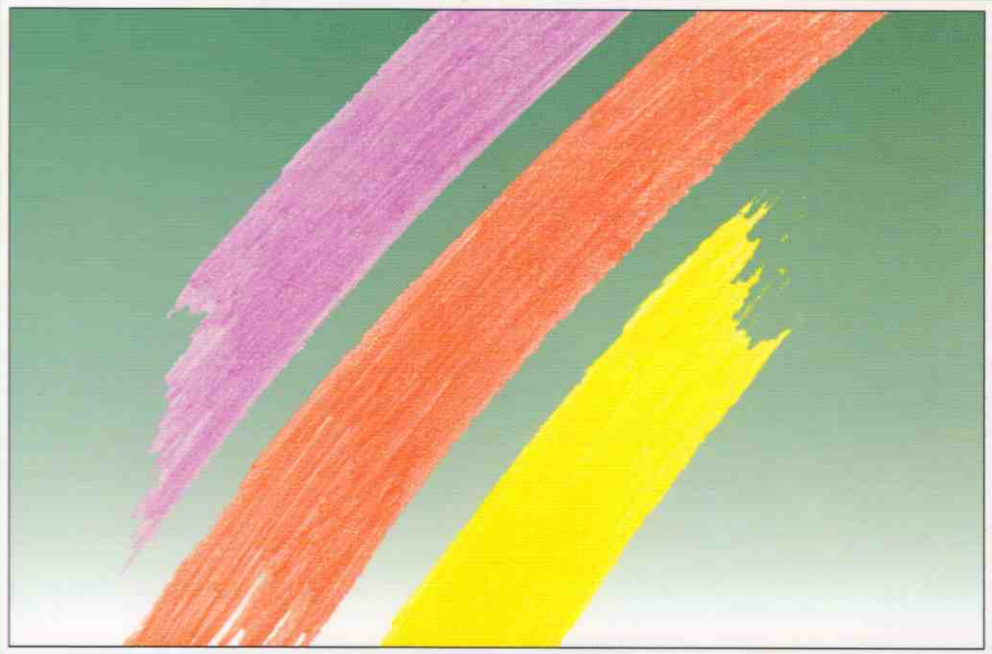
Quantitative Methods in Economics
(Multiple Criteria Decision Making XIV)



Proceedings of the International Conference
June 5 – 7, 2008
High Tatras, Slovak Republic

Iura Edition
člen skupiny Wolters Kluwer

The Slovak Society for Operations Research
Department of Operations Research and Econometrics
Faculty of Economic Informatics
University of Economics in Bratislava

Proceedings of the International Scientific Conference

QUANTITATIVE METHODS IN ECONOMICS
Multiple Criteria Decision Making XIV

5th - 7th June 2008
High Tatras, Slovakia

Programme Committee:
Brezina Ivan, University of Economics, Bratislava
Bobro Peter, IURA Edition, Bratislava
Fendek Michal, University of Economics, Bratislava
Fiala Petr, University of Economics, Praha
Hozlár Eduard, B.C. Consulting, Bratislava
Ivaničová Zlatica, University of Economics, Bratislava
Luptáčik Mikuláš, Technische Universität, Wien
Mlynarovič Vladimír, Comenius University, Bratislava
Turnovec František, Charles University, Praha

Organizing Committee:
Ivaničová Zlatica, University of Economics, Bratislava (chair)
Pekár Juraj, University of Economics, Bratislava
Surmanová Kvetoslava, University of Economics, Bratislava
Reiff Marian, University of Economics, Bratislava
Lukáčik Martin, University of Economics, Bratislava
Szomolányi Karol, University of Economics, Bratislava

Referees:
Brezina Ivan, University of Economics, Bratislava
Fendek Michal, University of Economics, Bratislava
Goga Marián, University of Economics, Bratislava
Husár Jaroslav, University of Economics, Bratislava
Ivaničová Zlatica, University of Economics, Bratislava
Koval' Ladislav, University of Economics, Bratislava
Mlynarovič Vladimír, Comenius University, Bratislava

5th - 7th June 2008
High Tatras, Slovakia

Technical Editor: *Reiff Marian*, University of Economics, Bratislava
Web Editor: *Lukáčik Martin*, University of Economics, Bratislava
Web: http://www.fhi.sk/sk/katedry/kove/ssov/papers

ISBN 978-80-8078-217-7

CONTENT

A QUEUEING THEORY APPLICATION TO THE DESIGN OF A COMMUNICATION SYSTEM FOR DATA TRANSMISSION IN URBAN MASS TRANSPORTATION

Ondrej Bartl, University of Žilina, Slovakia

Abstract: A queueing theory approach can help in revealing the answer to the question on the sufficiency of a base transceiver station capacity required for efficient data transmission in communication systems for urban mass transportation. Markovian queueing models have been built to obtain both an optimistic and a pessimistic estimate of the level of efficiency at data transmission for a considered capacity of a base transceiver station.
Keywords: Markovian queueing models, reneging of customers, losses of customers, capacity of a server, level of service efficiency

1. INTRODUCTION

Information on a current situation in urban mass transportation in comparison with the prescribed time-table is quite important for the control centre of a public transport company. A communication system is therefore built to transmit position data from vehicles of public transport to the control centre. The infrastructure of such a communication system is usually based on the TETRA (Terrestrial Trunked Radio) network composed of a set of base transceiver stations. A base transceiver station (BTS) receives digital radio signals with data from public transport vehicles moving inside its attraction area and transmit them further to the control and data processing centre of the public transport company. The data from vehicles are sent in the form of a short data service (SDS) message. A base station can only receive a limited number of SDS messages per unit time, which states the projected disposable capacity of the base station. When a communication channel for data transmission is busy by transmitting data from a vehicle to the base station, a new short data message from another vehicle cannot be transmitted and has to wait in the queue to be transmitted later. Waiting time in the queue is limited by a fixed limit and the message is cancelled as soon as the time limit is reached. It is important to know the fraction of short data messages that are transmitted. This ratio between the transmitted and all short data messages generated in the attraction area of the base station determines the level of efficiency at data transmission. If the level is high enough to be accepted, the disposable capacity of the base station is sufficient for the needs of the communication network. Queueing models can help at the analysis of the problem with the base station capacity. Queueing models should be tractable and they should describe the reality with some reasonable level of agreement. Markovian queueing models have been built to analyse the behaviour of the communication system for data transmission between the vehicles of a public transport company and the base transceiver station. As the models are some approximations to the reality, two types of queueing models have been employed, the one with reneging of customers and the other with losses of customers.

To classify queueing models, modified extended Kendall's shorthand notation [3] in the form of a six-part code $a/b/c/d/e/f$ can be used. The first symbol *a* specifies the interarrival-time distribution, the second symbol *b* specifies the service-time distribution, the third symbol *c* specifies the number of servers, the forth symbol *d* specifies the limit for the number of customers staying in the system at the same time, the fifth symbol *e* specifies the size of the source of population of potential customers and the sixth symbol *f* specifies the time limit for the waiting of a customer until service. When items of the shorthand code are associated with time, symbol *D* is used for deterministic time and symbol *M* for random time equipped with the memoryless property related to the exponential probability distribution. Memorylessness of all time distributions in a queueing model implies the Markovian nature of the model.

2. QUEUEING MODEL WITH RENEGING OF CUSTOMERS

A queueing model with reneging of customers analyses the activity of a queueing system with impatient customers. Any customer in such a system waits in the queue for a limited length of time. If the service of a customer does not start within the limited period of time, the customer leaves the queue without being served. That corresponds with the reality of data transmission between vehicles and the base transceiver station. Short data messages are customers, the base station (its communication channel for receiving data from vehicles) is a server and the transmission of a short data message from a vehicle to the base station is service. In order to obtain a tractable model, the deterministic fixed time limit for the waiting of a message in the queue has been replaced by a random exponentially distributed time limit.

When the board computer in a vehicle generates a new short data message containing identification, time and position data about the vehicle, a request-to-send the message is sent to the corresponding base transceiver station. If the communication channel for transmitting data messages is idle, the data message is transmitted from the vehicle to the base station. Otherwise the data message joins the queue to wait for being transmitted later. New data messages are generated according to some scenario, for example, when public transport vehicles leave the stops. The Poisson process can represent the arrival process of requests-to-send short data messages to the base transceiver station from the vehicles in the attraction area of the base station. It means that interarrival times are independent and identically distributed random variables V_k, $k = 1,2,\ldots$, each having the same exponential probability distribution with mean $1/\lambda$. An exponentially distributed random variable, say V, is equipped with the memoryless property $P\{V \le t + \Delta t | V > t\} = P\{V \le \Delta t\}, \forall \Delta t \ge 0, \forall t \ge 0,$ as shown below:

$$\begin{aligned} P\{V \le t + \Delta t | V > t\} &= P\{V \le t + \Delta t, V > t\}/P\{V > t\} = P\{t < V \le t + \Delta t\}/P\{V > t\} \\ &= [P\{V \le t + \Delta t\} - P\{V \le t\}]/P\{V > t\} = \left[\left(1 - e^{-\lambda(t+\Delta t)}\right) - \left(1 - e^{-\lambda t}\right)\right]/e^{-\lambda t} \\ &= 1 - e^{-\lambda \Delta t} = P\{V \le \Delta t\}, \quad \forall \Delta t \ge 0, \ \forall t \ge 0. \end{aligned} \tag{1}$$

If variable V represents interarrival time, the equation above says that the probability of the occurrence of a new arrival within the next short time interval Δt is the same when the time period t elapsed up to the current time point since the last arrival is either long or short. That corresponds to arrivals of request-to-send announcements at the base transceiver station, because the requests are sent by various public transport vehicles from various parts of the area covered by the base station and the number of the vehicles moving simultaneously inside the base station area is relatively high. Random variable $N(t)$ representing the number of arrivals in the time interval $[0,t]$ is Poisson distributed with mean $E[N(t)] = \lambda t$ for $t \ge 0$ and $\lambda > 0$. The probability mass function is given by $P\{N(t) = k\} = e^{-\lambda t} (\lambda t)^k / k!$, $t \ge 0$, $k = 0,1,\ldots$. The expected number of arrivals per unit time, $E[N(1)]$, is the rate of the Poisson arrival process $\{N(t), t \ge 0\}$. As $E[N(1)] = \lambda$, the requests-to-send a short data message arrive at the base transceiver station with rate λ requests per unit time. The size of parameter λ depends on the scenario for sending short data messages and it can be obtained from the time-table of urban mass transportation. The value typical of peak hours is taken into account for the analysis of the communication system inside the base station area.

Service time is the time period of transmitting a short data message from a public transport vehicle to the base station together with the accompanying communication operations to start and complete data transmission. Some random fluctuations in the whole transmission time are quite natural. Service time can therefore be considered as a random variable (with prevailing short time values). An exponentially distributed random variable, say U, with mean $1/\mu$ fits these features. The expected number of short data messages that the base transceiver station is able to receive per unit time is given by $1/E[U] = \mu$, which states the disposable capacity of the base station.

As there is one communication channel for data transmission between public transport vehicles and the base station, the queueing system to model the communication consists of one server. The length of a queue for short data messages waiting to be transmitted later is not limited. Thus, the number of customers concurrently occupying the queueing system is unlimited. The source of population of potential customers can be considered infinite, because many public transport vehicles are moving inside the attraction area of the base station at the same time and they often send short data messages to the base station. So, the number of requests-to-send a data message arriving at the base station per unit time is relatively large.

The time to wait until service is limited for each data message by a fixed limit. This deterministic waiting time limit would require to have included the history of customer arrivals in the state information on the current situation in the queueing system. To build a tractable queueing model, it is convenient to replace the fixed limit by an exponentially distributed random variable, say *T*, with mean $1/\nu$. Due to the memoryless property of the exponential probability distribution the resulting queueing system to model the reality possesses the Markovian property, because all the time distributions in the model are exponential. The customers that wait in the queue and do not renege are supposed to be served in accordance with the first-come-first-served queue discipline. Hence, the queueing system to analyse the operation of the communication system inside the attraction area of a base station is a Markovian queueing system denoted by the code $M/M/1/\infty/\infty/M$ in the shorthand notation.

