c^3, c^4]$. A fuzzy number C is said to be defined on $[a,b]$, if $Supp\ C \subseteq [a,b]$.

For a fuzzy number C, let us denote $C_\alpha = [\underline{c}(\alpha), \overline{c}(\alpha)]$ for any $\alpha \in (0,1]$, and $Cl(Supp\ C) = [\underline{c}(0), \overline{c}(0)]$. Then it can be seen that each fuzzy number C is determined not only by its membership function $C(x)$, $x \in \Re$, but also by the couple of real functions $(\underline{c}(\alpha), \overline{c}(\alpha))$, $\alpha \in [0,1]$. The functions $\underline{c}$ and $\overline{c}$ are the pseudo-inverse functions to the functions C_L and C_R that represent restrictions of the membership function C to the closed intervals $[c^1, c^2]$ and $[c^3, c^4]$, respectively.

If $\underline{c}(\alpha) = c = \overline{c}(\alpha)$ for all $\alpha \in [0,1]$, then C represents a real number; $C = c$, $c \in \Re$. A fuzzy number C is called symmetric, if $\frac{\underline{c}(\alpha)+\overline{c}(\alpha)}{2}$ is a constant function of α, $\alpha \in [0,1]$. A fuzzy number C is called linear, if $\underline{c}(\alpha)$, $\overline{c}(\alpha)$, $\alpha \in [0,1]$, are linear functions. Any linear fuzzy number C is fully determined by its significant values $c^1 \leq c^2 \leq c^3 \leq c^4$; its functions $\underline{c}$, $\overline{c}$ are given as follows: $\underline{c}(\alpha) = c^1 + \alpha(c^2 - c^1)$ and $\overline{c}(\alpha) = c^4 - \alpha(c^4 - c^3)$ for any $\alpha \in [0,1]$. For $c^2 \neq c^3$ the linear fuzzy number C is often called trapezoidal, for $c^2 = c^3$ triangular.

For any fuzzy number C, the non-increasing, non-negative function $\sigma_C(\alpha) = \overline{c}(\alpha) - \underline{c}(\alpha)$, $\alpha \in [0,1]$, is called a span function of C. It describes in details, for any membership degree α, the uncertainty of C. The real numbers $\sigma_C(1)$, $\sigma_C(\alpha)$ and $\sigma_C(0)$, are called the span of the kernel, the span of the α-cut and the span of the support, respectively. Obviously, $\sigma_C(\alpha) = 0$ for all $\alpha \in [0,1]$ if and only if C is a real number. It holds $\int_0^1 \sigma_C(\alpha)\, d\alpha = \int_\Re C(x)\, dx$; the second member in the equality represents the usual aggregated measure of imprecision of the fuzzy number C.

Calculations with fuzzy numbers are based on a so-called extension principle. Let $y = f(x_1, x_2, \ldots, x_n)$ be a continuous function on $\Re^n$. For any fuzzy numbers $X_1, X_2, \ldots, X_n$, $f(X_1, X_2, \ldots, X_n)$ is defined as a fuzzy number Y whose membership function is given for each $y \in \Re$ as follows

$$Y(y) = \begin{cases} \max\{\min_{i=1,\ldots,n}\{X_i(x_i)\} \mid y = f(x_1, \ldots, x_n)\} & \text{if such } x_1, \ldots, x_n \text{ exist,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In the fuzzy models of decision making, special structures of fuzzy numbers are used. The first of them is a fuzzy scale that enables a finite representation of an infinite interval by fuzzy numbers; the second, normalized fuzzy weights, will be a topic of the next section. A fuzzy scale on $[a,b]$ (see [7]) is a finite set of fuzzy numbers $A_1, A_2, \ldots, A_n$ that are defined on $[a,b]$, form a fuzzy partition on $[a,b]$, i.e. $\sum_{i=1}^{n} A_i(x) = 1$ holds for any $x \in [a, b]$, and are numbered according to their linear ordering. If a fuzzy scale represents a mathematical meaning of a natural linguistic scale, then it is called a linguistic fuzzy scale.

3 Normalized Fuzzy Weights

3.1 Definitions and Basic Properties

In decision making models, weights of criteria as well as probabilities of states of the world are usually set subjectively, i.e. they are uncertain. Therefore the models are more realistic if the weights of criteria and the probabilities of states of the world are expressed by means of the tools of the fuzzy sets theory. In this section, a special structure of fuzzy numbers, so called normalized fuzzy weights, will be introduced and its general properties will be studied. From the general point of view, the structure makes it possible to model mathematically an uncertain division of a unit into fractions.

Normalized fuzzy weights represent a fuzzification of the structure of normalized weights; m-tuple of real numbers $v_1, v_2, \ldots, v_m$ forms normalized weights if $v_i \geq 0$, $i = 1, 2, \ldots, m$, and $\sum_{i=1}^{m} v_i = 1$. Normalized weights can be used for describing a division of a unit into m uniquely specified fractions.

Fuzzy numbers V_i, $i = 1, 2, \ldots, m$, defined on $[0, 1]$ are called normalized fuzzy weights (see [6]) if for all $\alpha \in (0, 1]$ and for all $i \in \{1, 2, \ldots, m\}$ the following holds: for any $v_i \in V_{i\alpha}$ there exist $v_j \in V_{j\alpha}$, $j = 1, 2, \ldots, m$, $j \neq i$, such that

$$v_i + \sum_{j=1, j\neq i}^{m} v_j = 1 \ . \qquad (2)$$

The purpose of this definition is a generalization of the condition of normalization from the case of real numbers into the case of fuzzy numbers. Obviously, normalized weights $v_1, v_2, \ldots, v_m$ are a special case of normalized fuzzy weights. The membership degree $V_i(v_i)$ can be interpreted as the possibility of the fact that the weight of i-th criterion or the probability of i-th state of the world equals v_i; from the general point of view $V_i(v_i)$ means the possibility that the share of i-th subject in the whole is $v_i \cdot 100\%$.

The structure of normalized fuzzy weights was introduced in [6] in order to fuzzify the weights of criteria in the method of weighted average of partial fuzzy evaluations that was described in [7]. Later on, the normalized fuzzy weights were applied to modeling uncertain probabilities in fuzzy probability spaces with a finite set of elementary events (see [8], [9]). The same structure of fuzzy numbers was introduced for modelling fuzzy probabilities also in [5] (it was called "a tuple of fuzzy probabilities") as a generalization of the structure of interval probabilities developed in [1].

The fact whether fuzzy numbers V_i, $i = 1, 2, \ldots, m$, defined on $[0, 1]$ form normalized fuzzy weights or not can be verified in the following way (see [5] or [6]): For $i = 1, 2, \ldots, m$ let us denote $V_{i\alpha} = [\underline{v_i}(\alpha), \overline{v_i}(\alpha)]$, $\sigma_{V_i}(\alpha) = \overline{v_i}(\alpha) - \underline{v_i}(\alpha)$ for any $\alpha \in [0, 1]$. Furthermore, let us denote $\omega_{V_1,\ldots,V_m}(\alpha) = 1 - \sum_{i=1}^{m} \underline{v_i}(\alpha)$, $\alpha \in [0, 1]$. Fuzzy numbers V_i, $i = 1, 2, \ldots, m$, defined on $[0, 1]$ represent normalized fuzzy weights if and only if for any $\alpha \in [0, 1]$ and for $i^*(\alpha) \in \{1, 2, \ldots, m\}$, such that $\sigma_{V_{i^*(\alpha)}}(\alpha) = \max_{i=1,2,\ldots,m}\{\sigma_{V_i}(\alpha)\}$, the following holds

$$\sigma_{V_{i^*(\alpha)}}(\alpha) \leq \omega_{V_1,\ldots,V_m}(\alpha) \leq \sum_{j=1, j\neq i^*(\alpha)}^{m} \sigma_{V_j}(\alpha). \qquad (3)$$

Obviously, for $m = 2$ from (3) follows that $V_2 = 1 - V_1$. In the case of linear fuzzy numbers $V_i = (v_i^1, v_i^2, v_i^3, v_i^4)$, $i = 1, 2, \ldots, m$, it is sufficient to verify only the following two conditions

$$\sigma_{V_{i^*(1)}}(1) \quad \leq \omega_{V_1,\ldots,V_m}(1) \quad \leq \sum_{j=1, j\neq i^*(1)}^{m} \sigma_{V_j}(1) \tag{4}$$

$$\sigma_{V_{i^*(0)}}(0) \quad \leq \omega_{V_1,\ldots,V_m}(0) \quad \leq \sum_{j=1, j\neq i^*(0)}^{m} \sigma_{V_j}(0). \tag{5}$$

The uncertainty of the structure of normalized fuzzy weights can be characterized by two functions defined on $[0, 1]$. The first of them, so called expansibility function $\omega_{V_1,\ldots,V_m}(\alpha)$, $\alpha \in [0, 1]$, describes for each membership degree α the total value which is available for expansion of $v_i(\alpha), i = 1, 2, \ldots, m$. The second, so called overall span function $\sigma_{V_1,\ldots,V_m}(\alpha) = \sum_{i=1}^{m} \sigma_{V_i}(\alpha)$, $\alpha \in [0, 1]$, reflects for any α the overall variability of the possible vectors of normalized weights. In [6], the following relation between these functions was proved

$$\frac{m}{m-1} \cdot \omega_{V_1,\ldots,V_m}(\alpha) \leq \sigma_{V_1,\ldots,V_m}(\alpha) \leq m \cdot \omega_{V_1,\ldots,V_m}(\alpha), \text{ for all } \alpha \in [0, 1], \tag{6}$$

where the first equality holds if and only if $\sigma_{V_i}(\alpha) = \frac{\omega_{V_1,\ldots,V_m}(\alpha)}{m-1}$ for each $i \in \{1, 2, \ldots, m\}$, and the second equality holds if and only if $\sigma_{V_i}(\alpha) = \omega_{V_1,\ldots,V_m}(\alpha)$ for each $i \in \{1, 2, \ldots, m\}$.

3.2 Procedures of Setting Normalized Fuzzy Weights

In the practical applications, it is useful to set normalized fuzzy weights as linear fuzzy numbers $V_i = (v_i^1, v_i^2, v_i^3, v_i^4)$, $i = 1, 2, \ldots, m$, where $[v_i^2, v_i^3]$ represents the interval of fully possible values of i-th weight, while outside the interval $[v_i^1, v_i^4]$ the i-th weight cannot lie.

The easiest way of setting normalized fuzzy weights is following: First, an expert sets crisp normalized weights $v_1, v_2, \ldots, v_m$, $v_i \geq 0, i = 1, 2, \ldots, m$, $\sum_{i=1}^{m} v_i = 1$. Then he/she fuzzifies them into triangular fuzzy numbers $V_i = (v_i - s, v_i, v_i, v_i + s)$, $i = 1, 2, \ldots, m$, where $s \geq 0$ satisfies $v_i - s \geq 0$ and $v_i + s \leq 1$ for all $i \in \{1, 2, \ldots, m\}$.

A more general method of setting normalized fuzzy weights results from the conditions (4) and (5). Let $v_1, v_2, \ldots, v_m$, $v_i \geq 0$, $i = 1, 2, \ldots, m$, $\sum_{i=1}^{m} v_i = 1$, represent again a crisp estimation of normalized weights. Let real numbers k_i, s_i, $0 \leq k_i \leq s_i$, $v_i - s_i \geq 0$, $v_i + s_i \leq 1$, $i = 1, 2, \ldots, m$, characterize the width of kernels and supports of the particular fuzzy weights. Then linear fuzzy numbers $V_i = (v_i - s_i, v_i - k_i, v_i + k_i, v_i + s_i)$, $i = 1, 2, \ldots, m$, represent normalized fuzzy weights, if and only if the numbers k_i and s_i, $i = 1, 2, \ldots, m$, satisfy the following conditions

$$\sum_{i=1, i\neq i^*}^{m} k_i \geq k_{i^*} \quad \wedge \quad \sum_{j=1, j\neq j^*}^{m} s_j \geq s_{j^*}, \tag{7}$$

where $k_{i^*} = \max_{i=1,\ldots,m}\{k_i\}$ and $s_{j^*} = \max_{j=1,\ldots,m}\{s_j\}$.

The normalized fuzzy weights, which express either the importance of criteria or the probabilities of states of the world in the decision making models, can be set also linguistically. For the purpose, a special linguistic fuzzy scale has to be used. It can consists e.g. of the following linguistic terms: "negligible minority", "minority", "approximately half", "majority", "absolute majority". The mathematical meanings of these terms have to form a symmetrical fuzzy scale on $[0, 1]$, i.e. a fuzzy scale $T_1, T_2, \ldots, T_5$ satisfying $T_j(x) = T_{6-j}(1-x)$ for $x \in [0, 1]$ and $j = 1, 2, 3$. For example it can be set: $T_1 = (0, 0, 0, \frac{1}{6})$, $T_2 = (0, \frac{1}{6}, \frac{1}{3}, \frac{1}{2})$, $T_3 = (\frac{1}{3}, \frac{1}{2}, \frac{1}{2}, \frac{2}{3})$,

$T_4 = (\frac{1}{2}, \frac{2}{3}, \frac{5}{6}, 1)$, $T_5 = (\frac{5}{6}, 1, 1, 1)$. Each couple of fuzzy numbers T_j, T_{6-j}, $j = 1, 2, 3$, represents a couple of normalized fuzzy weights, i.e. it describes an uncertain division of a unit into two fractions; the situation also corresponds with the linguistic descriptions. Let us consider now the problem of setting criteria weights by means of such linguistic terms. Let us suppose that the criteria are linearly ordered according to their importance, i.e. $V_1 > V_2 > \ldots > V_m$. We express the share of the partial evaluation according to the most important criterion K_1 in the total evaluation, i.e. evaluation according to $K_1, K_2, \ldots, K_m$, by an adequate linguistic term T_{j_1} (e.g. "majority"). Then the share of the evaluation according to $K_2, K_3, \ldots, K_m$ in the total evaluation is T_{6-j_1} ("minority"). We repeat this procedure for each $i = 2, \ldots, m-1$, denoting by T_{j_i} the share of the evaluation according to K_i in the total evaluation according to $K_i, K_{i+1}, \ldots, K_m$. Then the normalized fuzzy weights of the criteria $K_1, K_2, \ldots, K_m$ are given in the following way

$$\begin{array}{lcl} V_1 & = & T_{j_1} \\ V_2 & = & T_{6-j_1} \cdot T_{j_2} \\ V_3 & = & T_{6-j_1} \cdot T_{6-j_2} \cdot T_{j_3} \\ \ldots & & \ldots \\ V_m & = & T_{6-j_1} \cdot T_{6-j_2} \cdot T_{6-j_3} \cdot \ldots \cdot T_{6-j_{m-1}} \; . \end{array} \tag{8}$$

Normalized fuzzy weights defined in this way are not linear, but they can be approximated by linear fuzzy numbers.

Normalized fuzzy weights of criteria can be set by means of the linguistic terms even if the criteria are not linearly ordered according to their importance but form a structure of linearly ordered classes of "approximately equal" elements. In the case, we apply first the above mentioned procedure to the classes of criteria. Then, the total fuzzy weight assigned to the class of k "approximately equally important" criteria is divided into k "approximately equal" fractions; for example by means of triangular fuzzy numbers $(\frac{1}{k+1}, \frac{1}{k}, \frac{k+2}{k \cdot (k+1)})$, $k = 2, 3, \ldots$.

For setting uncertain probabilities of states of the world we can use the linguistic terms analogously; e.g. we can say that the state S_1 occurs in the "approximately half" of all cases, otherwise in the "majority" of cases a state S_2 occurs, and if not, then a state S_3 comes.

4 The Fuzzy Weighted Average of Fuzzy Numbers

A fuzzy weighted average U of fuzzy numbers $U_1, U_2, \ldots, U_m$ with non-negative fuzzy weights $W_1, W_2, \ldots, W_m$ is defined in [2] as a fuzzification according to the extension principle (1) of the operation $\sum_{i=1}^m w_i \cdot u_i / \sum_{i=1}^m w_i$, $w_i \geq 0$, $i = 1, 2, \ldots, m$, $\sum_{i=1}^m w_i \neq 0$. The membership function $U(u)$ is given for each $u \in \Re$ in the following way

$$U(u) = \max\{\min\{U_1(u_1), U_2(u_2), \ldots, U_m(u_m), W_1(w_1), W_2(w_2), \ldots, W_m(w_m)\} \mid$$

$$u = \frac{w_1 u_1 + w_2 u_2 + \ldots + w_m u_m}{w_1 + w_2 + \ldots + w_m}, \; \sum_{i=1}^m w_i \neq 0\}. \tag{9}$$

The algorithm for computing the fuzzy weighted average given by (9) is studied in [3].

The fuzzy weighted average given by (9) cannot be used for aggregating partial fuzzy evaluations in multiple-criteria decision making models where uncertain weights of criteria express the shares of partial objectives of evaluation in the overall one (see [7], [8]). The same holds also for computing expected fuzzy evaluations in models of decision making under risk with uncertain probabilities of states of the world (see [8]). The problem is illustrated by the following example:

Let S_1, S_2 and S_3 be states of the world, their uncertain probabilities be, for simplicity, described by intervals $P_1 = [0.1, 0.2]$, $P_2 = [0.4, 0.6]$ and $P_3 = [0.3, 0.4]$ (P_1, P_2 and P_3 form normalized fuzzy weights) and the evaluations of an alternative x under states of the world S_1, S_2, S_3 be given by $U_1 = [0.1, 0.2]$, $U_2 = [0.4, 0.5]$ and $U_3 = [0.8, 0.9]$. Then, according to (9) where we replace W_i by P_i for $i = 1, 2, 3$, the fuzzy expected evaluation of x is $U = [0.45, 0.64]$. Let us notice that for computing the minimum value of U, i.e. 0.45, the third weight $\frac{p_3}{p_1+p_2+p_3} = \frac{0.3}{0.2+0.6+0.3} = 0.27$ and $0.27 \notin P_3$. So this value lies outside the expertly set range of possible probabilities of S_3.

In our case, a fuzzy weighted average of fuzzy numbers has to be defined as a fuzzification of an operation $\sum_{i=1}^{m} v_i \cdot u_i$, $\sum_{i=1}^{m} v_i = 1$, $v_i \geq 0$, $i = 1, 2, \ldots, m$. The fuzzy weighted average (see [6]) of fuzzy numbers U_i, $i = 1, 2, \ldots, m$, defined on $[a, b]$ with normalized fuzzy weights V_i, $i = 1, 2, \ldots, m$, is a fuzzy number U on $[a, b]$ whose membership function is given for each $u \in [a, b]$ by the following formula

$$\begin{aligned} U(u) = \quad & \max\{\min\{V_1(v_1), \ldots, V_m(v_m), U_1(u_1), \ldots, U_m(u_m)\} | \\ & \textstyle\sum_{i=1}^{m} v_i \cdot u_i = u,\ \sum_{i=1}^{m} v_i = 1,\ i = 1, 2, \ldots, m\}\ . \end{aligned} \tag{10}$$

The following denotation will be used for the fuzzy weighted average given by (10)

$$U = (\mathcal{F}) \sum_{i=1}^{m} V_i \cdot U_i\ . \tag{11}$$

In the case of normalized weights the fuzzy weighted averages given by (9) and (10) coincide. But for normalized fuzzy weights it does not hold. A fuzzy number representing the result of the fuzzy weighted average operation given by (10) is a subset of that one given by (9). For example, in the above mentioned example, if $U = (\mathcal{F}) \sum_{i=1}^{m} P_i \cdot U_i$, then $U = [0.46, 0.63]$. In the following text the fuzzy weighted average given by (10) will be considered.

The following algorithm of computing the fuzzy weighted average of fuzzy numbers is based on an algorithm that was originally developed for computing the expected value of a discrete random variable with interval probabilities (see [1]).

For each $\alpha \in [0, 1]$ let $\{i_k\}_{k=1}^{m}$ be such a permutation on an index set $\{1, 2, \ldots, m\}$ that it holds $\underline{u}_{i_1}(\alpha) \leq \underline{u}_{i_2}(\alpha) \leq \ldots \leq \underline{u}_{i_m}(\alpha)$. For $k \in \{1, 2, \ldots, m\}$ let us denote $v_{i_k}(\alpha) = 1 - \sum_{j=1}^{k-1} \overline{v}_{i_j}(\alpha) - \sum_{j=k+1}^{m} \underline{v}_{i_j}(\alpha)$. Let $k^* \in \{1, 2, \ldots, m\}$ be such an index that it holds $\underline{v}_{i_{k^*}}(\alpha) \leq v_{i_{k^*}}(\alpha) \leq \overline{v}_{i_{k^*}}(\alpha)$. Then

$$\underline{u}(\alpha) = \sum_{j=1}^{k^*-1} \overline{v}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha) + v_{i_{k^*}}(\alpha) \cdot \underline{u}_{i_{k^*}}(\alpha) + \sum_{j=k^*+1}^{m} \underline{v}_{i_j}(\alpha) \cdot \underline{u}_{i_j}(\alpha). \tag{12}$$

Let $\{i_h\}_{h=1}^{m}$ be such a permutation on an index set $\{1, 2, \ldots, m\}$ that it holds $\overline{u}_{i_1}(\alpha) \geq \overline{u}_{i_2}(\alpha) \geq \ldots \geq \overline{u}_{i_m}(\alpha)$. For $h \in \{1, 2, \ldots, m\}$ let us denote $v_{i_h}(\alpha) = 1 - \sum_{j=1}^{h-1} \overline{v}_{i_j}(\alpha) - \sum_{j=h+1}^{m} \underline{v}_{i_j}(\alpha)$. Let $h^* \in \{1, 2, \ldots, m\}$ be such an index that it holds $\underline{v}_{i_{h^*}}(\alpha) \leq v_{i_{h^*}}(\alpha) \leq \overline{v}_{i_{h^*}}(\alpha)$. Then

$$\overline{u}(\alpha) = \sum_{j=1}^{h^*-1} \overline{v}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha) + v_{i_{h^*}}(\alpha) \cdot \overline{u}_{i_{h^*}}(\alpha) + \sum_{j=h^*+1}^{m} \underline{v}_{i_j}(\alpha) \cdot \overline{u}_{i_j}(\alpha). \tag{13}$$

The span function of the fuzzy weighted average $\sigma_U(\alpha)$, $\alpha \in [0, 1]$, depends on the span functions $\sigma_{U_1}, \sigma_{U_2}, \ldots, \sigma_{U_m}$ and $\sigma_{V_1}, \sigma_{V_2}, \ldots, \sigma_{V_m}$ on the one hand and on the variability of the fuzzy numbers $U_1, U_2, \ldots, U_m$ on the other hand. It is illustrated by the following example:

Let K_1 and K_2 be a couple of criteria, corresponding normalized fuzzy weights V_1 and V_2 be set as triangular fuzzy numbers $V_1 = (0.4, 0.6, 0.8)$ and $V_2 = (0.2, 0.4, 0.6)$, and the partial evaluations of alternatives x_1, x_2 and x_3 according to K_1, K_2 be, for siplicity, given as real numbers $u_{11} = 0.9$, $u_{12} = 0.1$; $u_{21} = 0.1$, $u_{22} = 0.9$; $u_{31} = 0.5$, $u_{32} = 0.5$.Then the fuzzy weighted averages $U_i = (\mathcal{F}) \sum_{j=1}^{2} V_j \cdot u_{ij}$, $i = 1, 2, 3$, expressing overall evaluations of the alternatives, are $U_1 = (0.42, 0.58, 0.74)$, $U_2 = (0.26, 0.42, 0.58)$ and $U_3 = 0.5$.

We can see practically in the above example that the overall fuzzy evaluations of alternatives that are evaluated very differently according to the particular criteria are much more uncertain than the overall fuzzy evaluations of alternatives whose partial evaluations are almost uniform in multiple-criteria decision making models. Similarly, in decision making under risk, the expected fuzzy evaluations of alternatives that depend strongly on states of the world are more uncertain than expected fuzzy evaluations of alternatives which are relatively stable.

5 Applications of the Fuzzy Weighted Average in the Decision Making Models

The operation of a fuzzy weighted average can be applied to multiple-criteria decision making and decision making under risk in the following way:

First, let us consider the problem of multiple-criteria decision making, where the best of alternatives $x_1, x_2, \ldots, x_n$ is to be chosen. The alternatives will be evaluated with respect to a given objective that is partitioned into m partial objectives associated with criteria $K_1, K_2, \ldots, K_m$. Let the uncertain information concerning the shares of the partial objectives in the overall one be given by normalized fuzzy weights $V_1, V_2, \ldots, V_m$. Let $U_{i,j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, be uncertain partial fuzzy evaluations of alternatives x_i, $i = 1, 2, \ldots, n$, with respect to the criteria K_j, $j = 1, 2, \ldots, m$. $U_{i,j}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, are supposed to be the fuzzy degrees of fulfilment of the corresponding partial objectives and to be expressed by fuzzy numbers defined on $[0, 1]$. Then the overall fuzzy evaluation U_i of the alternative x_i will be calculated for $i = 1, 2, \ldots, n$ as the fuzzy weighted average of the partial fuzzy evaluations $U_{i,j}$ with the normalized fuzzy weights V_j, $j = 1, 2, \ldots, m$, i.e.

$$U_i = (\mathcal{F}) \sum_{j=1}^{m} V_j \cdot U_{i,j} \ . \tag{14}$$

The overall fuzzy evaluations U_i, $i = 1, 2, \ldots, n$, express again the fuzzy degrees of fulfilment of the overall objective of evaluation. The best alternative is either the first alternative in an ordering of the fuzzy numbers U_i, $i = 1, 2, \ldots, n$, or the closest to the ideal alternative whose evaluation is equal to 1. The overall fuzzy evaluations of alternatives can also be approximated linguistically by linearly ordered elements of a proper linguistic evaluation scale defined on $[0, 1]$. For more details about metrics, ordering of fuzzy numbers and linguistic approximation see [7].

Second, let us consider a problem of decision making under risk, where the best of risk alternatives $x_1, x_2, \ldots x_n$ is being chosen according to their degrees of fulfilment of an objective associated with a criterion K. The degrees of the objective fulfilment depend not only on the alternatives themselves but also on the fact which of the states of the world $S_1, S_2, \ldots, S_r$ occurs. Uncertain probabilities of the states of the world are expertly set by normalized fuzzy weights $P_1, P_2, \ldots, P_r$. The fuzzy evaluations of the alternatives x_i, $i = 1, 2, \ldots, n$, under the states of the world S_k, $k = 1, 2, \ldots, r$, are described as fuzzy numbers $U_{i,k}$, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, r$, defined on $[0, 1]$, that express to which extend the alternatives fulfil the given objective under the given states of the world. Expected fuzzy evaluations FEU_i of the alternatives x_i, $i = 1, 2, \ldots, n$, will be calculated according to the formula

$$FEU_i = (\mathcal{F}) \sum_{k=1}^{r} P_k \cdot U_{i,k} \ . \tag{15}$$

The best alternative will be chosen in an analogical way as in the previous case.

Examples of practical applications of such decision making models, namely in finance and banking, are described in [8] and [9].

References

[1] De Campos, L.M., Heute, J.F., Moral, S.: Probability inrtervals: a tool for uncertain reasoning. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems. 1994 Vol. 2(2), pp. 167-196.

[2] Dong, W.M., Wong, F.S.: Fuzzy weighted averages and implementation of the extension principle. Fuzzy Sets and Systems. 1987 Vol. 21, pp. 183-199.

[3] Guh, Y.-Y. et al.: Fuzzy weighted average: The linear programming approach via Charness and Cooper's rule. Fuzzy Sets and Systems. 2001 Vol. 117, pp. 157-160.

[4] Klir, G.J., Yuan, B.: Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall, New Jersey 1996.

[5] Pan, Y., Yuan, B.: Bayesian Inference of Fuzzy Probabilities. Int. J. General Systems. 1997 Vol. 26(1-2), pp. 73-90.

[6] Pavlačka, O.: Normalized fuzzy weights (in Czech). Diploma thesis, UP Olomouc 2004.

[7] Talašová, J.: Fuzzy Methods of Multiple-Criteria Evaluation and Decision Making (in Czech). VUP, Olomouc 2003.

[8] Talašová, J.: Fuzzy approach to evaluation and decision making (in Czech). Proceedings of Aplimat 2006, SUT, Bratislava 2006, pp. 221-236.

[9] Talašová, J., Pavlačka, O.: Fuzzy Probability Spaces and Their Applications in Decision Making. Austrian Journal of Statistics, to appear.

Analysis of the Czech Real Business Cycle Model[1]

Jiří Polanský and Osvald Vašíček
Masaryk University
Faculty of Economics and Administration
Department of Applied Mathematics and Computer Science
Lipová 41a, 602 00 Brno
polansky@econ.muni.cz osvald@econ.muni.cz

Abstract

The paper analyses the behavior of the two-sector real business cycle model. The model consists of the representative household's expected utility function, separate Cobb-Douglas production functions for consumption and investment goods, and capital accumulation constraints. It evaluates impacts of productivity shocks to the economy. The preference shock affects the marginal rate of substitution between consumption and leisure in the household utility function and the next two shocks are sector specific technology shocks. Each shock contains a separate autoregressive component governing its level and growth rate. The Kalman filter algorithm is used to estimate the model's structural parameters via maximum likelihood method on quarterly data series of the Czech economy. The final part investigates estimation results and impulse responses for the Czech economy.

Keywords

Real business cycle, two-sector model, economic growth, Kalman filter with likelihood function, Bootstrap method
JEL: C15, C51, E32, O41, O47

1 Introduction

The one-sector neoclassical growth models have been widely used in macroeconomic long-run modeling since the 1950s. They have central role in macroeconomics textbooks and in business cycle research. These models treat output in the economy as deriving from a single aggregate production function. The most important feature of one-sector models is the fact that they imply a balanced growth path, where the real consumption, investment, capital stock, and output all grow by the same average rate in the long-run. In other words, the ratios of any of these variables will be stationary stochastic processes. The hypothesis of balanced growth and stationary ”great ratios” has been held as a crucial stylized macroeconomic fact. Business cycles are commonly characterized as deviations from the model's long-run balanced growth path.

In recent years, some research papers have emerged, which try to overcome this approach. The reason can be found in the US economy where the real investment growth has outpaced the real consumption growth since the 1990s. Moreover, Whelan [8] proves that the post-1991 increase in real investment relative to real consumption does not appear as a particularly cyclical phenomenon but it reflects long-running trend. When the ratio of real investment to real consumption exhibits an upward trend, the traditional balanced growth path is undermined, and hence, the traditional one-sector models are inconsistent with these data. To model the long-run upward trend in real variables, we need to extend the ”traditional” real business cycle model to allow a faster pace of technological progress in one sector relative to the other. As possible solution to this problem can be multisectoral models. These models work with more than one production technology, and hence are able to deal with the ”new balanced growth path”, where the aggregates for consumption and investment grow at steady but different rates.

This paper deals with the extended two-sector real business cycle model and tries to analyze the behavior of the Czech economy. Our objective is to describe preference and technology shocks within the Czech economy and analyze their relative importance. For our purposes, we use a model of Peter N. Ireland and Scott Schuh [5], which extends some former models. This approach is relatively new in macroeconomics and allows to distinguish between different improvements within sector technologies which should be relevant for the Czech economy. The second part of the paper introduces the model. The third part describes the estimation procedure,

[1]This paper has been worked as a part of research activities at the grant project of GA CR 402/05/2172.

presents the results of the Czech economy and analyses the behavior of the model. The final part concludes the paper.

2 The Model

2.1 Preferences and Technologies

The infinitely-lived representative household has preferences described by the expected utility function

$$E_0 \sum_{t=0}^{\infty} \beta^t [ln(C_t) - (H_{ct} + H_{it})/A_t], \tag{1}$$

where C_t denotes consumption, H_{ct} and H_{it} denote labor supplied to produce consumption and investment goods, β is the discount factor with constraint $1 > \beta > 0$, and A denotes the preference shock. This shock competes with the various technology shocks in accounting for fluctuations in consumption, investment, and hours worked. With this specification of the utility function, A impacts the marginal rate of substitution between consumption and leisure.

The production structure of the Czech economy is divided into two sectors. The investment sector produces investment goods for both markets, while the consumption sector produces only the consumption good purchased by households. An important assumption is that it is costly to alter the composition of capital goods in an economy. For example, a positive technological shock in one sector will lead to increased investment and employment in this sector and increased output in subsequent periods. But it is unlikely that a significant quantity of existing capital would move from one sector to another. Within the economy, plenty of capital is sector-specific. Moreover, it takes time and resources to alter the production of new capital. For these reasons, the production structure includes capital adjustment costs. These costs apply to all investment I_{ct} or I_{it} allocated to the two sectors. In other words, they apply to both newly-created capital and existing capital that is reallocated across the two sectors. The two-sector production structure is described by production function and adjustment cost specifications for consumption and investment goods

$$[1 - \frac{\phi_c}{2}(\frac{I_{ct}}{K_{ct}} - \kappa_c)^2](u_{ct}K_{ct})^{\theta_c}(Z_{ct}H_{ct})^{1-\theta_c} \geq C_t \tag{2}$$

and

$$[1 - \frac{\phi_i}{2}(\frac{I_{it}}{K_{it}} - \kappa_i)^2](u_{it}K_{it})^{\theta_i}(Z_{it}H_{it})^{1-\theta_i} \geq I_{ct} + I_{it} \tag{3}$$

for all t. K_{ct} and K_{it} denote capital stocks allocated to the two sectors, u_{ct} and u_{it} are the corresponding rates of capital utilization, $1 > \theta_c > 0$ and $1 > \theta_i > 0$ are the Cobb-Douglas share parameters, and Z_{ct} and Z_{it} denote sector specific technology shocks. Capital adjustment costs subtract from output in each of the two sectors. The parameters $\phi_c \geq 0$ and $\phi_i \geq 0$ govern the magnitude of capital adjustment costs, and the parameters κ_c and κ_i can be set equal to the steady-state investment-capital ratios in the two sectors so that steady-state adjustment costs equal zero.

The capital accumulation constraints are described by

$$[1 - (1/\omega_c)u_{ct}^{\omega_c}]K_{ct} + I_{ct} \geq K_{ct+1} \tag{4}$$

$$[1 - (1/\omega_i)u_{it}^{\omega_i}]K_{it} + I_{it} \geq K_{it+1}, \tag{5}$$

where $\omega_c > 1$ and $\omega_i > 1$. These constraints suppose that higher rates of capital utilization are associated with faster rates of depreciation.

2.2 Driving Processes

The preference and the two technology shocks contain separate autoregressive components governing its level and its growth rate. This specification allows us to analyze not only the relative importance of these three shocks in the Czech economy and their impacts, magnitudes and persistences, but moreover we can recognize the exact essences of these shocks. The preference and technology shocks evolve according to

$$ln(A_t) = ln(a_t^l) + ln(A_t^g), \tag{6}$$

$$ln(a_t^l) = \rho_a^l ln(a_{t-1}^l) + \varepsilon_{at}^l, \tag{7}$$

$$ln(A_t^g/A_{t-1}^g) = (1-\rho_a^g)ln(a^g) + \rho_a^g ln(A_{t-1}^g/A_{t-2}^g) + \varepsilon_{at}^g, \tag{8}$$

$$ln(Z_{ct}) = ln(z_{ct}^l) + ln(Z_{ct}^g), \tag{9}$$

$$ln(z_{ct}^l) = \rho_c^l ln(z_{ct-1}^l) + \varepsilon_{ct}^l, \tag{10}$$

$$ln(Z_{ct}^g/Z_{ct-1}^g) = (1-\rho_c^g)ln(z_c^g) + \rho_c^g ln(Z_{ct-1}^g/Z_{ct-2}^g) + \varepsilon_{ct}^g, \tag{11}$$

$$ln(Z_{it}) = ln(z_{it}^l) + ln(Z_{it}^g), \tag{12}$$

$$ln(z_{it}^l) = \rho_i^l ln(z_{it-1}^l) + \varepsilon_{it}^l, \tag{13}$$

$$ln(Z_{it}^g/Z_{it-1}^g) = (1-\rho_i^g)ln(z_i^g) + \rho_i^g ln(Z_{it-1}^g/Z_{it-2}^g) + \varepsilon_{it}^g, \tag{14}$$

where the autoregressive parameters ρ_a^l, ρ_a^g, ρ_c^l, ρ_c^g, ρ_i^l, and ρ_i^g all lie between zero and one and the innovations ε_{at}^l, ε_{at}^g, ε_{ct}^l, ε_{ct}^g, ε_{it}^l, and ε_{it}^g are serially and mutually uncorrelated and normally distributed with zero means and standard deviations σ_a^l, σ_a^g, σ_c^l, σ_c^g, σ_i^l, and σ_i^g.

2.3 The Model Equilibrium

The representative household chooses contingency plans for C_t, H_{ct}, H_{it}, I_{ct}, I_{it}, u_{ct}, u_{it}, K_{ct+1}, and K_{it+1} for all t to maximalize the utility function (1) subjected to the constraints imposed by (2)-(5).

The first order conditions are

$$1 = \Lambda_{ct} C_t, \tag{15}$$

$$H_{ct} = (1-\theta_t)\Lambda_{ct} A_t C_t, \tag{16}$$

$$H_{it} = (1-\theta_i)\Lambda_{it} A_t I_t, \tag{17}$$

$$\Xi_{ct} = \Lambda_{it} + \phi_c \Lambda_{ct}(I_{ct}/K_{ct} - \kappa_c)(1/K_{ct})(u_{ct}K_{ct})^{\theta_c}(Z_{ct}H_{ct})^{1-\theta_c}, \tag{18}$$

$$\Xi_{it} = \Lambda_{it}[1 + \phi_i(I_{it}/K_{it} - \kappa_i)(1/K_{it})(u_{it}K_{it})^{\theta_i}(Z_{it}H_{it})^{1-\theta_i}], \tag{19}$$

$$\theta_c \Lambda_{ct} C_t = \Xi_{ct} u_{ct}^{\omega_c} K_{ct}, \tag{20}$$

$$\theta_i \Lambda_{it} I_t = \Xi_{it} u_{it}^{\omega_i} K_{it}, \tag{21}$$

$$\begin{aligned} \Xi_{ct} \quad = \quad & \beta E_t\{\Xi_{ct+1}[1-(1/\omega_c)u_{ct+1}^{\omega_c}]\} + \beta\theta_c E_t(\Lambda_{ct+1}C_{t+1}/K_{ct+1} + \\ & +\beta\phi_c E_t[\Lambda_{ct+1}(I_{ct+1}/K_{ct+1} - \kappa_c)(I_{ct+1}/K_{ct+1})(1/K_{ct+1})(u_{ct+1}K_{ct+1})^{\theta_c}(Z_{ct+1}H_{ct+1})^{1-\theta_c}]), \end{aligned} \tag{22}$$

$$\begin{aligned} \Xi_{it} \quad = \quad & \beta E_t\{\Xi_{it+1}[1-(1/\omega_i)u_{it+1}^{\omega_i}]\} + \beta\theta_i E_t(\Lambda_{it+1}I_{t+1}/K_{it+1} + \\ & +\beta\phi_i E_t[\Lambda_{it+1}(I_{ct+1}/K_{it+1} - \kappa_i)(I_{it+1}/K_{it+1})(1/K_{it+1})(u_{it+1}K_{it+1})^{\theta_i}(Z_{it+1}H_{it+1})^{1-\theta_i}]), \end{aligned} \tag{23}$$

and (2)-(5) with equality for all t. In these equations, Λ_{ct} and Λ_{it} denote the nonnegative multipliers on the production possibility constraints (2) and (3), and Ξ_{ct} and Ξ_{it} denote nonnegative multipliers on the capital accumulation constraints (4) and (5). The aggregate investment and hours worked are defined as

$$I_t = I_{ct} + I_{it} \tag{24}$$

$$H_t = H_{ct} + H_{it}. \tag{25}$$

2.4 The Solution of the Model

Equations (2)-(14) describe the behavior of the 24 model variables C_t, H_t, H_{ct}, H_{it}, I_t, I_{ct}, I_{it}, u_{ct}, u_{it}, K_{ct}, K_{it}, Λ_{ct}, Λ_{it}, Ξ_{ct}, Ξ_{it}, A_t, a_t^l, A_t^g, Z_{ct}, z_{ct}^l, Z_{ct}^g, Z_{it}, z_{it}^l, and Z_{it}^g. In equilibrium, these variables grow at different average rates, and some inherit unit roots from the nonstationary components of the shocks. The transformed variables $c_t = C_t/[A_{t-1}^g(Z_{it-1}^g)^{\theta_c}(Z_{ct-1}^g)^{1-\theta_c}]$, $h_t = H_t/A_{t-1}^g$, $h_{ct} = H_{ct}/A_{t-1}^g$, $h_{it} = H_{it}/A_{t-1}^g$, $i_t = I_t/(A_{t-1}^g Z_{it-1}^g)$, $i_{ct} = I_{ct}/(A_{t-1}^g Z_{it-1}^g)$, $i_{it} = I_{it}/(A_{t-1}^g Z_{it-1}^g)$, u_{ct}, u_{it}, $k_{ct} = K_{ct}/(A_{t-1}^g Z_{it-1}^g)$, $k_{it} = K_{it}/(A_{t-1}^g Z_{it-1}^g)$, $\lambda_{ct} = [A_{t-1}^g(Z_{it-1}^g)^{\theta_c}(Z_{ct-1}^g)^{1-\theta_c}]\Lambda_{ct}$, $\lambda_{it} = A_{t-1}^g Z_{it-1}^g \Lambda_{it}$, $\xi_{ct} = A_{t-1}^g Z_{it-1}^g \Xi_{ct}$, $\xi_{it} = A_{t-1}^g Z_{it-1}^g \Xi_{it}$, $a_t = A_t/A_{t-1}^g$, a_t^l, $a_t^g = A_t^g/A_{t-1}^g$, $z_{ct} = Z_{ct}/Z_{ct-1}^g$, z_{ct}^l, $z_{ct}^g = Z_{ct}^g/Z_{ct-1}^g$, $z_{it} = Z_{it}/Z_{it-1}^g$, z_{it}^l, and $z_{it}^g = Z_{it}^g/Z_{it-1}^g$ remain stationary, as do the growth rates of consumption, investment, and hours worked, computed as

$$g_t^c = C_t/C_{t-1} = a_{t-1}^g(z_{it-1}^g)^{\theta_c}(z_{ct-1}^g)^{1-\theta_c}(c_t/c_{t-1}), \tag{26}$$

$$g_t^i = I_t/I_{t-1} = a_{t-1}^g z_{it-1}^g(i_t/i_{t-1}), \tag{27}$$

$$g_t^h = H_t/H_{t-1} = a_{t-1}^g(h_t/h_{t-1}). \tag{28}$$

Equations (2)-(28) imply that in the absence of shocks, the model converges to a steady-state balanced growth path, along which all of the stationary variables are constant. Moreover, the equations (26)-(28) imply that consumption, investment, and hours worked grow at different rates along the balanced growth path, with

$$g_t^c = a^g(z_i^g)^{\theta_c}(z_c^g)^{1-\theta_c}, \tag{29}$$

$$g_t^i = a^g z_i^g, \tag{30}$$

$$g_t^h = a^g \tag{31}$$

for all t.

In the next step, the equations are log-linearized around the steady state to describe how the variables respond to shocks. The detailed derivation can be seen in Ireland and Schuh [6]. The system now consists of linear expectational difference equations, that are driven by exogenous shocks. The solution to this system is using the method described in Klein [7], which relies on the complex generalized Schur decomposition.

3 The Estimation Procedure and the Results

The Czech real business cycle model has 24 parameters, which describe preferences, technologies and the stochastic behavior of the exogenous shocks. Before the estimation, some constraints are imposed on seven parameters: $\beta = 0.99$, $\theta_i = 0.15$, $a^g = 0.9984$, $z_c^g = 1.0146$, $z_i^g = 1.0142$, κ_c and κ_i are set to make steady-state capital adjustment costs equal to zero. The 17 remaining structural parameters are estimated. For the estimation, we used Czech economy quaterly data. All time series are seasonally adjusted. The estimation technique is the Kalman filter algorithm with maximum likelihood method outlined by Hamilton [3]. The standard errors come from a parametric bootstrapping procedure, which simulates the estimated model in order to generate 1,000 samples of simulated artificial data for aggregate consumption, investment, and hours worked. Each sample contains the same number of observations as the original data sample. Then, the procedure reestimates the model 1,000 times using these artificial data sets.

The estimation results are shown in Table 1. The estimate for the capital's share in the consumption goods-producing sector is $\theta_c = 0.11$ which lies below the corresponding share parameter for the investment goods-producing sector. This result is opposite to estimates for US economy where this parameter lies above the corresponding parameter for the investment-goods-producing sector. The value of this parameter is lower than we would expect and is caused partly by the undercapitalization of the Czech economy during transformation process and probably partly by short data series. Other structural parameter estimates are similar to those for the US economy. The estimates of the capital adjustment cost parameters $\phi_c = 67.1456$ and $\phi_i = 10.57753$ imply that these costs are much more important in the consumption goods-producing sector than in the investment goods-producing sector. Within the Czech economy, capital utilization is more elastic in the investment goods-producing sector which can be seen from the estimates $\omega_c = 2.79761$ and $\omega_i = 2.16789$. These estimates suggest the higher production flexibility in the investment goods-producing sector.

The Table 1 also presents the estimates of volatilities and persistences of each of the six shocks. The values $\sigma_c^g = 0.00783$ and $\sigma_i^g = 0.00619$ imply that disturbances to growth rates of the consumption-specific technology shock Z_{ct} and the investment-specific technology shock Z_{it} have played important roles in the Czech economy. The magnitudes of $\sigma_c^l = 0.00283$ and $\sigma_a^g = 0.00102$ are much more lower than the first two

Table 1: Maximum Likelihood Estimates and Standard Errors

Parameter	Estimate	Standard Error	Parameter	Estimate	Standard Error
θ_c	0.11328	0.05837	ρ_i^l	0.90281	0.02594
ϕ_c	67.14560	54.13550	ρ_i^g	0.44695	0.12993
ϕ_i	10.57753	1393.99810	σ_a^l	0.00000	0.00002
ω_c	2.79761	164.97670	σ_a^g	0.00102	0.00013
ω_i	2.16789	277.19570	σ_c^l	0.00283	0.00193
ρ_a^l	0.90000	0.00028	σ_c^g	0.00783	0.03779
ρ_a^g	0.90204	0.08288	σ_i^l	0.00000	0.00153
ρ_c^l	0.04435	0.18416	σ_i^g	0.00619	0.00174
ρ_c^g	0.53934	0.15783			

but still relevant. The values $\rho_c^g = 0.53934$ and $\rho_i^g = 0.44695$ indicate that the growth rate components of both shocks appear equally persistent. Interesting is the fact that the preference shock has played only minor role in the Czech economy. The reason may again lie in the transformation characteristics of the Czech economy where new technologies and knowledges were "imported" during the transformation process, and hence, the preference shock has lower importance than the technology shocks.

The behavioral analysis is described in the next two figures. Figure 1 depicts impulse responses to investment-sector technology shock. Each panel shows the percentage-point response of aggregates of C, I, and H to a one-standard-deviation shock to the level or growth rate of productivity Z_i in the investment goods-producing sector. As can be seen from the Figure, the investment-specific technology shock impacts simultaneously on all three variables. In accordance with economic theory, the response to investment is more important than the responses of C and H.

Figure 2 plots the estimates of shocks evolution. They are constructed using the Kalman smoothing algorithms described by Hamilton [3]. All of these estimates, hence, reflect information contained in the full data sample. This Figure reflects the importance of the growth-rate components of both sector-specific technology shocks and a low significance of the preference shock in the Czech economy. The both sector-specific productivity movements are assumed to be highly persistent. We can see the aggregate productivity slowdown after the 1997 in the consumption as well as in the investment sector and the subsequent catch-up of the factor productivity.

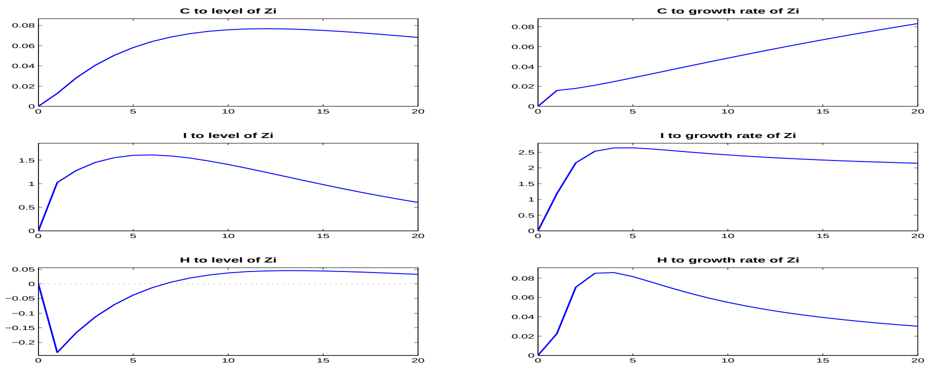


Figure 1: Impulse Responses to Investment-Sector Technology Shock

4 Conclusion

This paper presents the two-sector real business cycle model and estimates it with the Czech data. The estimation technique was the Kalman filter with maximum likelihood. The parameter estimates should represent

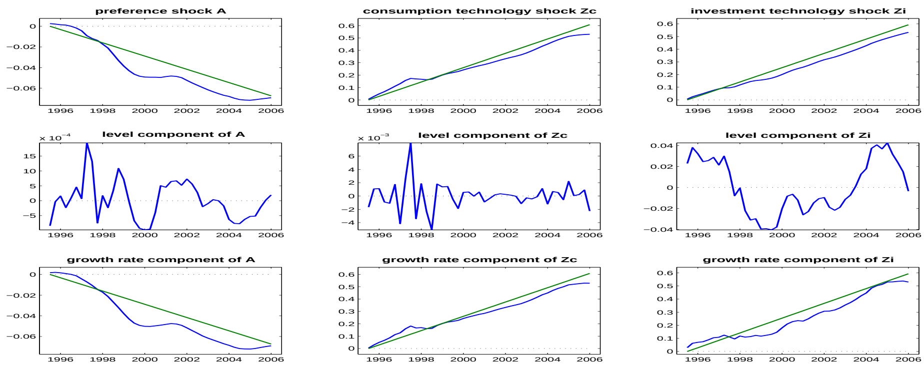


Figure 2: Smoothed Estimates of Preference and Technology Shocks

long-run features of the Czech economy after the 1995. The model identifies different types of technology shocks and analyses their impacts on aggregates of consumption, investment, and hours worked in the Czech economy. It suggests that the growth rate components of the technology shocks have played the prominent role. Moreover, the estimated model identifies the productivity slowdown after 1997 and the subsequent catch-up of the factor productivity.

References

[1] AADLAND, D. (2006): Advanced Topics in Economic Theory. Lecture Notes. University of Wyoming Personal Web Server. Access from internet `<http://www.uwyo.edu/aadland/classes/econ5110>` (cited to June, 2006).

[2] BLANCHARD, O. J. - KAHN, C. M. (1980): The Solution of Linear Difference Models under Rational Expectations. *Econometrica*, Volume 48, Issue 5 (Jul., 1980), 1305-1312.

[3] HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, 1994.

[4] HUFFMAN, G. W. - WYNNE, M. A. (1999): The role of intratemporal adjustment costs in a multisector economy. *Journal of Monetary Economics* 43 (1999), 317-350.

[5] IRELAND, P. N. - SCHUH, S. (2006a): Productivity and U.S. Macroeconomic Performance: Interpreting the Past and Predicting the Future with a Two-Sector Real Business Cycle Model. Boston College Personal Web Server. Access from internet <http://www2.bc.edu/˜irelandp/productivity.pdf> (cited to July, 2006).

[6] IRELAND, P. N. - SCHUH, S. (2006b): Data and Programs (zip): Productivity and US Macroeconomic Performance: Looking Backward and Forward with a Two-Sector Real Business Cycle Model. Boston College Personal Web Server. Access from internet <http://www2.bc.edu/˜irelandp/programs.html> (cited to February, 2006).

[7] KLEIN, P (2000): Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics & Control*, 24 (2000), 1405-1423.

[8] WHELAN, K. (2003): A Two-Sector Approach to Modeling U.S. NIPA Data. *Journal of Money, Credit, and Banking*, Vol. 35, No. 4 (August 2003).

[9] WHELAN, K. (2004): New Evidence on Balanced Growth, Stochastic Trends, and Economic Fluctuations. Research Technical Paper 7/RT/04, Central Bank and Financial Services Authority of Ireland, October 2004.

Fuzzy ANP – a New Method and Case Study

Jaroslav Ramík, Radomír Perzina
Silesian University Opava, School of Business Administration Karviná

e-mail: (ramik, perzina)@opf.slu.cz

Abstract

We propose a new decision model based on ANP for solving the decision making problem with fuzzy pair-wise comparisons and a feedback between the criteria. The evaluation of the weights of criteria, the variants as well as the feedback between the criteria is based on the data given in pair-wise comparison matrices. Extended arithmetic operations with fuzzy numbers are proposed as well as ordering fuzzy relations to compare fuzzy outcomes. An illustrating numerical example is presented to clarify the methodology.

Keywords

Multi-criteria decision making, analytic hierarchy process (AHP), pair-wise comparisons, systems with feedback, fuzzy numbers, analytic network process (ANP)
JEL classification: C44, C45

1 Introduction

When applying Analytic Hierarchy Process (AHP) in decision making, e.g. when you want to buy a best product for your personal use, say a car or digital camera, one usually meets two difficulties:

- when evaluating pair-wise comparisons on the nine point scale we do not incorporate uncertainty,
- decision criteria are not independent as it is required.

We solve these difficulties by proposing a new method which incorporates uncertainty adopting pair-wise comparisons by triangular fuzzy numbers, and takes into account interdependences between decision criteria.

The first difficulty is solved by the help of fuzzy evaluations: instead of saying e.g. "with respect to criterion C, element A is 3 times more preferable to element B" we say "element A is *possibly* 3 times more preferable to element B", where "*possibly* 3" is expressed by a (triangular) fuzzy number, similarly to the fuzzy number depicted in Fig. 3. In some real decision situations, interdependency of the decision criteria occur quite frequently, e.g. in the problem of choosing the best product the criterion "price of the product" is naturally influenced by other technical or esthetical criteria considered. Here, the influence is modeled by a feedback matrix the columns of which express the grades of influence of the individual criteria on the other criteria.