A continuous-time Markov chain $\{X(t), t \geq 0\}$ describes the evolution of the queueing system of the $M/M/1/\infty/\infty/M$ type. Random variable $X(t)$ represents the state of the queueing system at time *t*, where the size of the state is given by the number of customers in the system (in service plus in the queue) at time *t*. As interarrival times, service times and waiting times until reneging are memoryless random variables, they can be regarded as starting to pass at the time point when the current state of the system is observed. The Markovian property, when the future depends on the history only through the present state and the past history is not important, is therefore filled in accordance with the following equation [2]

$$P\{X(s+t) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\} = P\{X(s+t) = j \mid X(s) = i\}$$

$$\forall t \geq 0, \forall s \geq 0, \forall x(u), i, j \in S, 0 \leq u < s,$$

(2)

where S is the state space of the system, i.e. the set of all possible values of the state variable. The set S is given by $S = \{0,1,\ldots\}$. Since the operation of the communication system is analysed for peak hours, the stationarity of transition probabilities $P\{X(t+\Delta t)= j|X(t)=i\}$, $i, j \in S$, can be considered. Hence,

$$p_{i,j}(\Delta t)= P\{X(\Delta t)= j|X(0)=i\}= P\{X(t+\Delta t)= j|X(t)=i\},\ i, j \in S, \Delta t \geq 0, t \geq 0.$$

(3)

Probabilistic laws of motion describing the system evolution are based on the law of total probability. Using this law, the probability of the system state at time $t+\Delta t$ can be expressed for any state value i as follows

$$P\{X(t+\Delta t)=i\}=\sum_{j\in S} P\{X(t+\Delta t)=i|X(t)= j\}P\{X(t)= j\},\ i \in S, t \geq 0, \Delta t \geq 0.$$

(4)

Taking into account the stationarity of transition probabilities and employing the notation $p_i(t)= P\{X(t)=i\}$, $i \in S, t \geq 0$, for the system state probability, we have

$$p_i(t+\Delta t)=\sum_{j\in S} p_{j,i}(\Delta t)p_j(t),\ i \in S, t \geq 0, \Delta t \geq 0.$$

(5)

Equation (5) can be manipulated as shown below

$$p_i(t+\Delta t)= \sum_{j\in S, j\neq i} p_{j,i}(\Delta t)p_j(t)+ p_{i,i}(\Delta t)p_i(t) \quad \Leftrightarrow$$

$$p_i(t+\Delta t)= \sum_{j\in S, j\neq i} p_{j,i}(\Delta t)p_j(t)+\left[1-\sum_{j\in S, j\neq i} p_{i,j}(\Delta t)\right]p_i(t) \quad \Leftrightarrow$$

$$\lim_{\Delta t\to 0}\frac{p_i(t+\Delta t)- p_i(t)}{\Delta t} = \sum_{j\in S, j\neq i}\left(\lim_{\Delta t\to 0}\frac{p_{j,i}(\Delta t)}{\Delta t}\right)p_j(t)- \sum_{j\in S, j\neq i}\left(\lim_{\Delta t\to 0}\frac{p_{i,j}(\Delta t)}{\Delta t}\right)p_i(t) \quad \Leftrightarrow$$

$$\frac{dp_i(t)}{dt} = \sum_{j\in S, j\neq i} q_{j,i}\, p_j(t)- \sum_{j\in S, j\neq i} q_{i,j}\, p_i(t) \quad \Leftrightarrow$$

$$p_i'(t)= \sum_{j\in S, j\neq i} q_{j,i}\, p_j(t)+\left[-\sum_{j\in S, j\neq i} q_{i,j}\right]p_i(t)$$

to obtain the system of differential equations for system state probabilities

$$p_i'(t)= \sum_{j\in S, j\neq i} q_{j,i}\, p_j(t)+ q_{i,i}\, p_i(t), \quad i \in S, t \geq 0.$$

(6)

If the limits $p_i = \lim_{t\to\infty} p_i(t)$, $i \in S$, exist, we have that $p_i(t) = p_i$ and $p_i'(t) = 0$, $i \in S$, for time t approaching infinity. The system of differential equations (6) takes then the form of the system of linear algebraic equations for limiting system state probabilities

$$0 = \sum_{j\in S,\, j\neq i} q_{j,i}\, p_j + q_{i,i}\, p_i, \quad i \in S. \tag{7}$$

The system (7) requires to have added the equation

$$1 = \sum_{i\in S} p_i \tag{8}$$

stating that unknown variables p_i, $i \in S$, are probabilities. Limiting system state probabilities describe the probabilistic behaviour of a queueing system in the steady-state mode of operation. They are used to obtain performance measures of the queueing system.

To analyse the behaviour of the $M/M/1/\infty/\infty/M$ queueing system that approximately models the operation of the communication system inside the attraction area of the base transceiver station, the system of linear algebraic equations (7) must be built. When the transition rates $q_{i,j} = \lim_{\Delta t\to 0} p_{i,j}(\Delta t)/\Delta t$, $j \in S, j \neq i, i \in S$, are specified, the rates $q_{i,i}, i \in S$, are given by $q_{i,i} = -\sum_{j\in S,\, j\neq i} q_{i,j}, i \in S$. As $\sum_{j\in S} p_{i,j}(\Delta t) = 1, i \in S, \Delta t \geq 0$, it follows that rates $q_{i,i}, i \in S$, can also be expressed by $q_{i,i} = \lim_{\Delta t\to 0} -[1 - p_{i,i}(\Delta t)] \,/\, \Delta t$, $i \in S$. Since the aforementioned modelling queueing system is a Markovian queueing system, transition rates $q_{i,j}$, $j \neq i$, are determined by standard techniques employed at the analysis of such systems. All the time distributions considered in the analysed queueing system are exponential probability distributions. Then the probabilities of the concurrent occurrence of two or more changes of a current system state within a very short time interval, i.e. for $\Delta t \to 0$, are negligibly small in comparison with Δt. Therefore $q_{i,j} = \lim_{\Delta t\to 0} p_{i,j}(\Delta t)/\Delta t = 0$ for $i, j \in S, |j - i| > 1$. The only non-zero transition rates are those representing the change of the system state by one customer. We find out, that $q_{k,k+1} = \lambda$ for $k \in S$, which is associated with the probability $\lambda\Delta t + o(\Delta t)$ of one arrival of a customer during a very short time interval. The symbol $o(\Delta t)$ denotes any function of Δt such that

$\lim_{\Delta t \to 0} o(\Delta t)/\Delta t = 0$. Further we obtain, that $q_{k,k-1} = \mu + (k-1)\nu$ for $k \in S - \{0\}$, which is related to the probability that one customer leaves the queueing system within a very short time interval. This probability is a probability of the union of disjoint events. The current number k customers in the system can be decreased by one when either a customer has service completed with probability $\mu\Delta t + o(\Delta t)$ or a reneging customer leaves the queue with probability $\nu\Delta t + o(\Delta t)$ and there are $k-1$ waiting customers who could renege. The rates $q_{k,k}$ are then given by $q_{0,0} = -q_{0,1}$ for $k = 0$ and by $q_{k,k} = -(q_{k,k-1} + q_{k,k+1})$ for $k \in S - \{0\}$.

The resulting system of linear algebraic equations is specified as follows

$$0 = -\lambda\, p_0 + \mu\, p_1$$
$$0 = +\lambda\, p_{k-1} - [\lambda + \mu + (k-1)\nu] p_k + (\mu + k\nu) p_{k+1}, \quad k = 1,2,\dots \tag{9}$$

together with the condition (8). Using substitutions $z_k = -\lambda\, p_k + (\mu + k\nu) p_{k+1}$, $k = 0,1,2,\dots$, that lead from (9) to recurrent relations $p_{k+1} = [\lambda/(\mu + k\nu)] p_k$, $k = 0,1,2,\dots$, we obtain that

$$p_k = \frac{\lambda^k}{\prod_{i=0}^{k-1} (\mu + i\nu)} p_0, \quad k = 1,2,\dots \tag{10}$$

The probability p_0 is determined with the help of equation (8), which gives that

$$p_0 = \left[1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{\prod_{i=0}^{k-1} (\mu + i\nu)} \right]^{-1} \tag{11}$$

To get numerical results for limiting probabilities, it is convenient to use the technique of relative numbers $q_k = p_k / p_0$, $k = 0,1,2,\dots$, and calculate the values of q_k recurrently up to the value, say q_{m+1}, that is sufficiently small [1]. Then we have the following approximation obtainable from (8): $1/p_0 = \sum_{k=0}^{m} q_k + \sum_{k=m+1}^{\infty} q_k \approx \sum_{k=0}^{m} q_k$. Hence, $p_0 \approx 1/\sum_{k=0}^{m} q_k$, and subsequently $p_k \approx q_k / \sum_{k=0}^{m} q_k$, $k = 1,2,\dots,m$, and $p_k \approx 0$, $k = m+1, m+2, \dots$.

Recurrent equations $q_{k+1} = \left[\lambda/(\mu + k\nu)\right]q_k$, $k = 0,1,2,\dots$, to calculate relative numbers q_k follow from recurrent relations for limiting probabilities and the starting relative value is $q_0 = p_0/p_0 = 1$.

As soon as approximate values of limiting system state probabilities are available, (long-run) performance measures of the queueing system can be calculated. These measures reflect the behaviour of the system in the steady-state mode of operation. The important quantities are represented by random variables *K*, *L*, *S*. Variable *K* denotes the number of customers staying in the queueing system at any time point *t* (when *t* is large, i.e. $t \to \infty$). The number *K* comprises customers in the queue and a customer in service. The quantity *K* is thus the value of the system state $X(t)$ for *t* large. Variable *L* denotes the number of customers waiting in the queue at any time point *t* (when *t* is large). Variable *S* denotes the number of customers in service at any time point *t* (when *t* is large). Approximate values of the expectations of random variables *K*, *L*, *S* are obtained as follows

$$E[K] \approx \sum_{k=0}^{m} kP\{K = k\} = \sum_{k=0}^{m} k\, p_k \tag{12}$$

$$E[L] \approx \sum_{l=0}^{m-1} lP\{L = l\} = \sum_{l=0}^{m-1} lP\{K = l+1\} = \sum_{l=0}^{m-1} l\, p_{l+1} \tag{13}$$

$$\begin{aligned} E[S] &= \sum_{s=0}^{1} sP\{S = s\} = 0P\{S = 0\} + 1P\{S = 1\} \\ &\approx 0P\{K = 0\} + 1P\{1 \le K \le m\} = \sum_{k=1}^{m} P\{K = k\} = \sum_{k=1}^{m} p_k \end{aligned} \tag{14}$$

Now we can determine an approximate value for the level of efficiency at data transmission inside the attraction area of the base transceiver station. This level is given by the (long-run) ratio between the number of transmitted short data messages and the number of all short data messages generated by board computers in public transport vehicles in accordance with a considered scenario for sending data messages. The values of both quantities are those concerning data transmission within the area covered by the base station for peak hours. The level of efficiency denoted by η can be obtained using expected values of relevant random variables for the time period of one unit of time and the steady-state mode of system operation. The expected number of all short data messages generated per unit time inside the attraction area of the base station is given by the arrival rate λ. The data messages that cannot be transmitted immediately because the communication channel for data transmission is busy have to wait in the queue to be transmitted later. The expected number of data messages waiting in the queue at any point of time is given by the expected value $E[L]$ of customers staying in the queue. Waiting time of a customer until reneging is considered to be the exponentially distributed random variable T with mean $E[T] = 1/\nu$ in our approximate queueing model. Thus, any waiting place in the queue occupied by a customer can produce on average ν reneging customers per unit time. Since $E[L]$ waiting places are on average occupied by customers at any time instant, the expected number of reneging customers per unit time is given by the product $\nu\, E[L]$. The reneging of a customer is the same as the cancellation of a short date message because of its too long time of waiting in the queue. Hence, the expected number of short data messages that are transmitted not being cancelled per unit time is stated by the expression $\lambda - \nu\, E[L]$. The level of efficiency at data transmission is then determined by the fraction

$$\eta = \frac{\lambda - \nu\, E[L]}{\lambda} \tag{15}$$

The level η can also be interpreted as the probability of the transmission of a data message, whilst the number $1-\eta$ stands for the probability of the cancellation of a data message. If the level of data transmission efficiency seems to be acceptable as quite great, then the disposable capacity of the base transceiver station (given by the expected number μ of data messages that the base station is able to receive per unit time) can be evaluated as convenient for data transmission. The level of efficiency calculated with the help of the $M/M/1/\infty/\infty/M$ queueing model is an approximation to the efficiency level in a real communication system, the features of which more or less correspond with an $M/M/1/\infty/\infty/D$ queueing system. A fixed limit h for the duration of the waiting time of any short data message until its cancellation has been replaced by a random variable T exponentially distributed with mean $1/\nu$. It means that some messages in the model will wait until cancellation for a shorter period of time and some for a longer one in comparison with the given fixed limit h. The size of the mean $1/\nu$ to model the time to cancellation can equal h or it can be chosen on the basis of comparing numerical results of probabilities such as $P\{T > h\} = e^{-\nu h}$,

$P\{T \le g\} = 1 - e^{-\nu g}$, $P\{a \le T \le b\} = e^{-\nu a} - e^{-\nu b}$ for various values of the mean $1/\nu$ where, for example, $g = \frac{1}{3}h$, $a = \frac{2}{3}h$, $b = \frac{4}{3}h$. As the model allows for some data messages to wait in the queue longer then a fixed time limit h and the situation in a real communication system is influenced by many other factors not included in the model, the level of efficiency at data transmission obtained from the model can be considered to represent an upper limit of a real efficiency level. The level η given by (15) is therefore an optimistic estimate of a real level.