The interface between hierarchies, multiple objectives and fuzzy sets have been investigated by the author of Analytic Hierarchy Process (AHP) T.L. Saaty as early as in 1978 in [4]. Later on, Laarhoven and Pedrycz extended AHP to fuzzy pair-wise comparisons. In his books [5] and [6], T.L. Saaty extends the AHP to a more general process with feedback called Analytic Network Process (ANP). In this paper we extend the approaches from [1], [2] and [7] to the case of feedbacks between the decision criteria. We supply an illustrating case study to demonstrate the proposed method.

2 Multi-criteria decisions and AHP/ANP

In Analytic Hierarchy Process (AHP) we consider a three-level hierarchical decision system: On the first level we consider a decision goal G, on the second level, we have n independent *evaluation criteria*: $C_1, C_2,...,C_n$, such that $\sum_{i=1}^{n} w(C_i) = 1$, where $w(C_i) > 0$, $i = 1,2,...,n$, $w(C_i)$ is a positive real number – weight, usually interpreted as a relative importance of criterion C_i subject to the goal G. On the third level m variants (alternatives) of the decision outcomes $V_1, V_2,...,V_m$ are considered such that again $\sum_{j=1}^{m} w(V_j, C_i) = 1$, where $w(V_j,C_i)$, $j = 1,2,...,m$, is a non-negative real number - an evaluation (weight) of V_j subject to the criterion C_i, $i = 1,2,...,n$. This system is characterized by the *supermatrix* (see [7]):

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix},$$

the nonnegative matrix where $\mathbf{W_{21}}$ is the $n\times 1$ matrix (weighing vector of the criteria), i.e.

$$\mathbf{W}_{21} = \begin{bmatrix} w(C_1) \\ \vdots \\ w(C_n) \end{bmatrix},$$

and $\mathbf{W_{32}}$ is the $m\times n$ matrix:

$$\mathbf{W}_{32} = \begin{bmatrix} w(C_1,V_1) & \cdots & w(C_n,V_1) \\ \vdots & \cdots & \vdots \\ w(C_1,V_m) & \cdots & w(C_n,V_m) \end{bmatrix}.$$

The columns of this matrix are evaluations of variants by the criteria, $\mathbf{I}$ is the unit matrix. $\mathbf{W}$ is a *column-stochastic matrix*, i.e. the non-negative matrix and the sums of columns are equal to one. Then the limiting matrix $\mathbf{W}^{\infty} = \lim_{k\to+\infty} \mathbf{W}^k$ (see [3]) exists and is given as follows

$$\mathbf{W}^{\infty} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{32}W_{21}} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix}. \tag{1}$$

The decision system can be simplified as follows.

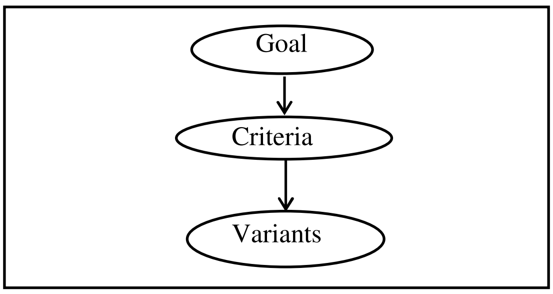


Fig. 1. Hierarchical system with 3 levels

Here, $\mathbf{Z} = \mathbf{W_{32}W_{21}}$ is the $m\times1$ matrix, i.e. the resulting *priority vector of weights of the variants*. The variants can be ordered according to these priorities.

In real decision systems with 3 levels there exist typical interdependences among individual elements of the decision hierarchy e.g. criteria or variants. Decision systems with dependences have been extensively investigated by Analytic Network Process (ANP), ([5], [6]). Consider now the dependences among the criteria, as is depicted in Fig. 2.

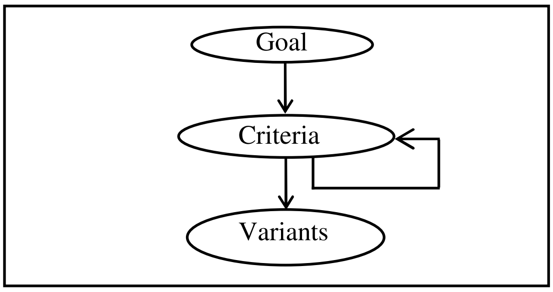


Fig. 2. Feedback amongst criteria

This system is given by the supermatrix:

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{21}} & \mathbf{W_{22}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix}, \tag{2}$$

where the interdependences of the criteria are characterized by $n\times n$ matrix $\mathbf{W_{22}}$:

$$\mathbf{W}_{22} = \begin{bmatrix} w(C_1, C_1) & \cdots & w(C_n, C_1) \\ \vdots & \cdots & \vdots \\ w(C_1, C_n) & \cdots & w(C_n, C_n) \end{bmatrix}.$$

In general, matrix (2) need not be column-stochastic, hence the limiting matrix does not exist. However, stochasticity of this matrix can be saved by additional normalization of the columns of the sub-matrix, to obtain a new sub-matrix $\begin{bmatrix} \mathbf{W_{22}^*} \\ \mathbf{W_{32}^*} \end{bmatrix}$. Then there exists a limiting matrix $\mathbf{W}^\infty$ such that

$$\mathbf{W}^\infty = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{32}}(\mathbf{I} - \mathbf{W_{22}})^{-1}\mathbf{W_{21}} & \mathbf{W_{32}}(\mathbf{I} - \mathbf{W_{22}})^{-1} & \mathbf{I} \end{bmatrix}. \tag{3}$$

Hence, the vector

$$\mathbf{Z} = \mathbf{W}_{32}(\mathbf{I} - \mathbf{W}_{22})^{-1}\mathbf{W}_{21} \tag{4}$$

is used for ranking the variants, i.e. the goal of the decision making process.

As the matrix $\mathbf{W_{22}}$ is close to the zero matrix and the dependences among criteria are usually weak, it can be approximately substituted by the first several terms of Taylor's expansion

$$(\mathbf{I} - \mathbf{W}_{22})^{-1} = \mathbf{I} + \mathbf{W}_{22} + \mathbf{W}_{22}^2 + \ldots \tag{5}$$

Then by substituting 4 terms from (5) to (4) we get

$$\mathbf{Z} = \mathbf{W}_{32}(\mathbf{I} + \mathbf{W}_{22} + \mathbf{W}_{22}^2 + \mathbf{W}_{22}^3)\mathbf{W}_{21}. \qquad (6)$$

In the next section, formula (6) will be used for computing fuzzy evaluations of the variants.

3 Fuzzy numbers and fuzzy matrices

When applying AHP in decision making, we usually meet difficulties in evaluating pair-wise comparisons on the 5 (or 9) point scale. In practice it is sometimes more convenient for the decision maker to express his/her evaluation in "words of natural language" saying e.g. "possibly 3", "approximately 4" or "about 5". Similarly, he/she could use the evaluations as "A is possibly weak preferable to B", etc. It is advantageous to express these evaluations by fuzzy sets of the real numbers, particularly, triangular fuzzy numbers, see Fig. 3.

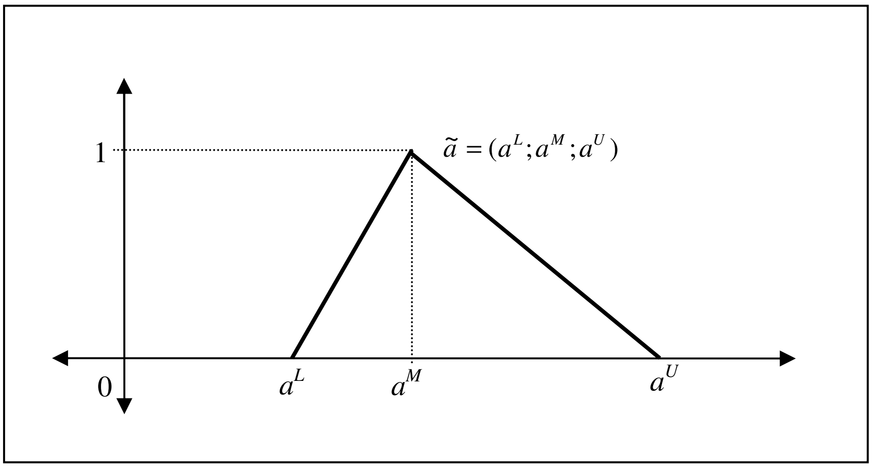


Fig. 3. A triangular fuzzy number

A *triangular fuzzy number a* is defined by a triple of real numbers, i.e. $a = (a^L; a^M; a^U)$, where a^L is the *Lower number*, a^M is the *Middle number* and a^U is the *Upper number*, $a^L \leq a^M \leq a^U$. If $a^L = a^M = a^U$, then a is said to be the *crisp number* (*non-fuzzy number*). Evidently, the set of all crisp numbers is isomorphic to the set of real numbers. In order to distinguish fuzzy and non-fuzzy numbers we shall denote the fuzzy numbers, vectors and matrices by the tilde above the symbol, e.g. $\tilde{a} = (a^L; a^M; a^U)$.

It is well known that the arithmetic operations +,-, * and / can be extended to fuzzy numbers by the Extension principle, see e.g. [2], in case of triangular fuzzy numbers $\tilde{a} = (a^L; a^M; a^U)$ and $\tilde{b} = (b^L; b^M; b^U)$, $a^L > 0$, $b^L > 0$, we obtain special formulae:

$$\tilde{a} \mathbin{\tilde{+}} \tilde{b} = (a^L + b^L; a^M + b^M; a^U + b^U),$$
$$\tilde{a} \mathbin{\tilde{-}} \tilde{b} = (a^L - b^L; a^M - b^M; a^U - b^U),$$
$$\tilde{a} \mathbin{\tilde{*}} \tilde{b} = (a^L * b^L; a^M * b^M; a^U * b^U),$$
$$\tilde{a} \mathbin{\tilde{/}} \tilde{b} = (a^L / b^U; a^M / b^M; a^U / b^L).$$

If all elements of an $m \times n$ matrix **A** are triangular fuzzy numbers we call **A** the *triangular fuzzy matrix* and this matrix is composed of triples as follows

$$\tilde{\mathbf{A}} = \begin{bmatrix} (a_{11}^L; a_{11}^M; a_{11}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ \vdots & \ddots & \vdots \\ (a_{m1}^L; a_{m1}^M; a_{m1}^U) & \cdots & (a_{mn}^L; a_{mn}^M; a_{mn}^U) \end{bmatrix}.$$

Particularly, if $\tilde{\mathbf{A}}$ is a triangular fuzzy matrix which is also pair-wise comparison one, we say that it is *reciprocal*, if $\tilde{a}_{ij} = (a_{ij}^L; a_{ij}^M; a_{ij}^U)$ then $\tilde{a}_{ji} = (\frac{1}{a_{ij}^U}; \frac{1}{a_{ij}^M}; \frac{1}{a_{ij}^L})$ for all *i,j* = 1,2,...,*n*. Then we have

$$\tilde{\mathbf{A}} = \begin{bmatrix} (1;1;1) & (a_{12}^L; a_{12}^M; a_{12}^U) & \cdots & (a_{1n}^L; a_{1n}^M; a_{1n}^U) \\ (\frac{1}{a_{12}^U}; \frac{1}{a_{12}^M}; \frac{1}{a_{12}^L}) & (1;1;1) & \cdots & (a_{2n}^L; a_{2n}^M; a_{2n}^U) \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{1}{a_{1n}^U}; \frac{1}{a_{1n}^M}; \frac{1}{a_{1n}^L}) & (\frac{1}{a_{2n}^U}; \frac{1}{a_{2n}^M}; \frac{1}{a_{2n}^L}) & \cdots & (1;1;1) \end{bmatrix}, \tag{7}$$

where $1 \le a_{ij}^L \le a_{ij}^M \le a_{ij}^M$, *i,j* = 1,2,...,*n*. Without loss of generality we assume that $1 \le a_{ij}^M \le a_{ik}^M$ whenever $i \le j \le k$.

4 Algorithm

The proposed decision support method of finding the best variant (or ranking the variants) can be formulated by an algorithm in the following three steps:

1. Calculate the triangular fuzzy weights from the fuzzy pair-wise comparison matrix or from fuzzy triangular fuzzy values.
2. Calculate the aggregating triangular fuzzy evaluations of the variants.
3. Find the "best" variant (eventually, rank the variants) defined as triangular fuzzy numbers.

Below we explain in details the individual steps of this algorithm.

Step 1: Calculate the triangular fuzzy weights from the fuzzy pair-wise comparison matrix or from fuzzy triangular fuzzy values

From now on we assume that the input data are uncertain and they are given by triangular fuzzy values. We distinguish two situations:

(A) the data are given by a triangular fuzzy pair-wise comparison matrix, or,
(B) the data are given in the form of triangular fuzzy values of the criteria.

Our purpose is to calculate the triangular fuzzy numbers – in this context we call them *fuzzy weights* - as evaluations of the relative importance of the criteria, and/or evaluations of the feedback of the criteria and/or evaluations of the variants according to the individual criteria.

(A) Given a fuzzy pair-wise comparison matrix $\tilde{\mathbf{A}}$ defined by (7). We assume that there exists a fuzzy vectors of triangular fuzzy weights $\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_r$, $\tilde{w}_i = (w_i^L; w_i^M; w_i^U)$, *i* = 1,2,...,*r*, such that the pair-wise comparison matrix (7) is an estimation of the fuzzy matrix

$$\tilde{\mathbf{W}} = \begin{bmatrix} \frac{\tilde{w}_1}{\tilde{w}_1} & \frac{\tilde{w}_1}{\tilde{w}_2} & \cdots & \frac{\tilde{w}_1}{\tilde{w}_r} \\ \frac{\tilde{w}_2}{\tilde{w}_1} & \frac{\tilde{w}_2}{\tilde{w}_2} & \cdots & \frac{\tilde{w}_2}{\tilde{w}_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\tilde{w}_r}{\tilde{w}_1} & \frac{\tilde{w}_r}{\tilde{w}_2} & \cdots & \frac{\tilde{w}_r}{\tilde{w}_r} \end{bmatrix}.$$

Here, $r = n$ if we look for the weights – i.e. the relative importance of the criteria, or $r = m$, in case we look for the weights – i.e. the evaluations of the variants according to some criterion. We shall find the fuzzy weights $\tilde{w}_1, \tilde{w}_2, ..., \tilde{w}_r$ by minimizing the fuzzy functional

$$\tilde{H} = \sum_{i,j} \tilde{\log}\left(\frac{\tilde{w}_i}{\tilde{w}_j} - \tilde{a}_{ij}\right)^2. \tag{8}$$

In (8), minimization of $\tilde{H}$ is understood in the sense of solving the optimization problem

$$\sum_{i,j} \max\left\{\log\left(\frac{w_i^L}{w_j^U} - a_{ij}^L\right)^2, \log\left(\frac{w_i^M}{w_j^M} - a_{ij}^M\right)^2, \log\left(\frac{w_i^U}{w_j^L} - a_{ij}^U\right)^2\right\} \longrightarrow \min; \tag{9}$$

subject to

$$w_i^U \geq w_i^M \geq w_i^L \geq 0, \; i = 1,2,...,r. \tag{10}$$

It can be shown, see [2], that there exists a unique explicit solution of problem (9),(10) as follows:

$$\tilde{w}_k = (w_k^L; w_k^M; w_k^U), \; k = 1,2,...,r,$$

where

$$w_k^S = \frac{\left(\prod_{j=1}^{r} a_{kj}^S\right)^{1/r}}{\sum_{i=1}^{r}\left(\prod_{j=1}^{r} a_{ij}^M\right)^{1/r}}, \; S \in \{L, M, U\}. \tag{11}$$

In [2], the method of calculating triangular fuzzy weights by (11) from the triangular fuzzy pair-wise comparison matrix (7) is called the *logarithmic least squares method*. This method can be applied both for calculating the triangular fuzzy weights of the criteria and for eliciting relative triangular fuzzy values of the criteria for the individual variants. Moreover, it can be used also for calculating feedback impacts of criteria on the other criteria.

(B) Now we assume that the evaluations of the importance of the criteria or evaluations of variants according to some criteria are uncertain, particularly expressed by triangular fuzzy numbers.
Let $\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_r$, $\tilde{v}_i = (v_i^L; v_i^M; v_i^U)$, i = 1,2,...,r, be set of triangular fuzzy numbers, e.g. fuzzy evaluations of variants according to some criterion. We assume that $0 < v_i^L \leq v_i^M \leq v_i^U$. For the purpose of aggregation of partial evaluations we "normalize" the values to obtain triangular fuzzy weights. We calculate the normalized fuzzy values as follows:

$$\tilde{w}_k = (w_k^L; w_k^M; w_k^U), \; k = 1,2,...,r,$$

where

$$\tilde{w}_k = \left(\frac{v_k^L}{S}; \frac{v_k^M}{S}; \frac{v_k^U}{S}\right) \tag{12}$$

and $S = \sum_j v_j^M$.

Step 2: Calculate the aggregating triangular fuzzy evaluations of the variants.

Having calculated triangular fuzzy weights and evaluations as it was mentioned above, we calculate the synthesis: the aggregated triangular fuzzy evaluation of the individual variants. For this purpose we use the formula (4), eventually, the approximate formula (6), applied to triangular fuzzy matrices, i.e. matrices with the elements being triangular fuzzy numbers given by triples of positive numbers:

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{32}\left(\mathbf{I} \simeq \tilde{\mathbf{W}}_{22}\right)^{-1}\tilde{\mathbf{W}}_{21}, \tag{4*}$$

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{32}(\mathbf{I} \tilde{+} \tilde{\mathbf{W}}_{22} \tilde{+} \tilde{\mathbf{W}}_{22}^{2} \tilde{+} \tilde{\mathbf{W}}_{22}^{3})\tilde{\mathbf{W}}_{21}. \tag{6*}$$

Here, for addition, subtraction and multiplication of triangular fuzzy numbers we use the fuzzy operations defined earlier.

Step 3: Find the "best" variant, order the variants.

In Step 2 we have found the variants described as triangular fuzzy numbers, i.e. by (6*) we calculated the triangular fuzzy vector $\tilde{Z} = (\tilde{z}_1,...,\tilde{z}_n)^T = \left((z_1^L; z_1^M; z_1^U),...,(z_{m1}^L; z_m^M; z_m^U)\right)^T$.

The simplest method for ordering a set of triangular fuzzy numbers is the *center of gravity method.* This method is based on computing the x-th coordinates x_i^g of the center of gravity of every triangle given by the corresponding membership functions $\tilde{z}_i$, $i = 1,2,...,n$. Evidently, it holds

$$x_i^g = \frac{z_i^L + z_i^M + z_i^U}{3}. \tag{13}$$

By (12) the variants can be ordered from the best (with the biggest value of (12)) to the worst (with the lowest value of (13)).

There exist more sophisticated methods for ordering fuzzy numbers, see e.g. [4], for a comprehensive review of comparison methods see [2].

5 Case study

In this section we analyze an example of decision making situation with 3 decision criteria and 4 variants. The goal of this realistic decision situation is to find the "best" product (a car) from 4 pre-selected ones according to 3 criteria: economical, technical and esthetical.

First, we apply the algorithm based on ANP described in the previous section for solving the decision making problem. The evaluation of the weights of criteria and variants according to criteria as well as the feedback between the criteria is based on the data from pair-wise comparison matrices. Here, instead of classical Saaty´s 9-point scale we use triangular fuzzy numbers to evaluate preferences between alternatives. In the data given below we use only symmetric triangular fuzzy numbers, this is, however, not necessary, the evaluations could be also non-symmetric.

Second, we solve the same problem applying classical AHP, i.e. we use non-fuzzy evaluations in the pair-wise comparisons and do not consider the feedback. In fact, this approach is a particular case of the previous more general situation when taking crisp evaluations, i.e. $a^L = a^M = a^U$, and the feedback matrix is a zero matrix, i.e. $\mathbf{W}_{22} = \mathbf{0}$.

Finally, we compare the previous "crisp" situation without feedback with the situation again with crisp evaluations, however, with a non-zero feedback between the criteria expressed by crisp values.

Step 1: Evaluate the pair-wise comparison matrices and calculate the corresponding weights

The data for importance of the criteria are given by the following pair-wise comparison matrix **C**:

C =

	L	M	U	L	M	U	L	M	U	
	1,000	1,000	1,000	2,000	3,000	4,000	4,000	5,000	6,000	- criterion 1: C1
	0,250	0,333	0,500	1,000	1,000	1,000	3,000	4,000	5,000	- criterion 2: C2
	0,167	0,200	0,250	0,200	0,250	0,333	1,000	1,000	1,000	- criterion 3: C3
	- criterion 1: C1			- criterion 2: C2			- criterion 3: C3			

By (11) we calculate the corresponding triangular fuzzy weights, i.e. the relative fuzzy importance of the individual criteria:

W21=

L	M	U		
0,508	0,627	0,733	- criterion 1: (economical criterion	- v(C1)
0,231	0,280	0,345	- criterion 2: (technical criterion	- v(C2)
0,082	0,094	0,111	- criterion 3: (esthetical criterion	- v(C3)

The data for fuzzy evaluations of the variants according to the individual criteria are given by the following 3 pair-wise comparison matrices **A1**, **A2**, **A3**:

A1 =

L	M	U	L	M	U	L	M	U	L	M	U	
1,000	1,000	1,000	2,000	3,000	4,000	4,000	5,000	6,000	6,000	7,000	8,000	- variant 1: V1
0,250	0,333	0,500	1,000	1,000	1,000	3,000	4,000	5,000	5,000	6,000	7,000	- variant 2: V2
0,167	0,200	0,250	0,200	0,250	0,333	1,000	1,000	1,000	4,000	5,000	6,000	- variant 3: V3
0,125	0,143	0,167	0,143	0,167	0,200	0,167	0,200	0,250	1,000	1,000	1,000	- variant 4: V4
- variant 1: V1			- variant 2: V2			- variant 3: V3			- variant 4: V4			

A2 =

L	M	U	L	M	U	L	M	U	L	M	U	
1,000	1,000	1,000	1,000	2,000	3,000	2,000	3,000	4,000	3,000	4,000	5,000	- variant 3: V3
0,333	0,500	1,000	1,000	1,000	1,000	1,000	2,000	3,000	2,000	3,000	4,000	- variant 2: V2
0,250	0,333	0,500	0,333	0,500	1,000	1,000	1,000	1,000	1,000	2,000	3,000	- variant 1: V1
0,200	0,250	0,333	0,250	0,333	0,500	0,333	0,500	1,000	1,000	1,000	1,000	- variant 4: V4
- variant 3: V3			- variant 2: V2			- variant 1: V1			- variant 4: V4			

A3 =

L	M	U	L	M	U	L	M	U	L	M	U	
1,000	1,000	1,000	3,000	4,000	5,000	6,000	7,000	8,000	7,000	8,000	9,000	- variant 2: V2
0,200	0,250	0,333	1,000	1,000	1,000	4,000	5,000	6,000	6,000	7,000	8,000	- variant 3: V3
0,125	0,143	0,167	0,167	0,200	0,250	1,000	1,000	1,000	5,000	6,000	7,000	- variant 1: V1
0,111	0,125	0,143	0,125	0,143	0,167	0,143	0,167	0,200	1,000	1,000	1,000	- variant 4: V4
- variant 2: V2			- variant 3: V3			- variant 1: V1			- variant 4: V4			

The corresponding fuzzy matrix **W32** of fuzzy weights – evaluations of variants according to the individual criteria is calculated by (11) as

W32 =

L	M	U	L	M	U	L	M	U	
0,450	0,547	0,636	0,113	0,160	0,233	0,088	0,100	0,114	- variant 1: V1
0,238	0,287	0,349	0,191	0,278	0,393	0,518	0,598	0,674	- variant 2: V2
0,103	0,121	0,144	0,330	0,467	0,587	0,229	0,266	0,309	- variant 3: V3
0,040	0,045	0,052	0,076	0,095	0,135	0,033	0,036	0,041	- variant 4: V4
- criterion 1: C1			- criterion 2: C2			- criterion 3: C3			

The data for evaluations of fuzzy feedbacks between the criteria are given again by the following 3 pair-wise comparison matrices **B1**, **B2**, **B3**:

B1 =	L	M	U	L	M	U	
	1,000	1,000	1,000	2,000	3,000	4,000	- criterion 2: C2
	0,250	0,333	0,500	1,000	1,000	1,000	- criterion 3: C3
	- criterion 2: C2			- criterion 3: C3			

B2 =	L	M	U	L	M	U	
	1,000	1,000	1,000	1,000	2,000	3,000	- criterion 1: C1
	0,333	0,500	1,000	1,000	1,000	1,000	- criterion 3: C3
	- criterion 1: C1			- criterion 3: C3			

B3 =	L	M	U	L	M	U	
	1,000	1,000	1,000	3,000	4,000	5,000	- criterion 1: C1
	0,200	0,250	0,333	1,000	1,000	1,000	- criterion 2: C2
	- criterion 1: C1			- criterion 2: C2			

By using (11) we obtain again the corresponding fuzzy weights and arrange these weights into the fuzzy feedback matrix W22. We put zeros into the main diagonal as we do not expect an impact of the criterion on itself.

W22 =	L	M	U	L	M	U	L	M	U	
	0,000	0,000	0,000	0,471	0,667	0,816	0,693	0,800	0,894	- criterion 1: C1
	0,612	0,750	0,866	0,000	0,000	0,000	0,179	0,200	0,231	- criterion 2: C2
	0,217	0,250	0,306	0,272	0,333	0,471	0,000	0,000	0,000	- criterion 3: C3
	- criterion 1: C1			- criterion 2: C2			- criterion 3: C3			

Within the decision model we should also consider a relationship (i.e. importance) between the influence of the criteria on the variants, and, on the other hand, the impact of the criteria on themselves. In our decision model, this relationship has the form of "weights" w1, w2, i.e. two non-negative numbers summing up to one resulting again from pair-wise comparison matrix **D**:

D =			
	1,000	0,300	- variants
	3,333	1,000	- feedback
	- variants	- feedback	

Hence, we obtain the corresponding weights $w1$, $w2$:

w1 =	0,231	- variants
w2 =	0,769	- feedback

Now, we multiply each element of matrix **W32** by $w1$ and by the same way each element of matrix **W22** by $w2$. Consequently, we obtain new matrices **W32*** and **W22*** and put them into the supermatrix **W**. Evidently, it becomes the stochastic matrix:

W32* =	L	M	U	L	M	U	L	M	U	
	0,104	0,126	0,147	0,026	0,037	0,054	0,020	0,023	0,026	- variant 1: V1
	0,055	0,066	0,081	0,044	0,064	0,091	0,120	0,138	0,155	- variant 2: V2
	0,024	0,028	0,033	0,076	0,108	0,135	0,053	0,061	0,071	- variant 3: V3
	0,009	0,010	0,012	0,017	0,022	0,031	0,008	0,008	0,009	- variant 4: V4
	- criterion 1: C1			- criterion 2: C2			- criterion 3: C3			

	L	M	U	L	M	U	L	M	U	
W22* =	0,000	0,000	0,000	0,363	0,513	0,628	0,533	0,615	0,688	- criterion 1: C1
	0,471	0,577	0,666	0,000	0,000	0,000	0,138	0,154	0,178	- criterion 2: C2
	0,167	0,192	0,236	0,209	0,256	0,363	0,000	0,000	0,000	- criterion 3: C3
	- criterion 1: C1			- criterion 2: C2			- criterion 3: C3			

Step 2: Calculate the aggregating triangular fuzzy evaluations of the variants.

By computing triangular fuzzy weights and evaluations as it was mentioned earlier, we calculate the synthesis: the aggregated triangular fuzzy evaluation of the individual variants. For this purpose we use the approximate formula (6), applied for triangular fuzzy matrices, i.e. matrices with the elements being triangular fuzzy numbers – triples of positive numbers:

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{32}(\mathbf{I} \tilde{+} \tilde{\mathbf{W}}_{22} \tilde{+} \tilde{\mathbf{W}}_{22}^{2} \tilde{+} \tilde{\mathbf{W}}_{22}^{3})\tilde{\mathbf{W}}_{21} \,. \tag{6*}$$

Here, for addition, subtraction and multiplication of triangular fuzzy numbers in (6*) we use the fuzzy arithmetic operations defined earlier.

In our case we obtain the aggregating triangular fuzzy evaluations of the variants:

	L	M	U	
Z =	0,174	0,327	0,584	- V1
	0,178	0,341	0,651	- V2
	0,134	0,271	0,504	- V3
	0,033	0,061	0,119	- V4

(13)

In the last step we have to rank the evaluations of the above fuzzy variants resulting in the best decision – finding the "best" variant by using a proper way of ordering the triangular fuzzy numbers.

Step 3: Find the "best" variant, rank the variants.

In Step 2 we have found the variants described as triangular fuzzy numbers, i.e. by (6*) we calculated the triangular fuzzy vector $\tilde{Z} = (\tilde{z}_1, ..., \tilde{z}_4)^T = ((z_1^L; z_1^M; z_1^U), ..., (z_4^L; z_4^M; z_4^U))^T$ given in (13). Here we use the simplest method for ordering a set of triangular fuzzy numbers - the center of gravity method computing the x-th coordinates x_i^g of the center of gravity of every triangle given by the corresponding membership functions $\tilde{z}_i$, $i = 1,2,3,4$ by the formula $x_i^g = \dfrac{z_i^L + z_i^M + z_i^U}{3}$. Particularly, we get

xgi	variant	rank
0,362	V1 =	2
0,390	V2 =	1
0,303	V3 =	3
0,071	V4 =	4

By (12) the variants can be ordered from the best (with the biggest value of xgi) to the worst (with the lowest value of xgi). The situation is graphically depicted in Fig. 4.

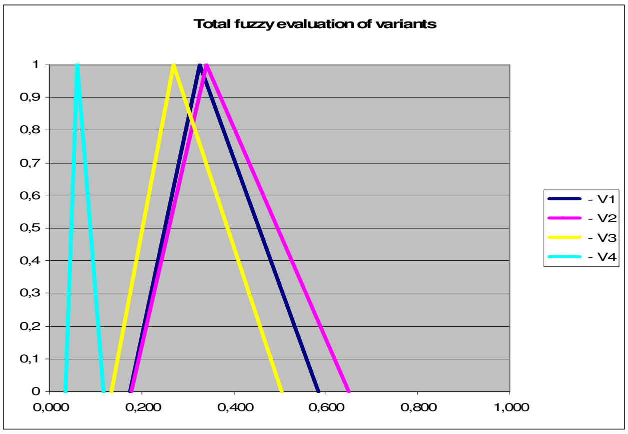


Fig. 4. Total fuzzy evaluation of variants

Moreover, we solve the same problem applying classical AHP, i.e. we use non-fuzzy evaluations in the pair-wise comparisons and do not consider the feedback. Such an approach is a particular case of the previous more general situation when considering crisp evaluations, i.e. $a^L = a^M = a^U$, and the feedback matrix is a zero matrix, i.e. $\mathbf{W}_{22} = \mathbf{0}$.
Hence, we get

W32 =

L	M	U	L	M	U	L	M	U
0,126	0,126	0,126	0,037	0,037	0,037	0,023	0,023	0,023
0,066	0,066	0,066	0,064	0,064	0,064	0,138	0,138	0,138
0,028	0,028	0,028	0,108	0,108	0,108	0,061	0,061	0,061
0,010	0,010	0,010	0,022	0,022	0,022	0,008	0,008	0,008

By (4*) we obtain the aggregating crisp evaluations of the variants:

Z =

L	M	U		rank
0,397	0,397	0,397	- V1	1
0,314	0,314	0,314	- V2	2
0,231	0,231	0,231	- V3	3
0,058	0,058	0,058	- V4	4

The situation is graphically depicted in Fig. 5.

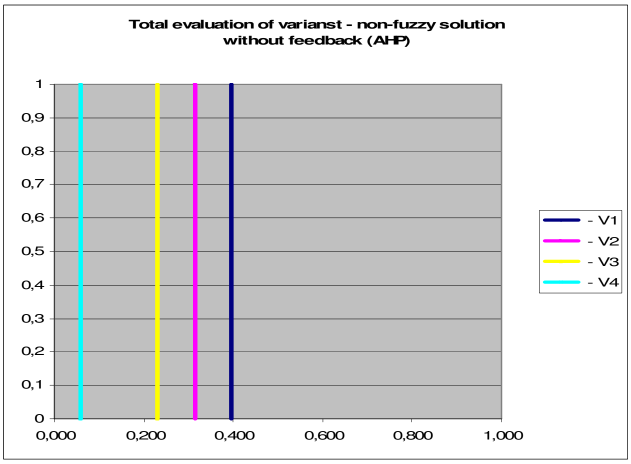


Fig. 5. Total evaluation of variants – AHP

Finally, we compare the previous "crisp" situation without feedback (i.e. classical AHP) with the situation with the same crisp evaluations, however, with a non-zero feedback between the criteria expressed by crisp values (i.e. classical ANP):

W22 =

	economical criterion			technical criterion			esthetical criterion			
	L	M	U	L	M	U	L	M	U	
	0,000	0,000	0,000	0,513	0,513	0,513	0,615	0,615	0,615	economical criterion
	0,577	0,577	0,577	0,000	0,000	0,000	0,154	0,154	0,154	technical criterion
	0,192	0,192	0,192	0,256	0,256	0,256	0,000	0,000	0,000	esthetical criterion

By (6*) we obtain the aggregating crisp evaluations of the variants:

Z =

L	M	U		rank
0,327	0,327	0,327	- V1	2
0,341	0,341	0,341	- V2	1
0,271	0,271	0,271	- V3	3
0,061	0,061	0,061	- V4	4

The situation is graphically depicted in Fig. 6.

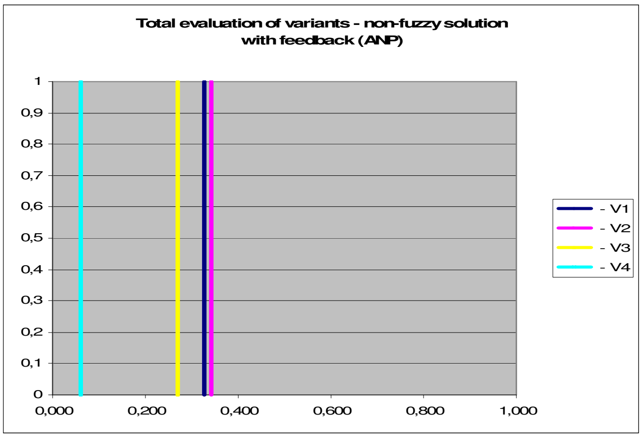


Fig. 6. Total evaluation of variants – ANP

Considering feedback dependences between the criteria the total rank of the variants can be changed as we have demonstrated in the above case study.

6 Conclusion

In this paper we have proposed a new decision model based on ANP for solving the decision making problem with fuzzy pair-wise comparisons and a feedback between the criteria. The evaluation of the weights of criteria, the variants as well as the feedback between the criteria is based on the data given in pair-wise comparison matrices. Extended arithmetic operations with fuzzy numbers have been proposed as well as ordering fuzzy relations to compare fuzzy outcomes. An illustrating case study has been presented to clarify the methodology. Based on the case study we can conclude that fuzzy (soft) evaluation of pair-wise comparisons may be more comfortable and appropriate for DM. Occurrence of dependences among criteria is more realistic. Dependences among criteria influence the final rank of variants and presence of fuzziness in evaluations change the final rank of variants.

References

[1] Buckley, J.J., Fuzzy hierarchical analysis. *Fuzzy Sets and Systems* 17, 1985, 1, p. 233-247, ISSN 0165-0114.

[2] Chen, S.J., Hwang, C.L. and Hwang, F.P., *Fuzzy multiple attribute decision making*. Lecture Notes in Economics and Math. Systems, Vol. 375, Springer-Verlag, Berlin – Heidelberg 1992, ISBN 3-540-54998-6.

[3] Horn, R. A., Johnson, C. R., *Matrix Analysis*, Cambridge University Press, 1990, ISBN 0521305861.

[4] Ramik, J., Duality in fuzzy linear programming with possibility and necessity relations. *Fuzzy Sets and Systems* 157, 2006, 1, p. 1283-1302, ISSN 0165-0114.

[5] Saaty, T.L., Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, p. 57-68, ISSN 0165-0114.

[6] Saaty, T.L., *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991, ISBN 0-9620317-2-0.

[7] Saaty, T.L., *Decision Making with Dependence and Feedback – The Analytic Network Process*. RWS Publications, Pittsburgh, 2001, ISBN 0-9620317-9-8.

[8] Van Laarhoven, P.J.M. and Pedrycz, W., A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems* 11, 1983, 4, p. 229-241, ISSN 0165-0114.

This research was partly supported by the grant project of GACR No. 402060431

Bayesian approach to change point detection of unemployment rate via MCMC methods

Soňa Reisnerová [1]
Institute of Information Theory and Automation, AS CR,
Pod vodárenskou věží 4, 182 08 Praha 8,
reisnero@utia.cas.cz

Abstract

Bayesian statistics becomes more popular in the latest days because of its ability to estimate parameters in models with quite complicated structure. The main problem connected with Bayesian methods always was to obtain posterior distribution. One way we can overcome this limitation is the application of random generation methods called Markov chain Monte Carlo (MCMC). In the present paper we use Bayesian approach with the help of MCMC methods for the change point detection in a hazard rate model. Concretely, we consider the discrete time process of numbers of events (counts) modeled as a series of Poisson distributions. Simultaneously, it is assumed that the intensities (Poisson parameters) depend on other observed factors. These factors can be expressed in the Cox's regression model form with parametric (e.g. linear in time) baseline hazard function, in which we expect an abrupt change at unknown time point. The estimation procedure for all parameters except parameter for the time of change is based on the Gibbs algorithm, for change point detection we use the Metropolis-Hastings algorithm. The significance of all model parameters, including the change magnitude, is also investigated. Finally, numerical demonstration of this model on numbers of unemployed people in Czech Republic is included.

Keywords

Bayesian approach; Change point problem; Cox's regression model; Hazard rate; MCMC methods.
JEL: C11, C15, J64

1 Bayes statistics and MCMC methods

Let us have a random vector X which has distribution depending on parameter vector θ. Our task is to estimate parameter θ based on observed realizations of vector X. Despite of clasical statistical point of view, where parameter vector is constant, Bayes approach works with parameters as vector of random variables. Marginal distribution of parameter vector θ is called prior distribution. As posterior distribution we consider conditional distribution of parameter θ under $X = x$, for x is observed value of vector X. According to the Bayes theorem, the density of posterior distribution is:

$$f(\theta \mid X = x) \sim f(\theta) \cdot f(X \mid \theta).$$

It means that the density of posterior distribution can be obtain as normalized product of appropriate part of likelihood function and prior density. The basics on Bayes statistics can be found in [5].

There are many situations in which we can use Bayes approach with advantage. If we have discretized version of spatial-temporal random point process model with mutually independent temporal part and parameters of spatial part have certain prior distribution with autoregressive dependence it is really helping to use this approach, for more details see [3] or [7].

Bayes statistics is often used together with well-known simulation methods, especially let us mention Markov Chain Monte Carlo methods. By such methods we can generate Markov chains with certain stacionary distribution and we use them for obtaining parameters estimation in difficult probability models, for example [8].

Concretely we apply Gibbs algorithm mentioned in [2] for generating random variables from posterior distribution and we also employ combination of Gibbs algorithm and sampling-rejecting method, for more detail see [9]. Obtained Markov chain convereges in distribution to the posterior distribution, in order to obtain the chain with approximately stationary distribution we remove initial part of the chain. The

[1] The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s.

problem how large size of chain should be removed is discussed for example in [1]. Median and others sampling quantiles obtained from the rest of chain should be close and converges to posterior distribution characteristics.

Such methods are also used with advantage for change point detection problem, see for example [6]. In present paper we deal with temporal random point process model with considered change in time. In the first part we specify the model, namely we formulate Poisson model with change in intensity of events, select prior distributions and derive conditional posterior distributions for all parameters. Finally, the numerical results on numbers of newly unemployed people in Czech Republic are presented.

2 Nonhomogeneous Poisson process

At first let us specify the data and the model suitable for them. If we observe occurence of some events in time, we can be interested in intensity with which the events happened. Such data are often represented by summary counts in discrete time. So we can considered them as discrete Cox's stochastic point process with random intensity changing in time. Randomness comes into the model from Bayes approach. Simultaneously we suppose that the process is driven by nonhomogeneous Poisson process with change at unknown time τ. Observed data follow the model:

$$\begin{aligned} N(t) &\sim Poiss(\lambda(t)), \qquad (1)\\ \text{where} \quad \lambda(t) &= N_0 \exp(a + ct + \beta_{j(t)}), \quad \text{for} \quad t = 1, \ldots, \tau \\ &= N_0 \exp(a + b + (c + d)t + \beta_{j(t)}), \quad \text{for} \quad t = \tau + 1, \ldots, T, \end{aligned}$$

where $j(t) = (t mod 12)+1$ and N_0 is some known constant connected with whole amount of subjects under the risk of the event. The use of such a constant keeps the value of parameters estimate in reasonable limits. Parameter $\beta = (\beta_1, \ldots, \beta_{12})$ explains the influence of months. We can see that the intensity is changed at unknown time τ, we consider both possible jump and change in slope of line. If $I(.)$ is the indicator function, we can rewrite the intensity function $\lambda(t)$ in the following way:

$$\lambda(t) = N_0 \exp(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)}) \quad \text{for} \quad t = 1, \ldots, T.$$

From (1) it follows that the probability of observed data, given the model (or the likelihood, given the data) has the following form:

$$\begin{aligned} L(N \mid a, b, c, d, \beta) &= \prod_{t=1}^{T} e^{-\lambda(t)} \frac{\lambda(t)^{N(t)}}{N(t)!} \qquad (2)\\ &\sim \prod_{t=1}^{T} \exp(-N_0 \exp(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)})) \cdot \\ &\cdot \quad \exp(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)})^{N(t)}. \end{aligned}$$

3 Estimation of model parameters

At first let us briefly describe the generating procedure. In every step we generate new candidate for τ, accept it or reject with certain probability, then generate new values for parameters a, b, β according to the following rules and compute parameters c and d.

As priors, we will consider gamma distribution, so let us remind the form of its density $\Gamma(u, v)$:

$$\begin{aligned} f(X \mid u, v) &= x^{u-1} e^{-x/v} \Gamma(u) v^u, \\ \text{then} \quad E(X) = uv \quad &\text{and} \quad var(X) = uv^2 \ . \end{aligned}$$

3.1 Estimation of parameters a and b

We assume that parameters a and b are both mutually independent and we suppose that both of them have log-gamma prior distribution. Let us consider parameter a first. For $A = e^a$ we obtain prior distribution:

$$\begin{aligned} \mathcal{L}_0(A) &= \Gamma\Big(\delta, \frac{\gamma_A}{\delta}\Big), \qquad (3)\\ f_0(A) &= A^{\delta-1} e^{-A\delta/\gamma_A} (\gamma_A/\delta)^{\delta} \Gamma(\delta)^{-1}. \end{aligned}$$

The posterior distribution is proportional to the product of the likelihood and the prior, in our case of (2) and prior density formulated in (3). In the sequel, we shall derive the form of conditional posterior distribution of model parameter, given the rest of parameters and data. Thus, for parameters a, when we multiply (2) and (3), we obtain:

$$\begin{aligned} f^*(A \mid .) &= e^{-TN_0A} A^{\sum_{t=1}^{T} N(t)} f_0(A) K_1 \\ &= e^{-A(TN_0+\delta/\gamma_A)} A^{\sum_{t=1}^{T} N(t)+\delta-1} K_2, \end{aligned}$$

K_1 and K_2 are normalizing constants. We can rewrite it as:

$$\mathcal{L}^*(A \mid .) = \Gamma\Big(\sum_{t=1}^{T} N(t)+\delta, \Big(TN_0+\delta/\gamma_A\Big)^{-1}\Big).$$

As a result we get gamma distribution, so we can use Gibbs algorithm and generate A from conditional posterior distribution directly. Finally we obtain a from relation $a = logA$.

For parameter b we will also assume log-gamma prior distribution. For $B = e^b$ we obtain prior distribution:

$$\begin{aligned} \mathcal{L}_0(B) &= \Gamma\Big(\delta, \frac{\gamma_B}{\delta}\Big), \\ f_0(B) &= B^{\delta-1} e^{-B\delta/\gamma_B} (\gamma_B/\delta)^{\delta} \Gamma(\delta)^{-1}. \end{aligned} \tag{4}$$

Similarly from (2) and (4) we can derive conditional posterior distribution of parameter B. We obtain

$$\begin{aligned} f^*(B \mid .) &= e^{-\sum_{t=1}^{T} N_0 B^{I(t>\tau)}} B^{\sum_{t=1}^{T} I(t>\tau)N(t)} f_0(B) K_3 \\ &= e^{-B(\sum_{t>\tau} N_0+\delta/\gamma_B)} B^{\sum_{t>\tau} N(t)+\delta-1} K_4, \end{aligned}$$

K_3 and K_4 are normalizing constants. We can rewrite it as:

$$\mathcal{L}^*(B \mid .) = \Gamma\Big(\sum_{t>\tau} N(t)+\delta, \Big(\sum_{t>\tau} N_0+\delta/\gamma_B\Big)^{-1}\Big).$$

As a result we get gamma distribution again, so we can use Gibbs algorithm and generate B from conditional posterior distribution directly too. Finally we obtain b from relation $b = logB$.

3.2 Estimation of parameters $\beta = (\beta_1, \ldots, \beta_{12})$

We suppose that seasonal parameters $\beta(j),\ j = 1, \ldots, 12$ have mutualy independant prior distribution driven by:

$$\begin{aligned} \mathcal{L}_0(\beta(j)) &= \mathcal{N}(0, \sigma^2), \\ f_0(\beta(j)) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}\beta_j^2\Big). \end{aligned} \tag{5}$$

From (2) and (5) we can derive conditional posterior distribution of parameter $\beta(j),\ j = 1, \ldots, 12$:

$$f^*(\beta_j \mid .) = h(e^{\beta_j}) \cdot f_0(\beta_j),$$

where

$$h(e^{\beta_j}) = \exp\{-\sum_{t:j(t)=j} N_0 e^{\beta_j}\} e^{\beta_j \sum_{t:j(t)=j} N(t)} K_5,$$

corresponds to gamma distribution for e^{β_j}, K_5 is normalizing constant. The posterior conditional distribution consists of two parts, the first is the gamma density for e^{β_j}, the second is nonnegative and bounded function. For the generation of a new representative of β_j the following procedure can be used: Generate a candidate e^{β_j} from

$$\Gamma\Big(\sum_{t:j(t)=j} N(t)+1, \Big(\sum_{t:j(t)=j} N_0\Big)^{-1}\Big),$$

take β_j and accept it with probability $p = f_0(\beta_j)/\max f_0(.)$. This second step is actually an application of "sampling-rejection" rule. Maximum may concern to all possible values of β_j or to certain pre-sampled population. Method "sampling-rejection" works more effectively than Metropolis-Hastings algorithm in general, for more details see [4].

3.3 Estimation of parameters c, d and τ

For parameters c and d we use maximum likelihood estimate. Log-likelihood function has following form:

$$\begin{aligned} logL(N \mid a, b, \beta, \tau) &= \sum_{t=1}^{T} N(t)(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)}) \\ &- N_0 \exp(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)}), \end{aligned}$$

we find the maximum of this function with respect to parameters c and d. Because analytical solution doesn't exist, we use a Newton-Raphson method as a helping tool for finding the extreme.

In every step we generate new candidate τ_1 from discrete uniform distribution on $\{1, \ldots, T-1\}$ and accept it with probability $p = L(\tau_1)/L(\tau_0)$, where τ_0 is last accepted candidate and L is likelihood function from (2). It means that we use Metropolis-Hastings algorithm for generating parameter τ.

4 Analysis of unemployment data

The model and methods were prepared for the analysis of the real data, the number of newly registered unemployed people. Data come from the Czech Statistical Office and contain numbers of unemployed people in Czech Republic during 66 month, from 1/1998 - 6/2003. We took N_0 equal to the number of active people in whole Czech Republic divided by 100.

All computations were made in Matlab 7.0. We generated the Markov chain from 10000 global iterations (sweeps), the first third of values was omitted and from the remaining part the samples representing approximately the posterior distributions were obtained. However, for every choosing of initial parameters value δ, σ, γ_a, γ_b we were not able to find statistical model with meaningful distribution of parameter τ. There wasn't possible to set enough strong acceptation rule for which the result would settle down.

Displayed results correspond to the models, where all parameters except parameter τ are estimated by maximum likelihood method and parameter τ is generated and accepted according to the Metropolis-Hastings algorithm. We mention two statistical models. Model 1 considers possible change at unknown time without jump, only with change in slope of line. Model 2 considers possible change at unknown time with jump and change in slope of line as well. We set the last parameter of vector β to zero and it means that the rest of parameters are estimated with relation to December.

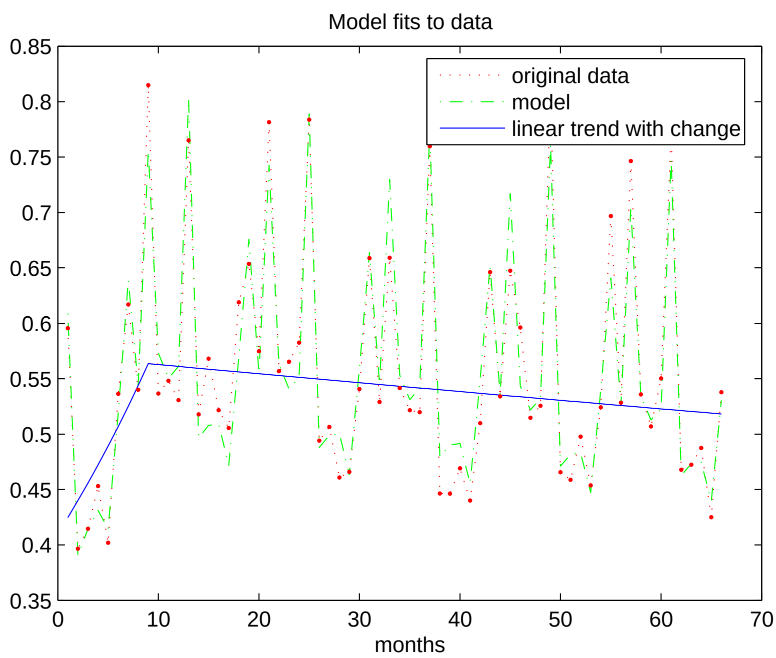


Figure 1 - model 1

Figure 1 displays the ratio N_t/N_0 as original data, the estimated hazard function

$$\hat{\lambda}(t) = N_0 \exp(a + (c + dI(t > \tau))t + \beta_{j(t)}) \quad \text{for} \quad t = 1, \ldots, 66$$

as a model and linear trend with change point, $\exp\{\hat{a} + (\hat{c} + \hat{d}\mathrm{I}(t > \hat{\tau}))t\}$.

Following Table 1 shows the estimated parameters a, c, d, β of model 1 with (asymptotic) 95 % confidence intervals.

parameters	estimates model 1	confidence interval model 1	
a	-0.8916	-0.8988	-0.8843
c	0.0353	0.0346	0.0361
d	-0.0368	-0.0376	-0.0360
January	0.3600	0.3552	0.3649
February	-0.1189	-0.1242	-0.1136
March	-0.0951	-0.1004	-0.0899
April	-0.0910	-0.0962	-0.0857
May	-0.1659	-0.1712	-0.1605
June	0.0236	0.0185	0.0287
July	0.1964	0.1913	0.2015
August	0.0007	-0.0046	0.0061
September	0.2936	0.2887	0.2986
Octomber	0.0183	0.0130	0.0236
November	-0.0209	-0.0262	-0.0155

Figure 2 displays the ratio N_t/N_0 as original data again, the estimated hazard function

$$\hat{\lambda}(t) = N_0 \exp(a + bI(t > \tau) + (c + dI(t > \tau))t + \beta_{j(t)}) \quad \text{for} \quad t = 1, \dots, T$$

as a model and linear trend with change point, $\exp\{\hat{a} + \hat{b}\mathrm{I}(t > \hat{\tau}) + (\hat{c} + \hat{d}\mathrm{I}(t > \hat{\tau}))t\}$.

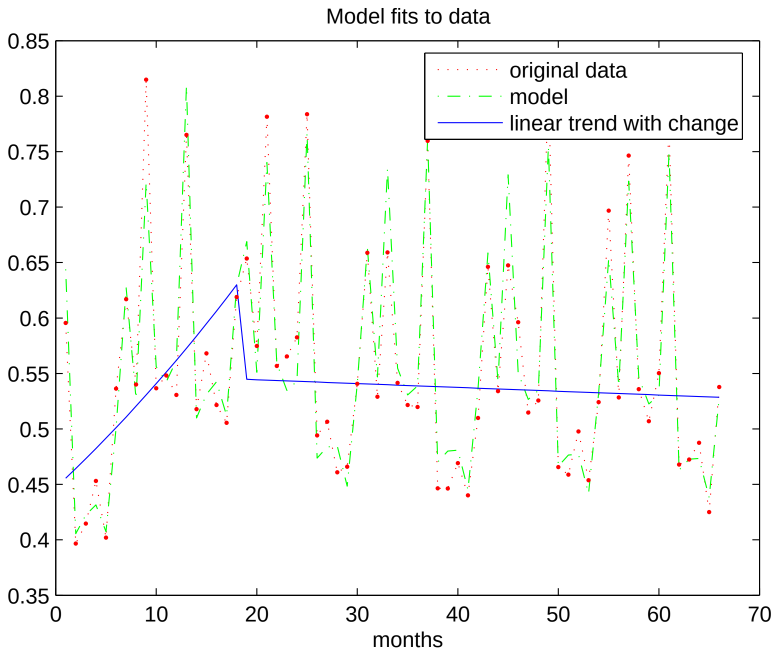


Figure 2 - model 2

Following Table 2 shows the estimated parameters a, b, c, d, β of model 2 with (asymptotic) 95 % confidence intervals.

parameters	estimates model 2	confidence interval model 2	
a	-0.8053	-0.8111	-0.7995
b	-0.1447	-0.1493	-0.1401
c	0.0191	0.0187	0.0195
d	-0.0197	-0.0201	-0.0193

parameters	estimates model 2	confidence interval model 2	
January	0.3460	0.3412	0.3508
February	-0.1351	-0.1404	-0.1298
March	-0.1135	-0.1187	-0.1082
April	-0.1113	-0.1166	-0.1061
May	-0.1882	-0.1935	-0.1828
June	-0.0006	-0.0057	0.0046
July	0.2054	0.2003	0.2105
August	0.0124	0.0070	0.0177
September	0.3081	0.3031	0.3131
Octomber	0.0280	0.0227	0.0333
November	-0.0160	-0.0213	-0.0106

In both cases there is possible to see quite good fit of the model to the data. The value of likelihood for model 1 is $-5.7695 \cdot 10^6$ and the value of likelihood for model 2 is $-5.7693 \cdot 10^6$, so from this point of view we can say, that the second model is better. We can support this statement by argument, that we does not expect the unemployment rate decreases so much as we can see in model 1. On the other hand we can not explain why there should be change with significant jump at 18^{th} month - in June 1999 according to the second model.

There are several interpretations of the change in the intensity of becoming unemployed between September 1998 and June 1999. Decreasing unemployment can be caused by change of the government, because it stabilized political and also economical situation in the Czech Republic. Another explanation can be connected with the economic background, there should be an upper boundary for possible amount of unemployed people in situation of relatively stable economy.

References

[1] Derman C. (1956) *Some asymptotic distribution theory for Markov chains with a denumerate number of states.* Biometrika **43**, 285 – 294.

[2] Gamerman D. (1997) *Markov Chain Monte Carlo.* Chapman and Hall.

[3] Gamerman D. and West M. (1998) *An Application of Dynamic Survival Models in Unemployment Studies.* The Statistician **36**, 269 – 274.

[4] Heikkinen J. and Arjas E. (1998) *Non-parametric Bayesian Estimation of a Spatial Poisson Intensity.* Scand. J. Statist. **25**, 435 – 450.

[5] Hušková M. (1985) *Baysovské metody.* Univerzita Karlova Praha.

[6] Legát D. (2004) *Metody MCMC.* Graduation theses. Univerzita Karlova Praha.

[7] Reisnerová S., Volf P. (2006) *Analysis of unemployment as discretized spatial-temporal process.* Proceedings of the ?rd Int. Conf. S4G 2006 , ??? – ???.

[8] Volf P. (2005). *Bayes analysis of time series with covariates.* Proceedings of the 23rd Int. Conf. MME 2005 , 421 – 426.

[9] Volf P., Linka A. (1997). *On MCMC methods in Bayesian regression analysis and model selection.* IASC Summer School, Padua, Italy.

Money Distribution as Network Flow Problem

Martina Šajtarová
Department of Probability and Mathematical Statistics,
Faculty of Mathematics and Physics, Charles University, Prague, sajtarova@centrum.cz

Abstract
We deal with money distribution to local branches of a bank in the presence of random restrictions on a total amount of units of money that can be transported on individual vehicles routes. This stochastic transportation problem is formulated as a network flow problem – as an N-stage stochastic transportation problem with random arc capacities. We introduce a simple approximation, one of possible approximations to full network recourse, and use it to solve this stochastic transportation problem.

Keywords
Money distribution, Network flow problem, Simple approximation
JEL: C61

1 The N-Stage Transportation Problem with Random Arc Capacities

Some stochastic transportation problems can be formulated as networks with random arc capacities. We present a special type of multistage stochastic programming problems – the *N-stage transportation problem with random arc capacities* that can be found for example in [2] and that will be further used for formulation of the money distribution problem.

In by us considered stochastic network problems there are flows between different points in space and time. We call these points in space cities and for simplicity suppose that sets of cities are identical for all N stages. Further we assume that in all cases each stage consists only of a single time period. Hence, all random variables in a given stage are realized simultaneously and after all random variables from all previous stages.

The set of all cities is denoted by $\mathbf{R}$ and $t = 1, \ldots, P$ are the time periods where P is the planning horizon. We do not consider problems related to the truncated horizon option. Next let (i, t) denote a node in the network representing a given city $i \in \mathbf{R}$ in time t. If t_{ij} is the travel time from city $i \in \mathbf{R}$ to city $j \in \mathbf{R}$ then we obtain notation (i, t, j) for the link from a point (i, t) to another point $(j, t + t_{ij})$.

Decision variables $x_{ij}(t)$ give the flows from city $i \in \mathbf{R}$ to city $j \in \mathbf{R}$ in time period t and random variables $\xi_{ij}(t)$ denote the random arc capacity for the link (i, t, j). Further we denote as $S_i(t)$ the flow through node (i, t) which is determined by our decisions made before time period t. Finally, $R_i(t)$ represents the exogenous demands on the network (flows entering or leaving the network) at point (i, t) and $c_{ij}(t)$ is cost for the route from city $i \in \mathbf{R}$ to city $j \in \mathbf{R}$ in the t-th period. In this paper we will also use the following vectors

$$\begin{aligned}
S(t) &= (S_i(t) : i \in \mathbf{R}) \\
x(t) &= (x_{ij}(t) : i, j \in \mathbf{R}) \\
c\,(t) &= (c_{ij}(t) : i, j \in \mathbf{R}) \\
\xi(t) &= (\xi_{ij}(t) : i, j \in \mathbf{R}).
\end{aligned}$$

Thus, the N-stage transportation problem with random arc capacities can be formulated as

$$\min_{x(1), S(1)} \{c\,(1)^T x(1) + \overline{Q}(S(1))\} \tag{1}$$

subject to

$$\sum_{j \in \mathbf{R}} x_{ij}(1) = R_i(1) \qquad \forall i \in \mathbf{R} \tag{2}$$

$$\sum_{i \in \mathbf{R}} x_{ij}(1) = S_j(1) \qquad \forall j \in \mathbf{R} \tag{3}$$

$$x_{ij}(1) \geq 0 \qquad \forall i, j \in \mathbf{R} \tag{4}$$

$$x_{ij}(1) \leq u_{ij}(1) \qquad \forall i, j \in \mathbf{R} \tag{5}$$

where

$$\overline{Q}(S(t-1)) = E_{\xi(t)}[Q(S(t-1),\xi(t))] \qquad t = 2,\ldots,P \tag{6}$$
$$\overline{Q}(S(P)) = 0. \tag{7}$$

The function $\overline{Q}(S(t-1))$ is the expected recourse function defined by the following recursion

$$Q(S(t-1),\xi(t)) = \min_{x(t),S(t)} \{c\,(t)^T x(t) + \overline{Q}(S(t))\} \tag{8}$$

subject to

$$\sum_{j\in\mathbf{R}} x_{ij}(t) = S_i(t-1) + R_i(t) \qquad \forall i \in \mathbf{R} \tag{9}$$
$$\sum_{i\in\mathbf{R}} x_{ij}(t) = S_j(t) \qquad \forall j \in \mathbf{R} \tag{10}$$
$$x_{ij}(t) \geq 0 \qquad \forall i,j \in \mathbf{R} \tag{11}$$
$$x_{ij}(t) \leq \xi_{ij}(t) \qquad \forall i,j \in \mathbf{R}. \tag{12}$$

Let us explain the individual equations in this model. In the first stage we have to move the whole supply $R_i(1)$ from all cities i. This is expressed in (2). The equations (3) and (10) define the flow $S_j(t)$ through node (j,t), $t = 1,\ldots,P$, as the sum of flows from all cities coming to this node during the t-th stage. The relation (9) gives that the sum of flows coming from node (i,t) to all cities in the t-th stage has to be equal to the sum of flows through this city i in the $(t-1)$-th stage, $S_i(t-1)$, and of an exogenous demand on the network in the t-th stage $R_i(t)$. Hence the right hand side of this relation states a flow which can be transported in the t-th stage from city i. Flows from city i to city j are non-negative variables in all P stages; moreover in this model they are limited by arc capacities $u_{ij}(1)$ for $t = 1$ and by random arc capacities $\xi_{ij}(t)$ for the rest $t = 2,\ldots,P$.

2 Simple Approximation

The simple approximation (see [2]) arises in cases where no recourse action is effective once the random vector ξ is realized. We only adjust our penalties.

The base of the simple approximation is the inclusion of a set of *recourse variables* which absorb the effect of randomness and therefore eliminate the interaction between random variables. The N-stage transportation problem with random arc capacities (1) - (12) can be approximated using the simple approximation by replacing the constraint (12) with the equation

$$x_{ij}(t) + x^-_{ij}(t) - x^+_{ij}(t) = \xi_{ij}(t) \tag{13}$$

where $x^+_{ii}(t)$ and $x^-_{ii}(t)$ are the *recourse variables* and are given by

$$x^+_{ij}(t) = \max\,[x_{ij}(t) - \xi_{ij}(t), 0] \tag{14}$$
$$x^-_{ij}(t) = \max\,[\xi_{ij}(t) - x_{ij}(t), 0]. \tag{15}$$

We assume that $x_{ij}(t)$ must be chosen before the realization of random variable $\xi_{ij}(t)$ is observed while $x^+_{ij}(t)$ and $x^-_{ij}(t)$ must be computed for a given realization of $\xi_{ij}(t)$ using relations (14) and (15). $x^-(t)$ can be interpreted as a lost demand and is associated with a penalty q^- while $x^+(t)$ represents a movement of flow which does not produce any income and is connected with a positive cost q^+.

Let $\overline{\phi}_{ij}(x(t))$ be the expected recourse function for stage $t = 2,\ldots,P$ for the flows between cities $i \in \mathbf{R}$ and $j \in \mathbf{R}$, defined by

$$\overline{\phi}_{ij}(x(t)) = E_{\xi_{ij}(t)}[q^-_{ij}\,x^-_{ij}(t) + q^+_{ij}\,x^+_{ij}(t)], \tag{16}$$

where $x^+(t)$ and $x^-(t)$ are given by (14) and (15).

If we use this relation we can rewrite the N-stage transportation problem with random arc capacities (1) - (12) as follows

$$\min_{x(1),S(1)} \{c\,(1)^T x(1) + \overline{\Phi}(S(1))\} \tag{17}$$

subject to (2) - (5) with

$$\overline{\Phi}(S(t-1)) = E_{\xi(t)}[\Phi(S(t-1),\xi(t))] \qquad t = 2,\ldots,P \tag{18}$$
$$\overline{\Phi}(S(P)) = 0. \tag{19}$$

The function $\overline{\Phi}(S(t-1))$ is the expected recourse function for the t-th stage and is for the simple approximation defined by the following recursion:

$$\Phi(S(t-1),\xi(t)) = \min_{x(t),S(t)} \left\{ \sum_{i\in\mathbf{R}}\sum_{j\in\mathbf{R}} \left\{ c_{ij}(t)\, x_{ij}(t) + \overline{\phi}_{ij}(x(t)) \right\} + \overline{\Phi}(S(t)) \right\} \tag{20}$$

subject to

$$\sum_{j\in\mathbf{R}} x_{ij}(t) = S_i(t-1) + R_i(t) \qquad \forall i \in \mathbf{R} \tag{21}$$

$$\sum_{i\in\mathbf{R}} x_{ij}(t) = S_j(t) \qquad \forall j \in \mathbf{R} \tag{22}$$

$$x_{ij}(t) \geq 0 \qquad \forall i,j \in \mathbf{R}. \tag{23}$$

We can notice that (20) - (23) are no longer functions of $\xi(t)$ because it has already been included into the functions $\overline{\phi}(x_{ij}(t))$. This implies that

$$\overline{\Phi}(S(t-1)) = \Phi(S(t-1),\xi(t)).$$

Thus the whole problem (17) - (23) can be rewritten as a single optimization problem

$$\min_{x(1),\,\ldots,\,x(P)} \left\{ c\,(1)^T x(1) + \sum_{t=2}^{P} \left(c\,(t)^T x(t) + \sum_{i\in\mathbf{R}}\sum_{j\in\mathbf{R}} \overline{\phi}_{ij}(x(t)) \right) \right\} \tag{24}$$

subject to

$$\sum_{j\in\mathbf{R}} x_{ij}(t) = S_i(t-1) + R_i(t) \qquad \forall i \in \mathbf{R} \tag{25}$$

$$\sum_{i\in\mathbf{R}} x_{ij}(t) = S_j(t) \qquad \forall j \in \mathbf{R} \tag{26}$$

$$x_{ij}(t) \geq 0 \qquad \forall i,j \in \mathbf{R} \tag{27}$$

$$x_{ij}(1) \leq u_{ij}(1) \qquad \forall i,j \in \mathbf{R}. \tag{28}$$

where $t = 1,\ldots,P$ and $S_i(0) = 0$ for all $i \in \mathbf{R}$.

Using this approach we can replace a complex nonseparable recourse function with a separable one. We get a classical convex, nonlinear network flow problem which can be solved using standard techniques (see [4]).

3 Money Distribution Problem

We will consider the following stochastic transportation problem – there is a security agency that delivers every day money from bigger branches of some bank to local branches of this bank. Our goal is to allocate vehicles on individual routes in the presence of random arc capacities to supply with money all local branches of the bank and to minimize our costs. We will define this problem as the N-stage transportation problem with random arc capacities and will use the simple approximation to solve it.

Our security agency has three depots ($\mathbf{I} = \{a,b,c\}$) and has to transport money to twenty local branches ($\mathbf{J} = \{dp1,\ldots,dp20\}$) of the bank. Thus the set of all cities $\mathbf{R}$ has been divided into two disjunctive sets $\mathbf{I}$ and $\mathbf{J}$. We assume that the distances between depots and bigger branches are insignificant in comparison with other distances and hence we will consider transportation costs only for the routes between depots and local branches (demand points) of the bank.

In depots there are three types of vehicles ($\mathbf{K} = \{veh1, veh2, veh3\}$) at our disposal of different capacities. We denote the capacity of the vehicle of the type $k \in \mathbf{K}$ on the route from the depot $i \in \mathbf{I}$ to the demand point $j \in \mathbf{J}$ or back as B_{ijk}. For simplicity we will suppose that a given vehicle has the same capacity on all routes.

We denote number of vehicles of the type $k \in \mathbf{K}$ in the depot $i \in \mathbf{I}$ as $A_{ik}(1)$. Thus the vector

$$A(1) = (A_{ik}(1) : i \in \mathbf{I},\ k \in \mathbf{K})$$

corresponds to an initial vector of supplies that has been in the past denoted as $R(1)$. However we decided to use different notation since it is evident that in this case no all vehicles have to leave depots.

For simplicity we suppose that N is only so high that we exactly know money demands of local branches - $m_j(t)$ denotes the money demands for the local branch $j \in \mathbf{J}$ in the t-th stage.

We can imagine this transportation problem very easily – individual routes from depots to local branches have restricted capacities because of security reason. Thus there exist restrictions on total amount of units of money that can be transported on the given arc in the given time period and these restrictions are random with discrete distributions. With respect to prior mentioned notation we denote these random arc capacities for the depot $i \in \mathbf{I}$ and the demand point $j \in \mathbf{J}$ in the t-th stage as $\xi_{ij}(t)$. For simplicity we will assume that the vectors $\xi(t)$ have the same distribution for all time periods $t = 2, \dots, P$ where P is the planning horizon. Because of this we will use only the notation $\xi = (\xi_{ij} : i \in \mathbf{I},\, j \in \mathbf{J})$ for random vectors without stating the stage in the rest of this section. Further we will assume that $\xi_{ij} = \xi_{ji}$. This implies that the same random restrictions will hold for routes of empty vehicles, too. Thus individual arcs from demand points to depots are limited with the same corresponding random arc capacities and in this case the random variable ξ_{ji} restricts the total capacity of all vehicles allocated to move from the demand point $j \in \mathbf{J}$ to the depot $i \in \mathbf{I}$ in the given stage.

We denote the set of possible scenarios for the future values of the random variables (ξ_{ij} and ξ_{ji}) as $\mathbf{H}$. Thus $\xi_{ij,h}$ denotes individual realizations of the random variable ξ_{ij} and states an amount of units of money that can be maximally transported from the depot $i \in \mathbf{I}$ to the local branch $j \in \mathbf{J}$ by the scenario $h \in \mathbf{H}$, possibly states the maximal total capacity of all vehicles allocated to route from the local branch $j \in \mathbf{J}$ to the depot $i \in \mathbf{I}$ by the scenario $h \in \mathbf{H}$ since again $\xi_{ij,h} = \xi_{ji,h}$ for all $h \in \mathbf{H}$. We denote as $\lambda_{ij,h}$ the probability of a realization of $\xi_{ij,h}$ and again for all $h \in \mathbf{H}$ we have $\lambda_{ij,h} = \lambda_{ji,h}$.

Further we will assume that the operating costs does not depend on the direction of the route of the given vehicle; hence $c_{ijk}(t) = c_{jik}(t)$. The decision variables $x_{ijk}(t)$ (for odd stages) represent the number of vehicles of the type $k \in \mathbf{K}$ that are allocated to route from the depot $i \in \mathbf{I}$ to the local branch $j \in \mathbf{J}$ and the decision variables $x_{jik}(t)$ (for even stages) represent the number of vehicles of the type $k \in \mathbf{K}$ that are allocated to move from the demand point $j \in \mathbf{J}$ to the depot $i \in \mathbf{I}$. For simplicity we suppose that in odd stages vehicles move on individual arcs fully loaded and that they leave their whole loads in the demand points they have been allocated to route to. We do not consider some additional costs connected with excessive supply. Further we will use the following vectors

$$x(t) = \begin{cases} (x_{ijk}(t) : i \in \mathbf{I},\, j \in \mathbf{J},\, k \in \mathbf{K}) & \text{for odd stages } t \\ (x_{jik}(t) : j \in \mathbf{J},\, i \in \mathbf{I},\, k \in \mathbf{K}) & \text{for even stages } t, \end{cases}$$

$$c(t) = \begin{cases} (c_{ijk}(t) : i \in \mathbf{I},\, j \in \mathbf{J},\, k \in \mathbf{K}) & \text{for odd stages } t \\ (c_{jik}(t) : j \in \mathbf{J},\, i \in \mathbf{I},\, k \in \mathbf{K}) & \text{for even stages } t, \end{cases}$$

$$S(t) = \begin{cases} (S_{ik}(t) : i \in \mathbf{I},\, k \in \mathbf{K}) & \text{for odd stages } t \\ (S_{jk}(t) : j \in \mathbf{J},\, k \in \mathbf{K}) & \text{for even stages } t. \end{cases}$$

where $S_{ik}(t)$ and $S_{jk}(t)$ state the flow of vehicles of the given type $k \in \mathbf{K}$ in the stage t through the depot $i \in \mathbf{I}$ and through the demand point $j \in \mathbf{J}$, respectively. These flows can be routed from nodes in the following stage.

We need to set up new variable $L_{ik}(t)$ for the odd stages t. This variable gives a number of vehicles of the given type $k \in \mathbf{K}$ that have not been allocated to route in the t-th stage from the depot $i \in \mathbf{I}$ and that can be used in the $(t+2)$-stage. In even stages all vehicles must return to some depot – these vehicles determine the flow through nodes and their number must be increased for the number of vehicles that have stayed in depots in the past in order to get the number of vehicles that can be allocated to routes in the following stage.

Now we can proceed to introduce the simple approximation for our stochastic transportation problem. First we need to define

$$y_{ij}(t) = \sum_{k \in \mathbf{K}} B_{ijk}\, x_{ijk}(t) \qquad \forall i \in \mathbf{I},\, \forall j \in \mathbf{J},\, t = 2k+1,\, k \in N$$

$$y_{ji}(t) = \sum_{k \in \mathbf{K}} B_{jik}\, x_{jik}(t) \qquad \forall j \in \mathbf{J},\, \forall i \in \mathbf{I},\, t = 2k,\, k \in N.$$

The variable $y_{ij}(t)$ states the total amount of units of money that are transported on the arc from the depot $i \in \mathbf{I}$ to the demand point $j \in \mathbf{J}$ in the t-th stage (we assume that all vehicles are moving fully loaded) and $y_{ji}(t)$ gives the total capacity of the vehicles allocated to move on the arc from the local branch $j \in \mathbf{J}$ to the depot $i \in \mathbf{I}$ in the t-th stage.

In this stochastic transportation problem we will have the following *recourse variables*:

$$y^+_{ij}(t) = \max\,[\, y_{ij}(t) - \xi_{ij}, 0\,] \qquad \forall i \in \mathbf{I},\, \forall j \in \mathbf{J},\, t = 2k+1,\, k \in N$$

$$y^+_{ji}(t) = \max\,[\, y_{ji}(t) - \xi_{ji}, 0\,] \qquad \forall j \in \mathbf{J},\, \forall i \in \mathbf{I},\, t = 2k,\, k \in N.$$

If we send in the t-th stage ($t \neq 1$) on the given arc vehicles with total capacity higher than the random capacity of this arc there arise an additional costs $q_{ij}^+ = q_{ji}^+$ (we assume that the additional costs do not depend on the stage and on the direction of the route) since we have to pay some extra security guards to guard this higher cash moving on the given arc. We do not suppose additional costs connected with total capacity of vehicles lower than is the random capacity for arc they are moving on.

Further we denote the individual possible values of the variables $y_{ij}^+(t)$ and $y_{ji}^+(t)$ for individual realizations of the random arc capacities $\xi_{ij,h}$ and $\xi_{ji,h}$ as $y_{ij,h}^+(t)$ and $y_{ji,h}^+(t)$ where

$$y_{ij,h}^+(t) = \max\left[\, y_{ij}(t) - \xi_{ij,h}, 0\,\right] \qquad \forall i \in \mathbf{I},\ \forall j \in \mathbf{J},\ \forall h \in \mathbf{H},\ t = 2k+1,\ k \in N$$
$$y_{ji,h}^+(t) = \max\left[\, y_{ji}(t) - \xi_{ji,h}, 0\,\right] \qquad \forall j \in \mathbf{J},\ \forall i \in \mathbf{I},\ \forall h \in \mathbf{H},\ t = 2k,\ k \in N.$$

These variables state how much we exceeded the random arc capacities in the t-th stage by the scenario $h \in \mathbf{H}$.

The expected recourse functions for the arc from the depot $i \in \mathbf{I}$ to the local branch $j \in \mathbf{J}$ and for the link from the demand point $j \in \mathbf{J}$ to the depot $i \in \mathbf{I}$ for the t-th stage, where $t \neq 1$, are denoted according to prior mentioned notation as $\overline{\phi}_{ij}(x(t))$ and as $\overline{\phi}_{ji}(x(t))$ and thanks to the discrete distribution of the random arc capacities ξ_{ij} and ξ_{ji} they can be written as

$$\begin{aligned}\overline{\phi}_{ij}(x(t)) &= E_{\xi_{ij}}[q_{ij}^+\, y_{ij}^+(t)] = q_{ij}^+ \sum_{h\in\mathbf{H}} \lambda_{ij,h}\, y_{ij,h}^+(t) = \\ &= q_{ij}^+ \sum_{h\in\mathbf{H}} \lambda_{ij,h} \max\left[\sum_{k\in\mathbf{K}} B_{ijk}\, x_{ijk}(t) - \xi_{ij,h}, 0\right] \qquad \forall i \in \mathbf{I},\ \forall j \in \mathbf{J},\ t = 2k+1,\ k \in N\end{aligned}$$

$$\begin{aligned}\overline{\phi}_{ji}(x(t)) &= E_{\xi_{ji}}[q_{ji}^+\, y_{ji}^+(t)] = q_{ji}^+ \sum_{h\in\mathbf{H}} \lambda_{ji,h}\, y_{ji,h}^+(t) = \\ &= q_{ji}^+ \sum_{h\in\mathbf{H}} \lambda_{ji,h} \max\left[\sum_{k\in\mathbf{K}} B_{jik}\, x_{jik}(t) - \xi_{ji,h}, 0\right] \qquad \forall j \in \mathbf{J},\ \forall i \in \mathbf{I},\ t = 2k,\ k \in N.\end{aligned}$$

For our transportation problem we will consider the planning horizon P equal to 4 (hence we want to plan two transports of money). With respect to the previous definition of the N-stage transportation problem with random arc capacities and to the introduced notation our stochastic transportation problem can be written as

$$\min_{x(1),\,\ldots,\,x(4)} \left\{ \sum_{t=1}^{4} c\,(t)^T x(t) + \sum_{i\in\mathbf{I}} \sum_{j\in\mathbf{J}} \left\{ \overline{\phi}_{ji}(x(2)) + \overline{\phi}_{ij}(x(3)) + \overline{\phi}_{ji}(x(4)) \right\} \right\} \tag{29}$$

subject to:

$$\sum_{j\in\mathbf{J}} x_{ijk}(1) \le A_{ik}(1) \qquad \forall i \in \mathbf{I},\ \forall k \in \mathbf{K} \tag{30}$$
$$\sum_{k\in\mathbf{K}} B_{ijk}\, x_{ijk}(1) \le u_{ij}(1) \qquad \forall i \in \mathbf{I},\ \forall j \in \mathbf{J} \tag{31}$$
$$\sum_{j\in\mathbf{J}} x_{ijk}(3) \le S_{ik}(2) + L_{ik}(1) \qquad \forall i \in \mathbf{I},\ \forall k \in \mathbf{K} \tag{32}$$
$$\sum_{i\in\mathbf{I}} x_{jik}(t+1) = S_{jk}(t) \qquad \forall j \in \mathbf{J},\ \forall k \in \mathbf{K},\ t = 1,3 \tag{33}$$
$$\sum_{i\in\mathbf{I}} x_{ijk}(t) = S_{jk}(t) \qquad \forall j \in \mathbf{J},\ \forall k \in \mathbf{K},\ t = 1,3 \tag{34}$$
$$\sum_{j\in\mathbf{J}} x_{jik}(2) = S_{ik}(2) \qquad \forall i \in \mathbf{I},\ \forall k \in \mathbf{K} \tag{35}$$
$$\sum_{i\in\mathbf{I}} \sum_{k\in\mathbf{K}} B_{ijk}\, x_{ijk}(t) \ge m_j(t) \qquad \forall j \in \mathbf{J},\ t = 1,3 \tag{36}$$
$$x_{ijk}(t) \ge 0 \qquad \forall i \in \mathbf{I},\ \forall j \in \mathbf{J},\ \forall k \in \mathbf{K},\ t = 1,3 \tag{37}$$
$$x_{jik}(t) \ge 0 \qquad \forall j \in \mathbf{J},\ \forall i \in \mathbf{I},\ \forall k \in \mathbf{K},\ t = 2,4. \tag{38}$$

The relation (30) gives a restriction on a number of vehicles of given type we can allocate to route from given depot in the first stage and the (31) states the limitation on the total amount of units of money that can be transported on the arc from the depot $i \in \mathbf{I}$ to the local branch $j \in \mathbf{J}$ in the first stage. The inequality (32) says that the number of vehicles of given type which can be allocated to route from given depot to all branches in the third stage is restricted with the sum of the vehicles which arrived to this depot during the second stage and of the vehicles which stayed here from the first stage. The equation (33) gives the number of vehicles of given type that have to be allocated to move from individual branches in the second respectively in the fourth stage and the equation (34) defines the total number of vehicles of the given type that arrived to the demand point $j \in \mathbf{J}$ during the first respectively the third stage from all depots and that can set off from this demand points at the beginning of the second respectively of the fourth stage. The relation (35) defines the total number of vehicles of the given type that arrived to the depot $i \in \mathbf{I}$ during the second stage from all demand points and the inequality (36) says that we have to fulfil or exceed money demands $m_j(1)$ and $m_j(3)$, respectively, of individual local branches. Of course, numbers of vehicles we allocate to move on individual routes between depots and demand points in all stages has to be non-negative variables.

It is obvious that

$$L_{ik}(1) = A_{ik}(1) - \sum_{j \in \mathbf{J}} x_{ijk}(1) \qquad \forall i \in \mathbf{I},\, \forall k \in \mathbf{K}.$$

The conditions (32) - (35) can be replaced with

$$\sum_{j \in \mathbf{J}} x_{ijk}(3) + \sum_{j \in \mathbf{J}} x_{ijk}(1) \leq \sum_{j \in \mathbf{J}} x_{jik}(2) + A_{ik}(1) \qquad \forall i \in \mathbf{I},\, \forall k \in \mathbf{K} \tag{39}$$

$$\sum_{i \in \mathbf{I}} x_{jik}(t+1) = \sum_{i \in \mathbf{I}} x_{ijk}(t) \qquad \forall j \in \mathbf{J},\, \forall k \in \mathbf{K},\, t = 1, 3 \tag{40}$$

in order to elimination of variables $S_{ik}(t)$, $S_{jk}(t)$ and $L_{ik}(1)$.

We used model system GAMS and MINOS 5.51 Solver to solving this stochastic transportation problem. The objective function value of the found solution is 6320.082 units - the individual vehicles routes in all stages form the cost to the extent of 4917.867 units and the expected value of penalty for overrunning the random arc capacities is 1402.215 units. See below on the GAMS output.

```
                       LOWER     LEVEL     UPPER    MARGINAL

---- VAR oc              .      4917.867    +INF       .
---- VAR ac              .      1402.215    +INF       .
---- VAR phi           -INF     6320.082    +INF       .

  oc  operating costs
  ac  additional costs
  phi  total expected costs
```

Acknowledgments

The participation in the conference MME 2006 was enabled due to grant-in-aid by ČSOB, a.s.

References

[1] Dupačová J., Hurt J., Štěpán J. (2002): Stochastic Modeling in Economics and Finance. Kluwer Academic Publishers, Preprint.

[2] Frantzeskakis L. F., Powell W. B. (1994): Restricted Recourse Strategies for Dynamic Networks with Random Arc Capacities. *Transportation Sci.* 28, 3-23.

[3] Kall P., Wallace S. W. (1994): Stochastic Programming. John Wiley & Sons, New York.

[4] Kennington J. L., Helgason R. V. (1980): Algorithms for Network Programming. John Wiley & Sons, New York.

Mathematical Modelling of Economic Cycles and Optimal Investment Strategy Working-Out

Mindia Salukvadze
A.Eliashvili Institute of Control Systems,
34, K. Gamsakhurdia Ave.,0160, Tbilisi, Georgia
msaluk@gw.acnet.ge

Nodar Jibladze
A.Eliashvili Institute of Control Systems,
34, K. Gamsakhurdia Ave.,0160, Tbilisi, Georgia
jibladze@gw.acnet.ge

Tamaz Obgadze
Georgian Technical University,
77, M. Kostava Str., 0175, Tbilisi, Georgia
dep71@gtu.ge

Nato Tushishvili
Georgian Technical University,
77, M. Kostava Str., 0175, Tbilisi, Georgia
dep71@gtu.ge

Abstract

On the basis of the Keynes theory and the Samuelson-Hiks acceleration principle is constructed a generalized, ordinary mathematical model of economic dynamics, which in special cases in the corresponding choice of the acceleration function and the consumption function gives the Duffing, Matye, Samuelson-Hiks models et al. The problem of Optimal Investment Strategy Working-Out is posed. The mathematical model for studying a foodstuff aggregate demand is made. The consumption function is formed by the method of multiple polynomial regression, and the investment policy is carried out on the basis of the Samuelson-Hiks principle. The problem of the investments optimal control for the purpose of avoiding inadmissible amplitudes of demand fluctuations is solved on the basis of the variational setting.

Keywords

model, dynamics, Keynes, elasticity, optimal, investment.
JEL: C51-Model Construction and Estimation.

The economic dynamics in the equilibrium economics is described by the equilibrium equation [1-8]:

$$Y(t) = C(t) + I(t), \tag{1}$$

where $C(t)$ - the consumption function; $I(t)$ - the investment policy.

$Y(t)$ represents the aggregate demand, which in the equilibrium economics coincides with the supply $X(t)$. Therefore, for the production volume we have the equation

$$X(t) = C(t) + I(t). \tag{2}$$

The consumption function is written by us in the form of [6-8]:

$$C(t) = \int_0^t F[X(\tau),\ \tau]\, d\tau, \tag{3}$$

where the dependence $F[X(\tau),\ \tau]$ is determined on the basis of the data regression analysis.

The investment policy, based on the Samuelson-Hiks acceleration principle, is written in the form of [6-7]

$$I(t) = \gamma(t)\dot{X}(t), \tag{4}$$

where $\gamma(t)$ - the acceleration function.

Substituting the correlations (3) and (4) in the equation (2) we get the integro-differential equation of economic dynamics

$$X(t) = \int_0^t F[X(\tau),\ \tau]\,d\tau + \gamma(t)\dot{X}(t). \tag{5}$$

In order to get rid of the right part of the equation (5), we differentiate it by the time parameter t, and then we get the ordinary mathematical model of economic dynamics in the form of

$$\gamma(t)\ddot{X}(t) + \left[\dot{\gamma}(t) - 1\right]\dot{X}(t) + F\left[X(t), t\right] = 0. \tag{6}$$

The constructed mathematical model is based on the Samuelson-Hiks acceleration concept. The investment policy is determined by the acceleration function $\gamma(t)$, which is the control parameter. The aim of the control is the stable development of the production $X(t)$ without resonance oscillations destroying the system.

For studying the proposed mathematical model let's consider some of its particular cases with different consumption functions and the Samuelson-Hiks acceleration function.

(i) Let's consider the case when

$$\gamma(t) = 0, \quad t > 0, \tag{7}$$

$$F\left[X(t), t\right] = t\left(\omega^2 + \varepsilon\cos 2t\right)X(t) - 0.9t. \tag{8}$$

Then from the equation (6) we get the Matye equation

$$\ddot{X}(t) + \left(\omega^2 + \varepsilon\cos 2t\right)X(t) = 0.9, \tag{9}$$

adding the initial conditions

$$X(0) = 1,\ \dot{X}(0) = 1. \tag{10}$$

With $\omega = 0.5$, and $\varepsilon = 0.2$, on the basis of the program system ″MATHCAD 2001 Professional″, we get the solution for the production volume $S^{\langle 1\rangle} = X(t)$, $S^{\langle 0\rangle} = t$ (Fig. 1), and the corresponding picture on the phase plane $\left(X(t), \dot{X}(t)\right)$, where $S^{\langle 1\rangle} = X(t)$, and $S^{\langle 2\rangle} = \dot{X}(t)$ (Fig. 2);

(ii) Let's consider the case when

$$\gamma(t) = \frac{p\,e^{pt} - 1}{p}, \quad p = const, \tag{11}$$

$$F\left[X(t), t\right] = \beta(t)\left[X(t)^3 - X(t) - A\cos\omega t - 0.3\right], \tag{12}$$

where

$$\omega = const, \quad A = const. \tag{13}$$

Then we get the Duffing equation from the equation (6)

$$\ddot{X}(t) + p\dot{X}(t) + X(t)^3 - X(t) - A\cos\omega t - 0.3 = 0, \tag{14}$$

adding the initial conditions

$$X(0) = 1,\ \dot{X}(0) = 1. \tag{15}$$

With $p = 0.2$, , $A = 0.25$ and $\omega = 1$, on the basis of the program system ″MATHCAD 2001 Professional″, we get the solution for the production volume $S^{\langle 1\rangle} = X(t)$, $S^{\langle 0\rangle} = t$ (Fig. 3), and the corresponding picture on the phase plane $\left(X(t), \dot{X}(t)\right)$, where $S^{\langle 1\rangle} = X(t)$, and $S^{\langle 2\rangle} = \dot{X}(t)$ (Fig. 4);

(iii) Considering the case, when

$$\gamma(t) = const\,, \tag{16}$$

$$\dot{X}(t) \approx \frac{X(t-h) - X(t-2h)}{h}, \quad h = 1\,, \tag{17}$$

$$F\left[X(t), t\right] = \alpha X(t-h), \qquad \alpha X(-h) = A\,, \tag{18}$$

where

$$A = (living\ wage) \times (a\ number\ of\ inhabitants)\,, \tag{19}$$

then from the equation (5) we get the recurrent Samuelson-Hiks model:

$$X(t) = (\alpha + \beta)\, X(t-1) - \beta X(t-2) + A\,.$$

Thus we've got the generalized ordinary mathematical model of economic dynamics, which in special cases can set the Samuelson-Hiks model, the Matye equation, the Duffing equation etc. And of the utmost importance is the fact that it gives an opportunity, in case of finding the corresponding consumption function and the acceleration function, to work out the optimal investment policy.

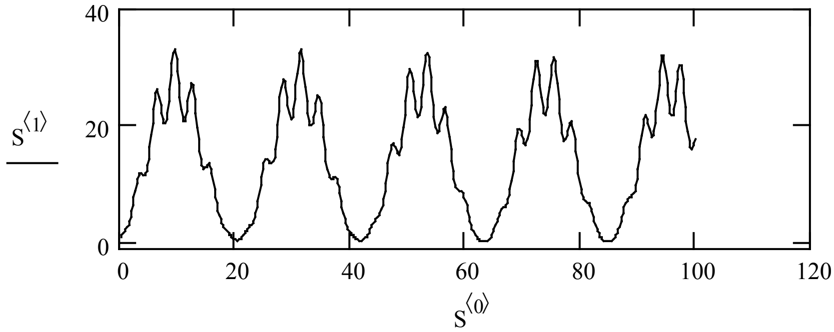


Fig.1

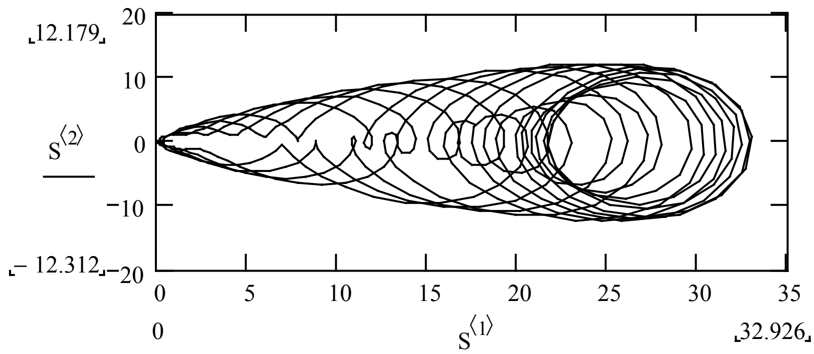


Fig.2.

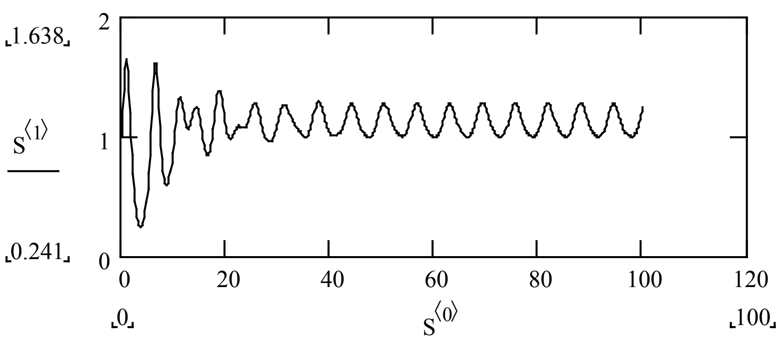


Fig.3

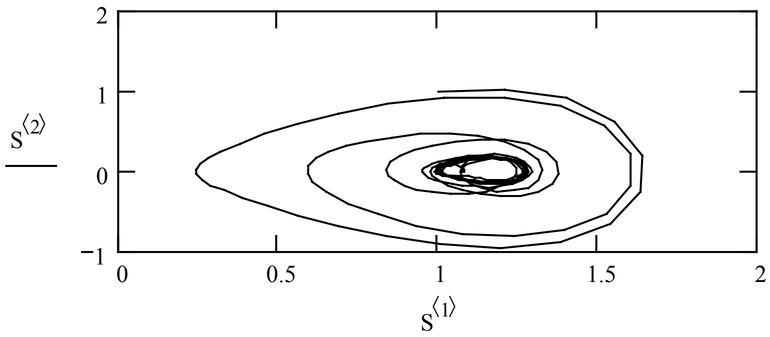


Fig.4

Let's consider the problem of the investments optimal control in order to avoid the resonance oscillations destroying the system.

The consumption function for the foodstuff composite demand $C(t)$ is determined on the basis of the multiple polynomial regression.

$$C(t) = m + \alpha \cdot X(t) + \beta \cdot P(t), \tag{20}$$

where $m = (a\ number\ of\ inhabitants) \times (living\ wage)$;

$\alpha > 0$ - the elasticity of demand by the family mean income;

$\beta > 0$ - the elasticity of demand by the foodstuff mean prices;

$P(t)$ - the function of the foodstuff mean price.

From the equation (20) we get

$$F\left[X(t), t\right] = \dot{C}(t) = \alpha \cdot \dot{X}(t) + \beta \cdot \dot{P}(t). \tag{21}$$

The foodstuff mean price $P(t)$ depends on the volume of supply $X(t)$, and on the rate of its change $\dot{X}(t)$, at that, the more is $X(t)$ or $\dot{X}(t)$, the less is the price, and vice versa, the less is $X(t)$, the more is the price, since we have a deficit. The equality to zero $P(t)$ corresponds to the market glut, i.e. we have the connection

$$P(t) = imp - \int_0^t \left(\zeta \cdot X(\tau) + \eta \cdot \dot{X}(\tau)\right) d\tau, \tag{22}$$

where imp - is the mean price of the imported foodstuff output when supplies are not laid in within the State, $\zeta \in [0,1)$, and $\eta \in [0,1)$.

Substituting the correlations (21) and (22) in the equation (6) we get the Prangishvili - Obgadze differential equation of economic dynamics in the form of:

$$\gamma(t) \cdot \ddot{X}(t) + \left(\dot{\gamma}(t) - \beta \cdot \eta + \alpha - 1\right) \cdot \dot{X}(t) - \beta \cdot \zeta \cdot X(t) = 0. \tag{23}$$

In other words, we get the equation of dynamics of the foodstuff output aggregate in the form of

$$\ddot{X}(t) + \frac{\dot{\gamma}(t) - \beta \cdot \eta + \alpha - 1}{\gamma(t)} \cdot \dot{X}(t) - \frac{\beta \cdot \zeta}{\gamma(t)} \cdot X(t) = 0, \tag{24}$$

where $\gamma(t)$ is the acceleration control parameter $\left(\gamma(t) \neq 0\right)$.

With $\gamma(t)=0$, from (23) we get that $X(t)\to 0$ approximates exponentially, i.e. if there is lack of investments, there wouldn't be a supply for foodstuff as well. Therefore, we consider only such investment policy, where $\gamma'(t)\neq 0$, i.e. we can write it down in the form of:

$$\gamma(t)=\gamma_0^{\,2}+\sum_i \delta_i \cdot t^i \tag{25}$$

where $\gamma_0 = const \neq 0$.

We should choose the values γ_0, δ_i from the condition of the investments optimality in the presented work [7]:

$$\sup\sqrt{\int_0^T X^2(t)\,dt-\frac{1}{T}\left(\int_0^T X(t)\,dt\right)^2}\leq \varepsilon, \tag{26}$$

where $\varepsilon>0$ is the admissible amplitude of the spread in values of the foodstuff output aggregate $X(t)$ from its mean value.

To solve the problem (24), (26), with the corresponding Cauchy conditions we write a program in Mathcad language, and correspondingly adapt our problem for using the block *Given* and the operator *Minimize*

We present the unknown function of the foodstuff composite demand $X(t)$ in the form of

$$X(t)=\sum_i a_i \cdot t^i \tag{27}$$

on the interval $t\in[0,T]$.

The equation (24) is satisfied on the time interval $[0,T]$, which is equivalent to the minimization of the left part of the equation by the standard $L_2[0,T]$, i.e. instead of the equation (24) we have the variational problem

$$I\left(X(t),\gamma(t)\right)=\sqrt{\int_0^T\left[\ddot{X}(t)+\frac{\dot{\gamma}(t)-\beta\cdot\eta+\alpha-1}{\gamma(t)}\cdot\dot{X}(t)-\frac{\beta\cdot\zeta}{\gamma(t)}\cdot X(t)\right]^2 dt}\to\min \tag{28}$$

provided (25), (27) and the restriction

$$\int_0^T X^2(t)\,dt-\frac{1}{T}\left(\int_0^T X(t)\,dt\right)^2\leq \varepsilon\cdot\frac{1}{T}\int_0^T X(t)\,dt. \tag{29}$$

As a result of the calculations, on the basis of MATHCAD 2001, we get the dependences $X(t)$, and $\gamma(t)$ (fig.5).

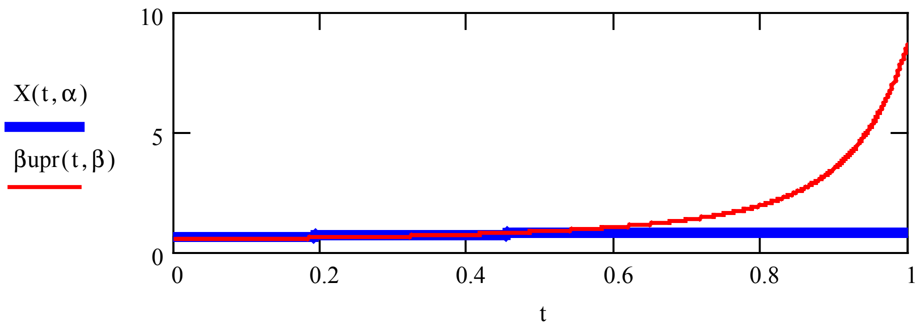


fig.5. The time dependence of the composite demand on the foodstuff and the acceleration ratio.

References

[1] T. Obgadze. *Higher mathematics for economists.* -Moscow: Institute of Liberal Education, -2002.

[2] M. Intriligator. *Mathematical optimization methods and economic theory.* Translation from English. -Moscow. -1975.

[3] U. Sharp, J.A.Gordon, D.Beili. *Investments.* Translation from English. -Moscow. -1992.

[4] E.Peters. *Chaos and order in the market of capitals. New analytical view on the cycles, prices and on the variability of the market.* Translation from English-Moscow. -2002.

[5] D. Errosmith, K. Place. *Ordinary differential equations. Qualitative theory with applications*. Translation from English. - Moscow. -1986.

[6] A.Prangishvili, L.Obgadze. *Generalised mathematical model of economic dynamics* //Georgian Scientific e-magazine: http://gesj.internet-academy.org.ge , -№3(7), 2005, pp. 55-58.

[7] A.Prangishvili, L.Obgadze. *Mathematical modeling of economic cycles and investments optimal control* // Georgian Scientific e-magazine: http://gesj.internet-academy.org.ge , -№3(7), 2005, pp. 59-61.

[8] N.Jibladze, T.Obgadze, L.Obgadze. *Modelling of economic cycles*.//Proceedings of the A.I.Eliashvili Institute of Control Systems of the Georgian Academy of Sciences, №9,Tbilisi, 2005, pp.45-49.

Approximations in Stochastic Growth Models

Karel Sladký

Department of Econometrics
Institute of Information Theory and Automation, AS ČR
Pod Vodárenskou věží 4, 182 08 Praha 8
sladky@utia.cas.cz

Abstract

In this note, we consider finite state approximations of the stochastic Ramsey type model in discrete-time version. Recalling standard procedure of stochastic dynamic programming we present explicit formulas for finding maximum global utility of the consumers (i.e. sum of total discounted instantaneous utilities) in the approximated model.

Keywords

Economic dynamics, stochastic Ramsey type model, Markov decision chains, discounted optimality, asymptotic behavior

JEL Classification: C61, E21, E22

1 Basic Models of Economic Growth

Formulation. There is no doubt that the seminal paper of F. Ramsey [10] on mathematical theory of saving stimulated most of the further research in macroeconomic dynamics. The heart of this paper is an economy producing output from labour and capital and the task is to decide how to divide production between consumption and capital accumulation to maximize the global utility of the consumers. Since in [10] the problem was considered in continuous-time setting, Ramsey suggested some variational methods for finding an optimal policy how to divide the production between consumption and capital accumulation. Ramsey's results were revisited and significantly extended only after almost thirty years by Cass, Koopmans and Samuelson (cf. [4], [7], [12]) and at present the Ramsey model can be considered, along with the Solow model and overlapping generations model (see e.g. [13]), as one of the three most significant tools for the dynamic general equilibrium model in modern macroeconomics.

Considering the Ramsey problem in discrete-time setting, the respective mathematical model can be formulated as follows:

We consider at discrete-time points $t = 0, 1, \ldots$ an economy in which at each time t there are L_t identical consumers with consumption c_t per consumer. We assume that the number of consumers grow at rate n, i.e. $L_t = L_0(1+n)^t$ for $t = 0, 1, \ldots$. The economy produces at times $t = 0, 1, \ldots$ gross output $\tilde{Y}_t$ using only two inputs: capital K_t and labour L_t. A production function $F(K_t, L_t)$ relates input to output, i.e.

$$\tilde{Y}_t = F(K_t, L_t) \quad \text{with} \quad K_0 > 0, \; L_0 > 0 \quad \text{given.} \tag{1}$$

In each period output must be split between consumption $c_t L_t$ and gross investment I_t, i.e.

$$c_t L_t + I_t \le \tilde{Y}_t = F(K_t, L_t), \tag{2}$$

investment I_t is used in whole (along with the depreciated capital K_t) for the capital K_{t+1}.

In addition, capital is assumed to depreciate at a constant rate $\delta \in (0,1)$, so capital related to gross investment at time point $t+1$ is equal to

$$K_{t+1} = (1-\delta)K_t + I_t. \tag{3}$$

Preferences over consumption of a single consumer are taken to be of the form

$$\sum_{t=0}^{T} \tilde{\beta}^t u(c_t) \quad \text{for a finite or infinite time horizon } T, \tag{4}$$

where $u(\cdot)$ is instantaneous utility function and $\tilde{\beta} < 1$ is a given discount factor.

The problem is to find the rule how to split production between consumption and capital accumulation that maximizes global utility of the consumers for a finite or infinite time horizon T, i.e. to maximize the function

$$L_0 \sum_{t=0}^{T} \tilde{\beta}^t (1+n)^t u(c_t) = L_0 \sum_{t=0}^{T} \beta^t u(c_t) \quad \text{where} \quad \beta = (1+n)\tilde{\beta}. \tag{5}$$

Assumptions and Notation. Denoting by $k_t := K_t/L_t$ the capital per consumer at time t, and similarly by $\tilde{y}_t := \tilde{Y}_t/L_t = F(k_t, 1)$ the output per consumer at time t (note that $F(\cdot,\cdot)$ is assumed to be homogeneous of degree one, i.e. $F(\theta K, \theta L) = \theta F(K, L)$ for any $\theta \in \mathbb{R}$), from (2), (3) we get

$$c_t + (n+1)k_{t+1} - (1-\delta)k_t \le \tilde{y}_t = F(k_t, 1), \tag{6}$$

and if we define the function $f(\cdot)$ by $f(k) := F(k,1) + (1-\delta)k$ then (6) can be written as

$$c_t + (n+1)k_{t+1} \le y_t = f(k_t) \tag{7}$$

where y_t is the total output at time t.

In the above formulation we assume that the production function $f(k)$ and the consumption function $u(c)$ fulfil some standard assumptions on production and consumption functions, in particular, that:

AS 1. The function $u(c) : \mathbb{R}^+ \to \mathbb{R}^+$ is twice continuously differentiable and satisfies $u(0) = 0$. Moreover, $u(c)$ is strictly increasing and concave (i.e., its derivatives satisfy $u'(\cdot) > 0$ and $u''(\cdot) < 0$) with $u'(0) = +\infty$ (so-called Inada Condition).

AS 2. The function $f(k) : \mathbb{R}^+ \to \mathbb{R}^+$ is twice continuously differentiable and satisfies $f(0) = 0$. Moreover, $f(k)$ is strictly increasing and concave (i.e., its derivatives satisfy $f'(\cdot) > 0$ and $f''(\cdot) < 0$) with $f'(0) = M < +\infty$, $\lim_{k\to\infty} f'(k) < 1$.

AS 3. Since $(1+n)$ is very close to 1, we can well approximate in (7) the expression $(1+n)k_{t+1}$ by k_{t+1} and assume that in (5) $\beta < 1$. Hence (7) can be replaced by

$$c_t + k_{t+1} \le y_t = f(k_t). \tag{8}$$

Dynamic Programming Formulation. Finding a sequence $(\boldsymbol{k}, \boldsymbol{c})^T = \{k_0, c_0, k_1, c_1, \dots, k_T, c_T\}$ with a given $k_0 > 0$ maximizing (5) under condition (8) can be formulated as:

Find

$$\hat{U}_{k_0}^{\beta}(T) := \max_{(\boldsymbol{k},\boldsymbol{c})^T} \sum_{t=0}^{T} \beta^t u(c_t) \quad \text{for a finite or infinite time horizon } T, \tag{9}$$

under the constraints (for $t = 0, 1, \dots, T$)

$$c_t + k_{t+1} \le f(k_t) \tag{10}$$

$$c_t \ge 0, \quad k_t \ge 0, \quad \text{with } k_0 > 0 \text{ given.} \tag{11}$$

Note that since $u(\cdot)$, $f(\cdot)$ are increasing (cf. assumptions AS 1 and AS 2) it is possible to replace the constraints (10), (11) by

$$c_t + k_{t+1} = f(k_t), \quad \text{with} \quad f(k_t) = y_t \tag{12}$$

$$c_t \ge 0, \quad k_t \ge 0, \quad k_0 > 0 \ \text{given, and if } \ T < +\infty \ \text{ also } \ k_{T+1} = 0, \tag{13}$$

and hence also (9) can be written as

$$\hat{U}^{\beta}_{k_0}(T) = \max_{\boldsymbol{k}^T} \sum_{t=0}^{T} \beta^t u(f(k_t) - k_{t+1}) \quad \text{for a finite or infinite time horizon } \ T, \tag{14}$$

where $\boldsymbol{k}^T = \{k_0, k_1, \ldots, k_T\}$ and $U^{\beta,\boldsymbol{k}^T}_{k_0}(T) = \sum_{t=0}^{T} \beta^t u(f(k_t) - k_{t+1})$.