3. QUEUEING MODEL WITH LOSSES OF CUSTOMERS

To obtain a lower limit for the level of efficiency at data transmission, a queueing model based on the $M/M/1/1/\infty/0$ queueing system can be built. This system has a Poisson arrival stream of customers with rate λ customers arriving per unit time, an exponentially distributed service time with mean $1/\mu$ time units, one server, no queue, an infinite source of population of potential customers and the zero limit for customer's waiting time. Since such a queueing system is not equipped with waiting places, the maximum number of customers staying in the system at the same time is equal to the number of servers that equals one. Customers cannot wait for service, i.e. the limit for the waiting time of a customer is equal to zero. Any customer arriving at the queueing system is rejected without being served, in case the server is busy when the customer arrives. Each short data message is thus immediately cancelled if the communication channel for data transmission is not idle when a request-to-send the message arrives at the base transceiver station. No additional attempt to transmit a data message rejected is permitted. It is clear that the number of short data messages transmitted according to the model is less than the number in a real communication system. Hence, the level of efficiency at data transmission obtained from such a model represents a lower limit, i.e. a pessimistic estimate, for a real level.

The $M/M/1/1/\infty/0$ queueing system is a Markovian queueing system with losses of customers. The corresponding mathematical model is built in a similar way as in the previous queueing model. The state $X(t)$ of the system at time *t* is represented by the number of customers staying in the system at time *t*. The set of possible values of the system state is given by $S = \{0,1\}$. The system of linear algebraic equations for limiting systems state probabilities has the form of two equations

$$0 = -\lambda\, p_0 + \mu\, p_1$$
$$0 = +\lambda\, p_0 - \mu\, p_1$$
(16)

together with the additional equation stating the condition for probabilities

$$1 = p_0 + p_1$$
(17)

The solution to the system (16), (17) is given by

$$p_0 = \frac{1}{1+\alpha}, \quad p_1 = \frac{\alpha}{1+\alpha}, \quad \text{where } \alpha = \frac{\lambda}{\mu}.$$
(18)

The limiting probability $p_0 = \lim_{t\to\infty} P\{X(t) = 0\}$ represents the probability for the occurrence of the situation when the queueing system is empty and the server is thus idle. An arriving customer is therefore received and served. Hence, the probability p_0 is the probability of transmitting a short data message and determines the (long-run) level of efficiency at data transmission. On the other hand, the limiting probability $p_1 = \lim_{t\to\infty} P\{X(t) = 1\}$ represents the probability for the occurrence of the situation when the queueing system is occupied by a customer in service. An arriving customer is therefore rejected having service denied forever. Hence, the probability p_1 is the probability of the cancellation of a short data message.

If the lower limit for the level of efficiency at data transmission represented by the probability p_0 obtained from the $M/M/1/1/\infty/0$ queueing model does not seem to be high enough, the scenario for sending short data messages can be modified. For example, data messages will be sent only when a public transport vehicle leaving a stop is behind the planned departure time prescribed by the time-table and the delay is greater than or equal to a given difference. The fraction of departures with such a delay is usually available, because public transport companies keep statistical data on the operation of urban mass transportation. The fraction can be interpreted as the probability for the occurrence of the aforementioned delay. Multiplying the arrival rate λ by this probability of delay, a modified arrival rate λ_{mod} can be obtained to represent the expected number of short data messages generated per unit time according to the modified scenario for sending data messages. Then the corresponding lower limit for the level of efficiency at data transmission can be calculated from the queueing model with losses of customers as well as the upper limit can be obtained from the model with reneging of customers. These estimates for the efficiency level should be interpreted as those corresponding to the transmission of more relevant data messages in comparison with the scenario when data messages are generated at each departure of a public transport vehicle from a stop inside the attraction area of the base transceiver station.

REFERENCES

1. Peško, Š., Smieško, J.: Stochastické modely operačnej analýzy. (Stochastic Models of Operational Research). University of Žilina, Žilina, 1999, (in Slovak).
2. Ross, S. M.: Stochastic Processes. 2nd edition. Wiley, New York, 1996.
3. Tijms, H.: Stochastic Models. An Algorithmic Approach. Wiley, Chichester, 1994.

This work was supported by the grant VEGA 1/3775/06.

Ondrej BARTL,
University of Žilina, Faculty of Management Science and Informatics,
Univerzitná 1, 010 26 Žilina, Slovakia,
e-mail: bartl@kst.fri.uniza.sk

APPLICATION OF CUSP CATASTROPHE THEORY TO U.S. STOCK MARKET CRASHES

Jozef Baruník[1,2] , Miloslav Vošvrda[1,2]

[1]Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague
[2]Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague

Abstract:

The paper is one of the first attempts to fit the cusp catastrophe theory to stock market data. We show that the cusp catastrophe model explains the crash of stock exchanges much better than alternative linear and logistic models. On the data of U.S. stock markets we demonstrate that the crash of October 19, 1987 may be better explained by cusp catastrophe theory, which is not true for the crash of Sept. 11, 2001. With the help of sentiment measures, such as index put/call options ratio and trading volume (the former models the chartists, while the latter the fundamentalists), we have found that the 1987 returns are clearly bimodal and contain bifurcation flags. The cusp catastrophe model fits these data better than alternative models. Therefore we may say that the crash may have been led by internal forces. However, the causes for the crash of 2001 are external, which is also evident in much weaker presence of bifurcations in the data. Thus alternative models may be used for its explanation.
JEL: C01, C53
Keywords: cusp catastrophe, bifurcations, singularity, nonlinear dynamics, stock market crash

1. Introduction

Unexpected stock market crashes has been a nightmare for the financial world ever since the capital market existed. The catastrophe theory attempts to unfold a part of information we might need to understand the crash phenomenon. It describes how small, continuous changes in control parameters, or independent variables influencing the state of the system, can have sudden, discontinuous effects on dependent variables. In the paper, we apply the theory to sudden stock market changes that are known as crashes. Zeeman (1974) was the first to qualitatively describe the *"unstable behavior of stock exchanges"* by Thom's (1975) catastrophe theory. We extend his ideas by incorporating quantitative analysis.

The article is rather empirical as it puts the theory to test on financial data.

As only a few papers deal with an empirical analysis of catastrophe theory, this paper may contribute to this research. We build on the Zeeman's qualitative description, and primary aim of the research is to answer the question of whether catastrophe models are capable of indicating the stock market crashes.

What we regard as the most significant aspect is testing on the real-world financial data. Our key assertion is that the cusp catastrophe model is able to fit the data more properly than an alternative linear regression model, and/or nonlinear (logistic) model. We fit the catastrophe model to the data of October 19, 1987 crash, known as Black Monday which was the greatest single-day loss (31%) that Wall Street has ever suffered in continuous trading. As for comparison, we use another large crash, that of September 11, 2001. The final part is devoted to the assumption that while in 1987 the crash was caused by internal forces, in 2001 it was external forces, namely 9/11 terrorist attack. Thus the catastrophe model should fit the data of 1987 well, as the bifurcations leading to instability are present. However, it does not seem to perform better than linear regression on the 2001 data. As the control variables we use the measures of sentiment, precisely OEX[1] Put/Call ratio which appears to be very good measure of the speculative money in the capital market, against trading volume as good proxy for large, fundamental investors.

2. The Cusp Catastrophe Model

Let us assume one dependent variable Y, and a set of n independent variables $\{X_1, X_2, \ldots, X_n\}$. Then y represents realization of a random variable Y, and x_i represents realizations of X_i. To obtain greater flexibility than using linear regression technique, $2n+2$ additional degrees of freedom are introduced. This could be done by defining control factors $\alpha_x = \alpha_0 + \alpha_1 x_1 + ... + \alpha_n x_n$ and $\beta_x = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$. These factors determine the predicted values of y given realizations of $\{x_1, x_2, \ldots, x_n\}$, meaning that for each value x there might be three predicted values of the state variable. The predictions will be roots of the following canonical form

$$0 = \alpha_x + \beta_x (y - \lambda)/\sigma - \left((y - \lambda)/\sigma\right)^3, \qquad (1)$$

which describes the cusp catastrophe response surface containing a smooth pleat. λ and σ are location and scale parameters. In the literature on catastrophe theory, α_x and β_x are so called *normal* and *splitting* factors,

[1]OEX are options with the Standard & Poor's 100 Index underlying

however, we prefer the notions a*symmetry* and *bifurcation* factors, respectively. Hence the statistical estimation problem is to find the estimates for the $2n+4$ parameters: $\{\lambda, \sigma, \alpha_0, ..., \alpha_n, \beta_0, ..., \beta_n\}$ from n observations of the $n+1$ variables $\{Y, X_1, \ldots, X_n\}$.

2.1. Stochastic Dynamics and Probability Density Function (PDF)

Let y_t be the function of time t for $t \in \langle 0, T\rangle$. From a dynamic system's point of view, Equation (1) can be considered as the surface of the equilibrium points of a dynamic system of the state variable y_t which follows the ordinary differential equation $dy_t = g(x, y_t)dt$, where $g(x, y_t)$ is the right hand side of Equation (1). For real world applications, it is necessary to add a non-deterministic behavior into the system, as the system usually does not determine its next states entirely. We may obtain a stochastic form by an adding of the Gaussian white noise term[2]. The system is then described by a stochastic differential equation of the form

$$dy_t = \left(\alpha_x + \beta_x \frac{(y_t - \lambda)}{\sigma_{y_t}} - \left(\frac{(y_t - \lambda)}{\sigma_{y_t}}\right)^3\right)dt + \sigma_{y_t} dW_t, \tag{2}$$

and $\sigma^2_{y_t}$ is an instantaneous variance of the process y_t. The W_t is a standard Wiener process and $dW_t \sim N(0, dt)$. Hartelman (1997) has established a link between a deterministic function of catastrophe system and a pdf of the corresponding stochastic process. He showed that the pdf $f(y_t)$ will converge in time to a pdf $f_S(y_\infty)$ corresponding to a limiting stationary stochastic process. This has led to a definition of stochastic equilibrium state and bifurcation which is compatible with their deterministic counterpart. Instead of fitting the deterministic process where the equilibrium points of the system are of a main interest, the attention is drawn to relative extremes of the conditional density function of y. Following Hartelman (1997) and Wangenmakers et al. (2005), the pdf of y is:

[2]Cobb, Watson (1980), Cobb (1981), Cobb (1985)

$$f_S\left(y_\infty \mid x\right) = \xi \exp\left[\alpha_x y_\infty + \frac{\beta_x}{2}\left(\frac{y_\infty - \lambda}{\sigma_{y_\infty}}\right)^2 - \frac{1}{4}\left(\frac{y_\infty - \lambda}{\sigma_{y_\infty}}\right)^4\right] \tag{3}$$

The constant ξ normalizes the pdf so it has unit integral over its range. The modes and antimodes of the cusp catastrophe pdf can be obtained by solving the equation $df_S(.|.)/dy_. = 0$, which will yield exactly implicit cusp surface equation - Equation (1). The parameters will be estimated by method of estimation developed by Hartelman (1997), Wagenmakers et al., (2005). As β_x changes from negative to positive, the pdf $f_S(y_\infty|x)$ changes its shape from unimodal to bimodal. It is also the reason why the β_x factor is called *bifurcation* factor. For $\alpha_x = 0$, the pdf is symmetrical, other values control asymmetry, thus α_x is *asymmetry* factor.