Observe that in virtue of assumption AS 2 and (10), (12) it holds:

Remark 1. i) If $f'(0) \le 1$ (and hence $f'(k) < 1$ for all $k > 0$), then by (12) every sequence $\{k_0, k_1, \ldots, k_t, \ldots\}$ must be decreasing and $\lim_{t\to\infty} k_t = 0$.
ii) If $f'(0) > 1$ (and hence, since $\lim_{k\to\infty} f'(k) < 1$, there exists some k' such that $f'(k) < 1$ for all $k > k'$), then there exists some $k^* > 0$ such that $f(k^*) = k^*$ and some $k_m \in (0, k^*)$ such that $f(k_m) - k_m = \max_k [f(k) - k]$.
Supposing that $k_0 > k^*$ then elements of the sequence $\{k_0, k_1, \ldots, k_t, \ldots\}$ must be decreasing for all $k_t > k^*$. Furthermore, if for some $t = t_\ell$ it holds $k_{t_\ell} < k^*$ then $k_t < k^*$ for all $t \ge t_\ell$, but $\{k_t, t \ge t_\ell\}$ need not be monotonous. However, in any case $k_t \le k_{\max} = \max(k_0, k^*)$ and $f(k_t) \le f(k_{\max}) =: y_{\max}$ for all $t = 0, 1, \ldots$.
iii) In case that $k'_0 > k_0 > 0$ then $\hat{U}^{\beta}_{k_0}(T) > \hat{U}^{\beta}_{k'_0}(T)$. This can be easily verified since if we start with capital $k'_0 > k_0$, selecting consumption at time 0 such that $c'_0 + k_1 = f(k'_0) > c_0 + k_1 = f(k_0)$ (recall that $f(k'_0) > f(k_0)$ and $u(\cdot)$ is increasing) and following for every $t > 0$ decisions given by $\boldsymbol{k}^T \equiv (k_0, k_1, \ldots, k_T)$ (the sequence of capital stocks yielding $\hat{U}^{\beta}_{k_0}(T)$ in (14)), then $u(f(k'_0) - k_1) > u(f(k_0) - k_1)$ and $u(f(k'_t) - k'_{t+1})$ with $k'_t \equiv k_t$ for all $t \ge 1$).

On employing separability occurring in (14), for finite T and a given $k_0 > 0$ we get

$$\hat{U}^{\beta}_{k_0}(T) = \max_{k_1} \left[u(f(k_0) - k_1) + \beta \hat{U}^{\beta}_{k_1}(T-1)\right], \ \hat{U}^{\beta}_{k_1}(T-1) = \max_{k_2} \left[u(f(k_1) - k_2) + \beta \hat{U}^{\beta}_{k_2}(T-2)\right],$$

$$\ldots \qquad \hat{U}^{\beta}_{k_{T-2}}(2) = \max_{k_{T-1}} \left[u(f(k_{T-2} - k_{T-1}) + \beta \hat{U}^{\beta}_{k_{T-1}}(1)\right], \ \hat{U}^{\beta}_{k_{T-1}}(1) = \max_{k_T} [u(f(k_T)],$$

and hence (using the celebrated Bellman's "principle of optimality", cf. [1])

$$\begin{aligned}\hat{U}^{\beta}_{k_0}(T) &= \max_{k_1}[u(f(k_1) - k_0) + \beta \max_{k_2}[u(f(k_2) - k_1) + \beta \max_{k_3}[u(f(k_3) - k_2) \\ &+ \beta \max_{k_4} [u(f(k_4) - k_3) + \ldots + \beta \max_{k_{T-1}} [u(f(k_{T-2}), k_{T-1}) + \beta \max_{k_T} u(k_T)] \ldots]]]].\end{aligned} \tag{15}$$

Now let us introduce the opposite time orientation, i.e. if T is fixed then for $n = 0, 1, \ldots, T$, let $c^n = c_{T-n}$, $k^n = k_{T-n}$. Then (15) can be rewritten as:

$$\begin{aligned}\hat{U}^{\beta}_{k^T}(T) &= \max_{k^{T-1}} [u(f(k^{T-1}) - k^T) + \beta \max_{k^{T-2}} [u(f(k^{T-2}) - k^{T-1}) + \beta \max_{k^{T-3}} [u(f(k^{T-3}) - k^{T-2}) \\ &+ \beta \max_{k^{T-4}} [u(f(k^{T-4}) - k^{T-3}) + \ldots + \beta \max_{k^1} [u(k^1, 1) + \beta \max_{k^0} u(k^0)] \ldots]]]]\end{aligned} \tag{16}$$

or

$$\hat{U}^{\beta}_{k^n}(n) = \max_{k^{n-1}} [u(f(k^{n-1}) - k^n) + \beta \hat{U}^{\beta}_{k^{n-1}}(n-1)] \qquad \text{for } \ n = 1, 2, \ldots, T, \tag{17}$$

where $\quad \hat{U}^{\beta}_{k^0}(0) = \max_{k^0} u(k^0) \quad$ for the selected value $\ k^0 > 0$.

2 Economic Growth Under Random Fluctuations

Interval Models. Up to now we have assumed that for a given k_t the total output $y_t = f(k_t)$ is determined by (12). To include random shocks or imprecisions into the model, we shall assume that for a given value of k_t we obtain the output y_t only with known probability $p(k_t) \equiv p(k_t; 0) < 1$, hence with probability $\bar{p}(k_t) = 1 - p(k_t)$ the total output will be different from y_t and can attain maximal and minimal possible values $f_{\max}(k_t)$ and $f_{\min}(k_t)$ respectively (of course, we assume that assumptions AS 2 also hold for $f_{\max}(\cdot)$ and $f_{\min}(\cdot)$). Since (cf. assumption AS 1) instantaneous utility function $u(\cdot)$ is increasing, on *replacing the production function $f(k_t)$ by $f_{\max}(k_t)$ and $f_{\min}(k_t)$ we obtain upper or lower bounds on the total output at time t and also, for fixed values of k_t, also the upper and lower bounds on the maximal global utility of the consumers respectively.* This approach is relatively simple, but ignores a lot of information and yields only a very rough bounds on optimal values.

Probabilistic Approaches. Obviously, significantly better results can be obtained if we replace the rough estimates of y_t generated by means of $f_{\max}(k_t)$ and $f_{\min}(k_t)$ by a more detailed information on the (random) output y_t generated by the capital k_t. Recall that by Remark 1ii the values of k_t, $y_t = f(k_t)$ (and hence also c_t) are bounded by $k_{\max}$, $y_{\max}$ respectively.

To this end we shall assume that in (8), (12)

$$y_t = Z_t f(k_t), \text{ where } Z = \{Z_t, t = 0, 1, \dots\} \text{ is a random process.} \tag{18}$$

Usually, we assume that Z is a Markov process (in general with state space $\mathbb{R}$) or an autoregressive process. Moreover, we assume that *the decision maker can observe the current values of the total output y_t and then select the value of k_{t+1}*. Such an extension well corresponds to the models introduced and studied in [14] and also in [6, 8]. Unfortunately, assuming that Z is a Markov process with compact state space $\mathbb{R}$ then a rigorous treatment of the model given by (17) requires a very sophisticated mathematics (see [3] or [14]) and is not suitable for numerical computation. To make the model computationally tractable we shall approximate our system governed by (12), (13) (with $f(\cdot) = f_{\max}(\cdot)$) by a discretized model with finite state space.

Discretized Markov Model. In what follows, we shall assume that the values of c_t, k_t, and y_t take on only discrete values. In particular, we assume that for sufficiently small $\Delta > 0$ there exists nonnegative integers $\bar{c}_t$, $\bar{k}_t$, and $\bar{y}_t$ such that for every $t = 0, 1, \dots$ it holds:
$\bar{c}_t\Delta = c_t$, $\bar{k}_t\Delta = k_t$, and $\bar{y}_t\Delta = y_t$ with $\bar{k}_t \le K := k_{\max}/\Delta$ and similarly $\bar{y}_t \le Y := y_{\max}/\Delta$.
Let elements of $\bar{k}_t$ be labelled by integers from $\mathcal{I}_K = \{0, 1, \dots, K\}$ and elements of $\bar{y}_t$ by integers from $\mathcal{I}_Y = \{0, 1, \dots, Y\}$. Hence for the total output y_t generated by the "randomized" production function we get for $\ell = 0, 1, 2, \dots, L$

$$\bar{y}_t = f(\bar{k}_t) - \ell\Delta \ \text{ with known probability } p(\bar{k}_t; \ell); \text{ obviously, } \sum_{\ell=0}^{L} p(\bar{k}_t; \ell) = 1, \tag{19}$$

and let $\boldsymbol{p}(\bar{k}_t) = [p(\bar{k}_t; 0), p(\bar{k}_t; 1), \dots, p(\bar{k}_t; L)]$.
We shall assume that $\boldsymbol{p}(\bar{k}_t)$ is "close" to $\boldsymbol{p}(\bar{k}_{t+1})$ for every $\bar{k}_t$, i.e. assume existence of some $\tilde{\Delta} > 0$ such that $|p(\bar{k}_{t+1}; \ell) - p(\bar{k}_t; \ell)| < \tilde{\Delta}$ for every $\ell = 1, \dots, L$ and $\bar{k}_t = 1, \dots, K$.

If the (random) total output at time t $\bar{y}_t = \bar{y}$ then the decision maker have option to invest for the next time point the capital $k_{t+1} = \bar{k}_{t+1}\Delta$ where $\bar{k}_{t+1} = \bar{g}_t, \dots, \bar{y}_t$ (with given $\bar{g}_t = 0, 1, \dots, f_{\max}(\bar{k}_t)$), and hence $\beta^t u((\bar{y} - \bar{k}_{t+1})\Delta)$ is the instantaneous utility accrued at time t to the global utility. In accordance with decision d taken at time t if the output $\bar{y}_t = \bar{y}$, at the next time $t+1$ capital $\bar{k}_{t+1}$ will be employed, see also the following diagram

$$\bar{k}_t \xrightarrow{p(\bar{k}_t, \bar{y}_t)} \bar{y}_t \xrightarrow{d} \bar{k}_{t+1}.$$

Using the above discretization and taking decisions with respect to the current states, the development of the economy over time can be well described by a (structured) Markov reward chain $X = \{X_\tau, \tau = 0, 1, \dots\}$ with finite state space $\mathcal{I} = \mathcal{I}_K \cup \mathcal{I}_Y$ (with $\mathcal{I}_K \cap \mathcal{I}_Y = \emptyset$), transition probabilities $p(\bar{k}_t;\bar{y}_t) = p_{ij}$, for $i = \bar{k}_t \in \mathcal{I}_K$, $j = \bar{y}_t \in \mathcal{I}_Y$, and a "non-random" transition from state $j = \bar{y}_t \in \mathcal{I}_Y$ to state $\ell = \bar{k}_t \in \mathcal{I}_K$ *associated with one-stage reward* $r_{j\ell} = u((\bar{y}_t - \bar{k}_{t+1})\Delta)$. Observe that actually "two transitions" of the Markov chain $X = \{X_\tau, \tau = 0, 1, \ldots\}$ occur within one-time period of the considered economy model and the *one-stage reward is accrued only in even transitions.* Hence the global utility (i.e. the total discounted reward of the Markov chain X) $U^\beta_{k_0}(T) = \mathsf{E}\{\sum_{t=1}^T \beta^t r_{X_{2t-1},X_{2t}} | X_0 = \bar{k}_0\}$ (the symbol E is reserved for expectation).

Further Extension of the Discretized Model. Up to now we have assumed that the probability vector $\boldsymbol{p}(\bar{k}_t)$ cannot be influenced by the decision maker. Now we extend the model in such a way that $\boldsymbol{p}(\bar{k}_t)$ will be replaced by a family of vectors $\boldsymbol{p}(\bar{k}_t, d(\bar{k}_t))$ for $d(\bar{k}_t) = 1, 2, \ldots, D$ depending on the decision taken in state $\bar{k}_t$. Moreover, some *cost,* denoted $c(d(\bar{k}_t))$, *will be accrued to this decision.*

Moreover, we shall assume that the decision d, taken if at time t the output $\bar{y}_t = \bar{y}$, assign the desired values of capital only with some probability, i.e. there is a set of feasible decisions $d(\bar{y}_t) = 1, 2, \ldots, D$ each of them assigns the value of capital $\bar{k}_{t+1}$ with known probability vector $\boldsymbol{p}(\bar{y}_t, d(\bar{y}_t)) = [p(\bar{k}_t; 1, d(\bar{y}_t)), p(\bar{k}_t; 2, d(\bar{y}_t)), \ldots, p(\bar{k}_t; \bar{y}_t, d(\bar{y}_t))]$.

We shall assume that $\boldsymbol{p}(\bar{k}_t, d)$ and $\boldsymbol{p}(\bar{y}_t, d)$ is "close" to $\boldsymbol{p}(\bar{k}_{t+1}, d)$ and to $\boldsymbol{p}(\bar{y}_{t+1}, d)$ for every $\bar{k}_t$ and $\bar{y}_t$ respectively, i.e. we assume existence of some $\tilde{\Delta} > 0$ such that $|p(\bar{k}_{t+1}; \ell, d) - p(\bar{k}_t; \ell, d)| < \tilde{\Delta}$ for every $\ell = 1, \ldots, L$ and $\bar{k}_t = 1, \ldots, K$ and $|p(\bar{y}_{t+1}; \ell, d) - p(\bar{y}_t; \ell, d)| < \tilde{\Delta}$ for every $\ell = 1, \ldots, D$ and $\bar{y}_t = 1, \ldots, K$. So the development over time is given by the following diagram

$$\bar{k}_t \xrightarrow[\;]{\substack{c(d(\bar{k}_t))\\ p(\bar{k}_t, \bar{y}_t; d(\bar{k}_t))}} \bar{y}_t \xrightarrow[\;]{\substack{u((\bar{y}_t - \bar{k}_{t+1})\Delta)\\ p(\bar{y}_t, \bar{k}_{t+1}; d(\bar{y}_t))}} \bar{k}_{t+1}$$

In contrast to the previous model transition from state $\bar{y}_t \in \mathcal{I}_Y$ to state $\bar{k}_t \in \mathcal{I}_K$ is random and given by a known probability vector $\boldsymbol{p}(\bar{y}_t, d(\bar{y}_t))$ depending on the selected decision.

3 Formulation in Terms of Stochastic Dynamic Programming

The above model can be considered as a structured standard Markov decision chain with finite state space $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$ (with $\mathcal{I}_1 \cap \mathcal{I}_2 = \emptyset$), finite set $\mathcal{D}_i = \{0, 1, \ldots, d(i)\}$ of possible decisions (actions) in state $i \in \mathcal{I}$ and the following transition and reward structure:

$p_{ij}(a)$: transition probability from $i \to j$ $(i, j \in \mathcal{I})$ if action $a \in \mathcal{D}_i$ is selected,

r_{ij} : one-stage reward for a transition from $i \to j$, with
$r_{ij} = u((i - j)\Delta)$ if $i \in \mathcal{I}_2$ and $j \in \mathcal{I}_1$
$r_{ij} = c(a)$ if $i \in \mathcal{I}_1$ and $j \in \mathcal{I}_2$, and action a is selected,

$r_i(a)$: expected value of the one-stage rewards incurred in state i if decision (or action) $a \in \mathcal{D}_i$ is selected in state i; in particular $r_i(a) = \sum_{j \in \mathcal{I}} p_{ij}(a) \cdot r_{ij}$.

A policy controlling the chain, say π, is a rule how to select actions in each state. Policy π is then fully identified by a sequence $\{d_\tau, \tau = 0, 1, \ldots\}$ of decision vectors (of dimension K and Y in odd and even steps respectively) whose ith element $d_\tau(i) \in \mathcal{D}_i$ identifies the action taken if $X_\tau = i$. Observe that decision vector d then completely identifies the transition probability matrix $\boldsymbol{P}(d)$ and the ith row of $\boldsymbol{P}(d)$ has elements $p_{i1}(d(i)), \ldots, p_{iN}(d(i))$.

Let the vector $U^{\beta,\pi}(\tau)$ denote expectation of the (random) global utility ξ_τ received in the τ next transitions of the considered Markov chain X if policy $\pi = (d_\tau)$ is followed, given the initial state $X_0 = i$, i.e., for the elements of $U^{\beta,\pi}(\tau)$ we have $U_i^{\beta,\pi}(\tau) = \mathsf{E}_i^\pi[\xi_\tau]]$ where

$\xi_\tau = \sum_{k=0}^{\tau-1} \beta^k r_{X_k,X_{k+1}}$ and E_i^π is the expectation if $X_0 = i$ and policy $\pi = (d_\tau)$ is followed. Then obviously

$$U_i^{\beta,\pi}(\tau+1) = r_i(d_i) + \beta \sum_{j \in \mathcal{I}} p_{ij}(d_i) \cdot U_j^{\beta,\pi}(\tau), \quad i \in \mathcal{I} \tag{20}$$

and for t (or τ) tending to infinity, i.e. when $\lim_{\tau\to\infty} U_i^{\beta,\pi}(\tau) = U_i^{\beta,\pi}$, (20) takes on the form

$$U_i^{\beta,\pi} = r_i(d_i) + \beta \sum_{j \in \mathcal{I}} p_{ij}(d_i) \cdot U_j^{\beta,\pi}, \quad i \in \mathcal{I}. \tag{21}$$

If $\hat{\pi}^T$ is (in general nonstationary) policy maximizing the values $U_i^{\beta,\pi}(T)$ for the fixed time horizon T then

$$U_i^{\hat{\pi}^\tau}(\tau) = \max_{d \in \mathcal{D}_i} [r_i(d) + \beta \sum_{j \in \mathcal{I}} p_{ij}(d) \cdot U_j^{\hat{\pi}^{\tau-1}}(\tau - 1)], \quad \text{for} \quad \tau = T, T-1, \ldots, 1, 0. \tag{22}$$

Furthermore, for T tending to infinity, i.e. when $\lim_{T\to\infty} U_i^{\hat{\pi}^T}(T) = U_i^{\hat{\pi}}$, then (21), (22) read

$$U_i^{\hat{\pi}} = \max_{d \in \mathcal{D}_i} [r_i(d) + \beta \sum_{j \in \mathcal{I}} p_{ij}(d) \cdot U_j^{\hat{\pi}}], \quad i \in \mathcal{I}. \tag{23}$$

Computation of Optimal Policies. In case that the time horizon T is finite, it is necessary to calculate (backwards) the dynamic programming recursion according to (22). Considering the infinite time horizon (i.e. if $\to \infty$), finding a solution of (23) is in some aspects much easier (optimal policy can be found in the class of stationary policies (i.e. policies selecting actions only with respect to the current state of Markov chain) and can be performed either by value iterations (successive approximations) or by policy iterations.

Acknowledgement. This research was supported by the Czech Science Foundation under Grants 402/05/0115 and 402/06/0990 and by the Grant Agency of the Academy of Sciences of the Czech Republic under Grant A 7075202.

References

[1] R. Bellman: Dynamic Programming. Princenton Univ. Press, Princenton, NJ. 1957.

[2] D. Blackwell: Discrete dynamic programming. Ann. Math. Statist. *33* (1962), 719–726.

[3] D. Blackwell: Discounted dynamic programming. Ann. Math. Statist. *36* (1965), 226–235.

[4] D. Cass: Optimum growth in an aggregate model of capital accumulation. Review of Economic Studies *32* (1965), 233–240.

[5] R.-A. Dana and C. Le Van: Dynamic Programming in Economics. Kluwer, Dordrecht 2003.

[6] B. Heer and A. Maußer: Dynamic General Equilibrium Modelling. Springer–Verlag, Berlin 2005.

[7] T. C. Koopmans: On the concept of optimal economic growth. In: The Econometric Approach to Development Planning. North Holland, Amsterdam 1965, pp. 225–287.

[8] M. Majamdar, T. Mitra and K. Nishinmura (eds.): Optimization and Chaos. Springer–Verlag, Berlin 2000.

[9] M. L. Puterman: Markov Decision Processes – Discrete Stochastic Dynamic Programming. Wiley, New York 1994.

[10] F. P. Ramsey: A mathematical theory of saving. Economic Journal *38* (1928), 543–559.

[11] S. M. Ross: Applied Probability Models with Optimization Applications. Holden–Day, San Francisco, Calif. 1970.

[12] P. A. Samuelson: A catenary turnpike theorem involving consumption and the Golden Rule. Amer. Econ. Review *55* (1965), 486–496.

[13] R. M. Solow: Growth Theory, An Exposition. Oxford Univ. Press, New York 1988.

[14] N.-L. Stokey and R. E. Lukas, Jr.: Recursive Methods in Economic Dynamics. Harvard Univ. Press, Cambridge, Mass. 1989.

Optimal Strategies at a Limit Order Market

Martin Šmíd

Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic
182 08 Prague, Pod Vodárenskou věží 4, Czech Republic

you@martinsmid.eu, http://www.martinsmid.eu

Abstract

We define a decision problem of an investor, trading continuously at a limit order market, maximizing a utility from his wealth at a random time horizon. We show that, in special cases (e.g. risk neutrality, quadratic or exponential utility function), the problem may be factorized and, given additional restrictions, it may even be solved.[1]

Keywords

limit order markets, optimal control, multistage decision problems, trading strategies, market microstructure

JEL Classification: C51,G10

Introduction

Recently, great attention has been payed to models of limit order markets. Existing works [2, 4, 5, 9, 10, and the references therein] assume the agents either to trade randomly or to solve one-stage (static) decision problems. We formulate a general multistage decision problem with random time horizon and we reformulate them as dynamic programming problems in two special cases - the quadratic and the exponential utility functions. We refer the reader, unfamiliar with functioning of limit order markets, to [4] or [8].

In particular, we model the behavior of an agent maximizing a utility from his wealth by trading at a frictionless limit order market with a single commodity such that the price process is continuous. The horizon of the agent is finite random, both the agent's commodity and money holdings are constrained from below by jump stochastic processes. We assume the market to be fully liquid, with zero bid-ask spread and without depth (any amount of the commodity may be purchased/sold for the current market price).

For the model to be general, the agent should be allowed to put any configuration of market and limit orders at any time instant. However, thanks to the continuity of the price process and the fact that we permit trading at any time instant, it suffices to assume the orders to be only market (putting a limit order with a limit price L is equivalent to submitting a market order at the time of the first hit of L by the price process). Moreover, since it makes no sense to put a buy market order and a sell market order at the same time, we may assume that only a single type of market order may be submitted at a single moment; therefore, a strategy of the agent may be described by a jump process denoting the commodity holding (jumps up of this proces mean buy market orders, its jumps down sell market orders). We equivalently describe those processes by collections of their jump times and their values at the jumps.

The paper is organized as follows: First, the model if formulated (Section 1). Then, the dynamic programming reformulation is derived (Section 2). Finally, note on solvability of the problem is made (Section 3).

[1]This work is supported by the grants no. 402/06/1417, 402/04/1294 and 402/03/H057 of the Czech Science Foundation.

1 Formulation of the Problem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ with a right continuous filtration $\mathcal{F}_t$, $t \geq 0$.[2] Let S be a continuous $\mathcal{F}_t$-adapted price process, let u be a utility function and let $\nu_0 \in \mathbb{R}$ be a $\mathcal{F}_0$-measurable random variable denoting an initial wealth of the agent (we assume the agent's initial holding of the commodity to be zero). Further, let $\underline{w}_t \in \mathbb{R} \cup \{-\infty\}$ and $\underline{c}_t \in \mathbb{Z} \cup \{-\infty\}$ be jump right continuous $\mathcal{F}_t$-adapted processes with finite number of jumps at each bounded interval denoting the cash holding constraint, and commodity holding constraint respectively. Furthermore, let Θ be a finite $\mathcal{F}_t$-optional time standing for a random investment horizon of the agent. Finally, for any finite $\mathcal{F}_t$-optional time τ, define the space of the strategies with the horizon τ as

$$\mathcal{X}^\tau = \{x : x = (\tau_i, c_i)_{i \in \mathbb{N}}\}$$

where $\tau_i [= \tau_i^x] \leq \tau$ is the ($\mathcal{F}_t$-optional) time of the agent's i-th action and $c_i [= c_i^x]$ is the $\mathcal{F}_{\tau_i}$-measurable[3] random variable denoting amount of the commodity owned by the agent at τ_i for each $x \in \mathcal{X}$ and $i \in \mathbb{N}$. Since, in reality, only a finite number of actions may be taken at a finite time, we assume that, for each strategy $x \in \mathcal{X}$, there exists a finite random variable $\varsigma [= \varsigma^x]$, such that

$$0 \leq \tau_1^x < \tau_2^x < \cdots < \tau_{\varsigma^x}^x = \cdots = \Theta \tag{1}$$

and, moreover, $v_i^x = 0$ whenever $i \geq \varsigma^x$.

Given this setting, the evolution of the agent's wealth given a strategy $x \in \mathcal{X}$ may be described by the jump process w with all its jumps belonging to the set $\{\tau_1, \tau_2, \dots\}$ such that

$$w_{\tau_i} [= w_{\tau_i}^x] = \sum_{0 \leq j < i} \nu_j - c_i S_{\tau_i}$$

where

$$\nu_i [= \nu_i^x] = c_i (S_{\tau_{i+1}} - S_{\tau_i})$$

for each $i \in \mathbb{N}$ (note that, thanks to our assumptions, we may write $w_\Theta = \sum_{i=0}^\infty \nu_i$).

The decision problem of the agent may now be formulated as

$$\hat{u} := \sup_{x \in X^\Theta} \mathbb{E} u(w_\Theta) \quad \text{subject to} \quad w_t^x \geq \underline{w}_t, c_t^x \geq \underline{c}_t, \quad 0 \leq t < \Theta,$$

where c_t^x is the process described by x. It is reasonable to suppose that $\hat{u} \geq C$ for some $C \in \mathbb{R}$. Moreover, without loss of generality, we may assume that, for each $x \in \mathcal{X}$, the set $\{\tau_1^x, \tau_2^x, \dots\}$ contains all the jumps of the process $(\underline{c}, \underline{w})$.[4]

2 Dynamic Programming Reformulation of the Problem

For any pair of finite optional times $t \leq s$, denote

$$\mathcal{X}^{t,s} = \{x \in \mathcal{X}^s : \tau_1^x = t\}.$$

Theorem 1 *Let u be exponential utility function, i.e. $u(x) = -\exp\{-ax\}$ for some $a > 0$. For any optional time $\tau \leq \Theta$ and any $\mathcal{F}_\tau$-measurable random variable z, define*

$$\phi_\tau(z) := \sup_{y \in X^{\tau,\Theta}} \mathbb{E}^{\mathcal{F}_\tau} u(w_\Theta^y) \quad \textit{subject to} \quad w_t + z \geq \underline{w}_t, c_t \geq \underline{c}_t, \quad \tau \leq t < \Theta$$

(ϕ is understood as a mapping from the space of all $\mathcal{F}_\tau$-measurable random variables into the same space). Then, for any pair of optional times $\sigma \leq \tau \leq \Theta$ and any $\mathcal{F}_\sigma$-measurable random variable w,

$$\phi_\sigma(w) = \sup_{x \in \mathcal{X}^{\sigma,\tau}} \mathbb{E}^{\mathcal{F}_\sigma} [u(w_\tau^x) \phi_\tau(w + w_\tau^x)] \quad \textit{subject to} \quad w_t + w \geq \underline{w}_t, c_t \geq \underline{c}_t, \quad \sigma \leq t < \tau.$$

[2] We require the filtration to be right continues for the hitting times to be optional (see [1, par. III.1.3]) so that our restriction to market orders is justified.

[3] For any optional time τ, the symbol $\mathcal{F}_\tau$ denotes the *pre-τ σ-field* defined as

$$\mathcal{F}_\tau = \{F \in \mathcal{F} : F \cup [\tau \leq t] \in \mathcal{F}_t, t \geq 0\}.$$

See [3] or [1] for more on this topic.

[4] Indeed, whenever $x = (\tau_i, c_i)_{i \in \mathbb{N}}$ is a strategy and $\theta \leq \Theta$ is an optional time, the strategy $y = (\tilde{\tau}_i, \tilde{c}_i)_{i \in \mathbb{N}}$ such that $\{\tau_1, \tau_2, \dots\} \cup \{\theta\} = \{\tilde{\tau}_1, \tilde{\tau}_2, \dots\}$ and that $\tilde{c}_i = c_{\max\{i : \tau_i \leq \tilde{\tau}_i\}}$ (we take $\tau_0 := 0$ and $c_0 := 0$) we have that $w^x = w^y$ and $c^x = c^y$.

Proof. Since $u < 0$ and since the optimal solution is bounded from below by our assumptions, the mappings $\phi_\bullet$ are well defined. Further, without loss of generality, we may assume that $\tau(\omega) \in \{\tau_i^x(\omega) : i \in \mathbb{N}\}$ for each $x \in X^\Theta$ and $\omega \in \Omega$ (see Section 2) and that $\nu_0 = 0$ (it is because the wealth constraint may be shifted and $\sup_x \mathbb{E}^{\mathcal{F}_\sigma} u(w_\Theta) = \exp\{-a\nu_0\} \sup_x \mathbb{E}^{\mathcal{F}_\sigma} u(w_\Theta - \nu_0)$).

For any σ, τ and w as in the Theorem, define

$$S_{\sigma,\tau,w} = \{x \in X^{\sigma,\tau} : w_t^x + w \geq \underline{w}_t, c_t^x \geq \underline{c}_t, \sigma \leq t < \tau\}.$$

Put $\vartheta = -u$ and fix σ, τ and w as in the Theorem. It is clear that

$$-\phi_\sigma(w) = \inf_{x \in X^{\sigma,\Theta}} \mathbb{E}^{\mathcal{F}_\sigma}[I_{S_{\sigma,\Theta,w}}(x)\vartheta(w_\Theta^x)]$$

where

$$I_\Xi(\xi) = \begin{cases} 1 & \text{if } \xi \in \Xi \\ \infty & \text{if } \xi \notin \Xi \end{cases}$$

and where, by definition,

$$\mathbb{E}^{\mathcal{F}} z = \infty \text{ whenever } z = \infty \text{ on some } A \in \mathcal{F} \text{ with } \mathbb{P}(A) > 0. \tag{2}$$

Further, for each $x \in X^{\sigma,\Theta}$, define

$$x_1 = (\tau_i^x \wedge \tau, \mathbf{1}_{\{\tau_i^x < \tau\}} c_i^x)_{i \in \mathbb{N}},$$

and

$$x_2 = (\tau_{j_i}^x, c_{j_i}^x)_{i \geq 0}.$$

where $j_i = \varsigma^{x_1} + i$. From our definitions it follows that

$$c_t^x = \begin{cases} c_t^{x_1} & \sigma \leq t < \tau \\ c_t^{x_2} & \tau \leq t \leq \Theta, \end{cases}$$

$$w_t^x = \begin{cases} w_t^{x_1} & \sigma \leq t < \tau \\ w_t^{x_2} + w_\tau & \tau \leq t \leq \Theta, \end{cases}$$

and

$$w_\Theta^x = w_\tau^{x_1} + w_\Theta^{x_2}$$

for each $x \in X^{\sigma,\Theta}$. Therefore and thanks to the facts that

$$\vartheta(r + s) = \vartheta(r)\vartheta(s)$$

for each $r, s \in \mathbb{R}$ and that

$$\Phi : X^{\sigma,\Theta} \to X^{\sigma,\tau} \times X^{\tau,\Theta}, \quad \Phi(x) = (x_1, x_2)$$

is a bijection, we may write

$$-\phi_\sigma(w) = \inf_{x \in X^{\sigma,\tau}} \inf_{y \in X^{\tau,\Theta}} \mathbb{E}^{\mathcal{F}_\sigma}[I_{S_{\sigma,\tau,w}}(x)\vartheta(w_\tau^x) I_{S_{\tau,\Theta,w+w_\tau^x}}(y)\vartheta(w_\Theta^y)]$$

$$\inf_{x \in X^{\sigma,\tau}} \mathbb{E}^{\mathcal{F}_\sigma}\left[I_{S_{\sigma,\tau,w}}(x)\vartheta(w_\tau^x) = \inf_{y \in X^{\tau,\Theta}} \left(\mathbb{E}^{\mathcal{F}_\theta} I_{S_{\tau,\Theta,w+w_\tau^x}}(y)\vartheta(w_\Theta^y)\right)\right]$$

which is nothing else but the assertion of the Theorem; we have the fact that $w^x|_{[0,\tau)}$ and $c^x|_{[0,\tau)}$ are constant in x^2 for each $x \in X^\Theta$, the chain rule and the pull out property of the conditional expectations (see [3, Theorem 6.1.] and note that these properties are preserved even after extension (2)).
$\square$

Theorem 2 *Let u be quadratic, i.e. $u(x) = x - \alpha x^2$ for some $\alpha \geq 0$, and assume that $\nu_0 < \max u$. For any finite optional time $\tau \leq \Theta$ and any $\mathcal{F}_\tau$-measurable random variables z, ζ, denote*

$$\psi_\tau(z, \zeta) := \sup_{y \in X^{\tau,\Theta}} \mathbb{E}^{\mathcal{F}_\tau} \tilde{u}(w_\Theta^y, \zeta) \quad \textit{subject to} \quad w_t + z \geq \underline{w}_t, c_t \geq \underline{c}_t, \quad \tau \leq t < \Theta.$$

where

$$\tilde{u}(w, \zeta) = u(w) - 2\alpha\zeta w.$$

Then, for any $\mathcal{F}_\tau$-measurable random variables w and η,

$$\psi_\sigma(w, \eta) = \sup_{x \in \mathcal{X}^{\sigma,\tau}} \mathbb{E}^{\mathcal{F}_\sigma}[\tilde{u}(w_\tau^x, \eta) + \psi_\tau(w + w_\tau^x, \eta + 2\alpha w_\tau^x)]$$

Proof. Since $u(\nu_0 + w^\Theta) = \beta + \gamma u'(w^\Theta)$ for some $\mathcal{F}_0$-measurable variables $\beta \in \mathbb{R}$ and $\gamma > 0$ and some quadratic utility function u' with an $\mathcal{F}_0$-measurable coefficient, we may assume that $\nu_0 = 0$. Fix σ, τ, w and η as in the Theorem and keep the notation of the proof of the previous Theorem. For each $x \in X^{\sigma,\Theta}$, we have

$$\mathbb{E}^{\mathcal{F}_\sigma}\tilde{u}(w^x_\Theta,\eta) = \mathbb{E}^{\mathcal{F}_\sigma}[u(w^{x_1}_\Theta)+u(w^{x_2}_\Theta)-2\alpha w^{x_1}_\Theta w^{x_2}_\Theta-\eta w^{x_1}_\Theta-\eta w^{x_2}_\Theta] = \mathbb{E}^{\mathcal{F}_\sigma}[\tilde{u}(w^{x_1}_\Theta,\eta)+\mathbb{E}^{\mathcal{F}_\tau}\tilde{u}(w^{x_2}_\Theta,\eta-2\alpha w^{x_1}_\Theta)].$$

The rest of the proof is analogous to the proof of the previous Theorem.
□

3 A Note on Solvability

In general, the decision problems studied in the present work may be very complicated. In special cases, however, a solution is possible: the case of risk neutrality (i.e. quadratic utility function with $\alpha = 0$) may be solved almost completely (see [7]). Further, if the problem may be factorized as in Section 2, if we restrict ourselves to strategies with N fixed (non-random) action times and if we allow the constraints to jump only at those times, our problem leads to a N-stage stochastic programming problem in a nested form whose solvability depends only on capabilities of the solving computer. For a solution of several simple problems with non-linear utility function, see [7] and [6].

References

[1] J. Dupačová, J. Hurt, and J. Štěpán. *Stochastic Modelling in Economics and Finance.* Kluwer, Dodrecht, 2002.

[2] R. L. Goettler, Ch. A. Parlour, and U. Rajan. Equilibrium in a dynamic limit order market. *The Journal of Finance*, 60(5):2149–2192, 2005.

[3] O. Kallenberg. *Foundations of Modern Probability.* Springer, New York, second edition, 2002.

[4] H. Luckock. A steady-state model of the continuous double auction. *Quantitative Finance*, 3:385–404, 2003. Preprint at `http://ej.iop.org/links/q19/Dal4WcNjf6rvjiylAiLexg/qf3505.pdf`.

[5] F. Slanina. Mean-field approximation for a limit order driven market. *Phys. Rev.*, 64:056136, 2001. Preprint at `http://xxx.sissa.it/abs/cond-mat/0104547`.

[6] M. Šmíd. Behavior of a risk averse EMH-believer at a limit order market, 2006. Preprint, Available at SSRN: `http://ssrn.com/abstract=885482`.

[7] M. Šmíd. Dynamic behavior of a rational agent at a limit order market, 2006. Submitted to QF Conference in Sydney, 2006, preprint available at `http://ssrn.com/abstract=901590`.

[8] M. Šmíd. On markov properties of limit order markets, 2006. Submitted to proceedings of Prague Stochastics 2006, preprint available at `http://ssrn.com/abstract=897917`.

[9] E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481–514, 2003. `http://www.santafe.edu/~baes/jdf/papers/doubleauctionfinal.pdf`.

[10] J. Wald and H. Horrigan. Optimal limit order choice. *The Journal of Business*, 78(2):597–619, 2005.

Growth models

K. Szomolányi, J. Lukáčik, Lukáčiková
Department of Operations Research and Econometrics, Faculty of Business Informatics, University of Economics, , Dolnozemská cesta 1B, 852 35 Bratislava, Slovak Republic
szomolan@euba.sk, lukacik@euba.sk, istvanik@euba.sk

Abstract

In the paper we provide an overview of the growth models. The models characterize real and monetary economies with exogenous and endogenous savings. The Solow model is characteristic for real economy with exogenous savings. Outcome of this model is golden rule. For monetary economy with exogenous savings is characteristic the Tobin model. There are no optimizing behavior of agents in both the Tobin and the Solow models. Optimizing behavior of agents implies that their savings are endogenous. The Ramsey model is characteristic for real economy with endogenous savings. Outcome of model is modified golden rule. For monetary economy with endogenous savings is the Sidrauski model. Outcome of the model combined with outcome of the Lucas welfare costs of inflation is the Friedman rule, or the Chicago rule respectively.

Keywords

Growth Models. Solow Model. Ramsey Model. Tobin Model. Sidrauski Model. Real Economy. Monetary Economy. Dynamic Optimization. Phelps's Golden Rule. Modified Golden Rule. Friedman Rule. Chicago Rule.
JEL: C02, E13

1 Introduction

To understand complicate economic process we need to analyze neoclassical growth models of real and monetary economy. One, who is home in such models, is able creating own views on macroeconomics and macroeconomic policy. The goal of the paper is provide complex view of the models. The hint for author is notes in [4].

2 Principles of Dynamic Optimization

"Consider a fish stock which has some natural rate of growth and which is harvested. Too much harvesting could endanger the survival of the fish, too little and profits are forgone. The obvious question is: 'what is the best harvesting rate, i.e., what is the optimal harvesting?'" [3].

This is typical example of dynamic optimization problem. Generally we can form the problem as:

$$\max J = \int_0^T U(x,u)\,dt$$
$$\dot{x} = f(x,u)$$
$$x(0) = x^0$$
$$x(T) = x^T$$

for continuous time and

$$\max J = \sum_{t=0}^{T} U(x_t,u_t)$$
$$x_{t+1} - x_t = f(x_t,u_t)$$
$$x_0 = x^0$$
$$x_T = x^T$$

for discrete time respectively; where t denotes time, u is control variable, x state variable.

There are three types of techniques relevant to economics that provide solution of such problem:

- Calculus of Variation
- Theory of Optimal Control
- Dynamic Programming

Suppose there is no control variable. In this case the Calculus of Variations (Euler equations) is the appropriate technique. Otherwise we can rely on the Lagrangean methods: Optimal Control using the Hamiltonian approach or the Dynamic Programming (Bellman equations) for discrete problems.

2.1 Calculus of Variations

Transform the continuous general problem into:

$$\max J = \int_0^T U(t, x, \dot{x})\,dt$$

$$x(0) = x^0$$

$$x(T) = x^T$$

Using perturbation intuition we can derive solution of problem – Euler equation:

$$\frac{dU(t, x, \dot{x})}{dx} = \frac{d\frac{dU}{d\dot{x}}}{dt}$$

2.2 Theory of Optimal Control

Define Hamiltonian function for general continuous problem as

$$H(x,u) = U(x,u) + \lambda f(x,u)$$

Lagrangean multiplier λ is costate variable. Optimality conditions are:

$$\frac{\partial H}{\partial u} = 0$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x}$$

$$x(0) = x^0$$

$$\lambda_T\left[x(T) - x^T\right] = 0$$

2.3 Dynamic Programming

Consider general discrete problem. Suppose that there is discount factor β valuating future values of the function U (widely applicable in economics), so we can rewrite utility function of the problem as:

$$\max J = \sum_{t=0}^{T} \beta^t U(x_t, u_t)$$

Define problem using value function as:

$$V(x_t) = \max\left\{U(x_t, u_t) + \beta V(x_{t+1})\right\}$$

Subject to:

$$x_{t+1} = f(x_t, u_t) - x_t$$

By substituting restriction to the value function we get optimality conditions:

$$\frac{dU}{du_t}+\beta\frac{df}{du_t}\frac{dV\left(x_{t+1}\right)}{du_t}=0$$

$$\frac{dU}{du_t}=\lambda_t$$

$$x_0=x^0$$

$$\lambda_T\left(x_T-x^T\right)=0$$

The value of slack parameter λ_t is given by envelope theorem (second optimality condition).

3 Real economy growth models

For real economy are characteristic the Solow and Ramsey models.
The continuous Solow model is characterized by well-known differential equation (see for example Gandolfo, 1997, or Shone, 2002):

$$\dot{k}=sf\left(k\right)-k\left(\delta+n\right)$$

Where k is per head capital, $f(k)$ is production function with the Inada conditions, s is marginal propensity of saving, n is population growth rate, δ is capital depreciation rate. If we consider that the production function $f(k)$ is Cobb-Douglass type, by solving given differential equation we get steady state, in which all variables growth by constant rate n. According to the Phelps's golden rule, if consumer maximize consumption-labor ratio, marginal productivity of capital (i.e. interest rate) must be equal to the growth rate n.

The Ramsey model is extension of the Solow model by consumer maximization of his welfare:

$$\max J=\int_0^\infty U\left(c\right)e^{-\beta t}dt$$

$$\dot{k}=f\left(k\right)-k\left(\delta+n\right)-c$$

$$k\left(0\right)=k_0$$

$$0\le c\le f\left(k\right)$$

To solve the Ramsey problem let's form Hamiltonian:

$$H=U\left(c\right)e^{-\beta t}+\mu\left(t\right)\left[f\left(k\right)-k\left(\delta+n\right)-c\right]$$

Define $\lambda(t)$ as:

$$\lambda\left(t\right)=\mu\left(t\right)e^{\beta t}$$

Then

$$H=\left\{U\left(c\right)+\lambda\left(t\right)\left[f\left(k\right)-k\left(\delta+n\right)-c\right]\right\}e^{-\beta t}$$

Optimality conditions are (see 2.2):

$$\frac{\partial H}{\partial c}=0,\ \text{so}\ \frac{\partial U\left(c\right)}{\partial c}=\lambda$$

$$\dot{\lambda}=\lambda\left[\beta+n+\delta-\frac{df\left(k\right)}{dk}\right]$$

$$0=\lim_{t\to\infty}k\left(t\right)\mu\left(t\right)$$

The second equation results from partial derivation of Hamiltonian (see 2.2). Combining the first two conditions we get:

$$\frac{\dfrac{\partial^2U\left(c\right)}{\partial c^2}}{\dfrac{\partial U\left(c\right)}{\partial c}}=\beta+n+\delta-\frac{df\left(k\right)}{dk}$$

Now define the Pratt's measure of relative risk aversion:

$$\sigma(c)=\frac{c\dfrac{\partial^2 U(c)}{\partial c^2}}{\dfrac{\partial U(c)}{\partial c}}$$

By substituting we get three key equations of the Ramsey model:

$$\dot{c}=\frac{c}{\sigma(c)}\left[\frac{df(k)}{dk}-(\beta+n+\delta)\right]$$

$$\dot{k}=f(k)-k(\delta+n)-c$$

$$0=\lim_{t\to\infty}k(t)\frac{\partial U(c)}{\partial c}$$

The first equation is an Euler condition. It is called Keynes-Ramsey rule. If productivity is high, consumption will be decreasing along the optimal path. In steady state motion of capital and consumption vanish, so interest rate equals to the economic growth rate plus discount factor β. It is called the *modified golden rule*. Equilibrium point [k^*, c^*] is saddle.

4 Monetary Economy Growth Models

For monetary economy are characteristic the Tobin and the Sidrauski models. The idea is that agents hold portfolios of real money and capital. Let resource constraints in discrete time be:

$$Y_t+\tau_t L_t+(1-\delta)K_{t-1}+\frac{(1+i_{t-1})B_{t-1}}{P_t}+\frac{M_{t-1}}{P_t}=C_t+K_t+\frac{M_t}{P_t}+\frac{B_t}{P_t}$$

At time t wealth of agents consists from Y_t units of production, $\tau_t L_t$ units of lump-sum money transfers, where L_t is population size; B_{t-1} units of inherited bonds that are evaluated by real interest rate $(1+i_{t-1})/P_t$; inherited M_{t-1} units of money evaluated by price index P_t and inherited K_{t-1} units of capital. For these resources they buy C_t units of goods and services; δK_{t-1} units of depreciated capital; K_t units of new capital; new real money balance and new bonds. Term τ_t is defined as:

$$\tau_t=\frac{M_t-M_{t-1}}{P_t L_t}$$

Now, let us express resource constraint in per capita units. Variables expressed in per capita units at time t we denote by small letters. Production of representative agent is $y_t = Y_t/N_t$, his consumption $c_t = C_t/N_t$, his capital $k_t = K_t/N_t$, money $m_t = M_t/(N_tP_t)$ and bonds $b_t = B_t/(N_tP_t)$, where N_t is population number. We assume that production function is linear homogenous of degree one, so:

$$F(\lambda K_{t-1},\lambda N_t)=\lambda Y_t$$

We know that $K_{t-1}/Nt = (K_{t-1}/N_{t-1})(N_{t-1}/Nt)$. Parameter of population growth $n = (Nt/N_{t-1}) - 1$, so:

$$y_t=f\left(\frac{k_{t-1}}{n}\right)$$

Initial resources of representative agent, i.e. his wealth denoted as ω_t at time t, is:

$$\omega_t=f\left(\frac{k_{t-1}}{1+n}\right)+\tau_t+\frac{1-\delta}{1+n}k_{t-1}+\frac{(1+i_{t-1})b_{t-1}+m_{t-1}}{(1+\pi_t)(1+n)}=c_t+k_t+m_t+b_t$$

where π_t is annual rate of inflation at time t; $\pi_t = P_t/P_{t-1}$. According to Sidrauski, the welfare of agent is given not only by his consumption, but also by his real money balance. And so he solves problem:

$$\max J=\sum_{t=0}^{T}\beta^t U(c_t,m_t)$$

$$c_t+k_t+m_t+b_t=f\left(\frac{k_{t-1}}{1+n}\right)+\tau_t+\frac{1-\delta}{1+n}k_{t-1}+\frac{(1+i_{t-1})b_{t-1}+m_{t-1}}{(1+\pi_t)(1+n)}$$

To solve the problem we use dynamic programming technique. So define the value function by:

$$V(\omega_t) = \max U(c_t, m_t) + \beta V(\omega_{t+1})$$

Knowing that $k_t = \omega_t - c_t - b_t$ we can rewrite problem as:

$$V(\omega_t) =$$

$$= \max U(c_t, m_t) + \beta V\left[f\left(\frac{\omega_t - c_t - m_t - b_t}{1+n}\right) + \tau_{t+1} + \frac{1-\delta}{1+n}(\omega_t - c_t - m_t - b_t) + \frac{(1+i_t)b_t + m_t}{(1+\pi_{t+1})(1+n)}\right]$$

We can maximize an unconstrained problem over c_t, b_t and m_t. The first order necessary condition for consumption tells that the marginal utility of holding additional capital must be equal to the marginal utility of consumption. The first order necessary condition for money gives that the marginal condition for holding bonds. The first order necessary condition for bonds gives the condition for marginal return on holding money. Using envelope theorem and first order conditions we can get:

$$\frac{\frac{dU(c_t, m_t)}{dm}}{\frac{dU(c_t, m_t)}{dc}} = \frac{i_t}{1+i_t}$$

All households are identical, so there are no mutual debt holdings.

There is monetary superneutrality in steady state: The growth rate of money does not affect the steady state values of the real variables. If we then assume that utility function is separable, such that: $U(m,c)=v(c)+w(m)$, optimality conditions are fullfiled if and only if the Friedman rule of zero nominal interest rates holds (known also as Chicago rule, see [2]).

References

[1] Gandolfo, G.: Economic Dynamics. Springer, Berlin 1997

[2] Lucas, R., E.: Inflation and Welfare. Econometrica 68, s. 247-274, 2000

[3] Shone, R.: Economic Dynamics. Cambridge University Press, 2002

[4] Notes in http://www.heijdra.org/ – website of Heijdra, B., J. (June, 2006):
Sterken, E.: Lecture Notes on Dynamic Programming, 2004
Sterken, E.: Monetary Macroeconomics
Sterken, E.: Money in the Utility Function, 2005

The Principle of Overcompleteness in Multivariate Economic Time Series Models*

Jaromír Tonner

May 30, 2006

Abstract

In this paper we apply the principle of overcompleteness to sparse parameter estimation in multivariate ARMA models (VARMA models). This new approach is based on the Basis Pursuit Algorithm originally suggested by Chen et al [1]. Overcompleteness means that we admit higher range of orders within which we are looking for lowest possible number of significant parameters (sparsity). A previous study [2] confirmed that this relaxation of the commonly used low-order assumption may yield more precise forecasts from ARMA models when compared with standard statistical estimation techniques. Here an analogical approach will be used for the analysis of multivariate economic time series. It is well-known that particular time series are strongly cross-correlated. That is why we expect our technique to be possibly successful for the multivariate case too.

Keywords

multivariate time series, sparse system, overcomplete system, VARMA models, ℓ_1 norm optimization, stationary time series

JEL: C32

1 Introduction

(Chen, Donoho, & Saunders, 1998) deal the problem of sparse representation of vectors (signals) by using special overcomplete (redundant) systems of vectors spanning this space.

In contrast with vectors which belong to a finite-dimensional space (Veselý,2002) formulates the problem of sparse representation within a more general framework of (even infinite - dimensional) separable Hilbert space.

In paper [2] we attacked the problem of sparse representation from overcomplete time series models using expansions in the Hilbert space L^2which we extend here for multivariate time series models.

2 Overcomplete VARMA model for stationary time series

Let $\{\boldsymbol{X_{t+1}} = [X^1_{t+1}, \dots, X^m_{t+1}]'\}$ be zero-mean stationary m - dimensional process with cross-covariance function $\gamma^{kl}(h) = \text{cov}(X^l_{t+1}, X^k_{t+1-h})$, $\boldsymbol{X_{t+1}} \sim$ VARMA(p,q)

$$\boldsymbol{X_{t+1}} = \sum_{j=1}^{p} \boldsymbol{\Phi^j X_{t+1-j}} + \sum_{k=0}^{q} \boldsymbol{\Theta^k Z_{t+1-k}},\ \boldsymbol{\Phi^j},\ \boldsymbol{\Theta^k}\ m \times m \, \text{matrix}, \tag{1}$$

*This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172.

in which case $\{\boldsymbol{Z_t}\}$ is a multivariate white noise, $\{\boldsymbol{Z_t}\} \sim \text{WN}(\boldsymbol{0}, \Sigma)$, Σ is positive definite. Dealing the process componentvise it is easy to extend the theory derived in [2] for multidimensional case as follows:

Let $H_t = \bar{\text{sp}}(\{X^r_{t+1-j}\}^{\infty}_{j=1}, {}^{m}_{r=1})$ be a separable closed space in L^2 spanned by historical components of $\{\boldsymbol{X_t}\}$ up to time t and $P_t : L^2 \to H_t$ the orthogonal projection operator. By ortogonalization of $\{X^r_t\}$ we get also $H_t = \text{sp}(\{Z^r_{t+1-j}\}^{\infty}_{j=1}, {}^{m}_{r=1})$ where $\{Z^r_t\}$ is time-uncorrelated, $Z^r_t = X^r_t - P_{t-1}X^r_t, t \in \mathbb{Z}$. Thus components of both $\{X^r_{t+1-j}\}$ and $\{Z^r_{t+1-j}\}$ are dictionaries in H_t. Merging both dictionaries, we get an new overcomplete dictionary $\{U^r_{t+1-j}\}^{\infty}_{j=1}, {}^{m}_{r=1}) = \{X^r_{t+1-j}\}^{\infty}_{j=1}, {}^{m}_{r=1}) \cup \{Z^r_{t+1-j}\}^{\infty}_{j=1}, {}^{m}_{r=1})$ in H_t. Fixing P, Q such that $0 \le p \le P \le \infty$, $0 \le q \le Q \le \infty$ we get an overcomplete but still finite atomic decomposition of $\hat{X}^r_{t+1}$ (i-th row of (1)):

$$\hat{X}^r_{t+1} = P_t X^r_{t+1} = \sum_{j=1}^{P} \sum_{s=1}^{m} \Phi^j_{rs} X^s_{t+1-j} + \sum_{k=1}^{Q} \sum_{s=1}^{m} \Theta^k_{rs} Z^s_{t+1-k} =: T^{P,Q}_t \boldsymbol{\xi_r} =: \sum_{l=1}^{P+Q} (\boldsymbol{U_{t+1-l}})' \xi^l_r, \tag{2}$$

with atoms $\boldsymbol{U_{t+1-j}} := \boldsymbol{X_{t+1-j}}$ for $j = 1, \ldots, P$ and $\boldsymbol{U_{t+1-P-k}} := \boldsymbol{Z_{t+1-j}}$ for $k = 1, \ldots, Q$ where $\boldsymbol{\xi} := \{\boldsymbol{\Phi}, \boldsymbol{\Theta}\}'$ is a $m(P+Q) \times m$ matrix of concatenation of coefficient matrix sequences $\boldsymbol{\Phi} := \{\boldsymbol{\Phi^j}\}^P_{j=1}$ and $\boldsymbol{\Theta} := \{\boldsymbol{\Theta^k}\}^Q_{k=1}$, more precisely $\boldsymbol{\xi^j} := \boldsymbol{\Phi^{j\prime}}$ for $j = 1, \ldots, P$ and $\boldsymbol{\xi^{k+P}} := \boldsymbol{\Theta^{k\prime}}$ for $k = 1, \ldots, Q$ Lower index marks corresponding column in the coefficient sequences or in the coefficient vectors, if $P < \infty$ and $Q < \infty$. Clearly $T_t := T^{P,Q}_t : \ell^2(J) \to H_t, J := \{1, \ldots, P+Q\}$, is bounded linear operator with closed range space $\mathcal{R}(T_t) = H_t$ of finite dimension. After changing the notation accordingly this model comprises all three commonly used representations, namely

- invertible representation $\boldsymbol{\hat{X}_{t+1}} = \sum_{j=1}^{\infty} (-\boldsymbol{\pi(j)}) \boldsymbol{X_{t+1-j}} =: T^{\infty,0}_t (-\boldsymbol{\pi})$;
- causal representation $\boldsymbol{\hat{X}_{t+1}} = \sum_{k=0}^{\infty} \boldsymbol{\psi(k)} \boldsymbol{Z_{t+1-k}} =: T^{0,\infty}_t \boldsymbol{\psi}$; $\boldsymbol{\psi(0)} = \boldsymbol{I_m}$ and $\boldsymbol{\psi(k)} = \boldsymbol{0_m}$ for $k < 0$;
- overcomplete ARMA(P, Q) representation $\hat{X}^i_{t+1} = T^{P,Q}_t \boldsymbol{\xi_i}$ with finite but sufficiently overestimated orders P, Q:
  $p \le P < \infty, q \le Q < \infty$; the choice $P = Q = 10$ being satisfactory in most cases.

Hereafter we shall deal with the third case in more detail, the sparse solution of which is expected to exclude redundant parameters which are nearly noughts allowing us to approach the original VARMA(p, q) model and its parameter estimates.
As $\mathcal{R}(T_t) = H_t$ is closed, the restriction of adjoint operator T^*_t onto H_t is a topological linear isomorphism $T^*_t : H_t$ onto closed subspace $H^{'}_t \subseteq \ell^2(J), \dim H_t = \dim H^{'}_t$. Thus instead of (2) we can solve the underdetermined system of $M := P + Q$ systems of linear equations (analogy to normal equations known from linear regression) obtained by applying T^*_t to both sides of (2):

$$\boldsymbol{b_t} = \boldsymbol{R_t \xi} \text{ or equivalently } b_i(t) = \sum_{j=1}^{M} R_{ij}(t) \xi^j \text{ for } i = 1, \ldots, M. \tag{3}$$

$b_t = [b_i(t)]^M_{i=1}$ is a column block matrix of type $M \times 1$ where each block $b_i(t) := \text{cov}(\boldsymbol{X_{t+1}}, \boldsymbol{U_{t+1-i}}) = [\langle X^r_{t+1}, U^s_{t+1-i} \rangle]^m_{r,s=1}$ is cross-covariance matrix of $\boldsymbol{X_{t+1}}$ and i-th atom $\boldsymbol{U_{t+1-i}}$. Similarly $R_t = [R_{ij}(t)]^M_{i,j=1}$ is $M \times M$ block matrix where $R_{ij}(t) := \text{cov}(\boldsymbol{U_{t+1-j}}, \boldsymbol{U_{t+1-i}}) = [\langle U^r_{t+1-j}, U^s_{t+1-i} \rangle]^m_{r,s=1}$ is cross-covariance matrix of j-th and i-th atom.

Lemma 1. *Let $\boldsymbol{X_t}$ be a stationary multivariate time series and $i, j \in \mathbb{Z}$ arbitrary. The following holds: If $\boldsymbol{X_t}$ is causal then cov$(\boldsymbol{X_{t-j}}, \boldsymbol{Z_{t-i}})$ = cov$(\sum_{k=0}^{\infty} \boldsymbol{\psi(k)} \boldsymbol{Z_{t-j-k}}, \boldsymbol{Z_{t-i}}) = \boldsymbol{\psi(i-j)} \Sigma$. Matrices Σ and $\boldsymbol{\psi(k)}$ can be obtained via multivariate Innovation algorithm (see [3]).*

Theorem 1.
If $\{\boldsymbol{X_t}\} \sim$ *VARMA*(p,q) *is zero- mean and causal with covariance (matrix) function* $\gamma = \{\boldsymbol{\gamma(h)}\}_{h=0}^{\infty}$, $\boldsymbol{\gamma(h)} := [cov(X_{t+h}^k, X_t^l)]_{k,l=1}^m = [EX_{t+h}^k X_t^l]_{k,l=1}^m = [\langle X_{t+h}^k, X_t^l \rangle]_{k,l=1}^m$*, then the equation (3) attains with* $0 \le p \le P < \infty$ *and* $0 \le q \le Q < \infty$ *the form (time index can be removed due to stationarity)*

$$\boldsymbol{b} = \boldsymbol{R\xi} \quad \text{with} \quad \boldsymbol{b} = \begin{bmatrix} \boldsymbol{\gamma_P} \\ \boldsymbol{\psi_Q}(I_Q \otimes \Sigma) \end{bmatrix}, \ \boldsymbol{R} = \begin{bmatrix} \boldsymbol{\Gamma_P} & (\boldsymbol{\Psi}(I_P \otimes \Sigma))' \\ \boldsymbol{\Psi}(I_P \otimes \Sigma) & (I_Q \otimes \Sigma) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\xi} = [\boldsymbol{\Phi}, \boldsymbol{\Theta}]' \tag{4}$$

where $\gamma_P := [\boldsymbol{\gamma(1)', \ldots, \gamma(P)'}]'$ *are covariance matrices and* $\psi_Q := [\boldsymbol{\psi(1)', \ldots, \psi(Q)'}]'$ *with* Σ *from Innovation algorithm.* I_P*,* I_Q*, is identity matrix of order* P*, resp* Q*,* $\boldsymbol{\Gamma}_P$ *and* $\boldsymbol{\Psi}$ *are Toeplitz matrices:*

$$\boldsymbol{\Gamma}_P := \boldsymbol{\gamma(i-j)}_{i,j=1}^P = \begin{bmatrix} \boldsymbol{\gamma(0)} & \boldsymbol{\gamma(1)} & \cdots & \boldsymbol{\gamma(P-1)} \\ \boldsymbol{\gamma(1)} & \boldsymbol{\gamma(0)} & \cdots & \boldsymbol{\gamma(P-2)} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{\gamma(P-1)} & \boldsymbol{\gamma(P-2)} & \cdots & \boldsymbol{\gamma(0)} \end{bmatrix} \quad \text{and} \tag{5}$$

$$\boldsymbol{\Psi} := \boldsymbol{\psi(i-j)}_{i,j=1}^{Q,P} = \begin{bmatrix} \boldsymbol{I_m} & \boldsymbol{0_m} & \cdots & \boldsymbol{0_m} \\ \boldsymbol{\psi(1)} & \boldsymbol{I_m} & \cdots & \boldsymbol{0_m} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{\psi(Q-1)} & \boldsymbol{\psi(Q-2)} & \cdots & \cdot \end{bmatrix} \text{of size } (Q \cdot m \times P \cdot m). \tag{6}$$

Now it is possible to find solution for all partial time series simultaneously by reconstructing $\boldsymbol{b} = \boldsymbol{R\xi}$

with $\boldsymbol{b} = vec(\boldsymbol{b}), \boldsymbol{R} = \boldsymbol{I_m} \otimes \boldsymbol{R}$ *and with* $\boldsymbol{\xi} = vec(\boldsymbol{\xi})$.

3 Design of the numerical simulation study

- simulated lengths: 1000 and 200 samples $(\boldsymbol{x})$, out of which the leading 800 and 160 $(\boldsymbol{x}_m)$, respectively, are used for parameter estimation; the remaining 200 and 40 for verification $(\boldsymbol{x}_v)$;
- simulations were done for several VARMA(p,q) models with varying orders and parameter matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ (see Tables below);
- 100 simulations were carried out for every pair (length,model);
- for each simulation the command `predict` from MATLAB's *System Identification toolbox* (IDENT) designed by (Ljung, 2002) was used to compute one-step predictions on $(\boldsymbol{x}_v)$ based on exact parameters and on two different estimation techniques:
  1. sparse method using BPA4 (algorithm described in [2]): P=10;
  2. maximum likelihood (ML) estimate using `arx` function from IDENT;
- for every triple (length,model,simulation) the quality of the prediction was evaluated using function `compare` from IDENT:
  1. standard deviation of one-step prediction errors,
  2. the percentage of the measured output $(\boldsymbol{x}_v)$ that was explained by the model.

  Their mean with sample std were summarized in Table 1 displayed below:
  $\Phi_1 = \begin{bmatrix} 0.5 & 0.7 \\ 0 & 0.5 \end{bmatrix}, \Theta_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1 & -0.7 \\ -0.7 & 1 \end{bmatrix},$

Type of estimate	metric	$\Phi_1\ \Theta_1\ \Sigma_1$ $n = 200$ X_t^1	X_t^2	$\Phi_1\ \Theta_1\ \Sigma_2$ $n = 200$ X_t^1	X_t^2
IDENT	$\widehat{\sigma}$	1.0580 ± 0.1192	2.7883 ± 0.3771	1.2659 ± 0.1010	1.2382 ± 0.0920
	%	68.7333 ± 3.8606	11.3014 ± 9.9543	21.4143 ± 6.0896	9.3848 ± 8.8330
EXACT	$\widehat{\sigma}$	1.0590 ± 0.1226	2.7579 ± 0.3363	1.2686 ± 0.0999	1.2396 ± 0.0962
	%	68.7727 ± 3.7684	12.8951 ± 7.9589	21.4255 ± 5.6535	9.5208 ± 8.2892
SPARSE	$\widehat{\sigma}$	1.0657 ± 0.1078	2.7422 ± 0.3755	1.2823 ± 0.1019	1.2436 ± 0.1010
	%	68.5538 ± 3.8148	12.7717 ± 10.7699	20.4376 ± 6.6134	9.2889 ± 8.1172

Type of estimate	metric	$\Phi_1\ \Theta_1\ \Sigma_1$ $n = 1000$ X_t^1	X_t^2	$\Phi_1\ \Theta_1\ \Sigma_2$ $n = 1000$ X_t^1	X_t^2
IDENT	$\widehat{\sigma}$	0.9909 ± 0.0506	3.0018 ± 0.1433	1.2205 ± 0.0551	1.2188 ± 0.0575
	%	73.1514 ± 2.7664	12.4174 ± 3.7509	21.1064 ± 2.2298	12.6711 ± 3.6243
EXACT	$\widehat{\sigma}$	0.9902 ± 0.0515	2.9975 ± 0.1430	1.2187 ± 0.0548	1.2171 ± 0.0577
	%	73.1707 ± 2.7807	12.5499 ± 3.7888	21.2162 ± 2.2028	12.7909 ± 3.6302
SPARSE	$\widehat{\sigma}$	0.9914 ± 0.0503	3.0017 ± 0.1452	1.2207 ± 0.0551	1.2179 ± 0.0576
	%	73.1354 ± 2.7593	12.4347 ± 3.7373	21.0886 ± 2.2597	12.7363 ± 3.6149

Table 1: Results for 200 resp. 1000 samples long simulations

4 Tests on real data

Let $\boldsymbol{X_t}$ be a four-dimensional time series of real national product, real consumption, real investments and state bonds 3 months nominal interest rate in the Czech republic between 1st quarter of 1995 and 4nd quarter 2006 (data source: CNB). Data was detrended (Figure 1).
Standard tests suggested VAR(1), so it was used for parameter estimation in IDENT toolbox. Sparse method estimated VAR(10). One-step ahead predictions for the last ten samples can be seen in Figure 2 (IDENT estimation) and Figure 3 (SPARSE estimation). The numerical comparison of predictions is shown in Table 2 where the same criteria as in the previous chapter were used.

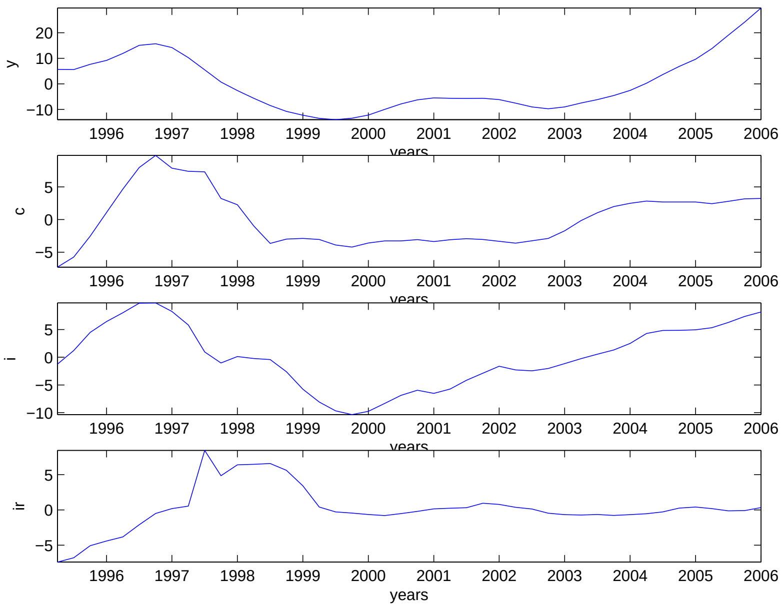


Figure 1: Real data y, c, i and nominal ir

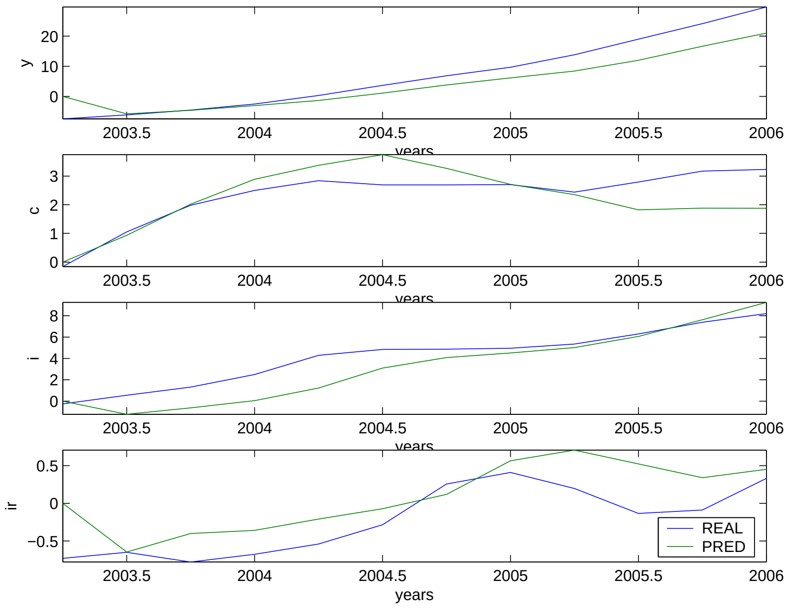


Figure 2: IDENT predictions

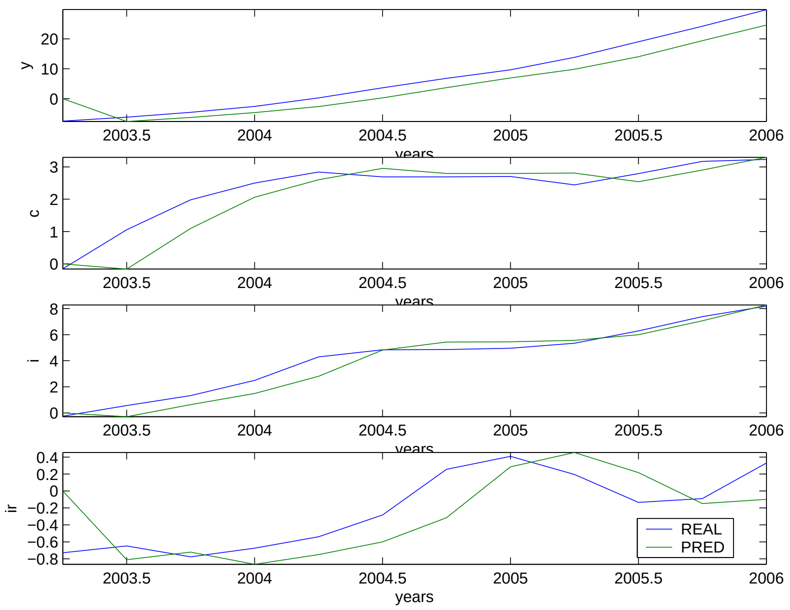


Figure 3: SPARSE predictions

Type of estimate	metric	y	c	i	ir
IDENT	$\widehat{\sigma}$	4.3797	0.7571	1.2456	0.2558
	%	61.9368	21.6084	40.4918	22.1080
SPARSE	$\widehat{\sigma}$	3.3413	0.4753	0.6441	0.3599
	%	71.4511	47.3018	73.9310	34.8296

Table 2: Results of one-step ahead predictions for the last 10 samples

5 Conclusions

From the table and other experiments we can draw the following conclusions:

- For larger sample sizes (roughly $n > 500$) the ML-estimate from IDENT and the sparse estimate produce practically equal predictions even though the parametrizations of both estimates are typically quite different.
- With decreasing sample size the sparse estimate tends to be superior to the ML-estimate from IDENT as to the precision of predictions (see Tests on real data).
- Sparse method significantly reduces the number of estimated parameters. In previous chapter sparse method reached a reduction from 160 ($= m^2 \cdot P$) to 4 parameters, but IDENT estimator of VAR(1) was full matrix (16 parameters).

References

[1] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998. reprinted in SIAM Review, **43** (2001), no. 1, pp. 129–159.

[2] V. Veselý and J. Tonner. Sparse Parameter Estimation in Overcomplete Time Series Models. *Austrian Journal of Statistics*, Special Issue devoted to the Workshop in: *Perspectives in Modern Statistical Inference III*, July 17–22, 2005, Mikulov, Czech Republic, 2006. (to appear).

[3] Brockwell, Peter J., Davis Richard A.: Time series: Theory and methods, 2 - nd edition., Springer - Verlag, New York, 1991.

[4] Ljung, L.: SYSTEM IDENTIFICATION: Theory for the User, Prentice Hall, Inc., Englewood Cliffs, NJ, 1987.

Testing nonlinear dependence in Czech stock index PX50 returns

Tran Van Quang
University of Economics in Prague
nám. W. Churchilla 4 130 67 - Praha 3
tran@vse.cz

Abstract

"Efficient Market Hypothesis"(EMH) postulates that asset prices are rationally connected to economic realities and always incorporate all the information available to the market. This concept has dominated quantitative capital market theory until recently. On one hand, a great number of studies support this hypothesis. On the other hand, there are plenty of evidence suggesting that financial markets are not efficient.

Although most of the empirical tests of the efficient markets hypothesis are based on linear models, a huge quantity of theoretical works around the world have been devoted to nonlinear processes (they can be both stochastic or deterministically chaotic) because these processes can generate output that is similar to the output of linear stochastic systems, hence it may offer an alternative explanation for the behaviour of asset prices.

In this paper, the existence of nonlinear dependence on Czech stock index PX50 returns will be examined. Nonlinear dependence may occur in a financial time series even though we have rejected the presence of linear dependence in it. Using daily observation for period 1997 to 2005 (exactly from January 1, 1997 to September 20, 2005, totally 2270 observations), some nonlinearity tests are carried out in order to decide if we can accept the weak form of efficient markets hypothesis.

First, the BDS test will be used to test the general presence of nonlinearity (either stochastic or deterministically chaotic). Secondly, several tests will be used to verify nonlinearity of a time series in its first (White test), second (Engle's test), and third moment (Hinich bispectrum). Finally, in order to test for chaos, two variables will be examined. The fractal nature of a possibly underlying strange attractor of the time series will be measured by estimating its correlation dimension. Computing the largest Lyapunov exponent, the sensitive dependence on initial conditions, a characteristic feature of a chaotic system, will be shown. The results of these tests seem to come to the conclusion that a nonlinear dependence does really exist in the daily Czech stock index PX50 returns.

Keywords:

market efficiency, nonlinearity (in)dependency, nonlinearity tests, Czech index PX50

JEL: C14, G14

1. Introduction

An efficient financial market hypothesis has dominated quantitative capital market theory for last decades. It states that asset prices are rationally connected to economic realities and always incorporate all the information available to the market, therefore no deterministic pattern can be detected. Market efficiency implies the absence of pure arbitrage opportunities and denies the profitability by the use of any investment strategy. Three forms of efficiency are distinguished according to the information set: weak form if the set includes only historical data, semistrong form if it obtains all publicly available data and strong form if it includes even private one.

Most papers concerning the issue of financial market efficiency used to test the weak form efficiency hypothesis by performing runs tests or autocorrelations tests to validate whether the returns generating process of a certain asset is deterministic (evidence against market efficiency) or stochastic (evidence for market efficiency). Usually we accept the existence of linear independence for a series P (for instance stock prices) when it is generated by a logarithmic random walk model given by

$$\log P_t = C + \log P_{t-1} + u_t$$

where u_t is an independent and identically distributed (iid) random variable with zero-mean and finite variance (often called white noise) and C is a constant drift. Evidence that log*P* follows a random walk means the weak form of the market efficiency hypothesis can be accepted, therefore returns, i.e. the log changes of prices, are unpredictable.

If the logarithmic price process follows a random walk, then the present and past returns $R_{(.)}$ where

$$R_t = \log P_t - \log P_{t-1}$$

are not associated with future returns and hence no predictability for future values of R_t exists. Therefore, the independence of R_t implies the existence of an efficient market and univariate time series methods won't succeed in capturing any returns process patterns.

However,even if there is no linear dependence, so it does not rule out nonlinear dependence. In fact, nonlinear dependence may exist in a time series even if we have already concluded for the lack of linear dependence. If present, nonlinear dependence would contradict the random walk model and the weak form of financial market efficiency hypothesis.

Checking if autocorrelation coefficients are not statistically different from zero is not sufficient. It is therefore necessary to test for the nonlinearity of returns. A set of methods were invented for determination of nonlinearity. They aim to detect the nonlinearity in mean, variance and higher moments of a time series. Because a chaotic process can generate a stochastic-like behaviour in a time series, it is necessary to assure if this chaotic proces is present in examined time series by calculating its quantificators as correlation dimension and the largest Lyapunov exponent, etc. In this paper, the BDS test, Engle's test, White test, Hinich test, Lyapunov exponent and correlation dimension test will be used to test the nonlinear dependence in the time series of Czech stock market index PX50.

2. Testing nonlinear dependence in a time series

In this section, a brief summary of the theoretical background of the above mentioned tests is given.

2.1. BDS test

Let us start with the one-dimensional series, $\{x_t\}_{t=1}^{n}$, which can be embedded into a series of *m*-dimensional vectors $X_t^m = (x_t, x_{t+1}, ..., x_{t+m-1})'$ giving the series $\{X_t\}_{t=m}^{n}$. The selected value of *m* is called the embedding dimension and each X_t is known as an *m*-history of the series $\{x_t\}_{t=1}^{n}$. This converts the series of scalars into a slightly shorter series of *m*-dimensional vectors with overlapping entries. For example, with *m* =2, the first three 2-histories will be: (x_t, x_{t+1}), (x_{t+1}, x_{t+2}), (x_{t+2}, x_{t+3}). Observe that since the first and the third history do not have any repeated element. It is obvious that from the sample size *n* , *N* = *n* - *m* + 1 *m*-histories can be made. A system which generated $\{x_t\}_{t=1}^{n}$ is ϑ-dimensional and provided $m \geq 2\vartheta + 1$, then the *N m*-histories recreates the dynamics of the data generation process and can be used to analyze the dynamics of the system. Based on the correlation integral $C(N,m,\varepsilon)$, which for a given embedding dimension m is given by

$$C(N,m,\varepsilon) = \frac{1}{N.(N-1)} \sum_{m \le t \ne s \le n}^{N-1} H\left(\varepsilon - \|X_t - X_s\|\right)$$

where ε is a sufficiently small number, $H(z)$ is the Heaviside function (which maps positive arguments into 1 and nonpositive arguments into 0) and $\|.\|$ denotes the distance induced by the selected norm. In other words, the correlation integral is the number of pair (t,s) such that each corresponding component of X_t and X_s are near each other, nearness being measured in terms of distance being less than ε. Intuitively, as a measure of the fraction of pairs of points $\left(x_t^m, x_s^m\right)$ in the series that are within a distance of ε from each other.

The BDS test (Brock, Dechert and Scheikman, 1991) is nonparametric test to assess the null hypothesis that a univariate time series $\{X_t,\, t = 1,2,...,n\}$ is independently and identically distributed against an unspecified alternative. This test is performed by examining the underlying probability structure of $\{X_t\}$ in order to search for any kind of dependence. Often referred as a nonlinear test, the BDS test statistic can be used to detect any deviation from independence even if due to the presence of nonlinear dependence in data.

The authors of BDS test show that $\{X_t\}$ is iid, then we have

$$C_m(\varepsilon) = C_1(\varepsilon)^m .$$

The BDS statistic is then given by

$$W_m^n(\varepsilon) = \frac{\sqrt{N}\left(C_m^n(\varepsilon) - \left(C_1^n(\varepsilon)\right)^m\right)}{\sigma_m(\varepsilon)}$$

where $W_m^n(\varepsilon)$ converges in distribution to a standard normal, $N(0,1)$, as n approaches infinity. The normal distribution is said to be found to be asymptotically, and a good approximation for the distribution of the BDS statistics when there are more than 500 observations.The BDS statistic depends, to a great extent, on the choice of values for ε and m. With large (small) ε the spatial correlation between the data points will tend to be high (low). The greater is the embedding dimension the smaller will be the number of non-overlapping histories and, as a consequence, the point defined by the embedding vectors will become closer and the value of the BDS statistic will tend to be higher. For a large absolute value of the test statistic we reject the null hypothesis of iid (randomness) since this provides evidence that the data are nonlinear.

The BDS test can only be used to produce indirect evidence about nonlinearity because the sampling distribution of the test statistic is unknown. The statistic is useful to test for patterns that occur more or less frequently than would be expected in independent data. The null hypothesis may be formulated as H_0 : pure whiteness, independent data, data generated by an iid stochastic process, H_1 : nonlinear dependence.

2.2. White test

In this test, the time series is fitted by a single hidden layer feed forward neural network, which is used to determine whether any nonlinear structure remains in the residuals of an autoregressive (AR) process fitted to the same time series. The null hypothesis is linearity in the mean relative to an information set. A process that is linear in the mean has conditional mean function that is a linear function of the elements of the information set, which usually contains lagged observations on the process.

The rationale for White's test (White, 1989) can be summarized as follows: under the null hypothesis of linearity in the mean, the residuals obtained by applying a linear filter to the process should not be correlated with any measurable function of history of the process. White's test uses a fitted neural net to produce the measurable function of the process's history and an AR process as a linear filter. White's method then tests the hypothesis that the fitted function does not correlated with the residuals of the AR process. The resulting test statistic has an asymptotic χ^2 distribution under the null of linearity in the mean.

2.3. Engle's test

Autoregressive conditional heteroscedasticity (ARCH) models were developed by Engle (1982) who also proposed a test that explicitly examines nonlinearity in the second moment. In its simplest form an ARCH (*p*) process can be written formally as

$$Y_t = \beta_1 + \beta_2 X_{2t} + ... + \beta_k X_{kt} + \varepsilon_t$$

where $\varepsilon_t \sim \sigma_t . v_t$ with $v_t \sim F(0,1)$ and

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + ... + \alpha_p \varepsilon_{t-p}^2$$

The null hypothesis of no autocorrelation in the error variance is :

H_0 : $\alpha_1 = \alpha_2 = ... = \alpha_p = 0$, which if accepted would lead us to deny the existence of an ARCH(*p*) model. The procedure which tests that hypothesis is as follows:

1. Regress Y_t linearly on X_t (if the information set, X_t, is restricted to the past observations of Y_t then we simply estimate an AR (*p*) process) and save the estimated residuals $\hat{\varepsilon}_t$.

2. Regress the squares of the estimated standard residuals $\hat{\varepsilon}_t^2$ on an intercept and *p* lagged values of $\hat{\varepsilon}_t^2$ as

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + ... + \alpha_p \hat{\varepsilon}_{t-p}^2 + \hat{\eta}_t$$

and save the estimated residuals $\hat{\eta}_t$.

3. Calculate the R^2 from the second regression and test the null hypothesis using the nR^2 statistic that follows a $\chi^2(p)$ distribution under the null of no ARCH dependence.

2.4. Hinich bispectrum test

The Hinich (Hinich, 1989) bispectrum test is used to estimate the bispectrum of a stationary time series and provides a direct test for nonlinearity and also a direct test for Gaussianity. If the process generating the data (in our case the rate of returns) is linear the skewness of the bispectrum will be constant. If the test rejects constant skewness then a nonlinear proces is implied.

Linearity and Gaussianity can be tested using a Hample estimator of the skewness function $\Gamma(\omega_1,\omega_2)$ with

$$\Gamma^2(\omega_1,\omega_2) = \frac{\left|B_{xxx}(\omega_1,\omega_2)\right|^2}{S_{xx}(\omega_1)S_{xx}(\omega_2)S_{xx}(\omega_1+\omega_2)}$$

where $S_{xx}(\omega)$ is the spektrum of $x(t)$ at frequency ω. The bispectrum at frequency pairs (ω_1,ω_2) is defined as

$$B_{xxx} = \sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} C_{xxx}(r,s)e^{\left[-i2\pi(\omega_1 r+\omega_2 s)\right]}$$

in the principle domain given by $\Omega = \left\{(\omega_1,\omega_2):0 \prec \omega_1 \prec 0,5; \omega_2 \prec \omega_1; 2\omega_1 + \omega_2 \prec 1\right\}$

The bispectrum is double Fourier transformation of the third order moments function and is the third order polyspectrum. The regular power spectrum is the second-order polyspectrum and is a function of only one frequency. It was proved that the skewness function $\Gamma(\omega_1,\omega_2)$ is constant over all frequencies $(\omega_1,\omega_2)\in\Omega$ if $x(t)$ is linear; while $\Gamma(\omega_1,\omega_2)$ is flat at zero over all frequencies if $x(t)$ is Gaussian. Linearity and Gaussianity can be tested using a sample estimator of the skewness function. Flatness conditions are necessary but not sufficient for third order nonlinear dependence for general linearity and Gausianity. But flatness of the skewness function is necessary nad sufficient for third order nolinear dependence. The null of the Hinich bispectrum test is actually given by:

H_0 : flat skewness function, absence of third order nonlinear dependence

H_1 : nonlinear dependence.

2.5. The correlation dimension test

Based on the correlation integral mentioned before, the correlation dimension can be defined as

$$D_c^m = \lim_{\varepsilon\to 0}\frac{\log C(N,m,\varepsilon)}{\log\varepsilon}$$

that is the slope of the regression of log *C(N,m,ε)* versus log *ε* for small values of *ε* and depends on the embedding dimension, *m*. As a practical matter one investigates the estimates values of D_c^m as m is increased. If as m increases, D_c^m continues to rise, then the system is stochastic. If, however, the data are generated by a deterministic process (consistent with chaotic behavior), then D_c^m reaches a finite saturation limit beyond some relatively small *m*. The correlation dimension can therefore be used to distinguish true stochastic processes from deterministic chaos (which may be low-dimensional or high-dimensional).

2.6. Lyapunov exponent

Lyapunov exponents measure the exponential rate at which two nearby orbits are moving apart. They provide an estimate of sensitive dependence on initial conditions, a defining feature of chaos. It basically means that if we allow for small changes in the state of a system it will grow at an exponential rate.

Consider two points, x_0 and $x_0 + \varepsilon$, apart from each other by only the infinitesimal difference ε and apply a map function to each of the two points n times. The difference between the results is given by

$$d_n = e^{n\lambda(x_0)}\varepsilon$$

and after solving for the convergence (or divergence) rate λ we have the Lyapunov exponent

$$\lambda = \lim_{n\to\infty} \frac{1}{n} \log\left|\frac{d_n}{\varepsilon}\right|$$

If a system has at least one positive Lyapunov exponent then the system is chaotic and trajectories, which start at two similar states, will diverge exponentially. The larger the dominant positive exponent the more chaotic the system is and the shorter the time span of system predictability is. A positive Lyapunov exponent is therefore viewed as an operational definition of chaotic behaviour.

3. The empirical results

In this paper, the existence of nonlinear dependence on Czech stock index PX50 returns was examined. Daily observations of Czech stock index PX50 for period from 1997 to 2005 (exactly from January 1, 1997 to September 20, 2005, totally 2270 observations) have been used. The price time series was transformed into a time series of logarithmic returns, and a set of already mentioned tests was carried out thereon.

3.1. BDS test

Tab.1: BDS statistics of the Czech stock index PX50

m	ε			
	0.5σ	σ	1.5σ	2σ
2	12.7977	11.9216	10.6162	9.5252
3	17.8809	15.8753	13.7975	12.364
4	24.1272	19.2022	16.0965	14.3627

Notes: m - embedding dimension, ε - distance between points, measured in terms of the standard deviation of the time series of the Czech stock index PX50

P-values of all cases are zero, which means that at all examined distances the null hypothesis is rejected, i.e. the logarithmic returns of the Czech stock index PX50 are not independent and identically distributed.

3.2. White test

The result of White test is stochastic, therefore a series of this test was conducted. The average P-value of all White test's atempts is less than 0,01 which means the null is rejected.

3.3. Engle's test

Tab. 2: The results of Engle's test of the Czech stock index PX50

lag	*p*-value	statistics	critical value	*H*
5	0	212.4187	11.0705	1
10	0	258.9766	18.3070	1
15	0	277.4383	24.9958	1
20	0	283.2374	31.4104	1

Notes: $H = 1$ means the null hypothesis H_0 is rejected.

3.4. Hinich test

Tab. 3: The results of Hinich Bispectral Tests

	statistics	*p*-value
the number of observations = 2270	-0.44289	0.67108
test of linearity hypothesis		

3.5. Correlation dimension

Tab. 4: Estimated correlation dimensions

m	CD	*m*	CD
1	0.695891	6	4.151413
2	1.431355	7	4.721593
3	2.171095	8	5.278817
4	2.882504	9	5.875056
5	3.542182	10	5.875056

Notes: *m* - embedding dimension, CD - correlation dimension

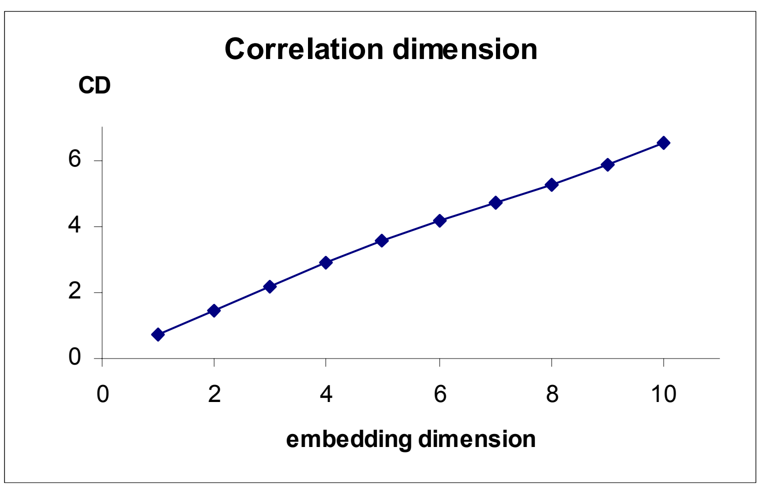


Fig. 1: Corelation dimension vs. embedding dimension
of the time series of the Czech stock index PX50

3.6. The largest Lyapunov exponent

The largest Lyapunov exponent is estimated as the slope of the regression of log *C(N,m,ε)* versus time for the upward section of the curve. (See fig. 2 below)

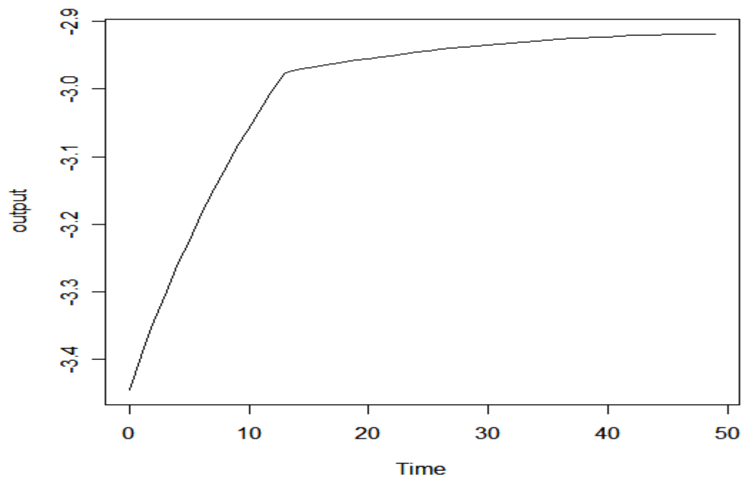


Fig. 2: Estimation of the largest Lyapunov exponent from correlation integral

Estimated value of the largest Lyapunov exponent of the Czech stock index PX50 is λ = 0.03861938.

4. Conclusions

BDS test of logarithmic returns of the Czech stock index PX50 show that they are not independent and identically distributed. White's test and Engle's test point to the existence if nonlinearity in the mean and the variance of the PX50 time series. The result of thr Hinich test does confirm the non-Gaussianity, but does not prove the presence of nonlinearity in the third moment of the examined time series. When quantifying the two descriptors of the deterministic chaos have been calculated. While the divergence of the calculated correlation dimensions may imply a stochastic behaviour of the time series, positiveness of the time series's Lyapunov exponent indicates its chaotically deterministic behaviour. This contradiction may be caused by noises present in the time series and in such cases chaotic behaviour should be considered. To sum up, the empirical results of tests carried out on the Czech stock index PX50 have strongly rejected the null hypothesis that returns of the PX50 are independent and identically distributed (except Hinich test) and also detected the presence of nonlinearity in the time series and this nonlinearity may result from a low-dimensional chaotic process and if confirmed, certain possibility of predictability of the behaviour of the Czech stock index PX50 may be revealed in future.

References

1. Brock W. A., Dechert W.D., Scheinkman J.A. (1991): *A test for independence based on the correlation dimension, Social System Research Institute*, University of Wisconsin.

2. White H.(1989): *Some asymptotic results for learning in single hidden layer feedforward network models*, Journal of the American Statistical Association 84, 1003 - 13.

3. Hinich M. J. (1982): *Testing for gaussianity and linearity of a stationary time series*, Journal of Time Series Analysis 6, 169-76.
4. Wolf A., et al. (1985): *Determining Lyapunov exponents from a time series*, Physica 16 D, 285-317.
5. Engle R. F. (1982): *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*, Econometrica, 50(4), 987-1007.
6. Grassberger P., Procaccia I. (1983): *Characterization of strange attractors*, Physical review letters 50, 349-369.

Wavelet Applications to Heterogeneous Agents Model

Lukáš Vácha and Miloslav Vošvrda *
Institute of Economic Studies, Faculty of Social Sciences, Charles University
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic, Prague
* vosvrda@utia.cas.cz

Abstract

Heterogeneous agents model with the WOA was considered for obtaining more realistic market conditions. The WOA replaces periodically the trading strategies that have the lowest performance level of all strategies presented on the market by the new ones. New strategies that enter on the market have the same stochastic structure as an initial set of strategies. This paper shows, by wavelets applications, strata influences of the trading strategies with the WOA.

Keywords

agents' trading strategies, heterogeneous agent model with stochastic memory, Worst out Algorithm, wavelets

JEL: C061; G014; D084

1 Introduction

Financial markets are considered as systems of the interacting agents processing new information immediately. Prices are driven by endogenous market forces. Agents adapt their predictions by choosing among a finite number of predictors (see [1]). Each predictor has a performance measure. Based on this performance, agents realize a rational choice among the predictors (see [2]). This approach relied on heterogeneity in the agent information and subsequent decisions either as fundamentalists or as chartists (see [4], [5]). A dynamics of the trading activities is more eventful by the WOA. For a simulation, an updated version of the WOA replacing from zero, to eight strategies with the lowest performance measure is used. A set of strategies is composed from fifteen different strategies, i.e., the replacement ratio of the market strategies is from 0% to 53.3%. The high replacement ratio is implemented for a simulation of dramatic changes in the mood on the market. From such conditions on market, there is a bigger chance of the price turbulence emergence. Discussed model is an application of an evolutionary selection of expectation rules in a financial market.

2 The Model

In this section we briefly discuss our model which is based on the papers of Brock and Hommes [1], and Vosvrda and Vacha [6]. The model is a standard discounted value asset pricing model derived from mean-variance maximization, extended to a case of heterogeneous beliefs. Agents are boundedly rational and select a forecasting or investment strategy based upon its recent performance. A convenient feature of the model is that it can be formulated in terms of deviations from a benchmark fundamental and therefore it can be used in experimental and empirical testing of deviations from the rational expectation (RE) benchmark.

* A support from the Grant Agency of Charles University under the grant 454 /2004/A –EK/FSV , the Czech Science Foundation under the grant 402/04/1294, and from the Ministry of Education of the Czech Republic under project MSM0021620841 is gratefully acknowledged.
* A support from the Grant Agency of Charles University under the grant 454 /2004/A –EK/FSV , the Czech Science Foundation under the grant 402/04/1294, and from the Ministry of Education of the Czech Republic under project MSM0021620841 is gratefully acknowledged.

2.1 Heterogeneous Beliefs

Beliefs about the conditional variance for all h, t are assumed to be equal and constant for all types. Beliefs about future dividends are assumed to be the same for all trader types and equal to the true conditional expectation. All traders are thus able to derive the fundamental price p_t^* that would prevail in a perfectly rational world. Traders nevertheless believe that in a heterogeneous world prices may *deviate* from their fundamental value p_t^*. It is convenient to use the *deviation* from the fundamental price $x_t = p_t - p_t^*$. Beliefs about the future price of the risky asset are of the form:

$$E_{h,t}\left[\mathbf{p}_{t+1}\right] = E_t\left[\mathbf{p}_{t+1}^*\right] + f_h^{\mathbf{L}}\left(x_{t-1},\ldots,x_{t-\mathbf{L}}\right) \tag{2.1}$$

where $\mathbf{L}$ is a random variable of lags. Each forecasting rule f_h represents a model of the market (e.g. a technical trading rule) according to which type h believes that prices will deviate from the fundamental price. The heterogeneous agent market equilibrium equation can be formulated in deviations from the benchmark fundamental as

$$\left(1 + r^{riskfree}\right)\cdot x_t = \sum_{h=1}^{H} n_{h,\,t-1}\cdot f_h^{\mathbf{L}}\left(x_{t-1},\ldots,x_{t-\mathbf{L}}\right) \equiv \sum_{h=1}^{H} n_{h,\,t-1}\cdot f_{h,\,t}^{\mathbf{L}} \tag{2.2}$$

Let a performance measure $\pi_{h,t}$ be defined by

$$\pi_{h,t} = E_t\left[\frac{\left(\mathbf{p}_{t+1} - \left(1 + r^{riskfree}\right)\cdot p_t\right)\cdot \rho_{h,t}}{\sigma_t^2}\right] \tag{2.3}$$

where

$$\begin{aligned}\rho_{h,t} &= E_{h,t}\left[\mathbf{p}_{t+1} - \left(1 + r^{riskfree}\right)\cdot p_t\right] = f_{h,t}^{\mathbf{L}} - \left(1 + r^{riskfree}\right)\cdot x_t = \\ &f_{h,t}^{\mathbf{L}} - \sum_{j=1}^{H} n_{j,t}\cdot f_{j,t}^{\mathbf{L}} = f_{h,t}^{\mathbf{L}}\cdot\left(1 - \sum_{j\neq h}^{H} n_{j,t}\cdot f_{j,t}^{\mathbf{L}}\right)\end{aligned}\quad. \tag{2.4}$$

So the π_h-performance is given by the realized performance for the h-investor. Let the updated fractions $n_{h,t}$ be given by the discrete choice probability (Gibb's distribution)

$$n_{h,t} = \exp\left(\beta\cdot\pi_{h,t-1}\right)/Y_{t-1} \tag{2.5}$$

where

$$Y_t = \sum_{j=1}^{H}\exp\left(\beta\cdot\pi_{j,t}\right) \tag{2.6}$$

2.2 Evolutionary Selection of Strategies

The parameter β is the intensity of choice measuring the amount of uncertainty in choice. We can say the more uncertainty the lesser the parameter β. The parameter β is a measure of investor's rationality. If the intensity of choice is infinite ($\beta = +\infty$), the entire mass of investors uses the strategy that has the highest performance. If the intensity of choice is zero, the mass of investors distributes itself evenly across the set of available strategies. All forecasts will have the following form

$$f_t^{\mathbf{L}} = g\cdot\left(x_{t-1} + \cdots + x_{t-l}\right) + b \tag{2.7}$$

where the g denotes the trend of investor, and the b denotes the bias of investor. If $b = 0$, the investor is called a pure trend chaser if $g > 0$, and a contrarian if $g < 0$. If $g = 0$, investor is called purely biased. Investor is upward (downward) biased if $b > 0$ ($b < 0$). In the special case $g = b = 0$, the investor is called fundamentalist, i.e., the investor believes that price return to their fundamental value. Fundamentalists strategy is based on all past prices and dividends in their information set, but they do not know the fractions $n_{h,t}$ of the other belief types.

3 Simulations and the Worst Out Algorithm

The WOA replaces periodically the trading strategies that have the lowest performance level of strategies presented on the market by the new ones. The new strategies that enter on the market are taken from the set that has the same stochastic parameters as the initial strategies, i.e., the trend $g \sim N(0,0.16)$, the bias $b \sim N(0,0.09)$, the memory length **L**$\sim U(1,100)$. Simulations are performed with fifteen agents or beliefs represented by trading strategies, the intensity of choice, β, is set to 120. The WOA makes the replacement after 40 iterations. For example, when we want to replace four strategies with the lowest performance (4WOA, replacement ratio is 26.6%) the algorithm after every 40 iterations evaluate and arrange in descending order the performance of all fifteen strategies in the market and the last four replaces by the new ones. Number of observations in our simulations is 8192. For a better understanding of the evolution dynamics with the WOA we compare eight cases that differ in the replacement ratio. The first one is without WOA (0WOA, replacement ratio 0%), the last one replaces eight strategies (8WOA, replacement ratio 53.3%).

3.1 Analysis of the Price Returns Time Series

For estimating and analyzing of correlation structures on capital markets, a nonparametric method, which is called rescaled range, or R/S analysis that is used for estimating the Hurst exponent [4]. The R/S analysis was used for distinguishing random and non-random systems, the persistence of trends, and duration of cycles. This method is very convenient for distinguishing random time series from fractal time series as well. Starting point for the Hurst's coefficient was the Brownian motion as a primary model for random walk processes. If a system of random variables is an independently, identically distributed, then H = 0.5..The values of Hurst exponent belonging to 0 < H< 0.5 signifies anti-persistent system of variables covering less space than random ones. Such a system must reverse itself more frequently than a random process, and we can equate this behavior to a mean-reverting process. Values 0.5<H<1 show persistent process that is characterized by long memory effects. This long memory occurs regardless of time scale, i.e., there is no characteristic time scale, which is the key characteristic of fractal time series [4], [6].

For the 0WOA case (no replacement of strategies), presented market strategies are generated randomly and the Hurst exponent, as we expect, is close to the EMH case, i.e., 0.5. When the WOA is implemented, we can see a strong learning effect that is transformed to long-memory (persistent) behavior of price returns.

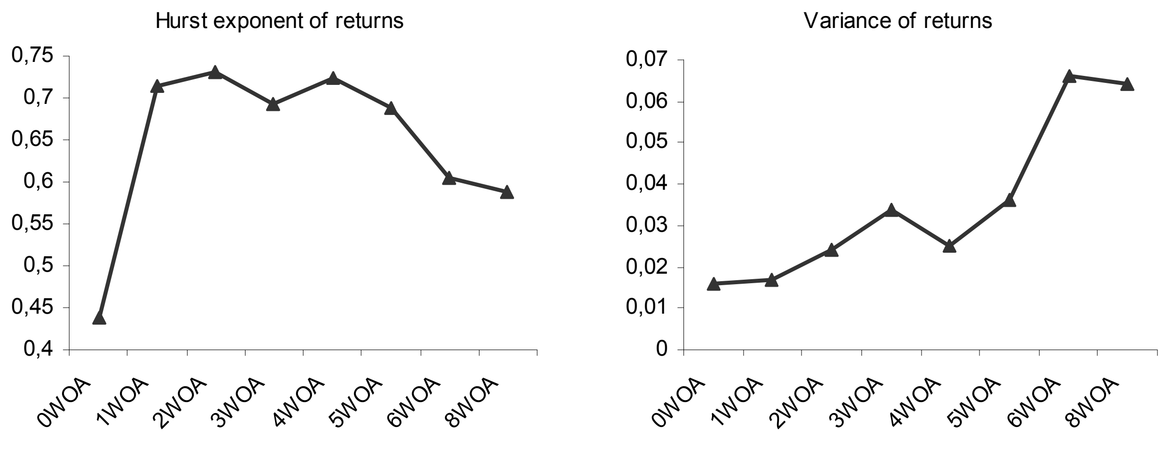


Figure 1 left: The value of the Hurst exponent with different replacement rate of the WOA, **right:** The value of the kurtosis of price returns time series with different replacement rate of the WOA.

The highest level of persistence is in the 2WOA case (13.3% replacement rate) where the market has enough time to learn. When the number of replaced strategies is higher the learning effect is weaken by the randomly chosen new strategies that appear on the market. With higher replacement ratio, the value of the Hurst exponent declines as the learning is "diluted" by new strategies that enter randomly

the market. This phenomenon takes place from the 5WOA case to the 8WOA case (33 % - 53 %replacement rate), see Figure 1 left panel and Table 1.
A variance of the price returns time series is depicted in Figure 1 right panel. We can observe rising trend as the replacement ratio increases. The higher is a number of incoming strategies the higher is price volatility. Such a high fluctuation of strategies causes that the dynamic system representing the simulated financial market has not time to stabilize.

Table 1: Hurst exponent, kurtosis and variance of the simulated price returns time series

	0WOA	1WOA	2WOA	3WOA	4WOA	5WOA	6WOA	8WOA
Hurst	0,438	0,714	0,732	0,693	0,724	0,687	0,605	0,589
Kurtosis	-1,2	56,6	27,7	7,9	21,7	10,0	4,8	4,2
Variance	0,016	0,017	0,024	0,034	0,025	0,036	0,066	0,064

3.2 Wavelet Decomposition of the Price Returns Time Series

The wavelet transform decomposes 1-dimensional time series into 2-dimensional time-scale (frequency) space. In particular, while Fourier analysis breaks down a time series into constituent orthogonal sinusoids of different frequencies (constant periodicities), wavelet analysis breaks down such a time series into constituent orthogonal wavelets of different scales. The wavelet transform replaces the basic sinusoidal waves by a family of basic wavelets generated by translations and dilatations of one particular wavelet atom $\psi_{\tau,a}(t)$ [7]. The higher scales correspond to the most dilated (stretched) wavelets. The more dilated the wavelet, the longer the portion of the time series with which it is being compared, and thus the coarser time series features being measured by the wavelet resonance coefficients *W*. There is an inverse relationship between scale and frequency (see [7] [8]):
low scale ↔ compressed wavelet ↔ rapidly changing time series details ↔ **high frequency**
high scale ↔ dilated wavelet ↔ slowly changing, coarser time series ↔ **low frequency**
We use a specific version of the discrete wavelet transform (DWT) that can be directly applied to a time series observed over a discrete set of times. Times series $x(t)$ can be completely decomposed in terms of approximations, provided by so-called scaling functions, and details, provided by the wavelets. The detailed wavelet resonance coefficients, which correlate wavelets with particular segments of the time series $x(t)$ as

$$W(j,n)=\int_{-\infty}^{+\infty} x(t)\psi_{j,n}(t)dt, \qquad j,n \in Z \tag{3.1}$$

where $\psi_{j,n}(t)$ is a dyadic wavelet basis,

$$\psi_{j,n}(t)=\frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j n}{2^j}\right) \tag{3.2}$$

The discrete dyadic scale parameter $a_j = 2^j$, while the translation interval is $\tau_n = 2^j n$. This procedure is also called wavelet multi-resolution analysis (MRA). In summary, the DWT is an important practical tool for financial time series analysis.
The basic reasons are: ability to re-express a time series in terms of coefficients that are associated with a particular time and a particular dyadic scale, we can also reconstruct a time series form its DWT coefficients; the DWT allows us to partition the energy in a time series into pieces that are associated with different scales and time [8].
As an illustrative example we compare the 1WOA and 8WOA case. In Figure 2 top panel we have a simulated price time series. The simulation were performed with the 1WOA i.e., low replacement ratio. At the left side of the Figure 2 top panel, around 2500 iteration there is a positive price jump. This price behavior is analyzed in a 6-scale wavelet resonance coefficient sequence decomposition, which shows the dynamic phenomena are identifiable at all six scales, see Figure 2 low panel. The 8WOA case (see Figure 3), have, in comparison to the lower replacement ratio case higher occurrence

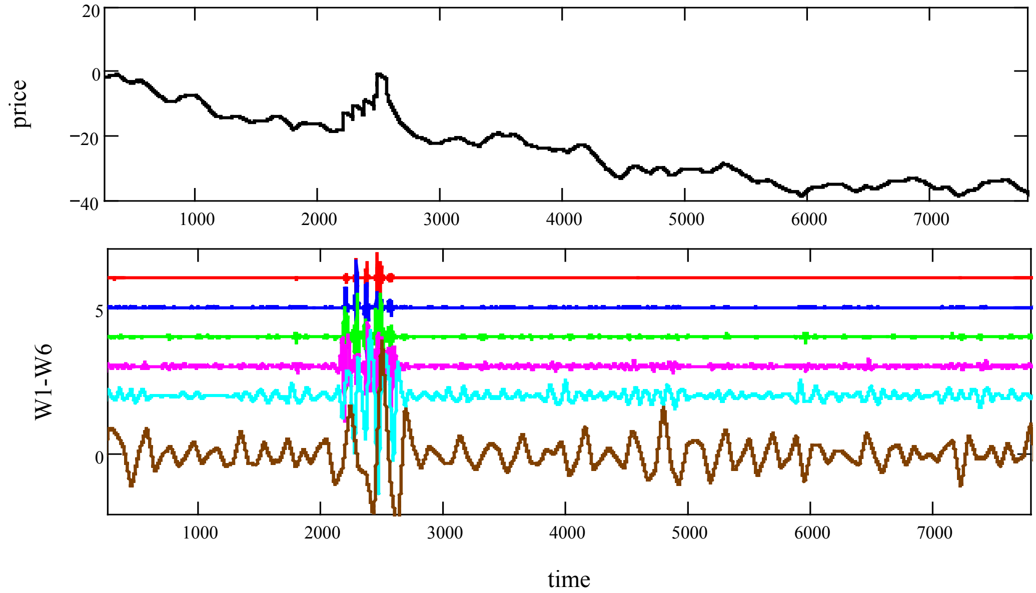


Figure 2: The 1WOA case, a 6-scale wavelet decomposition of the price time series, where the high frequency wavelet resonance coefficients are at the top, and the low frequency resonance coefficient are at the bottom.

of significant price swings. This represents higher standard deviation of the detail wavelet resonance coefficients depicted in Figure 4. This is also evident in Figures 2, 3.