Thoughtful reader has certainly noted that catastrophe theory models are an extension to traditional models, therefore they have to satisfy the requirement of the empirical testability. It should be remembered, that there is no single statistical test for acceptability of the catastrophe model. Due to the multimodality of cusp catastrophe, traditional measure for goodness of fit cannot be used. Considered residuals can be determined only if the probability density function at time t is one-peaked, and as the model generally offers more than one predicted value, it is difficult to find a tractable definition for a prediction error. In testing we follow Hartelman's (1987) approach. A comparison of the cusp and a linear regression model is made by means of a likelihood ratio test, which is asymptotically chi-squared distributed with degrees of freedom being equal to the difference in degrees of freedom for two compared models. As it may not be sufficient to reliably distinguish between catastrophe and non-catastrophe models, Hartelman (1987) compares catastrophe model also to a nonlinear logistic model. As the cusp catastrophe model and the logistic model are not nested, Akaike information criterion (AIC), and Bayesian information criterion (BIC) statistics are used in a testing routine to compare the models.

3. Empirical Testing

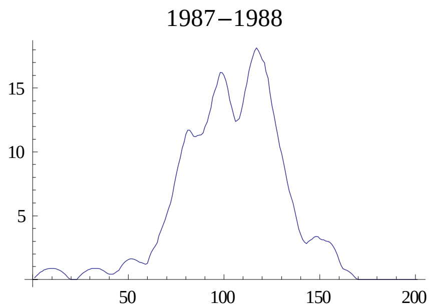
3.1. Data Description

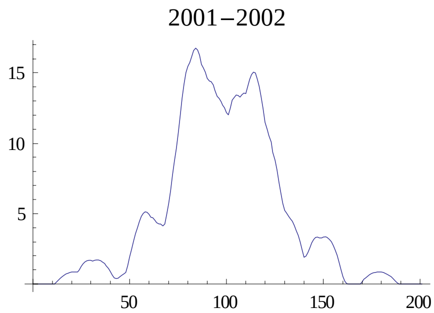
We primarily test the model on the set of daily data which contains most discussed stock market crash of October 19, 1987, known as Black Monday. The crash was the greatest single-day loss that Wall Street had ever suffered in continuous trading, 31%. The reasons for Black Monday have been widely discussed among professional investors and academics. However, not until today is there a consensus on the real cause. For comparison, we use another large crash, that of September 11, 2001. Our assumption is that while in 1987 the crash was caused by internal forces, the 2001 crash happened due to external force, namely the terrorist attack on the twin towers. Therefore the catastrophe model should fit the data of 1987 well as bifurcations leading to instability are present.

The data represents the daily returns of S&P 500 in the years 1987-1988 and 2001-2002 as the crashes took place inside these intervals. For the asymmetry side, we have chosen the daily change of down volume representing the volume of all declining stocks in the market. The trading volume represents good measure of the fundament, as it correlates with the volatility, and more importantly, good measure of what the large funds, representing fundamental investors, are doing. For bifurcation side OEX Put/Call ratio represents very good measure of speculative money. It is a ratio of daily put volume divided by daily call volume of the options with underlying Standard and Poor's 100 index. As financial options are the most popular vehicle for speculation, it represents the data of speculative money, while extraordinary biased volume or premium suggests excessive fear or greed in the stock market. These should be internal forces which causes the bifurcation.

3.2. Results

All the data are differenced once in order to gain stationarity. It can be seen that the data are leptokurtic, and much more interestingly, multimodal. For illustration of bimodality, we use kernel density estimation - see Figures 1 and 2 (we use Epanechnikov kernel which is of following form: $K(u)=\frac{3}{4}\left(1-u^2\right)\left(|u|\le 1\right)$ with smoother bandwidth so the bimodality can be seen):





Kernel density of the 2 year returns of 1987 and 1988 shows clear bimodality, and so does the kernel density of the second set of the data, i.e. years 2001 and 2002. The first test we consider is Hartelman's test for multimodality. It is evident from the previous figures that the returns are far from being unimodal. However as noted in Wangenmakers et al. (2005), there may occur inconsistencies between the pdf and the invariant function with respect to the number of stable states: examples of which can be found in Wangenmakers et al. (2005). Thus, we make use of the proposed Hartelman's kernel program to test for the multimodality and we have found that there is 75% probability that the 1987-1988 data contains at least one bifurcation point, and 26% probability that the the 2001-2002 data contains at least one bifurcation point. These results are also consistent with our assumption, that the first crisis was drawn by internal market forces (c.f. the presence of the bifurcations in the data), whereas the 2001 crash was caused mainly due to external forces, 9/11 attack.

Encouraged by the knowledge that bifurcations are present in our datasets we can now move to cusp fitting. As has been mentioned before, we use Hartelman's cuspfit software[3] for this purpose. The methodology is simple. First, the linear, nonlinear (logistic) and the cusp catastrophe models have been fitted to the data. Then we have tested whether the cusp catastrophe model fits the data better than the other two models by the procedure described at the beginning of the empirical part of this paper. We have obtained the following results:

In Table 1 there are the results of the cusp fit to the data of 1987-1988 which contains the crash of October 19, 1987. We can see that log likelihood is largest for the cusp catastrophe model. Chi-squared test, Akaike and Schwarz-Bayesian information criteria also favor the catastrophe model, and

[3]Applications are available at Han van der Maas's Website (http://users.fmg.uva.nl/hvandermaas/)

R^2 is again much better for the cusp catastrophe. Thus we can conclude that the cusp catastrophe model offers a more suitable explanation for the 1987 stock market crash. We believe that the quality of the fit arises from the choice of the variables. We have also tried other possible variables in order to explain the bifurcations, but none has proved as successful. The choice of the variables is logical as the tests for the bifurcations in the data confirmed their presence.

Table 1: Results of the fits to the 1987-1988 data

model	linear	logistic	**cusp**
R^2	0.1452	0.2558	**0.4025**
log likelihood	-6.09×10^3	-5.17×10^2	$\mathbf{-4.95\times10^2}$
AIC	1.23×10^3	1.05×10^3	$\mathbf{1.00\times10^3}$
BIC	1.24×10^3	1.07×10^3	$\mathbf{1.03\times10^3}$
parameters	4	5	**6**

Table 2: Results of the fits to the 2001-2002 data

model	linear	**logistic**	cusp
R^2	0.1128	**0.4682**	0.2023
log likelihood	-0.61×10^3	$\mathbf{0.45\times10^3}$	-0.55×10^3
AIC	0.12×10^4	$\mathbf{0.91\times10^3}$	0.11×10^4
BIC	0.12×10^4	$\mathbf{0.93\times10^3}$	0.11×10^4
parameters	4	**5**	6

Let us have a look at the second set of the data that of years 2001-2002. The results are in Table 2, and we can see, that the catastrophe model in this case is rather superfluous. The log likelihood is greater than in the linear model, but lower than in the logistic model. Also other information criteria favor the logistic model.

These results are in fact expected as of our earlier assumption, (i.e. that the 1987 crash was driven by internal forces, and the 2001 crash by external). While the 2001 data does have some bifurcations, the cusp catastrophe model clearly cannot fit the data significantly better than other models. This seems to be true, and for these data the catastrophe model did not perform better. However, for the 1987 crash the model seems to fit the data much better, and that is the sign, that the crash has occurred due to internal market forces.

4. Conclusions

Uncertain behavior of stock markets has always been on the leading edge of the research. Using the Cobb's (1985), Hartelman's (1997) and Wagenmakers's (2005) results we have managed to test cusp catastrophe theory on the financial data, and we have arrived at very interesting results which may help to move the frontier of understanding the stock market crashes further on. We may thus confirm, that the catastrophe models explains the stock market crash much better then alternative linear regression models, or nonlinear logistic model. We have fitted the data of the two stock market crashes, the first being the crash of October 19, 1987, and the second September 11, 2001. We have used the sentiment measures to model the proportion of technical and fundamental players in the market. OEX put/call ratio is a very good measure of the technical players and represents the speculative money in our model and the trading volume is the measure of fundamental players and represents the excess demand.

We have clearly identified the bimodality of the returns using the test for multimodality which confirms that there is 75% probability that there is at least one bifurcation point in the data. Finally, the cusp catastrophe model fits these data much better than other models that have been used. Hence we conclude that the internal processes of the first dataset led to the crash in 1987. On the other hand, the crash of the September 11, 2001 can be better explained by the alternative logistic model. We have also found only 26% probability that there is at least one bifurcation point in these data, which is also in line with our second assumption: that due to the fact that this crash was caused by external forces the presence of the bifurcations in the data is much weaker.

Our findings may contribute to the frontier of the research, as it is the first attempt to quantitatively explain stock market crashes by cusp catastrophe theory. The testing has been conducted only on the restricted datasets. Thus further work is to test on different data which describes the situations when the changes in speculative money in the stock market lead to a crash. The main significant question, that of whether cusp catastrophe theory may help with an early indication of the stock market crashes still remains to be answered.

Acknowledgement

The authors would like to thank Prof. van der Maas for supplying a copy of Hartelman's dissertation and for publishing invariant stochastic catastrophe theory tools on his web site. A support from the Czech Science Foundation under grants 402/06/0990 and 402/08/0107 and Centre of research MSMT LC06075 is gratefully acknowledged.

References

Cobb, L., Zacks, S., 1985. Applications of Catastrophe Theory for Statistical Modeling in the Biosciences. Journal of the American Statistical Association 392, 793-802.

Cobb, L., 1981. Parameter estimation for the cusp catastrophe model. Behavioral Science 26, 75--78.

Cobb, L., Watson,B., 1980. Statistical catastrophe theory: an overview. Math. Model 1, 311--317.

Hartelman, Pascal A. I., 1997. Stochastic Catastrophe Theory. Dissertatie reeks 1997-2, Faculteit Psychologie, Universiteit van Amsterdam.

Thom, R., 1975. Structural Stability and Morpohogenesis, New York: Benjamin.

Wagenmakers, E.J., Molenaar, P.C.M., Hartelman, P.A.I., van der Maas, H.L.J., 2005. Transformation invariant stochastic catastrophe theory. Physica D, 211, 263

Zeeman, E. C., 1974. On the unstable behaviour of stock exchanges. Journal of Mathematical Economics 1(1), 39-49.

A SIMPLE REAL BUSINESS CYCLE MODEL OF THE CZECH ECONOMY

Jaromír Baxa
jaromir.baxa@centrum.cz
ÚTIA AV ČR

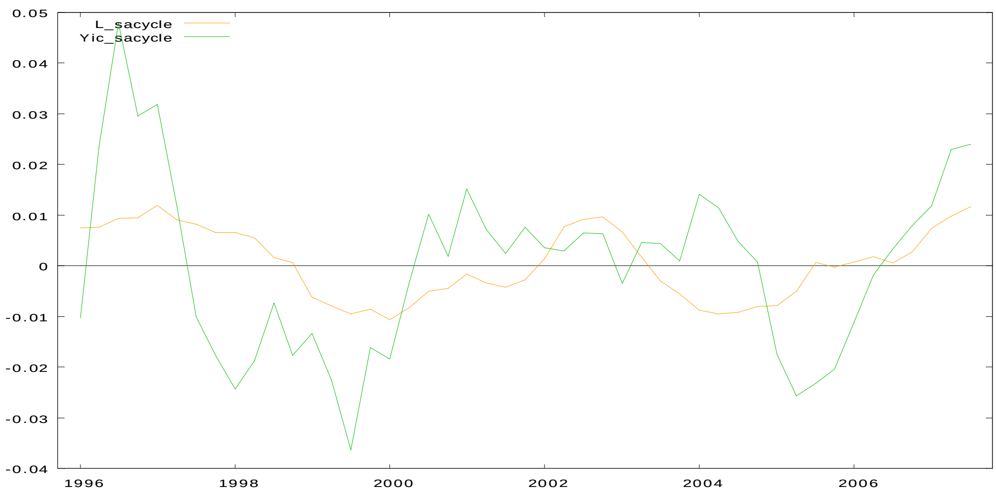
Introduction

Since early theoretical papers in the field of the real business cycles occurred in the eighties (Kydland-Prescott, 1982, 1986), it has been shown that many applications of these kind of models are able to capture surprisingly high proportion of economic fluctuations (see Cooley, 1995, or the Reader of Real Business Cycle by Hartley et al., 1998, for references and further extensions of the RBC model, application on the U.S. data in Kydland-Prescott, 1990, skeptical reply then in Hartley, 1999). Up today their applications and tests were usually performed using the data of stabilized economies of West, however their applications on transition or developing countries are quite rare (one of them is the Bergoegin's and Soto's article on tests of RBC in Chilean economy from 2002). Here we tried to apply very simple model on the Czech economy and our motivation was to explore how such a simple model behaves in this specific environment with strange behavior of employment which does not follow the cycle a lot, with recession in the mid of the period caused more less because of policy measures and as we will see later, with consumption insensitive on economic fluctuation. Several "stylized facts" are summarized in the first section, whereas the second and the following one presents the model and calibration exercise on the Czech economy. Not very surprisingly the performance of the calibrated model for the whole sample 1996-2007 was rather mixed: some characteristics were replicated relatively good, on the other hand the others, especially signs of some correlation coefficients seem to contradict the theory. However as the economy evolved over time the in-sample-fit increased to surprisingly good levels. Possible interpretations of that are sketched in the last section.