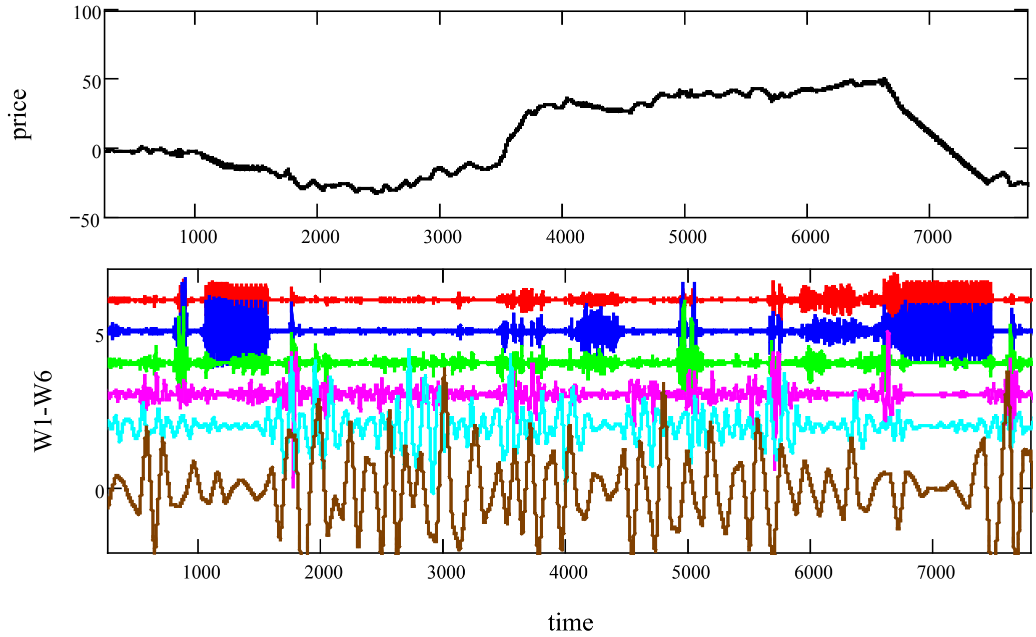


Figure 3: The 8WOA case, a 6-scale wavelet decomposition of the price time series, where the high frequency wavelet resonance coefficients are at the top, and the low frequency resonance coefficient are at the bottom.

3.3 Wavelet Variance

The wavelet variance decomposes a variance of stochastic processes on scale basis and hence is important in financial time series processing. The wavelet variance is a succinct alternative to the power spectrum based on the Fourier transform, yielding a scale-based analysis that is often easier to interpret than the frequency-based spectrum [8]. Such decomposition helps us to track an evolution of the energy contribution at various scales, which is related to traders' investment horizons. The 0WOA case has the lowest the wavelet variance (except W4) from all cases, Table 2. When we compare the 0WOA case with the 1WOA case, we see a significant increase in the wavelet variance.

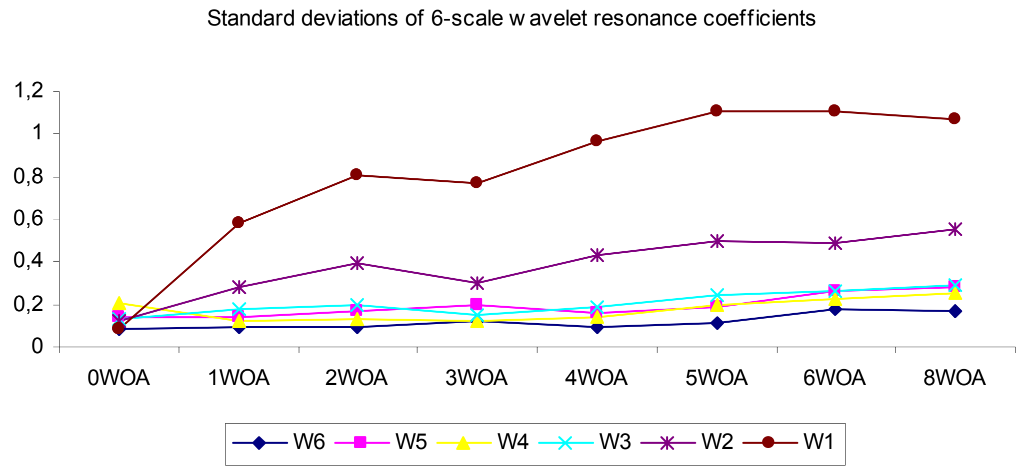


Figure 4: Standard deviations of 6-scale wavelet resonance coefficients of the price returns time series, where the highest frequency (the lowest scale) wavelet resonance coefficients are W6 and the lowest frequency W1

This refers to a considerable increase in energy at high scales W1, W2, see Figure 4. With higher replacement ratio the wavelet variance is raising, which indicates higher activity levels at all scales, except for the W1 where from the 5WOA case the wavelet variance of W1 drops slightly.

4 Conclusions

We demonstrate that the heterogeneous agent model considerably changes its behavior when we implement the Worst Out Algorithm (WOA). An implementation of the WOA increases a persistence of the price time series considerably, but when we are increasing the number of replaced strategies beyond some point, then the value of the Hurst exponent declines as the learning is "diluted" by new strategies that enter randomly the market. We can also observe higher price time series volatility as the replacement ratio rises.

Adoption of the wavelet variance is a very convenient tool for activity (energy) decomposition on scale basis. Simulations show, that the higher replacement ratio (it causes increment in the wavelet variance) the higher activity levels at all scales.

The R/S analysis and wavelet transforms enable to reformulate the EMH by a behavior of the Hurst exponent, the kurtosis of the price returns, the variance of the price returns, and the standard deviations of wavelet resonance coefficients of the price returns.

References

[1] Brock, W. A., Hommes, C.H.: *Heterogeneous beliefs and routes to chaos in a simple asset pricing model*. Journal of Economic Dynamics and Control 22. 1998.

[2] Chiarella, C., He, X.: *Heterogeneous beliefs, risk and learning in a simple asset pricing model with a market maker*. Quantitative Finance Research Group. University of Technology Sydney, Research Paper, No 35. 2000.

[3] Lucas, R.E.: *Asset prices in an exchange economy*. Econometrica 46. 1978.

[4] Peters, E., E.: *Fractal Market Analysis, Applying Chaos Theory to Investment and Economics*. Springer 1994

[5] Vacha L., Vosvrda M.: Dynamical *Agents' Strategies and the Fractal Market Hypothesis*, Prague Economic Papers, no. 2,vol. 14, 2005, pp. 172-179, ISSN 1210-0455

[6] Vosvrda M., Vacha L.: *Heterogeneous Agent Model with Memory and Asset Price Behaviour*, Prague Economic Papers, no.2, vol.12, 2003, pp. 155-168, ISSN 1210-0455

[7] Los, C. A.: *Financial Market Risk: Measurement and Analysis. Routlege International Studies in Money and Banking*. Taylor and Francis Books Ltd, London, UK. 2000.

[8] Percival, D. B., Walden, A. T.: *Wavelet Methods for Time series Analysis*. Cambridge University Press. 2000.

Interdependent consumer behavior – numerical investigation of stability

Ondřej Vácha [a], Ladislav Lukáš [b]
University of West Bohemia in Pilsen, Faculty of Economics,
[a] student, Dept. of Statistics and Operations Research, [b] Dept. of Statistics and Operations Research,
Husova 11, 30614 Plzeň
[a] kecin007@seznam.cz, [b] lukasl@kso.zcu.cz

Abstract

Paper discusses dynamic model describing interdependent consumer behavior, and presents some results regarding its stability obtained by numerical procedure coded in Matlab. First, the background for description of consumer behavior on the base of his/her utility function is presented, stressing that in various situations the utility experienced by a customer is strongly affected by other customers using the same resources. Such types of negative externalities lead to interdependent consumer behavior models presented in 2D state space. Description of consumer demand pattern is composed from two simultaneous effects, the own past consumption decisions and imposed externalities, in particular. The derived infinitesimal model is converted into discrete model by time differences technique, albeit an alternative approach using direct numerical integration of differential equations exists, e.g. by Runge-Kutta methods. Specific numerical values of the constitutive coefficients defining the consumer demand pattern play significant role when investigating stability of his/her behavior in time. Numerical procedure coded in Matlab enables both analysis of stability, and investigation of bifurcations occurred, as well. Numerical results are discussed in detail, especially those ones showing ranges of selected constitutive coefficients to produce early the bifurcation and later the coalescence of solutions, thus bringing system to a new globally stable fixed-point. Such results bring evidence that rather complicated evolution possibilities of consumer behavior in time may exist under interdependent influences upon the same resources.

Keywords

Consumer behavior, interdependent relations, dynamic model, numerical investigation of stability.

JEL: C63, C65, D11, D58.

1 Introduction

This paper attempts to analyze a model of interdependent consumer behavior of two individuals using the same resources. It is well known phenomena that the patterns of consumption are spreading within the individuals, who are related to each other in various ways. Any two consumers with similar interdependent utility functions, who observe the consumption of each other, adapt their own consumption in order to gain a positive social effect and to avoid a negative one. Each of those individuals is influenced by the past consumption of the other person as well as his (her) own past consumption decisions. Such externalities lead to interdependent consumer behavior models presented in 2D state space. In this case, we do not anticipate any influence of the rest of society on those two individuals. Let us consider the possibility of modifying a dynamic version of Cobb-Douglas utility function of the form

$$Q_{1,t} = \alpha_1 C_{1,t}^{a} \cdot C_{1,t+1}^{1-a} \tag{1}$$

The Cobb-Douglas utility function satisfies an important specification, it has constant elasticities with respect to consumption at present and in the future. Since we expect that a past consumption would not affect individual utility one for one, we consider weighted form of the first individual's consumption with a weight of α_1. We can express weighted consumption as follows.

$$\alpha_1 C_{1,t} = Q_{1,t}^{\frac{1}{a}} \cdot C_{1,t+1}^{\frac{a-1}{a}} \tag{2}$$

The influence of the other individual can be expressed by adding to the indifference curve equation a linear component, which would be a function of $C_{2,t}$. As before, we expect that a consumption of the other individual would not affect utility of the first individual one for one, so we consider weighted form of the first individual's consumption with a weight of D_{12} as

$$\alpha_1 C_{1,t} = Q_{1,t}^{\frac{1}{a}} \cdot C_{1,t+1}^{\frac{a-1}{a}} - D_{12} \cdot C_{2,t} \tag{3}$$

In order to gain a form expressing the present consumption, which is influenced by the individual's past consumption as well as by the consumption of the other person; we have to derive a utility function of the form

$$\text{Max. } Q_{1,t} = (\alpha_1 \cdot C_{1,t} + D_{12} C_{2,t})^a \cdot C_{1,t+1}^{1-a} \tag{4}$$

$$0 = C_{1,t+1}^{\frac{a}{a-1}} \left[(1-a)(\alpha_1 \cdot C_{1,t} + D_{12} C_{2,t})^a + a(\alpha_1 \cdot C_{1,t} + D_{12} C_{2,t})^{a-1} \right] \tag{5}$$

Which yields, following [1], a recurent expression

$$C_{1,t+1} = \frac{a-1}{a} \cdot (\alpha_1 \cdot C_{1,t} + D_{12} C_{2,t}) \tag{6}$$

This equation can be modified in order to consider the consumption of two commodities. In our numerical experiments we have used following two equations based on the formulation given in [1], which has the following structure expressing consumption of the first commodity by two consumers, measured in quantities $x_{i,t}$, i =1,2 :

$$x_{1,t+1} = \alpha_1 \frac{b_1}{p_x} x_{1t} \cdot y_{1t} + D_{12} \frac{b_1}{p_x} x_{2t} \cdot y_{2t} \tag{7}$$

$$x_{2,t+1} = D_{21} \frac{b_2}{p_x} x_{1t} \cdot y_{1t} + \alpha_2 \frac{b_2}{p_x} x_{2t} \cdot y_{2t} \tag{8}$$

Consumption of the second good is measured in quantities $y_{i,t}$ for i = 1,2. We assume the consumers to be rational and not to violate the budget constraint, therefore the consumption of the other commodity is determined via the stringent budget constraint, which is described by $p_x x_{it} + p_y y_{it} = b_i$ for i = 1,2, and for all t. We presume prices p_x, p_y and disposable income b_i to be constant over the time. After some technical manipulation we get the following quadratic difference equations:

$$x_{1,t+1} = \alpha_1 \frac{b_1}{p_x} x_{1t} \left(\frac{b_1}{p_y} - x_{1t} \frac{p_x}{p_y} \right) + \frac{b_1}{p_x} D_{12} x_{2t} \left(\frac{b_2}{p_y} - x_{2t} \frac{p_x}{p_y} \right) \tag{9}$$

$$x_{2,t+1} = \alpha_2 \frac{b_2}{p_x} x_{2t} \left(\frac{b_2}{p_y} - x_{2t} \frac{p_x}{p_y} \right) + \frac{b_2}{p_x} D_{21} x_{1t} \left(\frac{b_1}{p_y} - x_{1t} \frac{p_x}{p_y} \right) \tag{10}$$

Where D_{ij} and D_{ji}, i,j = 1,2, are external constitutive coefficients expressing the influence which the j's consumption does have on i's consumption pattern, and vice versa. On the contrary, experience parameters α_i express an ability of i-th consumer ability to learn from his (her) own past consumption.

The four constitutive coefficients define a four-dimensional vector of the system and play a significant role when investigating stability of consumer's behavior in time. In order not to violate the budged constraint the values of those coefficients are required to fulfill the following restrictions.

$$\frac{\alpha_1}{4}\frac{b_1^2}{p_x p_y}+\frac{D_{12}}{4}\frac{b_2^2}{p_x p_y}\le 1 \qquad \text{and} \qquad \frac{\alpha_2}{4}\frac{b_2^2}{p_x p_y}+\frac{D_{21}}{4}\frac{b_1^2}{p_x p_y}\le 1. \tag{11}$$

As mentioned before, prices p_x, p_y and disposable income b_i are constant over the time and their values have been set to b_1 = 10, b_2 = 20, p_x = 1/4 and p_y = 1 reducing restrictions (11) to the following:

$$\alpha_1 100 + D_{12} 400 \le 1 \qquad \text{and} \qquad \alpha_1 100 + D_{12} 400 \le 1. \tag{12}$$

Certain numerical values of the constitutive coefficients cause reduction in complexity of consumer behavior. This led us to analyze the stability of a system for various values of parameters using numerical procedures coded in Matlab. Results of this modeling are shown on Fig. 1, which presents a cycle diagram for values α_2 = 0.00052, D_{21} = 0.00792 and variable values of parameters α_1 and D_{12}.

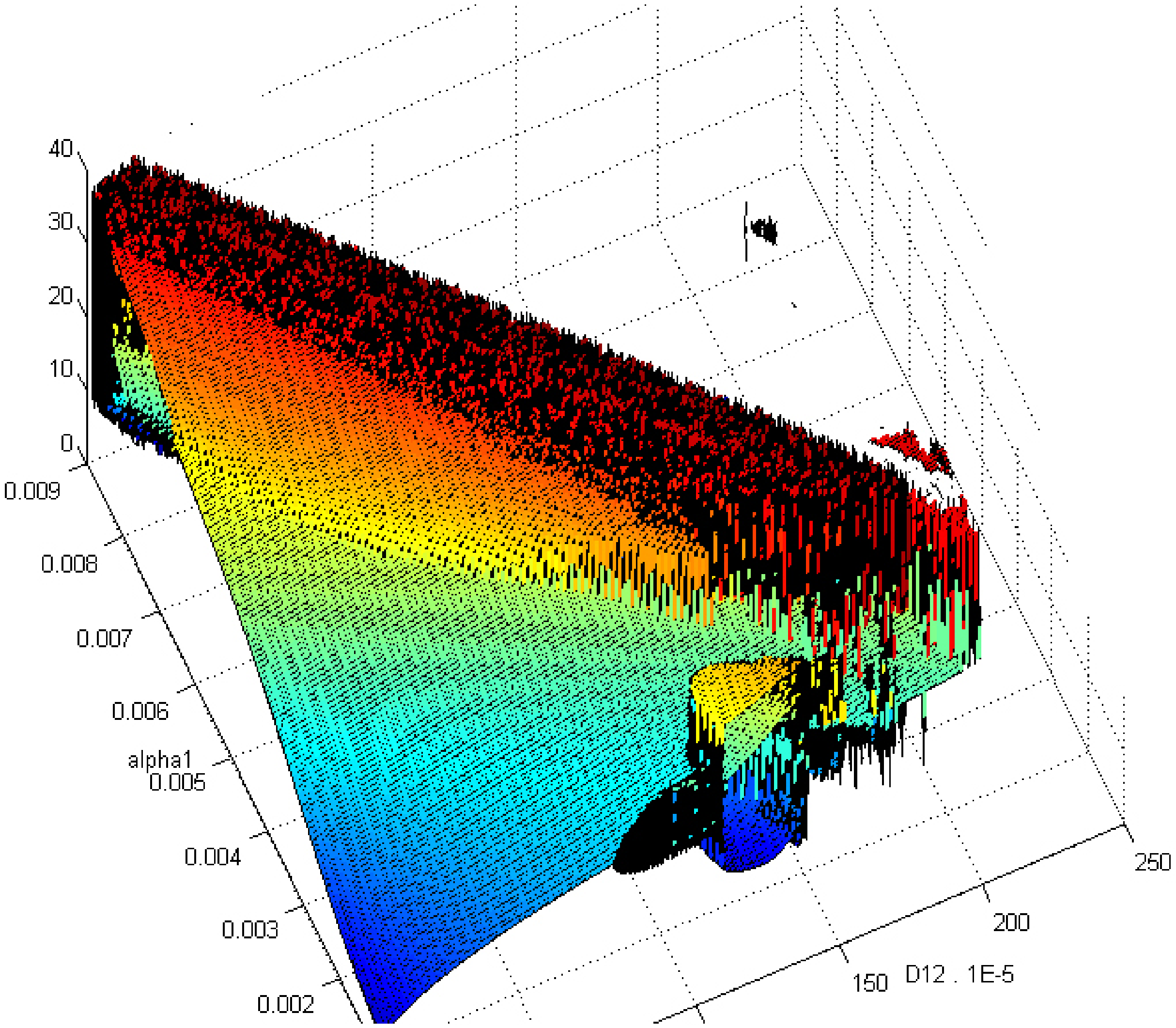
2 Appearance of the spindle structure and results of numerical experiments

We observed that certain constellation of coefficients produced early the bifurcation and later the coalescence of solutions, thus bringing system to a new globally stable fixed-point (see the cycle diagram). For example, when parameters are set to α_1=0.001 (α_2=0.00052, D_{21} = 0.00792) stability of the system is observed for values $D_{12} \in (0,0.00092)$ and both individuals consume fixed quantities of resources – see Fig.1. For larger values of D_{12} neither of individuals consumes fixed quantities of resources any longer and the trajectories start to behave irregularly. We couldn't designate a cycle order with exception of values $D_{12} \in (1.30,1.55)$, where trajectories converged to a cycle three periods. As D_{12} increases beyond 0.00179 the system is brought into a new globally stable fixed-point. Detailed graph for previous constellation of parameters with initials values $x_0 = (3,76)$ is shown on Fig.2.

Because of the unique appearance of observed phenomena is this structure referred to as a spindle structure, where the left-hand tip as well as the right-hand one announces so called Hopf-bifurcation and reversed Hopf-bifurcation, respectively. For lower values of constitution coefficients, the dynamics of the system is quite simple. As the level of mutual influence between the consumers is getting more intensive, the consumption is becoming more complex. Even though in some cases increasing a degree of interaction does not necessarily lead to higher complexity. Eventually the occurrence of a spindle structure approves that higher sensitivity to the consumption of another individual can result in a reduction of complexity of the consumption behavior. This event can be observed on Fig.2, where reduction of complexity, the convergence to a cycle of period three in particular, can be found inside of the spindle and at the point of reversed Hopf-bifurcation, where a globally stable fixed point reappears.

Figure 1

Cycle diagram for values α_2 = 0.00052, D_{21} = 0.00792 and variable values of parameters α_1 and D_{12}



To investigate the appearance of bifurcations from a global point of view a special numerical procedure in Matlab has been made for selected constellation of coefficients. Results obtained for values of experience parameters set to α_2 = 0.00052, α_1 = 0.001, and for values of external parameters D_{21}, D_{12} variable in range of $D_{12} \in (0.0008, 0.00225)$, $D_{21} \in (0.0075, 0.00792)$ shown on Fig.3 indicate, that some trajectories converge to a cycle of period three as the parameter D_{21} increases beyond 0.00760. As values of parameter D_{21} increases further beyond 0.00770 a spindle structure appears. Particular graphs for values $D_{21} = 0.00760$ and $D_{21} = 0.00770$ are displayed on the Fig.4 and Fig.5 below.

Figure 2
Example of the spindle structure for values $\alpha_2 = 0.00052$, $D_{21} = 0.00792$ and $\alpha_1 = 0.001$

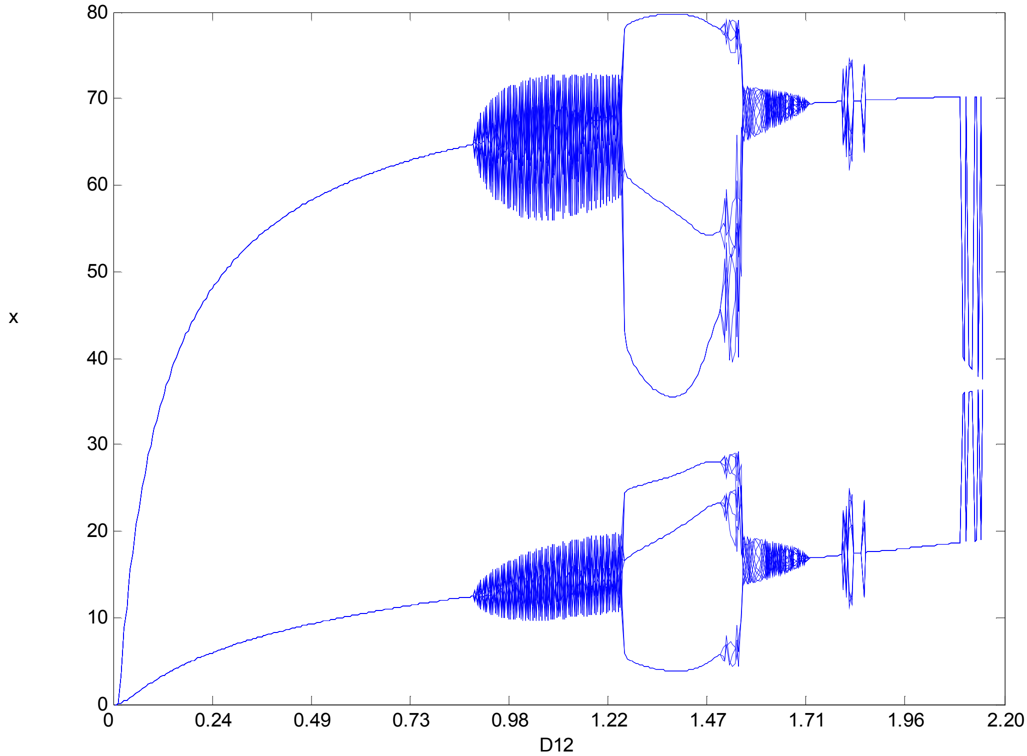


Figure 3
Diagram for values $D_{12} = \{0.0008, 0.00225\}$, $D_{21} = \{0.0075, 0079\}$, $\alpha_1 = 0.001$, $\alpha_2 = 0.00052$

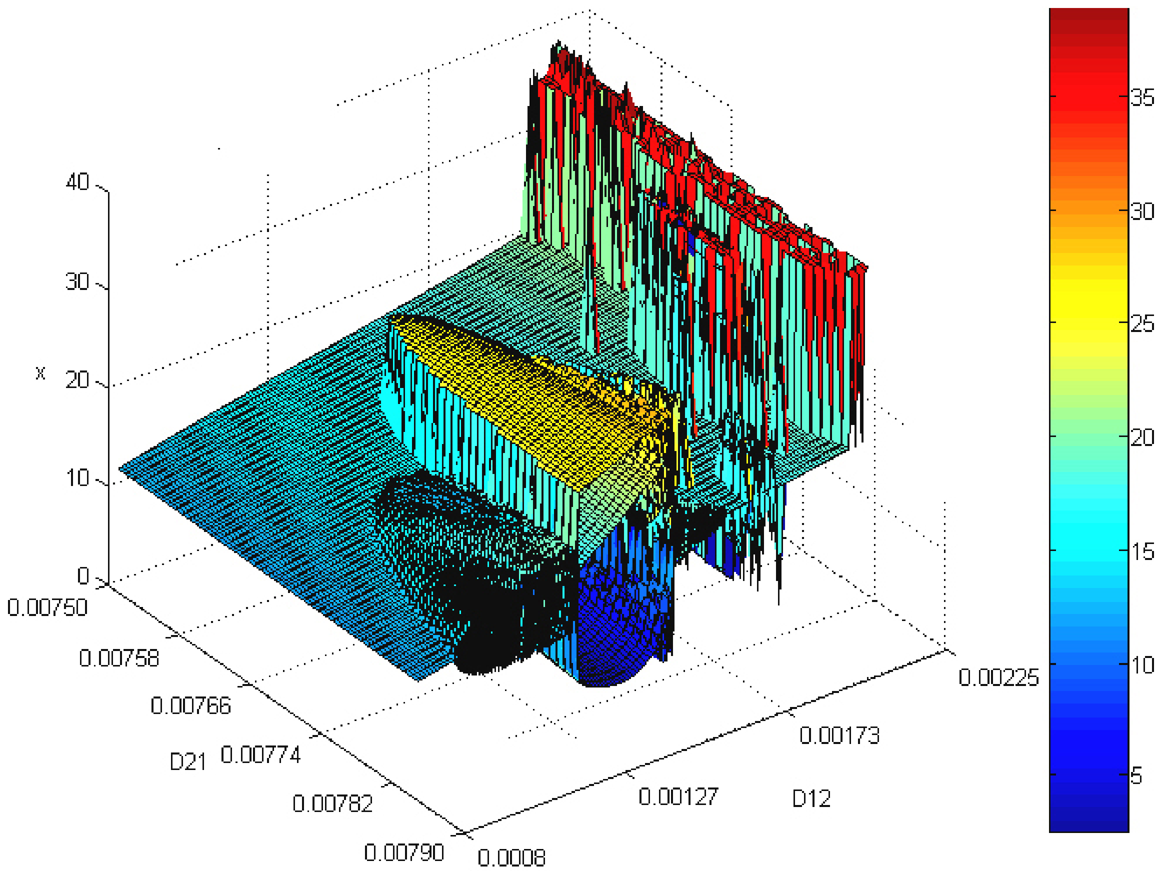


Figure 4
Graph for values $\alpha_2 = 0.00052$, $\alpha_1 = 0.001$ and $D_{21} = 0.00760$

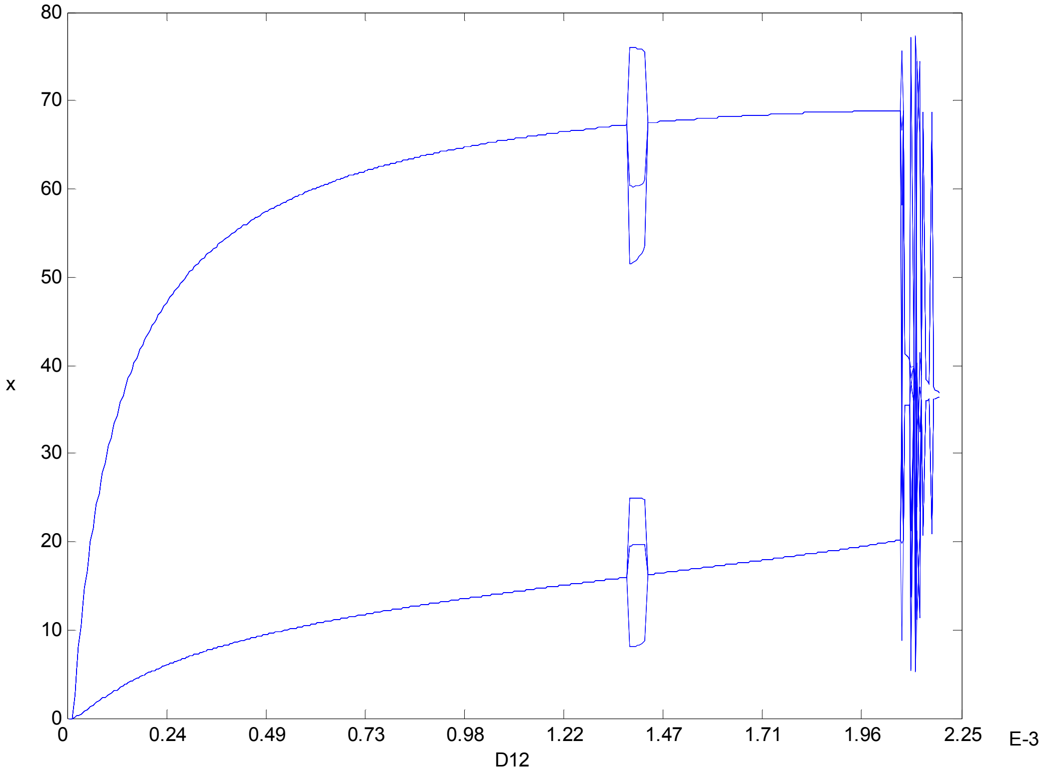
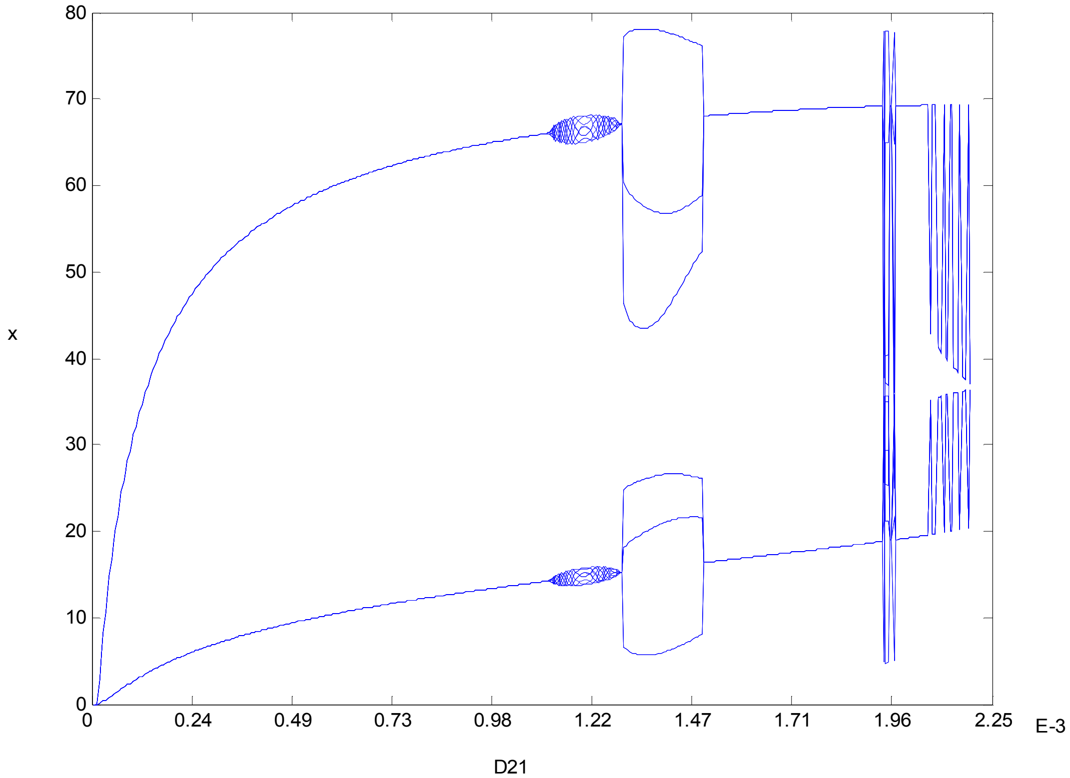


Figure 5
Graph for values $\alpha_2 = 0.00052$, $\alpha_1 = 0.001$ and $D_{21} = 0.00770$



3 Conclusion

Our model of interdependent consumer behavior led us to several intriguing conclusions about a behavior which is affected not only by individual's past consumption, but also by consumption of another individual. Any construction of economical model expects certain simplification of a reality. In this case we have considered consumption of only two resources by two individuals and we did not anticipate any influence of the rest of society on those two individuals. In spite of those simplifications our modeling has given us quite interesting results which exposes unexpected relations between economical magnitudes.

Acknowledgement: This work was funded by supported by the project LC06075 of the Ministry of Education of Czech Republic. The second author acknowledges gratefully also support by the grant No. 402/05/2392 of the Grant Agency of Czech Republic.

4 References

[1] Geartner, W., Jungeilges, J.: 'Spindles' and coexisting attractors in a dynamic model of interdependent consumer behavior. Journal of Economic Behavior and Org. 21 (1993), 223-231.

[2] H. Arsham: A Markovian model of consumer buying behavior and optimal advertising pulsing policy. (University of Baltimore, USA)

[3] Tamotsu Onozaki, Tatsuo Yanagita: Monopoly, Oligopoly and The invisible hand. (Department of Economics, Asahikawa University, Nagayama, Hokkaido, Japan)

[4] J.Benhabib, A.Rutichini: A note on a new class of solutions to dynamic programming problem arising in economic growth.(New York University, USA)

[5] Radoslav Raykov: A two-player game theoretical model of interdependent consumer choice (The Harvard College Economist, USA, Spring 2001)

[6] P. A. Samuelson, W.D. Nordhaus: Economics, in Czech, Svoboda Praha 1991

[7] R. G. D. Allen: Mathematical economics, in Czech, Academia Praha 1971

[8] D.Bauerová, L. Hrabáč: Mathematical economics I, Faculty of Economics, Ostrava 1996

5 Appendix A – example of Matlab code used for generating Fig.2

```
clear all;
close all;

alp1=0.001;
alp2=0.00052;
D21=0.00792;
b1=10;
b2=20;
px=0.25;
py=1;

n=10000;
poc=1;
m=1;

x1(1)=3;
x2(1)=76;
X1=[];
X2=[];

for D12=0.00001:0.000005:0.0022;

    for t=2:n,
        x1(t)=alp1*(b1/px)*x1(t-1)*(b1/py-x1(t-1)*px/py)+b1/px*D12*x2(t-1)*(b2/py-x2(t-1)*px/py);
        x2(t)=alp2*(b2/px)*x2(t-1)*(b2/py-x2(t-1)*px/py)+b2/px*D21*x1(t-1)*(b1/py-x1(t-1)*px/py);

    end;

X1=[X1;x1((n-10):n)];
X2=[X2;x2((n-10):n)];

end;

plot(X1,'b');hold on;
plot(X2,'b');hold on;
```

6 Appendix B – examples for another spindle structures for different initial values

Figure B1

Example of the spindle structure for values $\alpha_2 = 0.00052$, $D_{21} = 0.00792$ and $\alpha_1 = 0.001$, $x_0 = (3,80)$

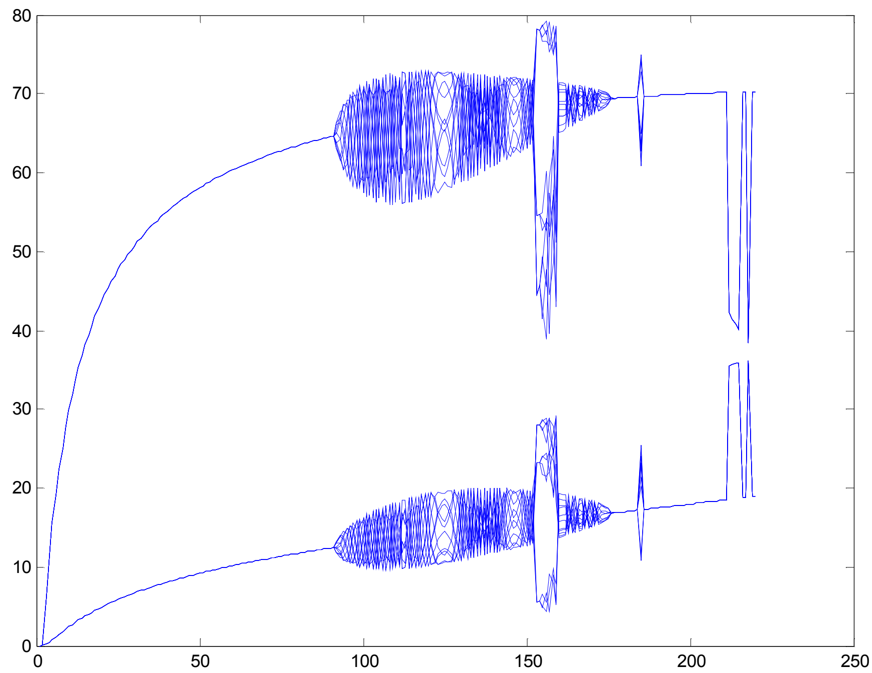
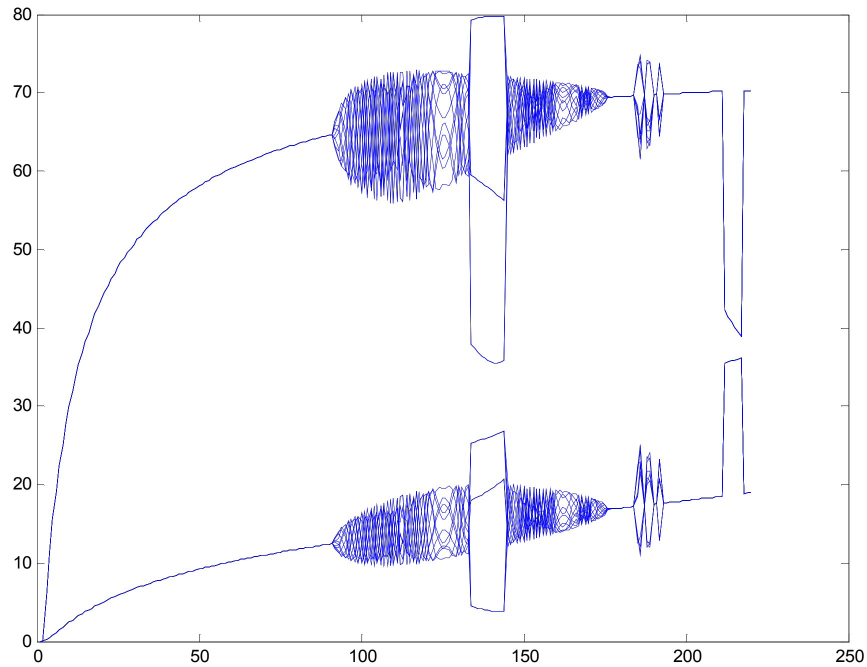


Figure B2

Example of the spindle structure for values $\alpha_2 = 0.00052$, $D_{21} = 0.00792$ and $\alpha_1 = 0.001$, $x_0 = (7,40)$



How to guarantee fair shares to everybody

Milan Vlach
The Kyoto College of Graduate Studies for Informatics
7 Monzen-cho, Tanaka, Sakyo-ku, Kyoto 606-8225 Japan
m_vlach@kcg.ac.jp
Charles University, Faculty of Mathematics and Physics
Malostranské náměstí 25, 118 00 Praha 1, Czech Republic
mvlach@ksi.ms.mff.cuni.cz

Abstract

We are concerned with situations in which the members of a group of individuals wish to divide a given divisible object in such a way that every member is satisfied with the piece he or she receives. The object in question may be heterogeneous, different members may value the same piece differently, and the meaning of satisfaction of a member may be defined in a number of different ways. Apparently the first modern framework for a rigorous analysis of this ever recurring problem was introduced by Steinhaus and his colleagues Banach and Knaster in the forties of the last century. Numerous results, some purely existential and non-constructive, some algorithmic, have been obtained over the past 60 years. Actually there are too many of them to be surveyed in one lecture. Instead we confine our discussion to one outstanding open question that seems to be of interest to both theorists and practitioners. It turns out that the so-called envy-free divisions exist under rather general assumptions. However these results have been established by means of highly non-constructive mathematical tools like fixed point theorems or Lyapunov's theorem on the ranges of vector measures. The question is how to find such a division by some satisfactory procedure in a finite number of steps.

Keywords

Fair division, Envy-freeness, Game-theoretic procedures
JEL: C69, C72, D63

1 Introduction

We are concerned with an old and ever recurring problem: *how to divide things fairly.* To make this question rigorous we have to know more about the nature of things to be divided and the notion of fairness. In accordance with a well-established custom we discuss the problem in terms of dividing a *cake* among two or more parties, which we will call *players*.

For economists interested in fair division, the cake is usually represented by a point in a space of perfectly divisible homogeneous commodities. A typical example is the case of m commodities and n players in situations without production. Then the cake is represented by a point

$$C = (c^1, c^2, \ldots, c^m)$$

in the nonnegative orthant $\mathbf{R}_+^m$ of the m-dimensional Euclidean space where the jth component c^j of C is interpreted as the amount of the jth commodity to be divided among n players. In other words, the cake C is a given total endowment to be divided into n commodity bundles

$$\begin{array}{lcl} x_1 & = & (x_1^1, x_1^2, \ldots, x_1^m), \\ x_2 & = & (x_2^1, x_2^2, \ldots, x_2^m), \\ \vdots & \vdots & \qquad \vdots \\ x_n & = & (x_n^1, x_n^2, \ldots, x_n^m), \end{array}$$

where the component x_i^j is interpreted as the amount of commodity j allocated to player i. Of course as feasible divisions of C are considered only the ordered n-tuples $(x_1, x_2, \ldots, x_n)$ of points in the commodity space $\mathbf{R}_+^m$ with the property

$$x_1 + x_2 + \cdots + x_n = C.$$

For example, in the case of a single commodity, we are given a positive number c^1 representing the initial endowment, and the feasible divisions of the cake $C = (c^1)$ among n players are ordered n-tuples $(x_1^1, x_2^1, \ldots, x_n^1)$ of nonnegative numbers satisfying

$$x_1^1 + x_2^1 + \cdots + x_n^1 = c^1.$$

Stated differently, in this simple case, the feasible divisions are represented by points belonging to an $(n-1)$-dimensional simplex in the n-dimensional space $\mathbf{R}^n$.

For comparison of feasible divisions, it is assumed that the preferences of player i are represented by a complete transitive binary relation $\succeq_i$ on the commodity space $\mathbf{R}_+^m$. Observe that preference relations are defined on the space of commodities and not on the set of feasible divisions. However, various relations between feasible divisions can be introduced on the basis of players' preferences on the space of commodities.

One of such relations between feasible divisions, frequently used in the economic literature, is that of *Pareto domination*:

> *A feasible division $(x_1, x_2, \ldots, x_n)$ is said to dominate a feasible division $(y_1, y_2, \ldots, y_n)$ if $x_i \succeq_i y_i$ holds for each player i and, at least for one player, this preference relation holds as strict preference.*

Maximal elements with respect to this partial order, that is, the undominated feasible divisions, are called *Pareto-optimal* or *Pareto-efficient*.

Since Pareto-efficient divisions can be extremely inequitable, sociologists, philosopher and politicians are interested also in notions related to equity. Especially, an extensive literature has been devoted to the notion of envy:

Player i is said to envy player j at feasible division $(x_1, x_2, \dots, x_n)$, if $x_j \succ_i x_i$. A feasible division $(x_1, x_2, \dots, x_n)$ is called envy-free if no player is envious at $(x_1, x_2, \dots, x_n)$.

Thus a feasible division is envy-free if no player strictly prefers a commodity bundle allocated to anyone else to the commodity bundle allocated to him or her. In contrast to the notion of Pareto-efficiency, the notion of envy-freeness does not require any comparison between feasible divisions.

Another class of problems arises when some of the commodities are not divisible. Particularly the case with one perfectly divisible homogeneous commodity (often called money) and a finite number of indivisible objects has been studied in the economic literature. The corresponding commodity space can be defined as the set X of all ordered pairs (α, K) where α is a nonnegative real number and K is a finite set of integers. Again it is assumed that the preferences of players are defined on the commodity space X. The cake to be divided among n players is again a given point $C = (c^1, c^2)$ in the commodity space, and the feasible divisions of C are ordered n-tuples $x = (x_1, x_2, \dots, x_n)$ of points in X with the properties

- $x_1^1 + x_2^1 + \cdots + x_n^1 = c^1,\ x_1^1 \geq 0, x_2^1 \geq 0, \dots, x_n^1 \geq 0;$
- the sets $x_1^2, x_2^2, \dots, x_n^2$ are pairwise disjoint and their union is equal to c^2.

The interpretation is obvious: x_i^1 represents the amount of the divisible commodity allocated to player i in the division $x = (x_1, x_2, \dots, x_n)$, and x_i^2 represents the set of indivisible objects assigned to player i. Note that the case in which c^2 is empty can be identified with the single commodity case of the previous model.

In what follows, we are concerned with still another model. Namely, we discuss the situations in which the cake may be heterogeneous. The cake is then represented by a nonempty set C, the set of feasible divisions is the set of ordered partitions of C into n subsets from some system $\mathcal{F}$ of subsets (feasible pieces) of C, and the players' preference relations are given on $\mathcal{F}$.

As an interesting illustration of this model, consider a situation where the time availability of some facility or service is beneficial only in intervals and where splitting up a given interval into subintervals with each player receiving several disjoint subintervals is not acceptable, see Berliant et al [4] for details. In this case, the set $\mathcal{F}$ is the system of all subintervals of a given non-degenerate bounded interval C of real numbers, and players' preference relations are defined on $\mathcal{F}$. Feasible divisions among n players are then ordered partitions of the given interval into n of its subintervals, and both efficiency and envy-freeness of feasible divisions are introduced by means of players' preferences on $\mathcal{F}$.

Stromquist [19] has shown that envy-free divisions exist in this model whenever players' preferences have numerical representations that are continuous in the subinterval endpoints (see also Woodall [24]). Berliant et al [4] have shown that each envy-free division is also efficient, provided that players' preferences are monotonous in the sense that if two different subintervals A and B are such that A is contained in B, then B is strictly preferred to A. Thus, for this one-dimensional model, the existence of envy-free efficient divisions has been established under very general assumptions. However, Gale [10] points out that it is not known whether a similar result holds for an analogous problem in two-dimensional case.

In the following sections, we are interested in problems where the cake divisibility, that is, the system of feasible pieces, satisfies the following natural requirements: The whole cake is a feasible piece, the complement of a feasible piece is also a feasible piece, and the union of two feasible pieces is also a feasible piece. In the mathematical terminology this means that the system $\mathcal{F}$ of feasible pieces is a finitely additive field (called also algebra) of subsets of the cake C.

2 Fairness

Before proceeding we need further notation and definitions. Recall that we denote the cake by C and the field of feasible pieces by $\mathcal{F}$; that is, $\mathcal{F}$ is a set of subsets of C such that $\mathcal{F}$ contains C and if $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$ and $A \setminus B \in \mathcal{F}$. An ordered n-tuple $D = (D_1, D_2, \ldots, D_n)$ of elements of field $\mathcal{F}$ is called a *feasible division* of C if the sets $D_1, D_2, \ldots, D_n$ are pairwise disjoint and their union is equal to C, that is, D is an n-partition of C. The informal meaning of D is that there are n players and each player i gets piece D_i of cake C.

Regarding preferences, we assume that the preferences of player i are represented by a transitive and complete binary relation $\succeq_i$ on field $\mathcal{F}$. The informal meaning of $A \succeq_i B$ is that player i finds A at least as good as B. The corresponding relation $\succ_i$ of strict preference and relation $\sim_i$ of indifference are defined on $\mathcal{F}$ as follows: $A \succ_i B$ if and only if $A \succeq_i B$ and *not*$(B \succeq_i A)$; and $A \sim_i B$ if and only if both $A \succeq_i B$ and $B \succeq_i A$.

A large portion of the literature is concerned with situations where each preference relation $\succeq_i$ is represented by a finitely additive measure μ_i on $\mathcal{F}$ as follows: $A \succeq_i B$ holds if and only if $\mu_i(A) \geq \mu_i(B)$. Often it is the case that all measures representing preference relations are probability measures.

Various notions of fairness have been introduced and studied within this framework. For example, Brams and Taylor [7], and Robertson and Webb [16] are concerned, among other topics, with the following notions of fairness.

A feasible division $D = (D_1, D_2, \ldots, D_n)$ among n players is said to be

- simple fair division if $\mu_i(D_i) \geq 1/n$ whenever $1 \leq i \leq n$,
- strong fair division if $\mu_i(D_i) > 1/n$ whenever $1 \leq i \leq n$,
- envy-free division if $\mu_i(D_i) \geq \mu_i(D_j)$ whenever $1 \leq i, j \leq n$
- super envy-free division if $\mu_i(D_j) < 1/n$ whenever $i \neq j$ and $1 \leq i, j \leq n$,
- exact division if $\mu_i(D_j) = 1/n$ whenever $1 \leq i, j \leq n$,

- dirty work fair division if $\mu_i(D_i) \leq 1/n$ whenever $1 \leq i \leq n$,
- dirty work envy-free division if $\mu_i(D_i) \leq \mu_i(D_j)$ whenever $1 \leq i, j \leq n$.

Moreover, further variations can be obtained by considering situations with players' entitlements. For example, in the case of simple fair division, we obtain the condition $\mu_i(D_i) \geq \alpha_i$ for each player i, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are numbers from the open unit interval $(0, 1)$ whose sum is equal to 1. In this case, we say that such a feasible division is α-fair where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$.

In what follows, we focus on the notion of envy-freeness; that is, on the feasible divisions $D = (D_1, D_2, \ldots, D_n)$ such that $D_i \succeq_i D_j$ for all i and j. In other words, we are interested in feasible divisions D such that no player wishes to receive (no player strictly prefers) a piece allocated to anyone else instead of the piece allocated to him.

3 Envy-Freeness: Existence

The existence of envy-free divisions cannot be guaranteed if no further restriction is imposed on the field of feasible pieces and players' preference relations. It turns out that most of the existence results in the literature assume that $\mathcal{F}$ is not only a field but also a σ-field of subsets of C; that is, a field which is closed under the formation of countably unions, and that the preference relation of each player i is represented by a non-atomic probability measure μ_i on $\mathcal{F}$. Here the non-atomicity of a measure μ means that if $\mu(A) > 0$ for some A in $\mathcal{F}$, then there exists a subset B of A such that B is also in $\mathcal{F}$ and $\mu(A) > \mu(B) > 0$. The existence of an envy-free division for the case of preferences given by non-atomic probability measures was established even before the notion of envy-freeness was introduced and investigated in the economic literature by Foley [9] and Varian [23], see Dubins and Spanier [8].

Obviously, some of these assumptions seriously limit possible applications. For example, some important preference relations have no numerical representations. Moreover, even if numerical representations exist, the requirement of additivity is too restrictive for many applications in economics because it corresponds to the assumption of constant marginal utility. It is therefore desirable to allow for representations that correspond to some form of decreasing marginal utility. This leads to considering capacities defined on $\mathcal{F}$ as follows.

A set function ν on $\mathcal{F}$ with values in $[0, 1]$ is called a capacity on $\mathcal{F}$ if

- $\nu(\emptyset) = 0$ and $\nu(C) = 1$,
- $\nu(A) \leq \nu(B)$ for all $A, B \in \mathcal{F}$ with $A \subset B$,
- $\nu(A_k) \downarrow 0$ for each monotone sequence $\{A_k\}$ with $A_k \in \mathcal{F}$ and $A_k \downarrow \emptyset$.

First we observe that a capacity ν is additive if and only if

$$\nu(A \cup B) = \nu(A) + \nu(B) - \nu(A \cap B)$$

for all (not necessarily disjoint) sets in $\mathcal{F}$. To reflect the wish of allowing for decreasing marginal utilities, it would be of interest to know whether envy-free divisions exist for preferences given by non-atomic capacities satisfying the following type of concavity requirement: for all $A, B \in \mathcal{F}$,

$$\nu(A \cup B) \leq \nu(A) + \nu(B) - \nu(A \cap B).$$

4 Envy-Freeness: Procedures

The existence of envy-free divisions is usually established by means of nonconstructive mathematical tools like fixed point theorems, the Borsuk-Ulam "sandwich" theorem, or Lyapunov's theorem on the ranges of vector measures. In spite of their depth, these results do not provide constructive means for finding the divisions in question.

Examination of possibilities for constructing envy-free divisions usually begins with recalling the widely known ancient rules "one cuts, the other chooses". Observe that these rules involve no advice to the cutter as to which pieces to create and no advice to the chooser as to which piece to choose. However, it is easy to understand that the cutter (player 1) can secure that he will not envy the chooser in the resulting division by dividing the cake into two parts that he considers to be equally valuable for him, and the chooser (player 2) can secure her no-envy by choosing the piece she prefers. Formally, if both players follow these strategies, then the resulting division (D_1, D_2) is envy-free because $D_1 \sim_1 D_2$ and $D_2 \succeq_2 D_1$, which implies both $not(D_2 \succ_1 D_1)$ and $not(D_1 \succ_2 D_2)$.

It should be pointed out that to ensure that the players can use the mentioned strategies, we have tacitly assumed that the field $\mathcal{F}$ and the players preference relations are such that

- there is a feasible division (A_1, A_2) with the property $A_1 \sim_1 A_2$,
- $B_1 \succeq_2 B_2$ or $B_2 \succeq_2 B_1$ for each feasible division (B_1, B_2) with $B_1 \sim_1 B_2$.

The latter is of course satisfied whenever the preference relation of the second player is complete, but it is easy to construct examples of transitive complete relations for which the former does not hold. It turns out that if $\mathcal{F}$ is a σ-field and $\succeq_1$ and $\succeq_2$ are representable by non-atomic probability measures on $\mathcal{F}$, then both conditions are satisfied. It is also useful to notice that the strategies described above guarantee the envy-freeness of the resulting division independently of player's information about preferences of the other player. However, from now on, we assume that players have no information about each other's preferences.