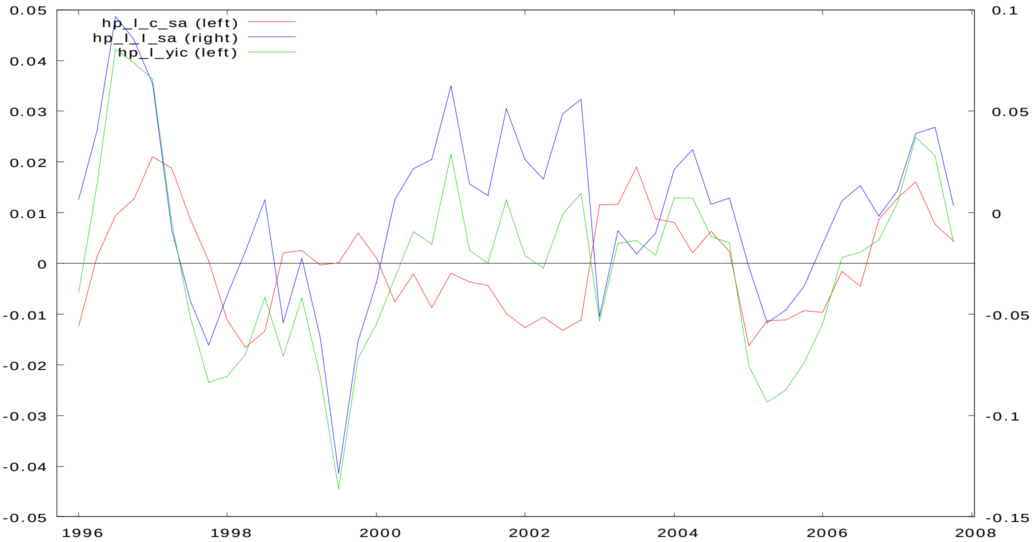
1. The Czech Business Cycle

Behavior of the main macroeconomic indicators of Czech economy since 1996 can be seen on plots 1.1-1.3. Plot 1.1. shows the development of the cyclical components of logarithms of the GDP (red) and employment (blue), extracted using the univariate Hodrick-Prescott filter on each variable. The next two plots depict cyclical components of time series of GDP, consumption and investment, estimated by univariate and multivariate Hodrick-Prescott filter (for details about the procedure see Mills, 2003). Labor input is not included in the common trend as the trend of employment has different pattern than the other variables. One point should be mentioned here: employment is not stationary as it is but it follows a stochastic trend which goes down till 1999, becomes stable for several periods and then becomes to increase during last years since 2005. Causes and effects of this pattern determined by the Czech way of transition are summarized in collection of papers in Flek et al. (2007). Basically the increase of unemployment in the late nineties caused more less by the progress in restructuring in the industry which has been rather limited in the mid of nineties due to the economic policy later on known as the banking socialism. However as the models was written in gaps originally, this long-term pattern is not captured by this model. Regarding other variables the GDP was constructed as if the economy were closed one and existed without governments as a sum of consumption and investment. Investment is the more volatile component of the GDP comparing to consumption, but there is one a little bit striking feature namely of the time series of consumption, which behaves more-less independently on the GDP, as the theory would suggest pro-cyclical behavior of consumption despite with lower volatility than investment reflecting smoothing nature of consumer expenditures. Tables with descriptive statistics and correlation coefficients are provided in Appendix.

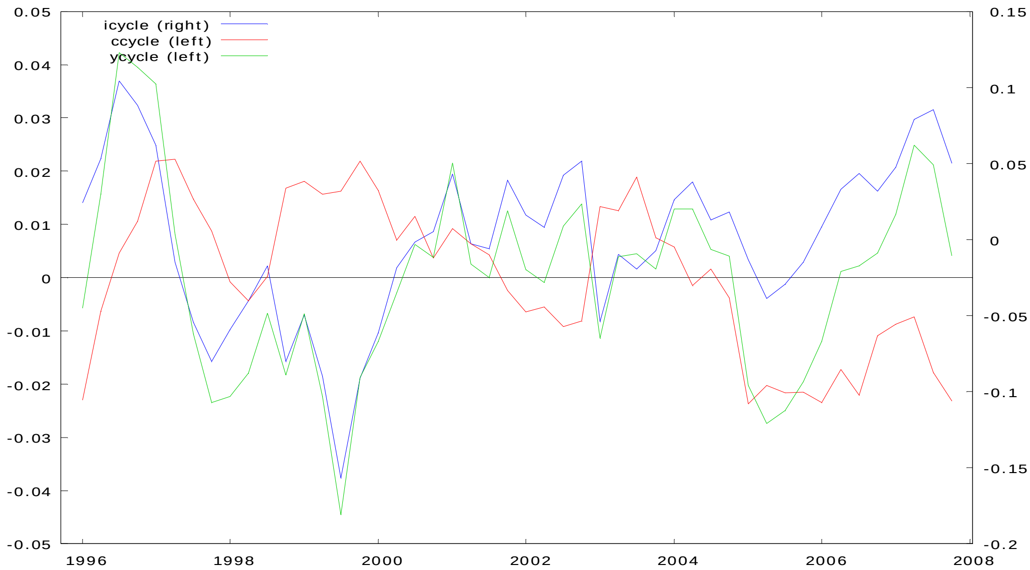
Plot 1.1 Cyclical Components of Y=C+I and Employment



Plot 1.2 Cyclical Components of Y, C, I (H-P Filter)



Plot 1.3 Cyclical Components of Y, C, I (H-P Filter, common trend assumed)



2. A Baseline Real Business Cycle Model

The model used here is the textbook form as in Dejong-Dave (2007). Households solve traditional maximizing problem: they maximize their life time utility U, given as the expected discount flow of utilities from consumption c_t and leisure l_t:

$$\max_{c_t, l_t} U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t). \quad 2.1$$

The term E_0 represents that the expectations are conditional upon information set available at time *0*, β is the discount factor determining the preference for current utility from consumption and leisure according to future utility and finally *u(.)* is the instantaneous utility function.
In our setting we used the constant relative risk aversion (CRRA) function

$$u(c_t, l_t) = \left(\frac{c_t^{\wp} l_t^{1-\wp}}{1-\Phi} \right)^{1-\Phi} . \quad 2.2$$

Households get their income from their firms equipped with a production technology that can be used to produce a single, non-differentiated product y_t. The inputs for the technology are capital k_t and labor of the members of the households n_t and the technology is approximated using traditional Cobb-Douglass production function

$$y_t = e^{z_t} k_t^{\alpha} n_t^{1-\alpha} , \quad 2.3$$

augmented for stochastic productivity shock z_t

$$z_t = \rho z_{t-1} + \varepsilon_t . \quad 2.4$$

Within each period households divide their one unit of time into labor and leisure, thus

$$n_t + l_t = 1. \quad 2.5$$

Similarly the output of the period can be used either for consumption and investment

$$y_t = c_t + i_t . \quad 2.6$$

And finally the capital increases each period for a margin of investment which is not used as a replacement for depreciated capital:

$$k_{t+1} = i_t + (1-\delta)k_t$$

2.7

where the δ is the deprecation rate.

The solution of the households problem might be obtained via dynamic programming (Stokey and Lucas, 1989, might represent a useful reference for this). The nonlinear system consisting from the Euler equation describing the intertemporal optimality condition for the consumption stream (2.9), the equation representing the choice of labor/leisure ratio upon observed productivity (2.8) and desired level of consumption and equations specifying variables follows (2.10-2.14):

$$\left(\frac{1-\wp}{\wp}\right)\frac{c_t}{l_t} = (1-\alpha)e^{z_t}\left(\frac{k}{n}\right)^{\alpha} \qquad 2.8$$

$$c_t^{\wp(1-\Phi)-1} l_t^{(1-\wp)(1-\Phi)} = \beta E_t \left\{ c_{t+1}^{\wp(1-\Phi)-1} l_{t+1}^{(1-\wp)(1-\Phi)} \left[1 + \alpha e^{z_{t+1}} \left(\frac{n_{t+1}}{k_{t+1}} \right)^{1-\alpha} - \delta \right] \right\} \qquad 2.9$$

$$y_t = e^{z_t} k_t^{\alpha} n_t^{1-\alpha} \qquad 2.10$$

$$z_t = \rho z_{t-1} + \varepsilon_t \qquad 2.11$$

$$y_t = c_t + i_t \qquad 2.12$$

$$n_t + l_t = 1 \qquad 2.13$$

$$k_{t+1} = i_t + (1-\delta)k_t \qquad 2.14$$

The system 2.8-2.14 has been both linearized and solved using DYNARE, the list of DYNARE code is attached in Appendix.

3. Data, Calibration and Estimation

The model has been parameterized using the Czech quarterly data from 1Q 1995 to 4Q 2007. All the variables were represented in 1995 real counterparts and they were first seasonally adjusted using the ARIMA X11 method and then detrended using the multivariate Hodrick-Prescott filter. The data are from the system of national accounts and they were provided by the Czech Statistical Office. The time series of y_t has been constructed as a sum of consumption and investment as neither the government nor another economies are present in the model. Thanks to this the resulting business cycle – or simply periods with lower growth than it would correspond to the trend – are a bit different in comparison to the cycle of the gross domestic product.

The way for calibration is following the procedure described in Dejong-Dave (2007, ch.6): parameter β was calculated from the corresponding average real short-term interest rate, which has been significantly lower than by Dejong and Dave proposed benchmark of 4-5%: the average was ranging from 2.4% to 2.8% per year depending on the subsample we used. Consequently corresponding value of discount factor $\beta = \frac{1}{1+\rho}$, where ρ is the quarterly discount rate, is 0.993. Usually the value of β is set to 0.99 in the literature, our discount factor would imply higher incentive to smooth the consumption over time. Such smoothing was not fully acknowledged by the data so we decided to lower the level and experiment with levels of β at 0.992 and 0.991.

Following Barro and Sala-I-Martin, 2004 the risk aversion parameter Φ in the models featuring intertemporal consumer optimization can be specified as

$$\Phi = \frac{1}{\mu}(r-\rho) \qquad 3.1$$

where the $\mu = 0.7\%$ is the average growth rate of the log of consumption, $r \in (0.6\%, 0.8\%)$ as the average quarterly return on physical capital derived from the short term interest rates on credits in 1999-2007 shock-free period and ρ, the discount rate, restricted to be positive, we received the appropriate range for the risk aversion parameter (1;1.5) from which we used the almost mid-value 1.2.

Parameters α, δ and φ are related due to the structure of the model. From the steady state conditions it can be shown that the relationship for these parameters is

$$\alpha = \left(\frac{\delta+\rho}{\delta}\right)\frac{\bar{i}}{\bar{y}} \qquad 3.2$$

where the $\bar{i}/\bar{y}$ is the average investment output ratio, which is approximately 0.36 in our dataset. Similarly the time allocation parameter φ equals to

$$\wp = \frac{1}{1+\frac{2(1-\alpha)}{\bar{c}/\bar{y}}} \qquad 3.3$$

with $\bar{c}/\bar{y}$ as consumption-output ratio (Dejong-Dave, 2007). Using the 3.2 and 3.3. we may generate a table of plausible combinations of α, δ and φ for previously defined ρ=0.007 (Table 5). Finally the variance of shocks was set to 0.008. After several experiments with calculations and Bayesian estimation of the parameters using the Metropolis-Hastings algorithm implemented in DYNARE we found the appropriate values highlighted in the table 3.1.

Table 3.1: Calibration of parameters α, δ and φ

delta	alpha	phi
0,0100	0,626	0,458
0,0125	0,575	0,426
0,0150	0,540	0,407
0,0175	0,516	0,395
0,0200	0,497	0,386

Table 3.2: Values of parameters used for solution and simulation

beta = 0.992;	phi = 0.41;	delta = 0.015;	alpha = 0.54;	pi = 1.2;	rho = 0.85;	sigma = 0.008;

4. The Results and Evaluation of the Model

The results of the estimated simple real business cycle model are reported in following tables:

Table 4.1a: Descriptive statistics of observed variables

Deviations from common trend

	mean	variance	var x/var y	1st autocov
y	0,0000000	0,0168	1,000	0,75
i	0,0000000	0,0500	2,967	0,8
c	0,0000000	0,0131	0,776	0,84
n	0,0005460	0,0070	0,414	0,89

Note: n (employment) is from the univariate Hodrick-Prescott filter.

Table 4.1b: Descriptive statistics of simulated variables

	mean	variance	var x/var y	1st autocov
y	0,0000000	0,0237	1,000	0,892
I	0,0000000	0,0532	2,245	0,845
c	0,0000000	0,0130	0,547	1,000
n	0,0000000	0,0115	0,484	0,828

Table 4.2: Correlations data – multivariate filter except n

y	**c**	**i**	**n**	
1	-0,024	0,848	0,411	**y**
	0,144	*0,022*	*0,087*	*s.e.*
	1	-0,496	-0,363	**c**
		0,074	*0,094*	*s.e.*
		1	0,432	**i**
			0,084	*s.e.*
			1	**n**

5% critical value (two-tailed) = 0,2876 for n = 47

thus => correlation 'c'-'y' insignificant

Table 4.3: Spearman's correlation coefficients

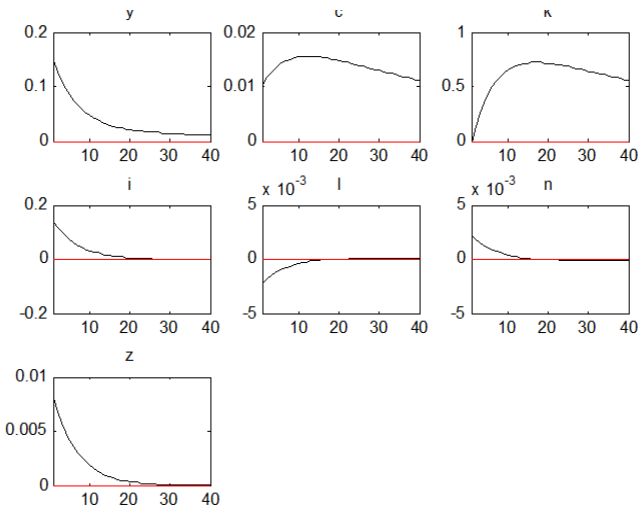
y	c	i	n	
1	0,004	0,833	0,407	y
	1	-0,471	-0,366	c
		1	0,445	i
			1	n

Table 4.4: Correlation coefficients of simulated variables

Correlations model

y	c	i	n	
1	0,61	0,96	0,93	y
	1	0,37	0,26	c
		1	0,99	i
			1	n

First we might observe the model predicts cyclical behavior of Y,C,I,N. The look on the data says something slightly different as reported in tables 4.1-4.4 – there is one striking feature: acyclical behavior of consumption which has not been captured by the model. Also, predicted autocorrelations suggest slightly higher persistence of shocks than predicted by the model. On the other hand the results are surprisingly good if we consider how turbulent period the years of 1995-2007 were as the model captures the behavior of investment and employment to the output relatively good.



Plot 4.1 Impulse responses of the technology shock

One might claim that some of the problems are caused by short time series.

That's for sure, due to that there are lots of uncertainties about true variances and correlation coefficients. Especially the correlation coefficients between consumption and other variables have large standard errors comparing to the values of correlation coefficients itself causing that we cannot conclude whether they are statistically significant or not. However the matrix of Spearman correlation coefficients which take into account order instead of actual values gives the same picture so that it looks the results are not biased by several extreme values.

Another question is whether the results are robust on the sample. That's really difficult to asses as we however cannot cut of one half of the data without a risk of losing some important information and with keeping in mind that also the confidence intervals increase incredibly[1]. Nevermind the results for the end of the sample, for the years 2004-2007 are reported in table 4.5.

Table 4.5: Results for subsample 2004-2007

Data 2004-2007

i	c	y	
1,000	0,270	0,931	i
	1,000	0,602	c
		1,000	y

Model

i	c	y	
1,000	0,374	0,963	i
	1,000	0,610	c
		1,000	y

The countercyclical behavior of consumption is lost and estimated correlation coefficients are very close to predicted ones. But the disclaimer pointed out in previous paragraph remains.

Table 4.6: Estimates of parameters

Parameters						
	prior m	post. mean	conf. interval		prior	pstdev
alpha	0.480	0.4960	0.4919	0.5004	beta	0.0400
delta	0.011	0.0054	0.0054	0.0054	beta	0.0010
phi	0.400	0.3837	0.3793	0.3904	norm	0.0500
rho	0.850	0.8479	0.8469	0.8488	beta	0.0500
standard deviation of shocks						
	prior m	post. mean	conf. interval		prior	pstdev
e	0.008	0.0165	0.0161	0.0168	invg	Inf

1 Best way will be to use a crystall ball to extract the future data if we ever had some.

Last but not least the table 4.6 reports estimates of parameters α, δ, φ, ρ and σ. A posterior estimates mostly confirmed our choice with two exceptions: the estimated variance is two times higher than the calibrated one, however using that value the variances of all variables increased to levels much higher than present in the data. The shape of impulse responses – their peaks and persistence – remained, only the scale changed. This lead to decision to keep the The second exception was the value of the depreciation rate which was estimated to the level of 0.005 per quarter, a level causing unstable behavior of capital and also implied steady state was not stable. We suspect this low estimate might be influenced by unstable behavior of investment at the beginning of our sample, but the mechanism is not clear.

Concluding Remarks

This very simple model performs relatively well, with respect to predicted variances, ratio of variances of variables according to the variance of the GDP series and also with respect to predicted autocorrelations. However the actual behavior, namely whether some variables follow the cycle or not (that's the case of consumption for example) are not well captured for the whole post 1995 period. The consumption generated by the calibrated model is still much less volatile then the GDP series and it tends to smooth the cycle, but when the common trend is introduced, the consumption behaves independently over the period. This strange behavior is confirmed by both Pearson and Spearman correlation coefficient which are negative or non-significant for Y-C deviations-of-trend relation, however recently the actual behavior is much closer to predictions of the model. Explanations of this phenomenon can be stated from different points of view: either we believe that the real business cycle theory is theoretically appropriate approximation of the most important economic relations, thus we can interpret the observed development of the variables as being caused by the transition period, or we may simply conclude that the actual behavior doesn't correspond to the model, so that the model is not a good approximation of the theory, at least for this period. In fact both views are about the same thing implying that for more precise conclusions we have to wait some more years.

Literature

Barro and Sala-I-Martin, 2003: *Economic Growth.* Cambridge: MIT Press.
Bergoegin – Soto 2002: *Testing Real Business Cycle Models in an Emerging Economy.* The Central Bank of Chile Working Paper No. 159.
Cooley, T. (ed.) 1995: *Frontiers of Business Cycle Research.* Princeton:

Princeton University Press.
Dejong, D. -Dave, Ch. 2007: *Structural Macroeconometrics*. Princeton: Princeton University Press.
Flek, V. (ed.) 2007: *Anatomy of Czech Labour Market.* Prague: Karolinum.
Hartley, J. et al. 1998: *Real Business Cycles: A Reader.* London and New York: Routledge.
Hartley, J. 1999: Real Myths and a Monetary Fact. In: *Applied Economics,* vol. 31, p. 1325 – 1329.
Kydland, F. E. and Prescott, E. C. 1982: Time to Build and Aggregate Fluctuations. In: *Econometrica*, vol. 50, no. 6, p. 1345-1370.
Kydland, F. E. and Prescott, E. C. 1990: Business Cycles: Real Facts and Monetary Myth. In: *Federal Reserve Bank of Minneapolis Quarterly Review,* vol. 14, no. 2, p. 3-18.
Prescott, E. C. 1986: Theory ahead of Business Cycle Measurement. In: *Federal Reserve Bank of Minneapolis Quarterly Review,* vol. 10, no. 4, p. 9-22.

Appendix 1: Dynare code used in this paper:

```
var y c k i l n z;
varexo e;
parameters beta phi delta alpha pi rho sigma;
beta = 0.992;
phi = 0.41;
delta = 0.015;
alpha = 0.54;
pi = 1.2;
rho = 0.85;
sigma = 0.008;
model;
c=phi/(1-phi)*l*(1-alpha)*exp(z)*k^alpha*n^(-alpha);
c^(phi*(1-pi)-1)*l^((1-pi)*(1-phi))=beta*(c(+1)^(phi*(1-pi)-
1)*l(+1)^((1-pi)*(1-phi))*(1+alpha*exp(z(+1))*k(+1)^(alpha-
1)*n(+1)^(1-alpha)-delta));
y=exp(z)*k^alpha*n^(1-alpha);
y=c+i;
k=i(-1)+(1-delta)*k(-1);
1=n+l;
z=rho*z(-1)+e;
end;
initval;
y = 1;
k = 9;
c = 0.7;
i = 0.3;
n = 0.5;
l = 0.5;
z = 0;
e = 0;
end;
shocks;
var e;
stderr sigma;
end;
steady;
check;
stoch_simul(periods=1000);
```

AN EVOLUTIONARY APPROACH FOR SOLVING VEHICLE ROUTING PROBLEM

Zuzana Čičková, Ivan Brezina

Abstract: This article describes the application of Self Organizing Migrating Algorithm (SOMA) to the Vehicle Routing Problem (VRP). SOMA is an optimization method based on Evolutionary Algorithms that are originally focused on solving non-linear programming tasks containing continuous variables; therefore the use of Evolutionary Algorithm requires some special approaches to guarantee feasibility of solutions. In this article, the approach for solving VRP will be presented. The results will be compared with optimal solution that was obtained by GAMS.

1. INTRODUCTION

The vehicle routing problem (VRP) is one of the well–known combinatorial optimization tasks. This problem consists in designing the optimal set of routes for a vehicle in order to serve a given set of customers. The vehicle has a certain capacity and it is located in a certain depot. Each customer has a certain demand. The interest in VRP is motivated by its practical relevance as well as by its computational complexity (NP–hardness). Further on there exist a distance (length, cost, time) matrix between the customers and the depot. The goal is to find optimal vehicle routes (usually minimum distance).

VRP is an important problem in the fields of transportation, distribution and logistics. Many methods have been developed for solving so that problem, as optimization methods (can be applied for small problem size), or heuristics with more or less success (e.g. Clarke–Wright algorithm, sweep algorithm etc.). Some approaches that belong to evolutionary techniques seemed to be effective for solving so that problem because of its considerable difficulty (NP–hardness).

2. PRINCIPLES OF SELF ORGANIZING MIGRATING ALGORITHM

Self organizing migrating algorithm (SOMA) was created in 1999 [4]. It can be classified as an evolutionary algorithm (EA), despite the fact that no new individuals are created during the computation, and only the position of individuals in the search place is changed. EA are optimization techniques that use mechanisms inspired by biological evolution, such as reproduction, mutation, recombination and natural selection. SOMA is based on the self–organizing behavior of groups of individuals in a "social environment".

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, $i = 1, 2, ..., np$ (*np*–number of individuals in the population). A population can be viewed as a $np \times (d+1)$ matrix (*d*–number of parameters of individual), where the columns represent individuals. Each individual represents one candidate solution for the given problem that is represented by parameters of individual $j = 1, 2, ..., d$. Associated with each individual is also the fitness $f_c(\mathbf{x}_j)$, $i = 1, 2, ..., d$ 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

SOMA, as other EA algorithms, is controlled by a special set of parameter. Recommended values for the parameters are usually derived empirically from experiments [2], [4]:

- *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. Insufficient value of *m* value could results in reaching not optimal solution.
- *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 (*mk*).