A natural question to ask is whether we can have something similar to the cut-and-choose rules for more than two players. By 'something similar', we mean rules within which each player has a strategy through which he can guarantee himself that he envies no other player in the resulting division, regardless of what strategies the other players follow. To define precisely what we mean by such a system of rules we borrow the idea of the extensive game form with perfect information from the theory of games, see Osborne and Rubinstein [14].

A *game-theoretic procedure* is defined by a set of players, a set of histories, and a player function, such that the following requirements are satisfied.

Players: The set of players is a nonempty finite set. We denote it by N and, without any loss of generality, assume that $N = \{1, 2, \ldots, n\}$ where n is a fixed positive integer.

Histories: The set of histories is a set H of sequences, some of which may be infinite. We call the components of a history *actions,* and assume that H has the following properties:

- The empty sequence belongs to H. We denote it by h_0 and call it the *initial history.*
- If a finite sequence $\langle a_1, a_2, \ldots, a_k\rangle$ belongs to H, then the sequence $\langle a_1, a_2, \ldots, a_l\rangle$ belongs to H for each integer l with $0 < l < k$.
- If an infinite sequence $\langle a_1, a_2, \ldots\rangle$ belongs to H, then, for each positive integer k, the finite sequence $\langle a_1, a_2, \ldots, a_k\rangle$ belongs to H.
- If an infinite sequence $h = \langle a_1, a_2, \ldots\rangle$ is such that, for each positive integer k, the finite sequence $\langle a_1, a_2, \ldots, a_k\rangle$ is in H, then h also belongs to H.

Player function: To introduce the notion of player function we need to distinguish between terminal and nonterminal histories. The set of *terminal* sequences is defined as follows: every infinite history is terminal, and a finite sequence $\langle a_1, a_2, \ldots, a_k\rangle$ from H is a terminal history if H does not contain a sequence $\langle b_1, b_2, \ldots, b_{k+1}\rangle$ with $b_1 = a_1, b_2 = a_2, \ldots, b_k = a_k$. The player function is a mapping of the set of nonterminal histories onto the set of players. We denote the player function by P and interpret $P(h)$ as the player who takes an action after nonterminal history h.

A game-theoretic procedure (N, H, P) will be called *finite* if the set H of its histories is finite. If the length of every history is finite, then we say that the procedure has a $finite\ horizon$. A strategy of a player in a procedure is a plan that specifies the action of the player for every history after which it is his turn to act. Put formally, a *strategy* of a player $i \in N$ in a procedure (N, H, P) is a mapping that assigns an action from the set $\{a : \langle a_1, a_2, \ldots, a_k, a\rangle \in H\}$ to each nonterminal history $h = \langle a_1, a_2, \ldots, a_k\rangle$ for which $P(h) = i$.

As an illustration, let us see how the cut-and-choose rules for two players can be considered as a game-theoretic procedure with a finite horizon. Obviously, we have $N = \{1, 2\}$. As the set of histories, we take the set consisting of

- the empty sequence h_0,
- the histories $\langle a_1\rangle$ where a_1 is the action of dividing the cake into two pieces belonging to the corresponding field $\mathcal{F}$,

- the histories $\langle a_1, a_2 \rangle$ where a_1 is the action of dividing and a_2 is the action of choosing one of the pieces created by action a_1.

Clearly, the set of terminal histories are all histories of the length two, and the player function can be any mapping of the set of nonterminal histories onto $\{1, 2\}$ with the property that $P(h_0) \neq P(h)$ for each $h \neq h_0$. For definiteness, we set $P(h_0) = 1$, that is, player 1 is the cutter and player 2 is the chooser. It can easily be seen that every pair of players' strategies determines a terminal history, and every terminal history determines a feasible division (D_1, D_2) of the cake where D_2 is the piece chosen by player 2. It depends on the underlying field $\mathcal{F}$ whether this procedure is finite or infinite.

Usually it is not necessary to describe a particular procedure in this form explicitly, because it can be well understood from a less formal description. However, it may be important to have a detailed description when we deal with problems for which it is not clear whether a procedure with some prescribed property exists.

For the case of three players a game-theoretic procedure for constructing an envy-free division has been established independently by Selfridge and Conway, see Woodal [24] or Brams and Taylor [6].

The Selfridge-Conway Procedure

Step 1: Player 1 cuts the cake into three pieces.

Step 2: Player 2 is given the choice of either passing (that is, doing nothing) or trimming a piece from one of the three pieces. The trimmings, if any, are set aside.

Step 3: Players 3, 2, and 1, in that order, choose a piece from among available pieces observing the following requirement: If Player 2 did not pass in step 2 then he must choose the piece he trimmed, if that piece was not chosen by Player 3 in step 2.

Step 4: If Player 2 passed at step 2, we are done. If Player 2 did not pass at step 2, then either Player 2 or Player 3 chose the untrimmed piece at step 3. Let us call this player the "cutter" and the other the "non-cutter". The cutter now divides the trimmings into three pieces.

Step 5: The three pieces into which the trimmings is divided are now chosen by the players in the following order: the non-cutter first, Player 1 second, the cutter third. Then we are done.

Again the rules do not offer any recommendation to the players as to which strategies to follow in order to be non-envious in the resulting division. To show that each player can follow the rules in such a manner that he is non-envious in the resulting division, regardless of which strategies the other players follow, we first specify such strategies for each player.

Player 1: In step 1, the player should cut the cake into three pieces that he considers to be equally good for him; in step 5, he should choose the piece he prefers.

Player 2: In step 2, the player should trim the most preferred piece to create a tie for the best, provided such a tie does not exist after step 1; otherwise he should pass. In step 3, the player should choose the piece he prefers, provided player 3 chose the trimmed piece; otherwise he has no choice, he has to take the trimmed piece. If the player becomes the cutter in step 4, then he should cut the trimmings into three pieces that he considers to be equally good; otherwise he should choose the best piece in step 5.

Player 3: In step 3, the player should choose the best piece. If he becomes the cutter in step 4, he should cut the trimmings into three pieces that he considers to be equally good; otherwise he should choose the best piece in step 5.

To verify that such strategies produce an envy-free division, we observe that, after the completion of the third step, either the whole cake or its part without trimmings is divided in envy-free way. Indeed, Player 3 envies no one because he chooses first, Player 2 can secure the best piece for himself by trimming, and the rules of step 2 guarantee that Player 1 does not receive the trimmed piece. It remains to examine the division of trimmings. The non-cutter envies no one because he is choosing first. The cutter is non-envious because he created equal pieces. Player 1 does not envy the cutter because he chooses before the cutter does, and he does not envy the non-cutter because the non-cutter gets the trimmed piece together with some part of the trimmings, which is certainly no better than the untrimmed piece of Player 1.

To make sure that the players can follow the recommended strategies, we have tacitly assumed that the underlying field $\mathcal{F}$ and preference relations satisfy the following conditions.

- There is a feasible division (A_1, A_2, A_3) of C such that $A_1 \sim_1 A_2 \sim_1 A_3$.
- For each A and B from $\mathcal{F}$ with $A \succ_2 B$, there is a subset X of A belonging to $\mathcal{F}$ such that $X \sim_2 B$.
- For every player $i \in \{2, 3\}$ and each $A \in \mathcal{F}$, there is a partition of A into three subsets B_1, B_2, B_3 belonging to $\mathcal{F}$ and such that $B_1 \sim_i B_2 \sim_i B_3$.
- For each player i and $A, B, Y, Z \in \mathcal{F}$,

  $(Y \succeq_i Z, A \succeq_i B, Y \cap A = \emptyset)$ implies
  $Y \cup A \succeq_i Z \cup B$, and

  $(Y \succ_i Z, A \succ_i B, Y \cap A = \emptyset)$ implies
  $Y \cup A \succ_i Z \cup B$.

Notice that both the cut-and-choose rules and the rules of the Selfridge-Conway procedure have the following nice property: the number of necessary cuts in the worst case (one cut for the former, five cuts for the latter) does not depend on players' preferences. It seems that Selfridge-Conway procedure has not been extended to a procedure with the same nice property for four or more players.

Brams and Taylor [6] show how to construct in a finite number of steps an envy-free division for an arbitrary finite number of players under the assumption that each player's preferences are given by a finitely additive measure. However, the number of necessary cuts in their procedure depends on players' preferences and can be arbitrarily great even in the case of four players. This leads to the following open question.

> *Is there a game-theoretic procedure for four players in the case of preferences given by finitely additive measures such that*
>
> - *it has a finite horizon,*
> - *the number of necessary cuts can be bounded by a number that is independent of players' preferences,*
> - *it is envy-free in the sense that each player can follow the rules in such a manner that he or she is non-envious in the resulting division, regardless of which strategies the other players follow?*

References

[1] A. Alkan, G. Demange, and D. Gale, *Fair allocations of indivisible goods and criteria of justice,* Econometrica 59 (1991), 1023–1039.

[2] J. B. Barbanel, *Game-Theoretic algorithms for fair and strongly fair cake division with entitlements,* Colloquium Mathematicum, LXIX (1995), 59–73.

[3] J. B. Barbanel and A. D. Taylor, *Preference relations and measures in the context of fair division,* Proceedings of American Mathematical Society 123 (1995), 2061–2070.

[4] M. Berliant, W. Thomson, and K. Dunz, *On the fair division of a heterogeneous commodity,* Journal of Mathematical Economics 21 (1992), 201–216.

[5] S.J. Brams and D.M. Kilgour. *Competitive Fair Division.* Journal of Political Economy 109 (2001), 418–443.

[6] S. J. Brams and A. D. Taylor, *An envy-free cake division protocol,* Amer. Math. Monthly 102 (1995), 9–18.

[7] S. J. Brams and A. D. Taylor, *Fair Division: From Cake-Cutting to Dispute Resolution,* Cambridge University Press, Cambridge, UK, 1996.

[8] L. E. Dubins and E. H. Spanier, *How to cut a cake fairly,* Amer. Math. Monthly 68 (1961), 1–17.

[9] D. C. Foley, *Resource allocation in the public sector,* Yale Economic Essays 7 (1967), 73–76.

[10] D. Gale, *Mathematical entertainments,* Math. Intelligencer 15 (1963), 48–52.

[11] F. Klijn, *An algorithm for envy-free allocations in an economy with indivisible objects and money,* Social Choice and Welfare 17 (2000), 201–215.

[12] H. W. Kuhn, *On games of fair division,* in *Essays in Mathematical Economics,* ed. by M. Shubik, Princeton University Press, Princeton, NJ, 1967, 29–37.

[13] F. Macheroni and M. Marinacci, *How to cut a pizza fairly: Fair division with decreasing marginal evaluations,* Social Choice and Welfare 20 (2003), 457–465.

[14] M. J. Osborne and A. Rubinstein, *A Course in Game Theory,* The MIT Press, Cambridge, Massachusetts, 1994

[15] E. A. Pazner and D. Schmeidler, *A difficulty in the concept of fairness,* Review of Economic Studies 41 (1974), 441–443.

[16] J. M. Robertson and W. A. Webb, *Cake Cutting Algorithms: Be Fair If You Can*, A K Peters, Ltd., Natick, Mass., 1998.

[17] D. Schmeidler and K. Vind, *Fair net trades,* Econometrica 40 (1972), 637–642.

[18] H. Steinhaus, *The problem of fair division*, Econometrica 16 (1948), 101–104.

[19] W. Stromquist, *How to cut a cake fairly*, Amer. Math. Monthly 87 (1980), 640–644.

[20] F.E. Su, *Rental Harmony: Sperner's Lemma in Fair Division*, Amer. Math. Monthly 106 (1999), 930–942.

[21] S.C. Sung and M. Vlach, *Competitive Envy-Free Division*, Social Choice and Welfare 23 (2004), 103–111.

[22] L.-G. Svenson, *Large indivisibles: an analysis with respect to price equilibrium and fairness,* Econometrica 51 (1983), 939–954.

[23] H. Varian, *Equity, envy, and efficiency,* Journal of Economic Theory 9 (1974), 63–91.

[24] D. R. Woodal, *Dividing a cake fairly,* Journal of Mathematical Analysis and Applications 78 (1980), 233–247.

Markov Chain Monte Carlo Methods in Computational Statistics and Econometrics

Petr Volf
ÚTIA AV ČR
Pod vodárenskou věží 4, Praha 8
volf@utia.cas.cz

Abstract

The present contribution deals with the MCMC methods, namely with the Gibbs algorithm, the Metropolis–Hastings algorithm and variants used for solution of optimization problems, namely the simulated annealing. The objective is to describe the schemes of the algorithms, to recall their theoretical foundation, and to show their use both in Bayes data analysis and in randomized optimization problem.

Key words

Markov chain Monte Carlo (MCMC), random search, optimization, computational statistics, Bayes statistics.
JEL classification: C15, C61

1 Introduction

Markov chain Monte Carlo methods are now quite frequently used both in the field of statistical data analysis and in the problems of search for optimal states of systems. MCMC procedures (similarly as other methods of sequential random search) generate a Markov chain leading to the solutions. They are constructed in such a way that the probability distribution of the chain converges to a desired distribution. This is guaranteed by the choice of the transition probabilities and by the theory of convergence of Markov chain to its invariant distribution. When MCMC is connected with the simulated annealing, we obtain the convergence to the maximal probable configurations of analyzed system, The method is convenient also for the cases of optimization with a constraint. That is why the MCMC algorithms should become standard tools in computational statistics and econometrics as well.

The MCMC sampling was first introduced by Metropolis and his collaborators in 1953 as the method for the efficient simulation of the energy levels of atoms in crystalline structure and was subsequently modified and generalized by Hastings (1970) and, lately, by many other authors. Today, MCMC is used especially in multidimensional problems with complicated dependency structure and also in image analysis and reconstruction.

2 Basic types of MCMC procedures

The idea of MCMC is rather simple. Let us consider a probability distribution $p(\boldsymbol{x}), \boldsymbol{x} \in E \subseteq R^p$, the target distribution of the chain. If $p(\boldsymbol{x})$ is rather complex, we cannot sample from it directly. MCMC offers an indirect method for obtaining samples from it, namely MCMC algorithms construct an aperiodic and irreducible Markov chain with state space E and with stationary (and limit) distribution $p(\boldsymbol{x})$.

2.1 Gibbs algorithm

Let $p_j(x_j|\boldsymbol{x}_{(-j)})$ be the conditional distributions $p(\boldsymbol{x})$, where $\boldsymbol{x}_{(-j)} = x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_p$. The algorithm starts with a chosen initial value $\boldsymbol{x}^{(0)}$ and subsequently updates one component of $\boldsymbol{x}$ after another. Namely, let $\boldsymbol{x}^{(m)}$ be the m-th member of the chain, the j-th component of the $m+1$-st member of the chain is obtained by sampling from

$$p_j(x_j^{(m+1)}|x_1^{(m+1)}, x_2^{(m+1)}, ..., x_{j-1}^{(m+1)}, x_{j+1}^{(m)}, ..., x_p^{(m)}).$$

The result is a Markov chain $\boldsymbol{x}^{(0)}, \boldsymbol{x}^{(1)}, ...,$ with invariant distribution $p(\boldsymbol{x})$. Moreover, the distribution of $\boldsymbol{x}^{(m)}$ converges to $p(\boldsymbol{x})$, under quite mild conditions (formulated for instance in [3]). The limitation of the method is that we need to know conditional distributions p_j, or we have to combine the Gibbs algorithm with another sampling scheme, e.g. the sampling–rejection method (this was used also in [5]).

2.2 Metropolis–Hastings algorithm

Metropolis–Hastings algorithm yields also the Markov chain of $\boldsymbol{x}$-s with chosen invariant distribution $p(\boldsymbol{x})$. Let $\boldsymbol{x}^m$ be the actual state of the chain. The candidate $\boldsymbol{x}^*$ for the next , $m+1$-st member, is drawn from the distribution (arbitrary for the moment) with density $q(\boldsymbol{x}|\boldsymbol{x}^{(m)})$ (i.e. it can, but need not, depend on the last state $\boldsymbol{x}^{(m)}$). The algorithm uses the following acceptance-rejection rule: Set

$$\alpha(\boldsymbol{x}^{(m)}, \boldsymbol{x}^*) = \begin{cases} \min\left\{\dfrac{p(\boldsymbol{x}^*)\, q(\boldsymbol{x}^{(m)}|\boldsymbol{x}^*)}{p(\boldsymbol{x}^{(m)})\, q(\boldsymbol{x}^*|\boldsymbol{x}^{(m)})}, 1\right\} & \text{if } p(\boldsymbol{x}^{(m)})\, q(\boldsymbol{x}^*|\boldsymbol{x}^{(m)}) > 0 \\ 1 & \text{if } p(\boldsymbol{x}^{(m)})\, q(\boldsymbol{x}^*|\boldsymbol{x}^{(m)}) = 0. \end{cases}$$

and accept new $\boldsymbol{x}^*$ (i.e. set $\boldsymbol{x}^{(m+1)} = \boldsymbol{x}^*$) with probability $\alpha(\boldsymbol{x}^{(m)}, \boldsymbol{x}^*)$. If $\boldsymbol{x}^*$ is not accepted, $\boldsymbol{x}^{(m+1)} = \boldsymbol{x}^{(m)}$.

Notice that we need not to know normalizing constant of the target distribution. Moreover, we may innovate one component of vector $\boldsymbol{x}$ after another, using (non-complete) conditional distributions. Notice also, that in the case when the proposal is the conditional target distribution, then MH algorithm becomes the Gibbs sampler (with probability of acceptance 1).

Symmetrical proposal distribution. In the case that the proposal distribution is symmetrical (we then also call it reversible), i.e. $q(\boldsymbol{x}^{(m)}|\boldsymbol{x}^*) = q(\boldsymbol{x}^*|\boldsymbol{x}^{(m)})$, the acceptance function reduces to

$$\alpha(\boldsymbol{x}^{(m)}, \boldsymbol{x}^*) = \min\{1, \frac{p(\boldsymbol{x}^*)}{p(\boldsymbol{x}^{(m)})}\}$$

This special case is the original **Metropolis algorithm**.

3 Optimization and simulated annealing

Simulated annealing offers a modification of the procedures mentioned above, yielding a method of random search for (global) optimum. Let us imagine a function $H(\boldsymbol{x}) : E \to R$, the objective is to find the minimum of H on E. In the present context this problem is re-formulated as the problem of looking for the maximum of the probability density $p(\boldsymbol{x}) \sim \exp(-H(\boldsymbol{x}))$. If we search for it randomly by utilizing Metropolis algorithm with the acceptance probability

$$\alpha(\boldsymbol{x}, \boldsymbol{x}^*) = \min\{\frac{\exp(-(H(\boldsymbol{x}^*))}{\exp(-(H(\boldsymbol{x}))}, 1\}$$

then the distribution of obtained chain converges to the target distribution $p(\boldsymbol{x}) = C.\exp(-H(\boldsymbol{x}))$.

If we, instead, use the acceptance probability $\alpha(\boldsymbol{x}, \boldsymbol{x}^*)^{\frac{1}{T(s)}}$, where s is the number of iteration, $T(s) > 0$, $lim_{s\to\infty} T(s) = 0$, then for $s \to \infty$ (and for a conveniently chosen $T(s)$, for example $T(s) = K/\log(1+s)$) we get the convergence of distribution of $\boldsymbol{x}^{(s)}$ to the uniform distribution on the points maximizing $p(\boldsymbol{x})$, i.e. points fulfilling $\{\ \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} H(\boldsymbol{x})\ \}$. Naturally, if there is only one optimal point, the convergence to it follows.

4 Conditions for convergence of Markov chain

Let us recall here certain facts concerning the limit behavior of Markov chains and the consequences to MCMC procedures. We shall consider the case of an uncountable (in general) state space $E \subset R$ (measurable, naturally). Let $X^{(0)}, X^{(1)}, X^{(2)}, \ldots$ be a Markov chain. Let us denote by $K(x, y)$ its transition kernel describing the transition from x to y. It may be substochastic (and in the MCMC algorithms it very often is), i.e.

$$\int_E K(x, y)dy \leq 1, \quad R(x) = 1 - \int_E K(x, y)dy \geq 0,$$

where $R(x)$ denotes the point probability that the chain remains at x. Further, denote

$$K^{(t)}(x, y) = \int_E K^{(t-1)}(x, u)K(u, y)du + K^{(t-1)}(x, y)R(y) + [R(x)]^{t-1}K(x, y)$$

the kernel characterizing the t-step transitions (again, it may be substochastic, there may be a positive probability that the state of the chain remains unchanged for t subsequent steps). Finally, let $p(x) > 0$ on E be the density of invariant distribution of the chain, i.e. satisfying

$$\int_A p(x)dx = \int_E P(X^{(1)} \in A|X^{(0)} = u)p(u)du,$$

for all sets (measurable) $A \subset E$. Here,

$$P(X^{(1)} \in A|X^{(0)} = u) = \int_A K(u,y)dy + R(u)I[u \in A].$$

The kernel is called **irreducible** if for all $x \in E$, for all $A \subset E$ there is $t > 0$ such that

$$P(X^{(t)} \in A|X^{(0)} = u) > 0.$$

The kernel is called **aperiodic** if, for no $n \geq 2$, there is a partition $\cup_{j=1}^{n} B_j = E$, such that chain jumps regularly (in a periodical manner), with probability 1, between sets B_j. These properties imply the convergence of the chain. Namely, the following holds (compare e.g. [3]):

Proposition 1.

If p is the density of invariant distribution of the chain $X^{(0)}, X^{(1)}, \ldots$ and if corresponding transition kernel K is irreducible and aperiodic, then for all $x \in E$, for $t \to \infty$ the following holds:

1. The distribution of the state of the chain at time t converges to p, i.e.

$$\int_E |K^{(t)}(x,y) - p(y)|dy \to 0.$$

2. For a real and p-integrable function f,

$$\frac{1}{t}\sum_{i=1}^{t} f(X^{(i)}) \to \int_E f(x)p(x)dx \quad a.s.$$

From the result 1. it immediately follows that the distribution of $X^{(t)}$ tends to the invariant distribution (the probability of trajectories without any transition is $R(x)^t \to 0$), the result 2. (ergodicity) confirms that the average of the chain members is a reasonable estimate of the mean of the invariant distribution and that it holds also for a transformation of the chain.

4.1 Conditions for convergence of Metropolis–Hastings algorithm

The MH algorithm requires the specification of two probability functions. The first is a conditional probability $q(y|x)$ for generating the candidates for the next state of the chain, the second is the acceptance probability $\alpha(x,y)$ which has predescribed form in order to fulfill the following condition of detailed balance.

Proposition 2.

Any transition kernel for which the conditions of **detailed balance**, namely

$$p(x)K(x,y) = p(y)K(y,x)$$

holds, has stationary distribution p.

It can be shown rather easily: Let the kernel K fulfills the property of detailed balance, then for p and for each $A \subset E$ it holds that

$$\begin{aligned}
&\int_E P\left(X^{(1)} \in A|X^{(0)} = u\right) p(u)\,\mathrm{d}u = \\
&= \int_E \left\{\int_A K(u,y)\,\mathrm{d}y + R(u)\,\mathbf{1}[z \in A]\right\} p(z)\,\mathrm{d}z = \\
&= \int_A \int_E K(y,u)\,p(y)\,\mathrm{d}z\mathrm{d}y + \int_E R(u)\,\mathbf{1}[u \in A]\,p(u)\,\mathrm{d}u = \\
&= \int_A (1 - R(y))\,p(y)\,\mathrm{d}y + \int_A R(u)\,p(u)\,\mathrm{d}z = \int_A p(y)\,\mathrm{d}y.
\end{aligned}$$

Although the detailed balance is more than is needed for the ergodicity and convergence to target distribution, it is a condition convenient especially for practical construction of samplers, because it can be easily checked.

Now, let $\alpha(x,y)$ be acceptance probability for MH algorithm, i.e.

$$\alpha(x,y) = \begin{cases} \min\left\{\dfrac{p(y)\,q(x|y)}{p(x)\,q(y|x)},\,1\right\} & \text{if } p(x)\,q(y|x) > 0 \\ 1 & \text{if } p(x)\,q(y|x) = 0. \end{cases}$$

The transition kernel of resulting Markov chain is given by the product $K(x,y) = q(y|x)\alpha(x,y)$ Then, it is seen that

$$p(x)K(x,y) = p(y)K(y,x),$$

i.e. that the detailed balance holds. Convergence properties of the chain follow from the corresponding properties of the generating probability (again, see [3]):

Proposition 3.

i) If q is aperiodic, then K is aperiodic,

ii) If q is irreducible (on E), and $q(y|x) = 0$ if and only if $q(x|y) = 0$, then K is irreducible.

Thus, if the proposal density q is chosen to be aperiodic a irreducible, the distribution of the chain converges to distribution $p(x)$ – the invariant distribution.

From the convergence of the chain it follows that if we generate a sufficiently long chain and cut out its initial part, the rest of the chain may be regarded as a sample generated approximately by desired target distribution. Further, it is due the ergodicity that the average of this sample can serve as a reasonable 'point estimate' of the mean value of target distribution. Naturally, the members of the chain are correlated. It is difficult to tell how long should be the chain in order to reach a certain proximity to target distribution, and how correlated are the members of the chain. There exist some practical approaches how to reduce the autocorrelation. For instance, we can select randomly only a subsequence of the resulting chain. We can also generate several independent parallel chains and take sub-chains from each of them. In such a way we obtain, in a reasonable time, a sufficiently rich final sample of less correlated items.

5 Application of MCMC methods to the Bayesian data analysis

Let us imagine that we observe the data $\boldsymbol{y}$, the realization of random variables $\boldsymbol{Y} = Y_1, ..., Y_n$, generated by a probability distribution with a density function $f(\boldsymbol{y}|\boldsymbol{\theta})$. Here $\boldsymbol{\theta} \in \Theta \subset R^p$ is an unknown parameter. In the Bayes scheme, the uncertainty of values of $\boldsymbol{\theta}$ is described by a prior distribution $g_0(\boldsymbol{\theta})$. Bayes rule then yields the posterior distribution of parameter given the observed data:

$$g(\boldsymbol{\theta}|\boldsymbol{y}) = C(\boldsymbol{y})f(\boldsymbol{y}|\boldsymbol{\theta})g_0(\boldsymbol{\theta}).$$

The objective is now to obtain a sample representing the posterior distribution and, eventually, to obtain the value of parameter maximizing it. Hence, in the context of Bayesian estimation problem, the target distribution of the MCMC methods is the posterior distribution of model parameters given the data.

If we are able to sample from conditional posterior distribution of one component θ_j, we can utilize the Gibbs sampler, In other case, the MH algorithm is available. It proceeds in the following way: In state $\boldsymbol{\theta}^{(m)}$, it first proposes a new value $\boldsymbol{\theta}^*$, drawing it from a distribution $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(m)})$. Then, $\boldsymbol{\theta}^{(m+1)}$ is set to $\boldsymbol{\theta}^*$ with probability

$$\min\left\{1,\, \frac{g(\boldsymbol{\theta}^*|\boldsymbol{y})\,q(\boldsymbol{\theta}^{(m)}|\boldsymbol{\theta}^*)}{g(\boldsymbol{\theta}^{(m)}|\boldsymbol{y})\,q(\boldsymbol{\theta}^*|\boldsymbol{\theta})}\right\},$$

otherwise $\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)}$. If the proposals generate an irreducible and aperiodic sequence, the convergence of distribution of $\boldsymbol{\theta}^{(m)}$ to the distribution given by $g(\boldsymbol{\theta}|\boldsymbol{y})$ is guaranteed. In the special case, when the prior distribution is used as a proposal distribution, it is easily seen from the Bayes rule that the acceptance probability reduces to $\min\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\theta}^*)}{f(\boldsymbol{y}|\boldsymbol{\theta}^{(m)})}\}$.

MH algorithm with change of dimension. In many cases of Bayesian parameter estimation, the parameter dimension is not known, so that it is also a part of unknown parameter and is changed during the process of sampling. The examples are for instance the spline regression model, where number of splines is unknown, or the problem of clustering with unknown number of clusters. It is then necessary to consider several different types of transitions. The problem is that now the transition probabilities may be defined in different spaces. Such a situation is discussed and cleared up for instance in [2], [4]. It is shown that two mutually reverse 'proposal' steps (here described by $q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ and $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$) have to be defined with respect to a symmetrical joint measure. Such a measure is actually the product of measures corresponding to both steps, restricted to subspaces of items which can be reached each from the other. The paper [4] offers also some practical recommendations.

6 Examples

One example of MCMC application can be found in [5]. The problem of modeling the time series of number of unemployed has been solved with the aid of the Gibbs sampler combined with the sampling-rejection method. Reisnerová (2006) in her contribution to present MME'06 utilizes a similar approach, moreover, the change point of the series trend is identified with the help of MH algorithm. Bothe cases show the use of MCMC in Bayes statistical analysis. The next simple example will demonstrate, on a case of transportation problem, the use of MH algorithm in optimization.

Let the inputs (sources) $n1(j), j = 1, ..., J$, and outputs (demands) $n2(k), k = 1, ..., K,$, be given, where $\sum_j n1(j) = \sum_k n2(k)$. Further, let $c(j,k)$ be costs of transport, the task is to find amounts $n(j,k)$ to fulfill the demands and minimize $C = \sum_j \sum_k n(j,k)c(j,k)$.

A quite simple random search strategy can be composed from the following steps:

1. Start from an arbitrary initial solution fulfilling the basal requirement $\sum_j n(j,k) = n2(k)$, $\sum_k n(j,k) = n1(j)$ for each j, k.

2. Select randomly indexes j_1, j_2 from inputs, k_1, k_2 from outputs. Let $m = min\{n(j_1,k_1), n(j_2,k_2)\}$.

3. Sample uniformly d from $1, .., m$. Propose the following changes:

   $$n^*(j_1,k_1) = n(j_1,k_1) - d,\ n^*(j_2,k_2) = n(j_2,k_2) - d,\ n^*(j_1,k_2) = n(j_1,k_2) + d,\ n^*(j_2,k_1) = n(j_2,k_1) + d.$$

4. This new configuration of transferred amounts will be accepted with the probability

   $$min\{1, exp(C0 - C1) \cdot \frac{q_{1,0}}{q_{0,1}}\},$$

   where C_0, C_1 are old and new total costs and q are probabilities of selection of certain change d of step 3, namely $q_{0,1} = 1/m$, $q_{1,0} = 1/min\{n^*(j_1,k_2), n^*(j_2,k_1)\}$.

5. Generate a long chain (several thousands configurations), last part of them (last half, say) take as representing the distribution of optimal solutions.

In the simulated annealing method the proposals may be the same, the acceptance probability is just

$$min\{1, exp(\frac{C0 - C1}{T(s)})\}$$

with $T(s)$ tending slowly to zero. Then the last configurations should converge to optimal solution.

From such a scheme, the way of solution of different modifications of the problem can be derived. For instance, the case with random inputs, or with simultaneously optimized profit of certain provider, again may be solved via the random search method.

Numerical example. We present here briefly the results of one standard transportation problem; it had 10 inputs and 20 outputs, we used both schemes described above (without and with simulated annealing), in both cases 5000 iterations. The example (and Figure 1) illustrates well the difference between the two cases, showing the convergence of S.A. method.

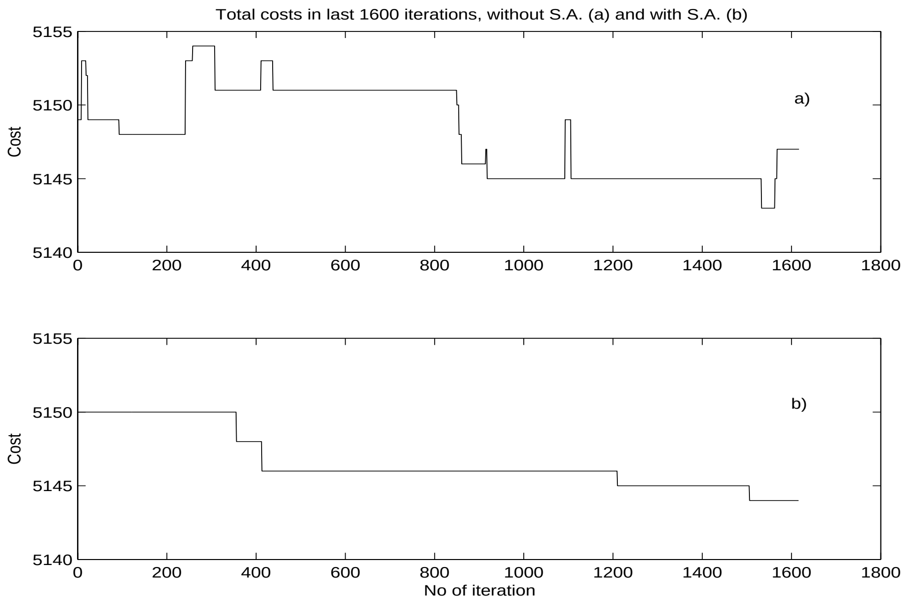


Figure 1: Development of total costs in last 1600 iterations, without simulated annealing (above) and with it (below)

7 Conclusion

In the paper we have shown that the MCMC methods help to solve both the problem of optimization (with constraints) and the problem of optimal model selection in the data analysis tasks. That is why the MCMC procedures are now highly popular and there exists a large number of applications. Therefore, it is very important to know more about the properties of these methods and about the conditions of their reliable performance. It was one of the aims of the present paper, to recall the background of MCMC methods and their relationship to the theory of probability and to the theory of Markov chains.

Acknowledgement: The research is supported by the GA ČR grant No 402/04/1294 .

References

[1] Gamerman D. (1997). Markov Chain Monte Carlo. Chapman and Hall.

[2] Green P.J. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika* 82, 711-732.

[3] Roberts G.O. and Smith A.F.M. (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms, *Stoch. Processes and Applic.* 49, 207-216.

[4] Richardson S., Green P. (1997). On Bayesian analysis of mixtures via an unknown number of components. JRSS 99, 731-792.

[5] Volf P. Bayes analysis of time series with covariates. In: Proceedings of MME'2005. (Skalská H. ed.). Gaudeamus, Hradec Králové 2005, 421-426.

Consumer Price Index and its Biases[1]

Renáta Wawerková, Dalibor Moravanský
Faculty of Economics and Administration
Masaryk University
Lipová 41a, 602 00 Brno
wawerkova@mail.muni.cz, dalibor@econ.muni.cz

Abstract

This paper offers some perspectives on the differences between the Consumer Price Index that statistical agencies produce and a theoretical cost of living index. It investigates the magnitude the differences and proposes elimination of these biases.

Keywords

Consumer Price Index, Cost of Living Index, Substitution Bias, Outlet Substitution Bias, Quality Change Bias, New Goods Bias, Reduction of Biases

JEL: C43, C82, E31

1 Methods of Measuring Inflation Impact on Consumer

The statistical agencies use the consumer price index based on the Laspeyres price index to measure the general price level. The Laspeyres price index reflects the cost of purchasing a fixed basket of goods representing the base period and the cost of buying the same basket in the present. For the construction of the Laspeyres price index it is sufficient to know the quantities in the base period and prices in the base and current period. Vector of current prices, included in the Laspeyres price index, is investigated and revised every month. Taken together, the statistical agencies are able to publicize the price level periodically. Nevertheless, using the base period's weights is not actual and its non-updating seems to be a problem.

Alternative approach to express how the inflation is understood by the economic subjects is the true cost of living. The true cost of living is based on the optimal consumer behavior, minimalizing its consumption cost. This approach is associated with A. A. Konüs, who first formulated the true cost of living index. The Konüs true cost of living index for a single consumer (household) is defined as the ratio of the minimum costs of achieving a certain reference utility level in a present period and in a base period. The prices in the present period can change and the prices in the base period are given (fixed).[2] This theoretic problem of the true cost of living can be approximated by the Fisher ideal price index, which is the geometric average of the Laspeyres and the Paasche price indices. The Paasche price index reflects the cost of purchasing a fixed basket of goods representing the present and then the cost of buying the same basket in the past. The main disadvantage of using the Paasche price index is finding of actual quantities, their measuring takes a long time and in the time of calculation of the index they are already not actual.

The Laspeyres price index tends to overstate the rise in the cost of living and is the maximum estimation of the price level, but in special cases it does not hold[3]. Conversely, the Paasche price

[1]This paper has been worked as a part of research activities at the grant project of GA ČR No. 402-05-2172.

[2]An appropriate generalization of the Konüs cost of living concept to the case of many households is Pollak's social cost of living index.

[3]The planned economies were shortage economies where consumer demand often went unsatisfied. The economic transition may result in increased purchases of goods whose prices are rising as a liberalization result that call forth additional supply to eliminate shortages. In these circumstances, price indices that use the fixed base Laspeyres index may understate true cost of living increases (see [7]).

index tends to understate the rise in the cost of living. The geometric average of the Laspeyres and Paasche price indices, the Fisher price index, reasonably approximates the cost of living assuming that the underlying utility function is homothetic (see [3], p. 48).

2 Biases in the Consumer Price Index

The statistical offices including the Czech Statistical Office use the Laspeyres price index to approximate the true price index. The fixed-base Laspeyres index does not allow substitution between goods, thus the higher product price (other prices are fixed) will increase the cost of living. Furthermore, it does not take into account substitution between outlets, quality changes of products and introduction of new goods.

The imperfection in the measurement of the inflation generates the biases between the measurable Laspeyres index and the theoretical cost of living and will be further investigated. We identify four important types of biases: the elementary substitution bias, the outlet substitution bias, the quality change bias and the new goods bias.

2.1 Elementary Substitution Bias

The fixed-base Laspeyres index does not allow substitution between goods, which can ameliorate the effect of price increases on household utility. When the price of good increases (other prices are fixed), the fixed-base index measures the higher cost of living than in reality because of not allowing substitution between goods.

Households can purchase a product at a variety of prices and this price heterogeneity has to be summarized as a single price so that it can be inserted into an index number formula. The single price is denoted as an „outlet unit value“, defined as the total value of the product sold during the time period divided by the corresponding quantity sold at an outlet (see [4], pp. 20-24). It concerns the aggregation at the elementary level. The aggregation at the commodity level comprises the product prices in the consumer basket and the aggregation is carried out by a weighted arithmetic mean of product prices, where weights are household expenditures for the given commodity. Substitution between goods is not allowed at the both aggregation levels and computations made at different levels will yield the same result as if made at a higher level directly (see [7], p. 5).

The Czech Statistical Office collects the prices of the goods and services in the second half of every month and the prices can be structured in the summary index groups[4]. Then they are aggregated into the index of the entire set (see [1]).

This substitution bias between goods, by which we suppose substitution at the elementary level, is denoted as the elementary substitution bias.

Elementary substitution bias was defined by Diewert (1998) as the difference between the fixed-base Laspeyres index (denoted as P_L) and the corresponding Fisher index (denoted as P_F), which approximates true cost of living index. Elementary substitution bias B_E can be defined as:

$$B_E \equiv P_L - P_F \cong \frac{1}{2}(1+i)var(\epsilon), \tag{1}$$

$$P_F = \sqrt{P_L P_P},\ P_L = \frac{\sum_{i=1}^{N} \frac{p_i^1}{p_i^0}\frac{p_i^0}{q_i^0}}{\sum_{i=1}^{N} p_i^0 q_i^0},\ P_P = \frac{\sum_{i=1}^{N} \frac{p_i^1}{p_i^0}\frac{p_i^1}{q_i^1}}{\sum_{i=1}^{N} p_i^1 q_i^1}, \tag{2}$$

where i is the inflation rate measured by the Laspeyres price index and $var(\epsilon)$ is the variance of the inflation adjusted percentage changes in prices among the goods. Elementary substitution

[4]The summary index structures in the 12 elementary groups in accordance with the international classification of the final consumption of households COICOP.

bias B_E will be approximately equal to one-half the Laspeyres price index P_L times the variance of the inflation-adjusted percentage changes in prices among the goods.

Elementary substitution bias in the consumer price index might be about 1,04 percentage points for 1996 in the Czech Republic and fall down in further years (except the year 1998 when the inflation was 10,6 percent) (see table below). The average elementary substitution bias during the years 1996 - 2000 was 0,72 percentage points and prices rose by 5,8 percent. Hanousek and Filer showed that elementary substitution bias was 1,24 percentage points for 1996 and 0,77 percentage points for 1997 (see [5], p. 6).

Table 1: Elementary substitution bias in the Czech Republic

Period	1996	1997	1998	1999	2000
Inflation (%)	8,7	8,5	10,6	2,1	4,0
$Var(\epsilon)$ (%)	1,91	1,86	2,37	0,73	0,71
B_E (p.p.)	1,04	1,01	1,31	0,37	0,37

Source: own calculation of CSO data

In the case of commodity substitution bias, this calculation can be repeated on a higher level, except that, in this case, the aggregation is happening across prices of different commodities instead of prices for the same commodity. This elementary and commodity substitution biases will continue as long as there is a dispersion in the relative prices. The substitution bias can be particulary eliminated by more frequent revision of the fixed consumer basket.

2.2 Outlet substitution bias

In the case of outlet substitution bias, we replace the substitution between the goods by a substitution between outlets. The outlet substitution bias relates to the consumers movement from the traditional high cost retailers to the discount outlets, which move into a market area and capture market share of that traditional high cost retailers.

The consumers have tendency to buy in the cheaper outlets and so they may decrease their cost of living. This outlet substitution is impossible to register by the fixed-base Laspeyres price index. This outlet substitution bias is similar to the elementary substitution bias, except that, in this case, we are just intent on question of where the consumers buy instead what the consumers buy.

Diewert (1998) defined outlet substitution bias B_O with the proviso that the differences in the services provided by discount and traditional retailers can be neglected, then a reasonable concept of the true price index is the average price paid by consumers over all outlets (see [4], p. 28):

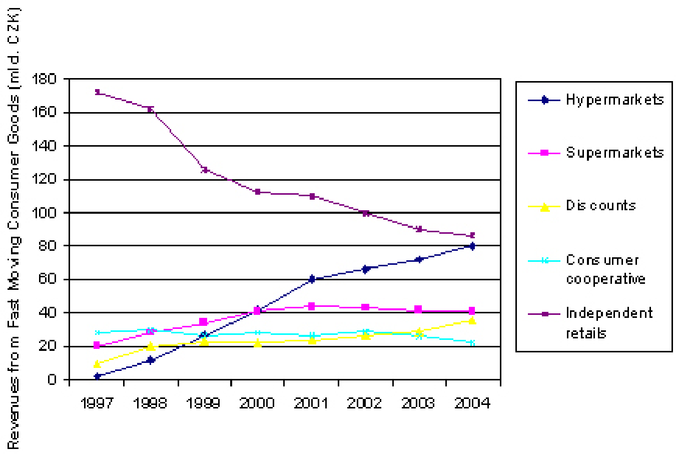
$$B_O \equiv P_L - P_T = (1+i)sd, \tag{3}$$

$$P_T \equiv (1-s)(1+i) + s(1+i)(1-d), \tag{4}$$

where $(1+i) = P_L$ is, again, the Laspeyres price index for the high cost retailers in the current period, s is the market share captured by low cost retailers (discount outlets) in the current period and d is percentage discount of the low cost retailers over traditional retailers.[5] Thus, the outlet substitution bias is the gap between the Laspeyres index and the „true“ price index.

[5]This formula implicitly assumes that the discount is constant in the two periods and the period-to-period trend is discount retails prices in the same as the traditional retailers trend.

Figure 1: Revenues of Fast Moving Consumer Goods according to trading Channels (in current prices (brutto))



Source: MAKRO Cash & Carry ČR and INCOMA Research (2005)

Table 2: Revenues according to trading Channels in the Czech Republic (annual change in p.p.)

Period	1998	1999	2000	2001	2002	2003	2004
Hypermarkets	3,90	6,63	5,75	5,50	2,26	2,89	2,39
Supermarkets	2,49	3,23	2,39	-0,13	-0,38	-0,01	-0,74
Discounts	3,63	1,77	-0,73	0,08	1,13	1,01	2,39
Consumer cooperative	-0,16	-0,51	0,04	-1,24	0,75	-0,90	-1,74
Independent retails	-9,85	-11,12	-7,45	-4,20	-3,77	-2,99	-2,30
Low cost retailers	10,02	11,63	7,41	5,44	3,02	3,89	4,03

Source: own calculation of INCOMA data

Table 3: Outlet substitution bias in the Czech Republic

Period	1998	1999	2000	2001	2002	2003	2004
Inflation (%)	10,6	2,1	4,0	4,7	1,8	0,1	2,8
s (p.p.)	10,02	11,63	7,41	5,44	3,02	3,89	4,03
d (%)	15,0	15,0	15,0	15,0	15,0	15,0	15,0
B_O (p.p.)	1,66	1,78	1,16	0,85	0,46	0,58	0,62

Source: own calculation of CSO data

We have found that the low cost retailers increased their share of the market by 5,80 % annually on an average and sold for 15 % less than traditional retailers in the Czech Republic during 1998–2004. We have measured the annual outlet substitution bias about 0,90 percentage points at average Czech inflation rates during 1998–2004 of 2,1%.

There is a question how to reduce the outlet substitution bias. The outlet substitution bias disappears when the low cost retailers capture the entire market, in which case the parameters are $s = 1$ and $d = 0$. In another way, the outlet substitution bias can be reduced by introducing the new weight, that will measure the shares of sales of the given product with regard to the kind of outlet.

2.3 Quality change bias

Quality change bias occurs when statistical agencies attribute to inflation the part of a price increase due to improved quality instead.

Statistical agencies could simply replace the old product by the new one, the agencies looks at the price of the new model subsequently. This approach works if any quality difference between the old and the new product is reflected by the price difference between them. But more typically, the new product has a higher quality, which is not fully offset by its price.

For a rough measure of the quality change bias B_Q, Diewert (1998) suggested this formula:

$$B_Q \equiv P_L - P_T = \frac{1+i}{1+e} se, \tag{5}$$

$$P_T \equiv (1-s)(1+i) + s\frac{1+i}{1+e}, \tag{6}$$

where $1+i$ is, again, the Laspeyres index, s is the share of commodities that have been replaced by new products and e is the percentage increase in the efficiency of a new product that is missed when the new product is linked into the index. Thus, the quality change bias B_Q is the difference between the Laspeyres price index P_L and the true price index P_T.

Quality improvements are more expressive in transition economies, because of low initial quality levels in these economies. The quality changes have occurred continously in even the most basic product. When we consider the Czech Republic, the upward bias in inflation is about 1,0 precentage point annually on the assumption that 10 % of the consumer basket improves in quality by 10 % in a given year (see [5], p. 7). It is very complicated to measure the quality improvement bias, we can only presume the values of the parameters s and e. Values in the table below are proposed for the period 1998 - 2004 in the Czech Republic with respect to data designed by Hanousek, Filer (for 1990s in the Czech Republic) and Diewert (for 1990s in USA).

Table 4: Quality change bias in the Czech Republic

Period	1998	1999	2000	2001	2002	2003	2004
Inflation (%)	10,6	2,1	4,0	4,7	1,8	0,1	2,8
s (%)	12,0	12,0	12,0	9,0	9,0	9,0	9,0
e (%)	9,0	9,0	8,0	8,0	7,0	7,0	6,0
B_Q (p.p.)	1,10	1,01	0,92	0,70	0,60	0,59	0,52

Source: own designed data and data CSO

This bias could be eliminated by the simultaneous observations of changes in prices for both on old and a new version of the product at a time when both are available. The price change associated with quality change is estimated as the difference in the market prices in this overlapping period. When a new product is not comparable with an old product, the price index can be computed excluding the new product. But in the consumer basket similar good is calculated for the period when the substitution was made, ignoring the product which is being replaced. For specific products, mostly durables, hedonic regression is used for elimination of the quality change bias. The last method is the direct quality adjustment. However, none of these techniques is appropriate for capturing the quality improvements in basic commodities that were common in transition economies (see [6], pp. 3-6).

2.4 New goods bias

New goods bias is a result of imputing new products. This new product is not included in the fixed base or is incorporated in it with some time delay. It often takes years before the new good is actually included in the basket. Thus, the new goods bias occurs because of the delay between when new goods enter a market and when they are included in the consumer price index.

When a new product is introduced into the market, it generally has a high price which is reduced in subsequent periods. Since statistical agencies do not introduce new goods into their commodity baskets until the new product has become important in the market, they often miss this early decline in price.[6] The price of a new product that makes its first appearance in the market, is called imputed price. In the case of not incorporating or late incorporating the new product into the basket, the inflation is overvaluated.

Diewert (1998) designed for a new goods bias B_N the following formula:

$$B_N \equiv P_L - P_T = \frac{1}{2}(1+i)sd, \tag{7}$$

$$P_T \equiv (1 - \frac{1}{2}s)(1+i) + \frac{1}{2}s(1+i)(1-d), \tag{8}$$

where $(1+i)$ is the Laspeyres estimate of overall price change, s is the market share of new goods, which have not yet been introduced into the basket of comodities and d is the percentage decline in prices of the new goods from their initial imputed prices. Thus, we can say the new good bias is the difference between the Laspeyres price index P_L and the true price index P_T.

For the Czech economy, the new goods constitute 5 % (s=5 %) of the basket and sell on an average for 20 % less than their imputed prices (d=20 %). The new goods bias, at Czech inflation rates, is measured about 0,55 percentage points during 1990s (see [5], p. 8). For the period 1998 - 2004 we considered that the percentage decline in prices of the new goods from their initial imputed prices will be lower then Hanousek and Filer assumed for 1990s.

Table 5: New goods bias in the Czech Republic

Period	1998	1999	2000	2001	2002	2003	2004
Inflation (%)	10,6	2,1	4,0	4,7	1,8	0,1	2,8
s (%)	5,0	5,0	5,0	5,0	5,0	5,0	5,0
e (%)	15,0	15,0	15,0	15,0	15,0	15,0	15,0
B_Q (p.p.)	0,41	0,38	0,39	0,39	0,38	0,38	0,39

Source: CSO data

The new good bias can be reduced by early introducing the new product in the basket, in which we observe product price movements. Another method is using imputed prices. The product imputed price is estimated econometricaly when the demand for that product is zero, the price of this product is equal to the imputed price. Then we can observe the product price decline when we compare the imputed price with real price of product.

3 Conclusion

This paper dealing with measuring the inflation described in the first section the construction of the Consumer Price Index that is based on the Laspeyres price index. Some disadvantages of this approach were discussed. Consequently, we described the true cost of living index, using alternative approach to express how is the inflation understood by the economic subjects. This theoretic problem of true cost of living could be reasonably approximated by the Fisher ideal price index.

[6] This phenomenon observed already Alfred Marshall years ago (see [2], p. 373).

In the second section, we focused on the differences between the previously introduced Consumer Price Index and the exact cost of living index, approximated by the Fisher ideal price index. We have identified four important types of biases: the elementary substitution bias, the outlet substitution bias, the quality change bias and the new goods bias. The biases were investigated in a greater detail and for each such a bias, the formula, which roughly measures it, have been introduced. We have shown extents of the biases for some Czech economy data.

References

[1] Czech Statistical Office. *Indexy spotřebitelských cen (metodická příručka pro uživatele).* [cit. 2006-05-22]. Available at < http://www.czso.cz/csu/redakce.nsf/i/indexy_spotrebitel-skych_cen_metodicka_prirucka/$File/manual_isc.pdf >

[2] Marshall, A. *Remedies for Fluctuations of General Prices.* Contemporary Review, 1887, pp. 355-375.

[3] Diewert, W., E. *Index Number Issues in the Consumer Price Index.* Journal of Economic Perspectives, Vol. 12, No. 1, 1998, pp. 47-58.

[4] Diewert, W., E. *Axiomatic and Economic Approaches to Elementary Price Indexes.* Discussion Paper No. 95-01, Department of Economics, University of British Columbia, Vancouver, Canada, 1995, pp. 1-60.

[5] Filer, R., K., Hanousek, J. *Output Changes and Inflationary Bias in Transition.* [cit. 2006-05-5]. Available at < http://www.cerge-ei.cz/pdf/wp/Wp167.pdf >

[6] Filer, R., K., Hanousek, J. *Consumers' Opinion of Inflation Bias Due to Quality Improvements in Transition in the Czech Republic.* [cit. 2006-05-11]. Available at < http://www.cerge-ei.cz/pdf/wp/Wp184.pdf >

[7] Filer, R., K., Hanousek, J. *Evaluating Imperfections and Biases in Price Indexes during Transition.* [cit. 2006-05-11]. Available at < http://www.cerge-ei.cz/pdf/wp/Wp186.pdf >

[8] MAKRO Cash & Carry ČR and INCOMA Research. *MAKRO–INCOMA Index.* [cit. 2006-08-10]. Available at < http://www.makro.cz/files/incoma/0508_makro_incoma_index_celk.ppt >

Some optimization problems with synchronization constraints. A survey of recent results.*

Karel Zimmermann
Faculty of Mathematics and Physics
Charles University Prague
Karel.Zimmermann@MFF.CUNI.CZ

Abstract

The aim of the contribution is to present a survey of various synchronization problems, in which the synchronization or coordination of events (e.g. train arrivals and departures, release times or deadlines of activities, providing services) is required. The requirements are represented by a set of feasible solutions (i.e. feasible arrivals, departures, deadlines and so on). The problems, which have to be solved in connection with such situations are the following: (1) to find out whether the set of feasible solutions is non-empty; (2) if the set of feasible solutions is non-empty, to find in some sense the optimal solution (i.e. a feasible solution, which minimizes or maximizes a given objective function); (3) if the set of feasible solutions is empty, to propose an approach how to manage such incorrectly posed problems e.g. an appropriate change of input data, which leads to a solvable problem); (4) to provide the corresponding post-optimal stability and sensitivity analysis. Examples of such synchronization problems and corresponding solution methods will be presented. The set of feasible solutions of the problems is in general a non-convex compact set in a finite-dimensional Euclidean space. The set of feasible solutions is described by a finite set of equations and inequalities with special max- or min-separable functions (i.e. functions, which are equal to the maximum or minimum of a finite set of functions of one variable). The objective functions are max- or min-separable too. The proposed methods give global optimal solutions and have a polynomial complexity.
Keywords: non-convex optimization, synchronization of dynamic events.
JEL: 90C26, 90C30, 49K40

---

*Under the support of 202-03/2060982 VZ, Czech Science Foundation # 402/06/1071

1 Introduction.

Operations research problems occuring e.g. in technology or economics contain interdependent processes with a given duration. The interdependency means that some processes or activities must be finished before other processes or activities can begin. Such situations were studied in the recent literature ([1], [4]). If we want to determine release times and/or deadlines of the interdependent processes satisfying certain additional conditions, it is necessary to synchronize the rerlease times and deadlines in such a way that the interdependency is taken into account. In these situations we have to find out whether the set of feasible release times and deadlines satisfying given technological conditions and taking into account the interdependency restrictions is nonempty and if the answer is positive find one element of such set.
In the present paper we study systems of equations and inequalities,which make possible to solve a special class of such activity synchronization problems, in which special so called max-separable or min-separable functions can be made use of to describe the synchronization requirements. A function $f : R^n \to R^1$ is called max-separable if $f(x_1, ..., x_n) = max_{1\leq j\leq n} f_j(x_j)$, and in a similar way the concept of a min-separable function is defined. The synchronization problems considered here are described by a finite systems of equations and inequalities with special max- or min-separable functions on both sides of the equations. The set of solutions of such systems is in general a non-convex set in R^n. Originally such systems with variables only on one side of such equations and inequalities were studied (e.g. [2], [3], [5]). Recently problems with variables on both sides of equations and inequalities became the subject of research in some papers as an appropriate tool for solving various types of activity synchronization problems ([1], [4]). In [1] systems of equations with special max-separable functions, so called $(max, +)$-linear functions, in which $f_j(x_j) = c_j + x_j$, where c_j are real numbers are considered. In [4] systems of inequalities with (max,+)-linear functions on one side and $(min, +)$-linear functions (i.e. functions of the form $f(x) = min_{1\leq j\leq n}(c_j + x_j)$) on the other side of the inequalities are investigated. In [6], unlike to [4], systems of equations with $(max, +)$-linear functions on one side and $(min, +)$-linear functions on the other side of the equations are studied and a finite algorithm with complexity $O(m^2n^2)$ for solving such systems is proposed. The aim of the present paper is to provide a survey of the recent results and to propose a procedure combining the the algorithms from [1] and [6] for solving of $max, min, +$ systems with $(max, +)$- or $(max, +)$-linear functions on any of the two sides of any equation.

2 Examples of synchronization problems.

Example 2.1 Let us have n railway stations S_j, $j = 1, \ldots, n$, from which passengers are delivered to m railway stations C_i, $i = 1, \ldots, m$. Let us assume that traveling times from S_j to C_i are equal to a_{ij}, so that if x_j denotes a departure time from S_j, the arrival time to C_i will be equal to $a_{ij} + x_j$. Under these assumptions the last train coming from stations $S_1, S_2, \ldots, S_n$ will come to C_i at a time $a_i(x) \equiv max_{1\leq\ j\leq n}(a_{ij}+x_j)$. Let us assume that all passengers coming from $S_j,\ j=1,\ldots,n$ must have the possibility to change at C_i for other transport means. Let $T_1,\ldots,T_p$ be another group of stations, from which passengers are transported to C_i by different transport means (e.g. by buses). Let travelling times from T_k to C_i be equal to b_{ik} for $i=1,\ldots,m,\ k=1,\ldots,p$ and let $y_1,\ldots,y_p$ be departure times from stations $T_1,\ldots,T_p$. Let $b_i(y) \equiv max_{1\leq k\leq p}(b_{ik})+y_k)$. Then the earliest time for departure from station C_i is equal to $max\{a_i(x), b_i(y)\}$. If $a_i(x) \neq b_i(y)$ for some i, it is evident that the passengers who come earlier to C_i must wait for those who will come later. To minimize the waiting times at stations C_i we require to synchronize (if possible) the departure times in such a way that $a_i(x) = b_i(y)$ for all $i=1,\ldots,m$. We assume further that the departure times cannot be chosen arbitrarily, but within prescribed time intervals, i.e. that $\underline{x} \leq x \leq \overline{x}$, $\underline{y} \leq y \leq \overline{y}$ for given $\underline{x}$, $\overline{x}$, $\underline{y}$, $\overline{y}$. Therefore we have to find out whether such synchonized departure times exist and if the answer is positive, find a feasible (x, y), which lies within the prescribed bounds and satisfies the equalities $a_i(x) = b_i(y)\ \forall i \in \{1,\ldots,m\}$. Let us note that if there is no connection between some stations C_i and S_j or T_k or if it is not necessary to wait for the passengers from some station S_j at some station C_i, we can include formally such situations in the model by setting the corresponding travelling times a_{ij} , b_{ik} equal to $-\infty$ (note that for practical calculations, it is evidently possible to choose the corresponding coefficients as sufficiently low finite negative numbers). For similar reasons, we can assume w.l.o.g. that $n = p$.

Example 2.2 Let us have a similar situation as in the preceding example, but the passengers from stations $S_j,\ j=1,\ldots,n$ must change at C_i for transport means (e.g. buses) leaving from C_i for stations T_k. Let arrival times of these transport means to stations T_k be denoted by z_k for $k=1,\ldots,p$. Let the corresponding travelling time from C_i to T_k be equal to c_{ik}, so that the departure times from C_i to T_k are equal to $z_k - c_{ik}$. Theferore if all passengers from stations S_j must have the possibility to change for stations T_k at station C_i, the latest arrival time to C_i, which is equal to $a_i(x) = max_{1\leq\ j\leq n}(a_{ij}+x_j)$ must be synchronized (less or equal) with the earliest departure time from C_i to stations T_k, which is under our assumptions equal to $c_i(z) \equiv min_{1\leq k\leq p}(z_k - c_{ik})$. If $a_i(x) \neq c_i(z)$, then some passangers either wait at C_i for the connection or miss the connection. The avoid these undesirable cases, we may require that $a_i(x) = c_i(z)\ \forall i \in \{1,\ldots,m\}$. Further we will require similarly as in the preceding example that $\underline{x} \leq x \leq \overline{x}$, $\underline{z} \leq z \leq \overline{z}$ for given $\underline{x}$, $\overline{x}$, $\underline{z}$, $\overline{z}$. Therefore we have to find a solution of the system of equations $a_i(x) = c_i(z), \forall i = 1,\ldots,m$,which lies within the prescribed upper and lower bounds.

Example 2.3 We can unify the requirements fom the preceding two examples by solving the system of equations $a_i(x) = b_i(y),\ a_i(x) = c_i(z)$ for all $i = 1,\ldots,m$ under the additional requirements that that $x,\ y,\ z$ lie within the prescribed bounds (we assume that $a_i(x),\ b_i(y),\ c_i(z)$ have the same meaning as in the preceding examples).

3 Problem formulation

. Let $I = \{1, \dots, m\}$, $J = \{1, \dots, n\}$. We will consider the following system of equations and inequalities:

$$a_i(x) = b_i(y), i \in I, \tag{1}$$

$$a_i(x) = c_i(z), i \in I, \tag{2}$$

$$\underline{x} \leq x \leq \overline{x},\ \underline{y} \leq y \leq \overline{y},\ \underline{z} \leq z \leq \overline{z}, \tag{3}$$

where $a_i(x) \equiv max_{j \in J}(a_{ij}+x_j)$, $b_i(y) \equiv max_{k \in J}(b_{ik}+y_k)$, $c_i(z) \equiv min_{k \in J}(c_{ik}+z_k)$, $x,\ y,\ z \in R^n$, and $\underline{x},\ \overline{x},\ \underline{y},\ \overline{y},\ \underline{z},\ \overline{z} \in R^n$ are given lower and upper bounds. In the next section we will propose a procedure based on previous results from the literature, which either finds a solution of (1), (2), (3) or finds out that no such solution exists.

4 Solution of the problem.

It follows from [1] that if the set of solutions of subsystem (1), (3) is nonempty, it has the maximum element (x^*, y^*) in the sense that if (x, y) is an arbitrary solution of system (1), (3), then $(x, y) \leq (x^*, y^*)$ holds. The maximum element (x^*, y^*) can be found with the aid of an $O(m^5n^5)$ algorithm proposed in [1]. Further it follows from [6] that the solution of system (1), (2), (3) can be found as the optimal solution of the following minimization problem:

$$f(x, z) \equiv max_{i \in I}\, |a_i(x) - c_i(z)| \longmapsto min \tag{4}$$

subject to

$$a_i(x) = b_i(y),\ \forall i \in I, (\underline{x}, \underline{y}, \underline{z}) \leq (x, y, z) \leq (x^*, y^*, \overline{z}). \tag{5}$$

The optimal solution $(x^{opt}, y^{opt}, z^{opt})$ of this problem can be found by an $O(m^2n^2)$ combining ALGORITHM I from [1] and the algorithm presented in [6]. If $f(x^{opt}, z^{opt}) = 0$, then evidently $(x^{opt}, y^{opt}, z^{opt})$ is a solution of system (1), (2), (3). If $f(x^{opt}, z^{opt}) > 0$, then no solution of the system exists. Therefore the procedure solving system (1), (2), (3) can be summarized as follows:

$\boxed{1}$ Find the maximum solution (x^*, y^*) of system (1), (3) using the algorithm from [1];
$\boxed{2}$ If no solution of system (1), (3) exists go to $\boxed{5}$;
$\boxed{3}$ Find the optimal solution $(x^{opt}, y^{opt}, z^{opt})$ of minimization problem (4), (5) using a combination of algorithms from [1], [6] ;
$\boxed{4}$ If $f(x^{opt}, z^{opt}) = 0$, then $(x^{opt}, y^{opt}, z^{opt})$ is a solution of system (1), (2), (3), STOP.
$\boxed{5}$ No solution of system (1), (2), (3) exists, STOP.

Remark 4.1 Let us note that if e.g. $\underline{y} = \overline{y}$ and $\underline{z} = \overline{z}$, the system (1), (2), (3) is in fact "one-sided" (i.e. it has only variables x on the left-hand sides and constants on the right-hand sides) and can be solved by making use of other methods designed for the "one-sided" systems, which are presented in other references (e.g. [2], [3], [5]).

Remark 4.2 Let us note that the solution method proposed above can be easily applied also for systems with the same variables on both sides of the system. We can simply add to the equations (3) n further equations $max_{j\in J}(a_{ij} + x_j) = max_{j\in J}(b_{ij} + y_j)$ for $i = m+1, \ldots, m+n$ with $a_{m+jj} = b_{m+jj} = 0$ for $j = 1, \ldots, n$ and $a_{ik} = -\infty$, $b_{ik} = -\infty$ if $i = m + j$, $k \neq j$. In a similar way system $a_i(x) = c_i(z),\ \forall i \in I$ can be adjusted to the case that $x = z$. Similarly, we can by an appropriate choice of additional coefficients include in the systems also additional precedence requirements like $x_j \leq x_k + \alpha$, $x_j \geq x_k + \beta$, $x_j \geq y_k + \gamma$ etc., where $\alpha,\ \beta,\ \gamma$ are given constants. Note that including an additional requirement that some or all variables x_j, y_j must be integer needs also only a slight technical modification of the proposed algorithm.

Remark 4.3 On the other hand, system $a_i(x) = b_i(y),\ i \in I$, can be considered as a special system with the same $2n$ variables (x, y) on both sides of the equations. For that purpose, it is sufficient to introduce new coefficients $a_{ij} = b_{ij} = -\infty$ for $i \in I,\ n+1 \leq j \leq 2n$ and consider the system $\tilde{a}_i(x, y) = \tilde{b}_i(x, y),\ i \in I$, where $\tilde{a}_i(x, y) = max_{1\leq j\leq 2n}(a_{ij} + x_j)$, $\tilde{b}_i(x, y) = max_{1\leq j\leq 2n}(b_{ij} + x_j)$. Similarly the system $a_i(x) = c_i(z),\ i \in I$ can be transformed to $\tilde{a}_i(x, z) = \tilde{b}_i(x, y),\ i \in I$ by an appropriate choice of additional coefficients $a_{ij} = -\infty,\ c_{ij} = +\infty$ for $n + 1 \leq j \leq 2n$.

References

[1] Butkovic, P., Zimmermann, K.: A Strongly polynomial algorithm for solving two-sided linear systems in max-algebra, Discrete Applied Mathematics 154(2006), pp. 437 - 446.

[2] Cuninghame-Green, R.A.: Minimax Algebra, Lecture Notes in Economics and Mathematical Systems, 166, Springer Verlag, Berlin, 1979.

[3] Vorobjov, N.N.: Extremal algebra of positive matrices, Elektronische Datenverarbeitung und Kybernetik, 3, 1967, pp. 39-71 (in Russian).

[4] Woude van der, J., Olsder, G.J.:On (min,max,+)-inequalities, Contemporary Mathematics, 377, 2005, pp. 353 - 361.

[5] Zimmermann, K.: Disjunctive optimization, max-separable problems and extremal algebras, Theoretical Computer Science 293, 2003, pp. 45-54.

[6] Zimmermann, K.: One class of separable optimization problems, Proceedings of the Int. Conf. on Appl. Math., Plovdiv, August 2006, to appear.

Real option applications based on the generalised multinomial flexible switch options methodology

Zdeněk Zmeškal[1]
Dpt. of Finance, Faculty of Economics, VŠB-TU
zdenek.zmeskal@vvsb.cz

Abstract

The real option methodology application with possibility of sequential multinomial decision-making is described in the paper. The stochastic dynamic Bellman's optimisation principle is explained and applied; moreover optimisation criterion of the present expected value is demonstrated and used. Likewise, option valuation approach on replication strategy and risk-neutral probability is described. Illustrative example of the application of the real multinomial flexible switch options methodology for three chosen modes is presented. The usefulness, effectiveness and suitability of application the generalized flexibility model in company valuation and evaluation projects were verified and confirmed.

Keywords

Financial option; Real options; Discrete Binomial Model; Pricing; Stochastic dynamic Bellman's Optimisation Principle; Switch Option

JEL: C6, C 44, C53, F2, F21, G1, G11, G15, G2, G21

1 Introduction

The real option is understood as a flexible approach in financial decision-making in strategy decisions of the non-financial companies, concerning the real assets (assets, debt, equity, investments, commodity, land, research cost, technology, process, production). Comparing with traditional passive strategies, active measures in the future are considered in real projects managing. Examples of flexible actions are: abandonment temporarily shut down, expansion, contraction, change of the technological process, production structure parameters, sale, purchase etc. Real option methodology is based on the financial option methodology and applied on real assets.

Real options topic is under consideration of the academic and managerial community for several years, and it is the basic novelty of corporate finance. As the basic references and books concerning the topic said should be introduced for instance: Trigeorgis (1998), Sick (1995), Dixit&Pindyck (1994), Brennan&Trigeorgis (1999), Ronn (2002).

Real option valuation approaches comes out the stochastic dynamic programming on the Bellman's optimisation principle. 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 is the replication strategy using the risk-neutral probability. General principle of options apprising is the martingale approach.

The intention of the paper is to apply the real option methodology in company valuation and project evaluation. Particularly, the generalised real multinomial flexible switch options methodology with possibility of sequential multinomial decision-making will be derived, explained, applied and verified. The stochastic dynamic Bellman's optimisation model will be applied on the present expected

[1] prof. Dr. Ing. Zdeněk Zmeškal, department of Finance, faculty of economics, VŠB-TU Ostrava, Sokolská 33, Ostrava 701 21, zdenek.zmeskal@vsb.cz
The paper is supported by the Grant Agency of the Czech Republic (GACR) 402/04/1357.

optimisation criterion. Option valuation approach on the replication strategy and risk-neutral probability will be applied.

2 Description of the replication strategy

The generalised principle of valuation is called the martingale principle. The principle is being defined so that a value has to be equal to expected future value, implying the random process is without trend. In the case of the risk neutral approach this category is ratio of random value and risk-free asset, so after rearranging

$$V_t = e^{-r\cdot dt} \cdot \widehat{E}(V_{t+dt}). \tag{1}$$

2.1 The replication strategy derivation

Deriving the replication strategy, we suppose a compact (effective) market, asset-bearing the incomes (dividends, coupons, etc.) proportional to an asset price. The replication strategy will be applied for the discrete binomial model and one risk (random) factor. The model is of discrete version and for the sake of simplicity an intra-interval continuous compounding is applied.

The replication strategy is based on creation a portfolio from underlying asset *S* and risk-free asset *B* so, for every situation derivative value is to be replicated; it means a derivative value equals a portfolio value.

Portfolio value in apprising a moment $t, \Pi_t \equiv a \cdot S_t + B_t = f_t$,

the portfolio value in a moment $t + dt$ for growing price, $\Pi_{t+dt} \equiv a \cdot S^u_{t+dt} + B_t \cdot e^{r\cdot dt} = f^u_{t+dt}$,

the portfolio value in a moment $t + dt$ for declining price, $\Pi_{t+dt} \equiv a \cdot S^d_{t+dt} + B_t \cdot e^{r\cdot dt} = f^d_{t+td}$,

where *S* is underlying asset value, *a* is a amount of underlying asset, *B* is risk-free asset value, *f* is derivative value, *r* is risk-free rate, *u* (*d*) are indexes of growth (fall) of underlying asset, S^u_{t+dt} $\left(S^d_{t+dt}\right)$ are their prices in up-movements (down-movements).

By solution of three equations for variables *a, B,* f_t, we can get a general formula for derivative price,

$$f_t = e^{-r\cdot dt} \cdot \left\{ f^u_{t+dt} \cdot \left[\frac{e^{r\cdot dt} \cdot S_t - S^d_{t+dt}}{S^u_{t+dt} - S^d_{t+dt}} \right] + f^d_{t+dt} \cdot \left[\frac{S^u_{t+dt} - e^{r\cdot dt} \cdot S_t}{S^u_{t+dt} - S^d_{t+dt}} \right] \right\}. \tag{2}$$

This is the general formula for derivative price valuation by the replication strategy, which should be written as follows,

$$f_t = e^{-r\cdot dt} \cdot \left[f^u_{t+dt} \cdot (\widehat{p}) + f^d_{t+dt} \cdot (1 - \widehat{p}) \right], \text{ or } f_t = e^{-r\cdot dt} \cdot \widehat{E}(f_{t+dt}). \tag{3}$$

Here $$\widehat{p} = \frac{e^{r\cdot dt} \cdot S_t - S^d_{t+dt}}{S^u_{t+dt} - S^d_{t+dt}} \tag{4}$$

is the risk-neutral probability of up-movement and $\widehat{E}(f_{t+dt})$ is the risk-neutral expected value.

The derivative price is determined as a present value of expected value in a following period. This probability can be considered neither a market growth nor a subjective probability. Due to (3) the derivative price is equal to the present value of risk-neutral expected value of subsequent period, which coincides with generalised martingale principle see (1).

Expressing the underlying ex-dividend asset price, under proportional continuous non-random income *c* and *u, d* random development due to Geometric Brown's process as follows $S^u_{t+dt} = S_t \cdot e^{u+c}$; $S^d_{t+dt} = S_t \cdot e^{d+c}$, substituting to (4) and after re-arranging we get

$$\widehat{p}=\frac{e^{(r-c)}-e^{d}}{e^{u}-e^{d}}. \tag{5}$$

This formula should be generalised after substituting for risk-neutral probability growth parameter $\widehat{g}=r-c$, as follows $\widehat{p}=\dfrac{e^{\widehat{g}}-e^{d}}{e^{u}-e^{d}}$.

2. 2 Valuation procedure of an American option

Option pricing using the discrete binomial model respecting a stochastic dynamic programming model and risk-neutral probability can be divided into following steps.

(i) Determination of the risk-neutral growth parameter $\widehat{g}$.

(ii) Underlying asset modelling

(a) Subjective approach by virtue of expert estimation and forecast

(b) Objective approach on the basis of statistical estimation of random underlying asset from market data (e. g. arithmetic, geometric Brown's process, mean-reversion process, Vasicek, CIR, Ito's process etc.).

In the case of the geometric Brown's process are firstly computed up-movement and down-movement indexes which characterise a volatility coinciding with market volatility, so under condition $u=-d$, then $e^{u}=e^{\sigma\cdot\sqrt{dt}}$, $e^{d}=e^{-\sigma\cdot\sqrt{dt}}$ and $S_{t+dt}^{u}=S_{t}\cdot e^{u}$, $S_{t+dt}^{d}=S_{t}\cdot e^{d}$.

(iii) At the maturity day, *T*, option price is equal to intrinsic value, $f_{T}^{u}=g_{T}^{u}$, or $f_{T}^{d}=g_{T}^{d}$. Computation of the intrinsic value (payoff function), *g*, depends on the option type. For example, in the case of a call option $g_{t}^{u}=\max\left(S_{t}^{u}-X;0\right)$, and a put option, $g_{t}^{u}=\max\left(X-S_{t}^{u};0\right)$ X being exercise price.

(iv) Working backwards from the end of the binomial tree to the beginning, price of an option is calculated in every node and at the initial node as well due to formulas.

For European option

$$f_{t}=e^{-r\cdot dt}\cdot\left[f_{t+dt}^{u}\cdot(\widehat{p})+f_{t+dt}^{d}\cdot(1-\widehat{p})\right].$$

An American option can be exercised whenever during pre-specified period and for its price can be written,

$f_{t}=\max\limits_{q\in S}\left(g_{t}^{q}\right)$, it means $f_{t}=\max\limits_{q\in S\ or\ q=S+1}\left\{g_{t}^{q};g_{t}^{S+1}=e^{-r\cdot dt}\cdot\left[f_{t+dt}^{u}\cdot(\widehat{p})+f_{t+dt}^{d}\cdot(1-\widehat{p})\right]\right\}$.

The functions g_{t}^{q} means exercise the option, $g_{t}^{\acute{S}+1}$ depicts non-exercise the option. This equation is the Bellman's optimal equation for stochastic dynamic programming. Parameter *q* represents choice (option) of process, generally called mode. Symbol $\widehat{p}$ depicts the risk-neutral probability defined previously. At the beginning of the period f_{0} is than the price of an option.

(v) Determination of the decision-making variables, Q_{t},

$$Q_{t}=\underset{q\in S\ or\ q=S+1}{\arg\max}\left\{g_{t}^{q};g_{t}^{S+1}=e^{-r\cdot dt}\cdot\left[f_{t+dt}^{u}\cdot(\widehat{p})+f_{t+dt}^{d}\cdot(1-\widehat{p})\right]\right\}.$$

The function argmax means the argument of the max function, so the decision parameter *q* corresponds to the maximum value of the objective function.

(vi) The sensitivity analysis concerning the input data.

3 Stochastic dynamic programming

Dynamic programming represents optimal management problem for finding the optimal decision-making trajectory. It is the way of multi-periods optimisation by virtue of the Bellman's optimisation principle. Stochastic dynamic programming in comparing with deterministic programming means that whole process is of a random type.

Optimisation of the whole process in the approach means that it is possible to optimise a particular period separately, which means simultaneously an optimisation of the whole process. Final system state depends on previous all states and also on beginning state. Optimal decision is made with respect to the future possible states and also future forward-looking decision-making.

The Bellman's optimality principle, considered as axiom, means that, let a beginning decision is whatever then the following decisions have to be of optimal strategy in respect with previous decision.

The application assumptions of the principle consist in division the whole process in periods and objective function has to be of separated type. Thus, optimisation objective function is to be expressed as the aggregation of particular period's optimisation functions. Calculation procedure is performed recurrently from the final period to beginning period, so those in opposite direction to the process flow.

Assumptions of the stochastic dynamic programming are: division of the process in particular periods; periods are characterised by possible random states; particular decisions are depicted by the mode (e. g. technology, equipment, process, stage of development); total objective function must be of separation type, so that expressed as the aggregation of particular objective functions.

The problem solved by the stochastic dynamic programming is formulated so that whatever the beginning state is introduced and it is necessary to determine such decision trajectory, which is of optimal total objective function. The basic point is division the whole process (N-period extreme process) in particular periods and for every period is found the optimal solution. So that, in the beginning every period the system is in some mode and according to the period optimisation criteria result follows a decision about the transition or keeping the mode. Solution procedure consists in the whole process is transformed in successive founding of the particular optimal solutions. Backward recurrent procedure is applied.

3.1 Derivation the recurrent formula for present value criteria

The present value optimisation criterion fulfils separation condition so as the dynamic programming due to Bellman's optimisation principle can be employed. Derived and explained is recurrent present expected value formula. Subsequently, maximisation a present expected value is applied for the optimal choice of mode and trajectory as well.

Present value of cash-flow under assumption that cash-flow of the period is paid in the beginning of the period is formulated in the way,

$$V_N = \widehat{E}\left[\sum_{t=0}^{N-1} \beta_t \cdot x_t\right] = \sum_{t=0}^{N-1} \beta_t \cdot \widehat{E}(x_t).$$

Here V_N is value for N periods to final period, $\beta_t = (1+R)^{-t}$ is discount factor, x_t is cash flow in the particular period beginning. It should be rewritten as follows,

$$V_N = x_0 + \sum_{t=1}^{N-1} \beta_t \cdot \widehat{E}(x_t) = x_0 + \beta \cdot \sum_{t=1}^{N-1} \beta_{t-1} \cdot \widehat{E}(x_t) = x_0 + \beta \cdot \left[x_1 + \sum_{t=2}^{N-1} \beta_{t-1} \cdot \widehat{E}(x_t)\right].$$

So that the value of particular period is expressed recurrently depending on subsequent period,

$$V_N = x_0 + \beta \cdot \widehat{E}[V_{N-1}] \text{ where } \widehat{E}[V_{N-1}] = x_1 + \sum_{t=2}^{N-1} \beta_{t-1} \cdot \widehat{E}(x_t).$$

Analogically for following period

$$V_{N-1} = x_1 + \sum_{t=2}^{N-1} \beta_{t-1} \cdot \widehat{E}(x_t) = x_1 + \beta \cdot \sum_{t=2}^{N-1} \beta_{t-2} \cdot \widehat{E}(x_t) = x_1 + \beta \cdot \left[x_2 + \sum_{t=3} \beta_{t-2} \cdot \widehat{E}(x_t) \right].$$

It is apparent; a value of particular period is again expressed in terms of a subsequent period

$$V_{N-1} = x_1 + \beta \cdot \widehat{E}(V_{N-2}), \text{ where } V_{N-2} = x_2 + \sum_{t=3} \beta_{t-2} \cdot \widehat{E}(x_t).$$

Commonly, a recurrent formula for every period is written

$$V_{N-k} = x_k + \sum_{t=k+1}^{N-1} \beta_{t-k} \cdot \widehat{E}(x_t) = x_k + \beta \cdot \sum_{t=k+1}^{N-1} \beta_{t-(k+1)} \cdot \widehat{E}(x_t) = x_k + \beta \cdot \left[x_{k+1} + \sum_{t=k+2}^{N-1} \beta_{t-(k+1)} \cdot \widehat{E}(x_t) \right]$$

and so, whatever period value is determined in terms of a subsequent period in this way,

$$V_{N-k} = x_k + \beta \cdot \widehat{E}[V_{N-k-1}].$$

Value of the last period is written as follows

$$V_1 = x_{N-1} + \beta \cdot \widehat{E}(V_0).$$

In the preceding paragraphs the recurrent formulae of the present expected value was showed. Now possibility of decision about the mode choice will be carry out under present expected value optimisation criteria. Firstly, example for two modes the A and B will be described. The beginning mode will be the A and switch cost from mode A to B are depicted $C_{A,B}$.

Recurrent formulas for problem solution are following,

$$V_N^A = \max_{A,B} \left[x_0^A + \beta \cdot \widehat{E}\left(V_{N-1}^A\right); x_0^B - C_{A,B} + \beta \cdot \widehat{E}\left(V_{N-1}^B\right) \right],$$

$$V_{N-k}^A = \max_{A,B} \left[x_k^A + \beta \cdot \widehat{E}\left(V_{N-k-1}^A\right); x_k^B - C_{A,B} + \beta \cdot \widehat{E}\left(V_{N-k-1}^B\right) \right] \quad ,$$

$$V_1^A = \max_{A,B} \left[x_{N-1}^A + \beta \cdot V_0^A; x_{N-1}^B - C_{A,B} + \beta \cdot V_0^B \right].$$

In the case of possibility to switching among greater number of modes under assumption that the beginning mode is mode *m*, a subsequent is mode *q*, which is chosen form modes set *S*. It is possible to proceed due to following recurrent formulas:

$$V_N^m = \max_{q \in S} \left[x_0^q + C_{m,q} + \beta \cdot \widehat{E}\left(V_{N-1}^q\right) \right], \tag{6}$$

$$V_{N-k}^m = \max_{q \in S} \left[x_k^q + C_{m,q} + \beta \cdot \widehat{E}\left(V_{N-1-k}^q\right) \right], \tag{7}$$

$$V_1^m = \max_{q \in S} \left[x_{N-1}^q + C_{m,q} + \beta \cdot V_0^q \right]. \tag{8}$$

4 Example of company valuation with dynamic flexibility by virtue of switch options

Intention of the part is to apply the generalised flexible approach with a dynamic choice possibility of particular modes (switch option) and finding the optimal trajectory on the expected present value objective function basis. Three variants will be investigated in compliance with the beginning mode (situation) of the company: Variant 1 – Mode A, Variant 2 – Mode B, Variant 3 – Mode C.

The model of the stochastic dynamic programming on the basis of the binomial models American option; replicate 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 non-random phase is of perpetuity version. Furthermore, non-symmetrical switch costs are supposed.

4. 1 Procedure of the stochastic dynamic programming valuation of multinomial options

Valuation procedure of multinomial options with non-symmetrical switch options in respect with the stochastic dynamic programming on the Bellman's principle under the discrete binomial model and risk-neutral probability is performed in following steps.

(i) The determination of the risk-neutral growth parameter $\widehat{g}$.

(ii) Cash flow modelling likes an underlying asset

(a) Subjective approach by virtue of expert estimation and forecast

(b) Objective approach on the basis of statistical estimation and forecasting of random process. In the case of Brown's geometrical process,

$$x^{u}_{t+1,s+u} = x_{t,s} \cdot u; \quad x^{d}_{t+1,s+d} = x_{t} \cdot d\,.$$

(iii) At the beginning of the second phase the value for the second phase is $V^{q}_{0,s}$, here s is state and q is mode.

(iv) Value computation is based on the Bellman's stochastic equation of the dynamic programming. Here N-k depicts number of periods to the end of the first phase, s a state, q a mode, and $\widehat{p}$ is risk-neutral probability. Backward recurrent procedure from the end of the first phase to the beginning for states and modes of particular period in accordance with the optimal value and decision is calculated.

Valuation formula for one period to the end of the first phase,

$$V^{m}_{1,s} = \max_{q \in S}\left[x^{q}_{N-1,s} - C_{m,q} + \beta \cdot V^{q}_{0,s}\right].$$

Valuation formula for other periods by virtue of the recurrent procedure,

$$V^{m}_{N-k,s} = \max_{q \in S}\left\{x^{q}_{k,s} - C_{m,q} + \beta \cdot \left[\widehat{p} \cdot V^{q}_{N-1-k,s+u} + \widehat{q} \cdot V^{q}_{N-1-k,s-d}\right]\right\}.$$

Valuation formula at the beginning of the whole first phase (the first period),

$$V^{m}_{N,s} = \max_{q \in S}\left\{x^{q}_{0,s} - C_{m,q} + \beta \cdot \left[\widehat{p} \cdot V^{q}_{N-1,s+u} + \widehat{q} \cdot V^{q}_{N-1,s-d}\right]\right\}.$$

(v) Identification of the decision variant, $Q_{t,s}$, Variant 1 – Mode A, Variant 2 – Mode B,

Variant 3 – Mode C,: $Q_{t,s} = \arg\max_{q \in S}\left\{x^{q}_{k,s} - C_{m,q} + \beta \cdot \left[\widehat{p} \cdot V^{q}_{N-1-k,s+u} + \widehat{q} \cdot V^{q}_{N-1-k,s-d}\right]\right\}.$

(vi) The sensitivity analysis concerning the input data.

4. 2 Computation procedure and results

Input data of the applied model: risk-free rate r = 10%, up-movement index u = 1,2; the value for the beginning of the second phase $V_{0,s}^{q}$ for the states s and modes q. There are in the table presented switch cost C_{ij} connected with switching among particular modes, the keeping the same mode is linked with no switch cost of course.

Switch cost $C_{i,j}$		Subsequent modus		
		A	B	C
Beginning modus	A	0	-44	44
	B	44	0	-10000
	C	-44	-10000	0

Computation procedure in coincidence with the procedure of the stochastic dynamic programming valuation of multinomial options for three beginning modes presents Fig. 1.

Fig. 1 Valuation procedure of dynamic flexible multinomial switch options methodology

Cash flow

	Mode A - NORMAL						Mode B - EXPANSION						Mode C - CONTRACTION					
n/t	0	1	2	3	4	V_0	0	1	2	3	4	V_0	0	1	2	3	4	V_0
0	10,0	12,0	14,4	17,3	20,7	41,5	11,5	13,8	16,6	19,9	23,8	47,7	8,5	10,2	12,2	14,7	17,6	35,3
-2		8,3	10,0	12,0	14,4	28,8		9,6	11,5	13,8	16,6	33,1		7,1	8,5	10,2	12,2	24,5
-4			6,9	8,3	10,0	20,0			8,0	9,6	11,5	23,0			5,9	7,1	8,5	17,0
-6				5,8	6,9	13,9				6,7	8,0	16,0				4,9	5,9	11,8
-8					4,8	9,6					5,5	11,1					4,1	8,2

$$x_{t,n} = x_0 \cdot u^{(t+n)}$$

Precalculation

n/t	A 0	A 1	A 2	A 3	A 4	B 0	B 1	B 2	B 3	B 4	C 0	C 1	C 2	C 3	C 4
0	62,4	61,9	58,7	51,5	37,7	66,9	66,5	63,2	56,0	43,4	57,7	57,0	53,3	45,1	32,0
-2		41,7	39,5	34,5	26,2		46,2	43,9	38,9	30,1		37,5	35,1	30,2	22,3
-4			27,0	23,7	18,2			30,7	27,0	20,9			23,3	20,1	15,5
-6				17,5	12,6				19,0	14,5				13,9	10,7
-8					8,8					10,1					7,5

$$\widehat{E}(V_{t,n}) = \beta \cdot \left[\widehat{p} \cdot V_{N-1-k,s+u}^{q} + \widehat{q} \cdot V_{N-1-k,s-d}^{q}\right]$$

Present value

n/t	A 0	A 1	A 2	A 3	A 4	B 0	B 1	B 2	B 3	B 4	C 0	C 1	C 2	C 3	C 4
0	73,4	75,3	74,8	70,9	62,2	78,4	80,3	79,8	75,9	67,2	68,4	69,9	69,1	64,7	54,4
-2		50,8	50,4	47,7	41,7		55,8	55,4	52,7	46,7		46,1	45,5	42,5	36,6
-4			33,9	32,1	28,2			38,7	36,6	32,4			29,9	28,1	24,2
-6				23,3	20,6				26,3	22,6				19,3	16,6
-8					15,6					16,6					11,6

$$V_{N-k,s}^{m} = \max_{q \in S}\left\{x_{k,s}^{q} - C_{m,q} + \beta \cdot \left[\widehat{p} \cdot V_{N-1-k,s+u}^{q} + \widehat{q} \cdot V_{N-1-k,s-d}^{q}\right]\right\}$$

Actual mode

n/t	A 0	A 1	A 2	A 3	A 4	B 0	B 1	B 2	B 3	B 4	C 0	C 1	C 2	C 3	C 4
0	B	B	B	B	B	B	B	B	B	B	A	A	A	A	A
-2		B	B	B	B		B	B	B	B		A	A	A	A
-4			A	A	A			B	B	B			A	A	A
-6				A	C				A	A				A	C
-8					C					A					C

$$Q_{t,s} = \arg\max_{q}\left\{x_{k,s}^{q} - C_{m,q} + \beta \cdot \left[\widehat{p} \cdot V_{N-1-k,s+u}^{q} + \widehat{q} \cdot V_{N-1-k,s-d}^{q}\right]\right\}$$

Calculated results show the company values with flexible modes (actions) for three mode versions; normal, expansion and contraction. Investigated and evaluated are due to beginning mode all three variants. Result values are for three variants following 78,4; 73,4 and 68,4 monetary units.

It is apparent, that if the beginning mode is the A mode, then it is optimal to switch to the B mode, and under unfavourable conditions to the A or C. If the beginning mode is of B then mode is maintained, only under unfavourable circumstances it is switched to the A. If the beginning mode is the C, then it is switched immediately to the A, and under unfavourable conditions to the C. We can conclude that the beginning mode influences distinctly the optimal decision (switching modes) trajectory.

It is possible to demonstrate having assumed the switch costs are null or symmetrical (switching to the mode and back), the optimal decision is to choose the mode with maximum cash flow. However, for unsymmetrical costs and more than two modes used, the optimal decisions are influenced and determined by the future options (choices), which is the consequence of the considerable inertia and hysteresis effect. For example, it should be optimal to postpone the project even if the NPV is positive. Or, continue the production process even though the production cash flow is temporarily under variable cost.

5 Conclusions

The purpose of the paper was to describe, explain and verify possibility to apply the real option methodology for company valuation and evaluation projects. It was intended to apply the flexible multinomial switch options methodology. Groundwork point of the paper was the application of the stochastic dynamic programming problem under Bellman's optimality principle. The replication strategy and risk-neutral approach were applied.

Firstly, the replication strategy and the risk-neutral approach were described and explained. Subsequently, the Bellman's optimality principle and its application in the valuation of company and projects on the basis of recurrent expected present value calculation were demonstrated. Consecutively, generalised recurrent valuation optimisation formula with multinomial decisions (options) by virtue of switch options and Bellman's stochastic dynamic programming formula was derived.

Introduced flexible real option methodology was applied in the example of company valuation with options to select and switch among three modes: normal, expansion and contraction. The same modes were supposed as the beginning ones.

It was explained and verified, that the multinomial flexible switch options approach suitably models the real decision-making and valuation conditions and that typical feature of the process is the considerable inertia and hysteresis effect. Likewise, the sensitivity analysis on the input data, especially switch cost is sufficient. Generalised approach of the sensitivity analysis application by virtue of the fuzzy sets methodology is presented for instance in Zmeškal (1999, 2001, 2004, 2005).

It arises from the preceding explanations and discussions, that it is possible and suitable to apply the real option methodology in a small open economy and transitive phase. Described flexible multinomial switch real option methodology should be considered to be the generalised methodology of a company valuation and project evaluation under flexible conditions and switch options variants. This feature allows getting more information about the economic efficiency of companies and projects.

References

Black, F., Scholes, M., 1973. The Pricing of Options and Corporate Liabilities. Journal of Political Economy 81, 637-659.

Boyle, P., Longstaff, F.A., Ritchken, P. (1995), *Advances in Futures and Options Research*, JAI Press, Vol. 8.

Brennan, M. J., Tigeorgis, L. (1999), *Project, Flexibility, Agency and Product Market Competition: New Development in the Theory and Application of Real Options Analysis*, Oxford university Press.

Damodaran, A. (1994), *Damodaran on Valuation, Security Analysis for Investment and Corporate Finance*, J. Willey & Sons, Inc., 350-351.

Dixit, A. K., Pindyck, R.S. (1994), *Investment under Uncertainty*, University Press.

Dluhošová, D. (2004) *Přístupy k analýze finanční výkonnosti firem a odvětví na bázi metody EVA – Economic Value Added,* Finance a úvěr- Czech Journal of Economics and Finance, 11-12 2004, roč. 54

Dluhošová, D. (eds) *Nové přístupy a finanční nástroje ve finančním rozhodování*, VŠB-TU Ostrava, ISBN 80-248-0669-X. 640s

Hull, J. C. (2000), *Options, Futures, and other Derivatives*. Prentice Hall.

Kulatilaka, N. (1988) Valuing the flexibility of flexible manufacturing system. IEEE Transactions in Engineering Management. 35, No. 4, pp. 250-257

Kulatilaka, N. (1995) The value of flexibility: General model of real options. In chapter 5, pp 89-109, In Trigeorgis, L. (1995) Real options in capital investments, models, strategies and applications. Praeger, Westport

Kulatilaka, N. Trigeorgis, L. (1994) The general flexibility to switch: real options revisited. International Journal of Finance 6, No. 2, pp. 778-798

Musiela, M., Rutkowski, M. (1997), *Martingale Methods in Financial Modelling*, Springer-Verlag, 48-50.

Ronn, E. I. (2002) *Real Options and Energy Management. Using Options Methodology to Enhance Capital Budgeting Decisions*. Risk Waters Group

Sick, G. (1995), Real Options. In Jarrow, R. et all, *Handbooks in OR and MS*, Vol. 9, Elsevier Science B.V., 631-691.

Trigeorgis, L. (1998), *Real Options - Managerial Flexibility and Strategy in Resource Allocation*, Harvard University.

Zmeskal, Z. (2001) Application of the fuzzy - stochastic Methodology to Appraising the Firm Value as a European Call Option, European Journal of Operational Research, Vol. 135/2, pp 303-310

Zmeskal, Z. at al (2004) *Financial models*. VSB-TU, Faculty of Economics, Ostrava

Zmeškal, Z. (1999) Možnosti stanovení hodnoty firmy jako bariérové americké call opce. Sborník Mezinárodní konference In: Ekonomika firiem 1999, Ekonomická univerzita Bratislava

Zmeškal, Z. (2005) Approach to Soft Binomial Real Option Model Application (fuzzy - stochastic approach), The 12th Global Finance Conference, Dublin, Ireland

Zmeškal, Z. a kol. (2004) *Finanční modely*. 2. upravené vydání Praha: Ekopress Praha, 2004

Zmeškal. Z.: (1999) Fuzzy-stochastický odhad hodnoty firmy jako call opce. Finance a úvěr, Economia a. s. Praha, No. 3

Czech Society for Operations Research
Czech Econometric Society
Faculty of Economics, University of West Bohemia in Pilsen

Conference Programme

24th International Conference
Mathematical Methods in Economics 2006
September 13-15, 2006, Pilsen, Czech Republic

Tuesday 2006-09-12
15.00 – 19.00 Registration – room HJ204

Wednesday 2006-09-13
08.00 – 18.00 Registration – room HJ204 (08.00 – 10.00) – room TY213 (13.30 – 17.30)
10.00 – 12.00 Opening & Plenary session – room HJ200
12.00 – 13.30 Lunch
13.30 – 17.10 Parallel sessions – rooms TY214, TY309, TY310, TY404
17.10 – 18.00 General meeting of the Czech Society for Operations Research – room TY214
19.30 – 22.00 Conference dinner – Restaurant ‘Na Spilce’, Pilsen Brewery Prazdroj

Thursday 2006-09-14
08.00 – 13.30 Registration – room TY213
08.30 – 11.45 Parallel sessions – rooms TY214, TY309, TY310, TY404, TY306
11.45 – 13.30 Lunch
13.30 – 20.00 Conference trip – Castle Švihov, Chudenice American Garden

Friday 2006-09-15
08.30 – 12.30 Registration – room TY213
09.00 – 11.50 Parallel sessions – rooms TY214, TY309, TY310, TY404
11.50 – 12.15 Closing session – room TY211

Conference topics (session abbreviations):
CEF – Computational Economics & Finance
D-M – Decision Making
EcM – Econometrics & Modelling
Fin – Finance & Modelling
Opt – Optimization
O-R – Operations Reseach
SMA – Stochastic Modelling & Applications

Wednesday 2006-09-13

10.00 – 12.00 **Opening & Plenary session** – room HJ200, chair **Jablonský Josef**

Průša Josef – Rector of the University of West Bohemia
Hrdý Milan – Dean of the Faculty of Economics

10.30 – 11.45 **Vlach Milan** (Kyoto Colledge of Graduate Studies for Informatics, Kyoto, Japan): **How to guarantee fair shares to everybody**

10.45 – 12.00 **Murgu Alexandru** (BT Networks Research Centre, Ipswich, United Kingdom): **Peer group games in economics of communication networks**

12.00 – 13.30 Lunch

13.30 – 15.10 **Parallel sessions**

TY214	TY309	TY310	TY404
W1 **SMA_1** chair: **Dupačová J.**	**W2** **Fin_1** chair: **Vašíček O.**	**W3** **EcM_1** chair: **Kočenda E.**	**W4** **D-M_1** chair: **Fiala P.**
Klicnarová J. Application of MartingaleApproximations to AR, MA and ARMA Processes	**Zmeškal Z.** Real Option Applications Based on the Generalized Multinomial Flexible Switch Options Methodology	**Hušek R., Švarcová R.** Forecasting Macroeconomic Variables after the Accession of CR into the EU	**Ramík J., Perzina R.** Fuzzy ANP – a New Method and Case Study
Hušková M., Marušiaková M. Change in Mean versus Random Walk: a Simulation Study	**Bena J., Hanousek J.** Rent Extraction by Large Shareholders: Evidence Using Dividend Policy in the Czech Republic	**Musil K., Vašíček O.** New Keynesian Model of the Small Open Czech Economy	**Kalčevová J., Fiala P.** IZAR – Multicriteria Decision Support System
Tonner J. The Principle of Over-completeness in Multivariate Economic Time Series Models	**Fukač M.** Heterogeneous Expectations, Adaptive Learning and Inflation Stabilization in Short Terms	**Hloušek M.** Czech Business Cycle – Stylized Facts	**Fendeková E., Fendek M.** Models of Regulation in Network Industries (in teh Field of Slovak National Electricity Markets)
Kaňková V. Stochastic Programming Problems with Linear Recourse; Applications to Problems of two Managers	**Kopa M.** Stability of Optimal Portfolios: Non-Smooth Utility Approach	**Polanský J., Vašíček O.** Analysis of the Czech Real Business Cycle Model	**Miťková V., Mikušová N.** Gender Gap: a Case of Some European Countries

15.10 – 15.30 Coffe break – room TY213

15.30 – 17.10 **Parallel sessions**

TY214	TY309	TY310	TY404
W5 **SMA_2** chair: **Vošvrda M.**	**W6** **Fin_2** chair: **Zmeškal Z.**	**W7** **EcM_2** chair: **Hušek R.**	**W8** **O-R_1** chair: **Zimmermann K.**
Sladký K. Approximation in Stochastic Growth Models	**Mlynarovič V.** Decision Support Systems for Portfolio Selection	**Hanousek J., Kočenda E., Švejnar J.** Origin and Concentration: Corporate Ownership, Control and Performance	**Fábry J.** Dynamic Traveling Salesman Problem
Houda M. Approximation in Stochastic and Robust Programming Problems	**Pánková V.** Investment under Monetary Uncertainties	**Kodera J., Vošvrda M.** Goodwin's Predator-Prey Model with Endogenous Technical Progress	**Kučera P., Houška M., Beránková M.** Methods for the Multiple-tours Traveling Salesman Problem Making the Final Solution in One Time
Fíglová Z. Analysis of Panel Data with Binary Dependent Variable	**Hrdý M.** Mathematical Methods in Accounting	**Szomolanyi K., Lukáčik J., Lukáčiková** Growth Models	**Horáčková L., Kořenář V., Pelikán J.** Time Limited Vehicle Routing Problem
Aragon F., Caridad J., Villamandos N. Short Term Equilibrium in Press Distribution with Random Demand	**Čulík M.** Real Option Application for Modular Project Valuation	**Salukvadze M., Jibladze N., Obgadze T., Tushishvili N.** Mathematical Modeling of Economic Cycles and Optimal Investment Strategy Working-out	**Melecký J.** A Simple Stock Market Model Involving Delay

17.10 – 18.00 General meeting of the Czech Society for Operations Research – room TY214

19.30 – 22.00 Conference dinner – Restaurant 'Na Spilce', Pilsen Brewery Prazdroj

Meeting point – monumental Main gate of Pilsen Brewery Prazdroj at 19.30

Thursday 2006-09-14

08.30 – 10.10 **Parallel sessions**

TY214	TY309	TY310	TY404	TY306
T1 Opt_1 chair: **Vlach M.**	**T2 D-M_2** chair: **Ramík J.**	**T3 Econometric Day SMA_3** chair: **Sladký K.**	**T4 O-R_2** chair: **Fábry J.**	**T5 sw Workshop**
Fiala P., Flusserová L., Kořenář V. Modeling of Network Competition	**Pavlačka O., Talašová J.** The Fuzzy Weighted Average Operation in Decision-Making Models	**Dupačová J.** Optimization under Exogenous and Endogenous Uncertainty	**Bartl D.** Some Notes on Simplex Method	
Zimmermann K. Some Non-convex Separable Optimization Problems under Synchronization Constraints	**Matsuhisa T., Strokan P.** Bayesian Belief Communication Leading to Nash Equilibrium II	**Reisnerová S.** Bayesian Approach to Change Point Detection of Unemployment Rate via MCMC Method	**Čičková Z., Brezina I.** SOMA Application to the Travelling Salesman Problem	09.00 **Vošvrda M.** Mathematica
Jablonský J. DEA Models with Random Inputs and Outputs	**Talašová J.** Fuzzy Evaluations in Decision-Making Models	**Šmíd M.** Optimal Strategies at a Limit Order Market	**Dlouhý M., Novosádková I.** Models of Efficiency Evaluation of Hospitals in the Czech Republic	
Lagová M., Kalčevová J. Computer Support of Courses of Linear Optimization Models	**Gavalec M., Mls K.** Cognitive Hierarchy Process – an Approach to Decision Making Support	**Čerbáková J.** Stability of Bayes Actions		

10.10 – 10.30 Coffe break, room TY213

10.30 – 11.45 **Parallel sessions**

TY214	TY309	TY310	TY404	TY306
T6 SMA_4 chair: **Kaňková V.**	**T7 Fin_3** chair: **Hrdý M.**	**T8 Econometric Day EcM_3** chair: **Kodera J.**	**T9 D-M_3** chair: **Talašová J.**	**T10 sw Workshop**
Fitzová H. Bayesian Estimation of Closed and Open Czech Economy Model	**Cahlík T., Hlaváček J., Chytilová J., Reichlová N., Švarc P.** Multi-agent Approaches in Economics	**Menez R., Bentes S., Mendes D.** Nonlinear Dynamic Volatility in Stock Markets: from Asymmetric Signals to Econophysics Modeling	**Kubát J.** Electricity Market Game	**Daněk J.**, Humusoft ® Matlab
Volf P. Markov Chain Monte Carlo Methods in Computational Statistics and Econometrics	**Škovránková L., Škovránková P.** Some Application of Statistical Information Theory	**Šedivá B.** Dynamic Economic Modeling with Time Delay Variables	**Charouz J.** Fuzzy Logic as Liquidity Solution in a Bank	
Gorka J., Osiń ska M. STUR Tests and their Sensitivity for Non-linear Transformations and GARCH. A Monte Carlo Analysis	**Marček D.** Application of Dynamic Models and a SV Machine to Inflation Modeling	Meeting of the Czech Econometric Society	**Matusik S.** Demographic Factors in Economic Development in Communes of the Malopolskie Voivodeship	

13.30 – 20.00 Conference trip – Castle Švihov, Chudenice American Garden

Meeting point – bus station (public transport) Husova street, direction to the Theatre, near the Faculty building, Husova 11

1-st bus (english speaking guide) departure time 13.30

2-nd bus (czech speaking guide only) departure time 14.00

Return – Chudenice American garden – departure time 19.00, both buses leave together – arrival to Pilsen, Husova street, about 20.00

Friday 2006-09-15

09.00 – 10.15 **Parallel sessions**

TY214	TY309	TY310	TY404
F1 **SMA_5** chair: **Volf P.**	**F2** **Fin_4** chair: **Mlynarovič V.**	**F3** **EcM_4** chair: **Moravanský D.**	**F4** **O-R_3** chair: **Dlouhý M.**
Hofman J., Lukáš L. Measurement of Supplier-Customer System Complexity Based upon Entropy	**Jindrová M.** Exchange Rate Disconnect in Model of Exchange Rate Indetermination	**Martinčík D., Šedivá B.** Extended IS-LM Model – Construction and Analysis of Behaviour	**Šajtarová M.** Money Distribution as Network Flow Problem
Kuncová M. Practical Application of Monte Carlo Simulation in MS-Excel and its Add-ons – The Optimal Mobile Phone Tariffs for Various Types of Consumers in The Czech Republic	**Bruzda J.** Testing for Logistic and Exponential Smooth Transition Cointegration	**Hančová M.** Kriging as a Prediction Tool in Economic Time Series	**Mikušová N., Mitková V.** Dynamic Production Inventory Model
Kuchynka A. Empirical Application of Threshold and Smooth Transition Models	**Lukáš L.** Derivation of Exchange Rate Computer-Agent Models Using Dynamic Clearing Conditions	**Hančlová J.** Minimum Wage Impact on Wage and Unemplozment in the CR	**Friebelová J., Friebel L.** Using the Weibull Distribution for Simulation of Machine Lifetime

10.15 – 10.35 Coffe break – room TY213

10.35 – 11.50 **Parallel sessions**

TY214	TY309	TY310	TY404
F5 **Opt_2** chair: **Pánková V.**	**F6** **Fin_5** chair: **Marček D.**	**F7** **EcM_5** chair: **Cahlík T.**	**F8** **CEF_1** chair: **Lukáš L.**
Hladík M. Separation of Convex Polyhedral Sets with Uncertain Data	**Frank J., Gavalec M.** Volatility of Prices in a Multiple Relation of Composite Commodities	**Němec D., Moravanský D.** Testing of Hysteresis in Unemployment	**Vácha L., Vošvrda M.** Wavelet Analysis in the Heterogeneous Agents Model
Hajduková J. Voting Location Problem on Some Special Graphs	**Tichý T.** Transaction Costs and Option Portfolio	**Wawerková R., Moravanský D.** Consumer Price Index and its Biases	**Tran Q.** Testing Nonlinear Dependence in Czech Stock Index PX-50 Returns
Pražák P. Green’s Theorem and Optimal Control in Economics	**Pígl J**. Sets of Admissible and Effective Portfolios in Time and their Modification by Term Contracts	**Műller L.** The Floods – a New Economical Problem	**Vácha O., Lukáš L.** Interdependent Consumer Behaviour – Numerical Investigation of Stability

11.50 – 12.15 **Closing session –** room TY211

Proceedings of the 24th International Conference

Mathematical Methods in Economics 2006

13th – 15th September 2006
Pilsen
Czech Republic

Editor: **Ladislav Lukáš**
Technical editors: **Ludmila Bokrošová , Kateřina Štruncová**
Publisher: **University of West Bohemia in Pilsen, Univerzitní 8, 306 14, Pilsen**
Printing: **TYPOS – Digital Print, spol. s r.o., Pilsen**

Issue: **130 copies**

ISBN 80-7043-480-5