Each individual is evaluated by cost function (fitness) and the individual with the highest fitness – *Leader* ($x_{L,j}^{mk}$) is chosen for the current migration loop. According to the *step*, other individuals begin to jump towards the *Leader* according the rule:

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

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:

$$x_{i,j}^{mk+1} = \min\left\{ f_c(x_{i,j}^{mk}), f_c(x_{i,j,start}^{mk}) \right\} \tag{2}$$

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 *d–k* dimensional subspace. This fact establishes a higher robustness of the algorithm. Parameter *prt* has the same effect as mutation for genetic algorithm. Vector **prt** is created before the individual begin to move in the search space.

$$\text{prt}_j = \begin{cases} 1, \text{ ak } rand_j \langle 0,1 \rangle > prt \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

The general convention used is known as *AllToOne* strategy. In literature [2], [4] 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 VRP

The presented model consists of 14 customer cities with certain demand: Bratislava (BA) – 6,94 units, Komárno (KN) – 6,97 units, Malacky (MA) – 2,63 units, Nitra (NR) – 5,49 units, Považská Bystrica (PB) – 3,77 units, Prievidza (PD) – 9,19 units, Partizánske (PE) – 3,42 units, Pezinok (PK) – 5,73 units, Piešťany (PN) – 5,02 units, Skalica (SI) – 1,85 units, Trenčín (TN) – 11,76 units, Topoľčany (TO) – 2,81 units, Trnava (TT) – 13,88 units, Žilina (ZA) – 8,69 units. The central depot is located in Sereď (SER). The vehicle capacity is set to 50 units. The distance matrix (shortest distances in km) between the customers and the depot was obtained from MS AutoRoute.

By solving, a natural representation of individual, known from genetic algorithm, was used. Using this representation, the cities are listed in the order in which they are visited. Each city is assigned with a different integer value 1 to *n* that represent sequence of visited cities (*n* represent number of customers). Then, number of parameters of individual is $d = n$. The initial population $P^{(0)}$ is generated as follows:

$$\mathrm{P}^{(0)} = x_{i,j}^{(0)} = \mathrm{randperm}\left(d\right) \;\; i = 1,2,...,np \;\; j = 1,2,...,d \tag{4}$$

This function is assigning each individual with a random permutation vector of integers sized *d* (random permutation of cities index on the vehicle route). If during the SOMA's migration loop the unfeasible individual is created, 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.

Fitness is the cost of corresponding tour. SOMA algorithm has a slight dependence on the setting of control parameter, so it was necessary to set the parameter efficiently. During the computation, the used set of parameters was chosen on basis the article [1] that describes the use of 128 experiments and employment of some statistical approaches (ANOVA, Kruskal–Wallis test etc.) to specify this parameters for traveling salesman problem: *prt*=0,8, *step*=0,9. The values of *d* is fixed according to problem size to 14, parameters *np*=50 and *mig*=200, *mass* = 3 were used during the simulation. For computation system Matlab7 was used. Final result of 10 simulation was the tour length 707 km in any case (e.g SER– PN – TN –PB – ZA – PD – PE –TO – SER –TT – SI – MA – PK – BA – KN – NR – SER).

This problem was also solved by optimization system GAMS used mathematical programming formulation as was published in [3]. The result was identical with tour length 707 km.

4. CONCLUSION

In this paper, we used SOMA for solving vehicle routing problem containing 14 customer cities in Slovakia. SOMA belongs to evolutionary techniques as alternative approach to classical optimization techniques (formulated as integer programming problem solved by GAMS). The results of simulation show that SOMA works efficiently for solving the VRP. Optimal solution obtained by GAMS was the same quality as results of all realized simulations.

REFERENCES

[1] Čičková, Z.: Solution of the traveling salesman problem via evolution methods. INPROFORUM 2007, České Budějovice, november 2007

[2] 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

[3] Pelikán, J.: Diskrétní modely operačního výzkumu. PROFESSIONAL PUBLISHING, Praha 2001

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

Acknowledgement: This work has been supported by research project VEGA 1/4588/07

Contact Adress:
Zuzana Čičková, Ivan Brezina
Katedra operačného výskumu a ekonometrie
Fakulta hospodárskej informatiky Ekonomickej univerzity v Bratislave
Dolnozemská 1/b
851 32 Bratislava
☎ +421 2 6729 5820
e-mail: cickova@euba.sk, brezina@euba.sk

NATURAL MONOPOLY COST-ORIENTED PRICE REGULATION

Michal Fendek
University of Economics, Bratislava, Slovakia

Abstract

In this paper we are dealing with the issue of the price regulation on the network industries market. This issue is particularly actual nowadays because questions about reasonable profit and objectively justified costs of network industries entities presently resonate in expert public. Besides the general characteristic of network industries regulation principles we will aim on presentation of type models of regulation, which regulate maximum reasonable profit margins for regulated firm on the basis of its total costs analysis.

State has powerful tools at its disposal for the regulation of the monopolistic market equilibrium conditions development – the regulatory policy. With certain level of simplification we can tell that the state enforces specific regulatory mechanisms influencing natural monopolies behavior.

Keywords

Network industries, regulated prices, reasonable profit in regulated industries, rate of cost regulation, Regulatory Office for Network Industries, natural monopoly regulation,.

1 Introduction

Modern market structures in economically developed countries are generally represented by imperfect competition. We need to realize, that monopoly, in consequence of its exclusive market position as a sole supplier of commodity or service on the relevant market, can set the market price and supply any level of production, which allows it to maximize profit above the level of capacity of competitive company.

Monopoly is, of course, acting within the legitimate market conditions and its position is determined by the real market situation. On the other side it is natural for such means to exist, which allow redistribution of this profit in a way that part of it would support nationwide economy development objectives of the country it is operating in.

In principle, there are two schemes, which can be applied by the state in regulation of monopoly price policy.

a) Taxation of monopoly

There is a potent tool at states disposal to regulate the development of market equilibrium conditions on monopolistic market. It is a tax policy. In a simplified way, we can tell that the state is redistributing the profit of monopolistic firm through its taxation. This process of taxation in the imperfect competition environment has, in comparison with perfect competition environment, certain specificity.

Firm operating in perfect t competition environment can respond to enforcement of particular scheme of tax burden only by modifying the amount of goods supplied on the market. Monopoly has, however, broader room for decision making because in its competence there is not only setting the amount of goods supplied on the market but also, after taking into account consumer behavior, setting the price of good on relevant market.

b) Application of specific regulatory mechanisms affecting behavior of natural monopolies

For the regulation of subjects of the chosen industries, which are in light of their market share monopolies or natural monopolies, the state is creating institution, so-called regulator, which task, under the state authority, is to create such legislative environment and regulatory mechanisms, which will ensure market equilibrium for observed commodities while granting fair profit for regulated subject. Among these industries particularly belong network industries, which subjects often share characteristics of monopolies and state has therefore desire to regulate them.

In Slovak conditions the role of the regulator is represented by Regulatory Office for Network Industries, which role is to issue licenses, regulate prices and quality norms for network industry products. Regulatory offices in relation with fulfillment of their main mission – technical and price regulation of enterprise in regulated activities of chosen industries – are facing the effective solution of two tasks:

(1) Preparation of working and, this needs to be stressed, competitive market environment for network industries goods while applying standard regulatory mechanisms, mainly in context of Slovak entry into the EU and gradual adaptation of Slovakia to the conditions on energy markets of united Europe;

(2) Preparation of such analytical apparatus for network industries price regulation, which would guarantee effective development of regulated subjects. In the first period

it was equally important to eliminate deformations in prices of network industries products.

2 Cost-oriented models of network industries price regulation

This part of the paper will deal with price regulation scheme on the basis of return over costs regulation. Return over costs is scheme of natural monopoly regulation, which is in principle different form the model of regulation on the basis of rate of return. It derives the barrier for the not exceedement of reasonable profit only from the part of the regulated entity's input activities, namely from the volume of investment. This undesirably motivated monopoly to disproportionate increase of capital investment, which was of course contra productive.

Return over costs regulation sets the maximum profit margin for regulated firm on the basis of its overall costs. We can see that there is certain analogy between this form of regulation and regulation on the basis of the rate of return. However the difference is that return over costs regulation does not prefer particular cost group, but uses the overall costs.

The idea, that firm's profit is in this case some kind of function of its costs is of course mystification. This scheme, however, effectively hinders natural monopoly from asserting such combination of its supply and monopolistic market price, which would allow it to make inappropriate profit in comparison with exerted costs.

In short, the keystone of return over costs regulation is that regulator as the base for regulated entity's reasonable profit definition sets its overall costs and defines reasonable profit as a certain allowed percentage *RoC* of its costs. Analytically we can express this condition as

$$RoC \times n(q) \geq \pi(q)$$

or

$$RoC \times (w \times L + r \times K) \geq \pi(q)$$

where
q - production
$n(q)$– function of the total costs of the firm, $n: R \to R$
$\pi(q)$ – profit function $\pi : R \to R$

L - labor (production factor)
K – capital (production factor)
w – labor price
r – capital price
RoC – reasonable profit margin set by the regulator corresponding to the unit costs.

Regulated output and regulated price in the return over costs environment are calculated by solving the following mathematical programming task

$$\pi(q) = \pi\big(f(K,L)\big) = p\big(f(K,L)\big) \times f(K,L) - w \times L - r \times K \to \max \tag{1}$$

subject to

$$p\big(f(K,L)\big) \times f(K,L) - w \times L - r \times K - RoC \times \big(w \times L + r \times K\big) \le 0 \tag{2}$$

$$K, L \in R_{\ge 0} \tag{3}$$

Solution of the optimization task (1) ... (3) is optimal consumption volume of production factors labor L^* and capital K^*, on the basis of which, with the help of the production function, the regulated optimal volume of output $q_{RoC}{}^*$ is quantified

$$q^*_{RoC} = f(K^*, L^*)$$

And regulated optimal price $p_{RoC}{}^*$ with the help of the price-demand and production function on the basis of the relation

$$p^*_{RoC} = p(q^*_{RoC}) = p(f(K^*, L^*)$$

While also in this regulation approach the rate of return on revenues defined by the parameter RoC is respected, i.e. exogenous control parameter set by the regulator.

Let us now transform the optimization task (1) ... (3) with two variables L, K into a task with one variable q in a following way

$$\pi(q) = p(q) \times q - n(q) \to \max \tag{4}$$

subject to

$$p(q) \times q - n(q) - RoC \times n(q) \le 0 \tag{5}$$

$$q \in R_{\ge 0} \tag{6}$$

where
$t(q) = p \times q$ – function of revenues of the firm, t: $R \rightarrow R$
$n(q) = nv(q) + n_F$ – function of the total costs of the firm, n: $R \rightarrow R$
$nv(q)$ – function of the variable costs of the firm, nv : $R \rightarrow R$
n_F - fixed costs of the firm, $n_F \in R$
RoC – reasonable profit margin set by the regulator corresponding to the unit costs.

On the basis of the substitution $t(q) = p(q) \times q$ we can reformulate the optimization task (4) ... (6) to

$$\pi(q) = t(q) - n(q) \rightarrow \max \qquad (7)$$

subject toPic. 1: Regulation on the return over costs basis – *ROC*

$$t(q) - n(q) - RoC \times n(q) \leq 0 \qquad (8)$$
$$q \in R_{\geq 0} \qquad (9)$$

TC(Q)

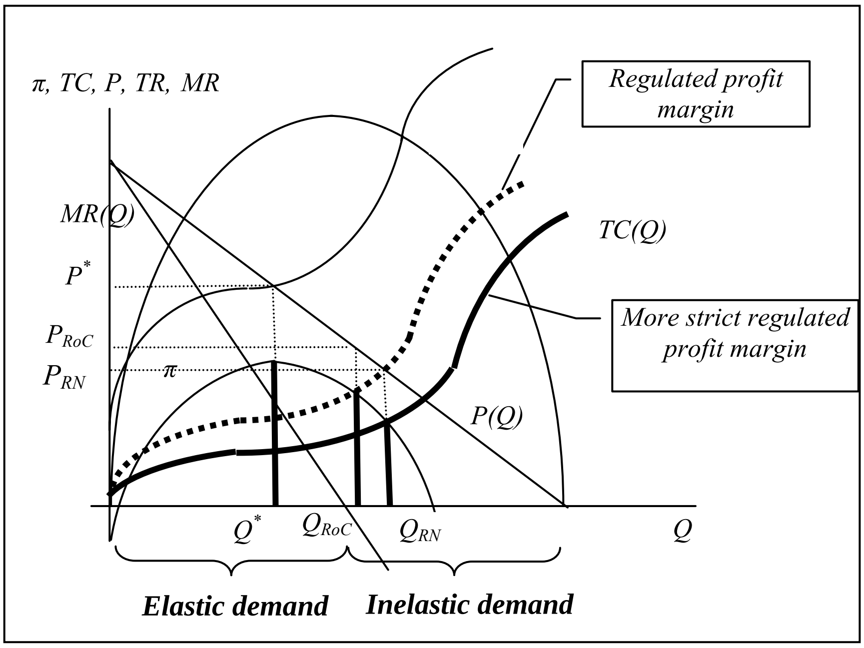
Task (7) ... (9), or task (4) ... (6) analytically describes the situation that is geometrically interpreted on the pic. 1. Graphs of the basic functions describing production and cost attributes of the regulated firm and the attributes of the relevant market as well are taken from the return on output regulation model.

Revenues function in the situation on the pic. 1. Reaches its maximum in the point q^*, which represents supply of the firm in the elastic demand zone. Firm is selling its goods for a relative high price p^* in the elastic demand zone, i.e. in the zone for positive values of the marginal revenues function $t(q)$.

Regulated firm has the tendency to set its decision making parameters in a way the limit set by the regulator would allow it to reach maximum profit. As we can seen on the pic. 1, margin of the regulated profit is represented by the curve, which is a transformation of the total cost curve. We get it by multiplying the function values by regulation parameter RoC , i.e. by the regulatory allowed portion of the profit from the costs.

Let us note that *RoC* parameter is not in percentage but in relative expression. So it has the value $RoC \in (0,1)$. In the situation presented on the pic. 1. the firm will have volume of the output q_{RoC} and with greater volume of the output by the lower price p_{RoC} it will reach lower profit.

Pic. 1: Regulation on the return over costs basis – *ROC*



If the regulator would decide to use more strict regulation conditions and enforce lower allowed rate of profit RoC_N , ($RoC > RoC_N$) upon the regulated entity, the regulated entity would henceforth increased its output on the volume q_{RN} by decreasing production price p_{RN} and by gradual decrease of the profit. However price decrease and volume of output increase at the same time encourages the growth of the social welfare.

The comparison of the products market price and production costs of the firm also leads to interesting conclusion while using this type of the regulation. Let us explore the reasonable profit margin in return over costs regulation from this aspect again.

$$RoC \times TC(Q) \geq \pi(Q) \tag{10}$$

In this condition we express the total cost function of the firm analytically. We get reformulated condition expression (10):

$$RoC \times (nv(q) + n_F) \geq p(q) \times q - (nv(q) + n_F)$$

and after further modification

$$p(q) \times q \leq RoC \times (nv(q) + n_F) + (nv(q) + n_F)$$

$$p(q) \times q \leq (1 + RoC) \times (nv(q) + n_F) \qquad (11)$$

After division of the equation by the supply volume $q>0$ we get

$$p(q) \leq \frac{(1 + RoC) \times (nv(q) + n_F)}{q}$$

$$p(q) \leq (1 + RoC) \times \frac{(nv(q) + n_F)}{q}$$

$$p(q) \leq (1 + RoC) \times np(q) \qquad (12)$$

where

$np(q) = \dfrac{(nv(q) + n_F)}{q}, \qquad q > 0$ are total average costs of the firm.

Relation (12) represents substantial feature of the return over costs regulation, which also explains already mentioned mystification about direct relation of the regulated reasonable profit margin and total costs of the natural monopoly.

3 Conclusion

On the basis of the relation (12) and from the geometrical interpretation of the optimization task solution (1) ... (3) we can formulate the following important conclusions about firm behavior in the condition of the return over costs regulation:

1. In general, return over costs regulation constructs reasonable profit margin for the regulated entity on the basis of the proportional portion of its total

expended costs. This proportional portion is defined by the *RoC* parameter. So primary it encourages the producer to produce greater volume of supply by the lower price, which is increasing social welfare.

2. The particular optimal position of the regulated firm is determined by the characteristics of the cost function, which is directly related to the character of the profit function on one side, because

$$\pi(q) = t(q) - n(q)$$

And on the other side by the characteristics of the price-demand function *p(q)*, which specifies elastic and inelastic demand zones.

3. From the relation (16) we can see that regulated firm can set its production parameters, production prices and consumption of production factors only in a manner for its production market price to be up to *RoC* percent greater than the average unit costs of production. We can see that ineffective cost increase of the firm, in accordance with regulatory relation of this method, would be albeit creating room for reasonable profit increase however the validity of the relation (16) needs to be ensured and such combination of supply production price found, that would ensure its consumption.

It is obvious that in elastic demand zone, i.e. by positive marginal revenues, regulated firm produces greater volume of output compared to nonregulated firm and tries not to waste the production factors.

4 References

1 Bien, F.: Systemwechsel im Europäischen Kartelrecht. In: Der Betrieb N' 46/2000. Düsseldorf : Verlagsgruppe Handelsblatt, 2000.

2 Fendek, M, - Fendeková, E: Mikroekonomická analýza. Bratislava: IURA Edition, 2008.

3 Fendeková, E.: Podmienky rovnováhy firmy v rôznych typoch trhových štruktúr. In: Transformácia ekonomiky a rozvoj podnikania v SR. Bratislava: FPM EU, 1995

4 Fendeková, E: Oligopoly a regulované monopoly. Bratislava: IURA Edition, 2006.

5 Mehr Wettbewerb auf allen Märkten. Zehntes Hauptgutachten der Monopolkommission 1992/1993. Baden-Baden: Nomos Verlaggeselschaft, 1993.

6 Martin, S.: Advanced Industrial Econmics. Cambridge: Blackwell, 1996.
7 O'Sullivan, A. - Sheffrin, S. - Perez , P.: Microeconomics: Principles, Applications, and Tools. New York: Prentice Hall, 2006.
8 Pepall, L. - Richards, D. J. – Norman, D.: Industrial Organization: Contemporary Theory and Practice (with Economic Applications). New York: South-Western College Publishing, 2004
9 Scherer, F.M.-Ross, D.: Industrial Market Structure and Economic Performance. Boston: Houghton Miffin Company, 1990.
10 Shy, Oz.: The Economics of Network Industries. Cambridge: Cambridge University Press, 2001
11 Tirole, J.: The Theory of Industrial Organisation. Massachusetts, Cambridge: The MIT Press, 1997.
12 Train, K., E.: Optimal Regulation. London: The MIT Press, 1995

Prof. Dr. Michal Fendek, Department of Operations Research and Econometrics, University of Economics Bratislava, Dolnozemská 1, 852 35 Bratislava, Slovakia, e-mail:fendek@dec.euba.sk

SUPPLY CHAIN FORMATION

Petr Fiala, Department of Econometrics, University of Economics, Prague

Abstract: Supply chain management has generated a substantial amount of interest both by managers and by researchers. Supply chain formation is the problem of determining the production and exchange relationships across a supply chain. The paper is devoted to modeling and solving the supply chain formation problem. The problem is modeled as the task dependency network. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items. Using of combinatorial auctions is promising for solving the supply chain formation problem.
Keywords: supply chain management, supply chain formation, task dependency network, combinatorial auction

1. INTRODUCTION

Supply chain management has generated a substantial amount of interest both by managers and by researchers. Supply chain management is now seen as a governing element in strategy and as an effective way of creating value for customers. There are many concepts and strategies applied in designing and managing supply chains (see Simchi-Levi, Kaminsky, Simchi-Levi, 1999). The expanding importance of supply chain integration presents a challenge to research to focus more attention on supply chain modeling (see Tayur, Ganeshan & Magazine, 1999). Supply chain management is more and more affected by network and dynamic business environment. The overall business environment is becoming increasingly dynamic. Demand and supply for custom products can be very dynamic. Supply chains operate in network environment as supply networks. Dynamic information and decision-making models are called to accommodate this new changes and uncertainties. Complex business negotiations often involve interrelated exchange relationships among multiple levels of production. Te paper describes an approach for modeling and solving the supply chain formation problem. The problem is formulated in Section 2. The model as the task dependency network is introduced in Section 3. In Section 4, using of combinatorial auctions for solving the supply chain formation problem is presented. Results and possible extensions are discussed in Section 5.

2. SUPPLY CHAIN FORMATION PROBLEM

Supply chain formation is the problem of determining the production and exchange relationships across a supply chain (Walsh, Wellman, Ygge, 2000, Walsh, Wellman, 2003). To respond to rapidly changing market conditions, companies must be able to dynamically form and dissolve business interactions, requiring automated support for supply chain formation. Agents in the supply chain are characterized in terms of their capabilities to perform tasks, and their interests in having tasks accomplished. A central feature of the model of the problem is hierarchical task decomposition. In order to perform a particular task, an agent may need to achieve some subtasks, which may be delegated to other agents. These may in turn have subtasks that may be delegated, forming a supply chain through a decomposition of task achievement. Constraints on the task assignment arise from resource contention, where agents require a common resource to accomplish their tasks. Tasks are performed on behalf of particular agents. If two agents need a task then it would have to be performed twice to satisfy them both. In this way, tasks are the same as any other discrete, rival resource. Hence, there is no distinction in the model, and use the term "good" to refer to any task or resource provided or needed by agents. The assumption that goods cannot be shared or reused is necessary for analysis.

3. TASK DEPENDENCY NETWORK

The problem, can be formulated as so called a task dependency graph (Walsh, Wellman, Ygge, 2000). A task dependency network is a directed acyclic graph $G = (V, E)$, representing dependencies among agents and goods. $V = G \cup A$, where G is the set of goods and $A = C \cup \Pi$ is the set of agents, comprised of consumers C, and producers Π. Edges, E, connect agents with goods they can use or provide. There exists an edge $<g, a>$ from $g \in G$ to $a \in A$ when agent a can make use of one unit of g, and an edge $<a, g>$ when a can provide one unit of g. When an agent can use or provide multiple units of a good, separately indexed edges represent each unit. The goods can be traded only in discrete quantities.

Figure 1 shows an example task dependency network for a supply chain problem. Here the goods are indicated by circles, and agents by boxes and triangles. Suppliers and consumers are indicated by boxes and producers by triangles. The numbers under agent boxes represent production costs and consumption values. An arrow from an agent to a good indicates that the agent can provide that good, and an arrow from a good to an agent indicates that the agent can make use of the good.

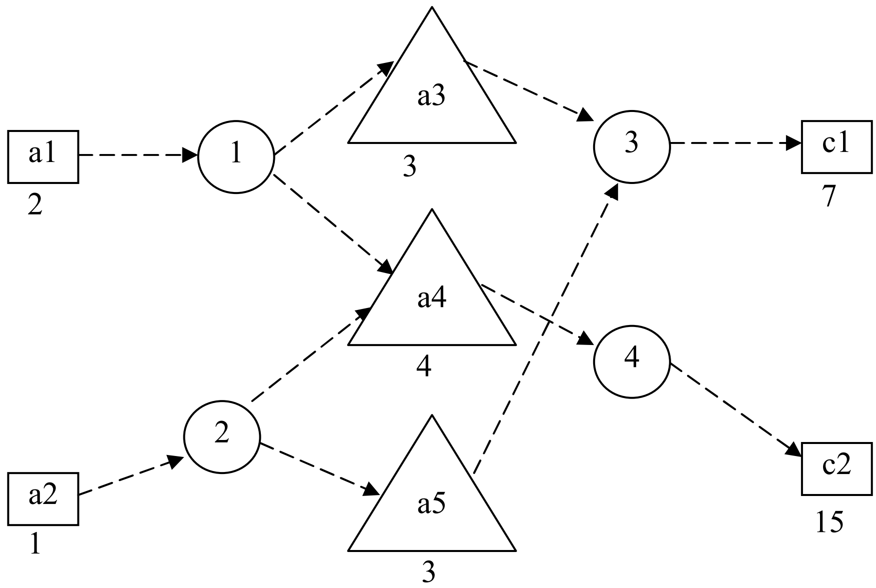


Fig. 1. Task dependency graph

An allocation is a subgraph $(V', E') \subseteq (V, E)$. For $g \in G$, an edge $<a, g> \in E'$ means that agent *a* provides *g*, and $<g, a> \in E'$ means a acquires *g*. A producer is active iff it provides its output. A producer is feasible iff it is inactive or acquires all its inputs. Consumers are always feasible. An allocation is feasible iff all agents are feasible and all goods are in material balance, that is the number of edges into a good equals the number of edges out.

The value of allocation (V',E') is:

$$value((V',E')) \equiv \sum_{c \in C} v_c(E') - \sum_{\pi \in \Pi} k_\pi(E'),$$

where

v_c is consumption value for the consumer *c*, k_π is production cost for the producer π.

The set of efficient allocations contains all feasible allocations (V*, E*) such that

$value((V^*, E^*)) = \max_{(V',E')}\{\text{value}((V', E')) \mid (V', E') \text{ is feasible}\}$.

A solution is a feasible allocation such that one or more consumers acquire a desired good. If $c \in C \cap V'$ for solution (V', E'), then (V', E') is a solution for consumer.

Figure 2 presents a suboptimal solution and Figure 3 presents an efficient solution of the supply chain formation problem.

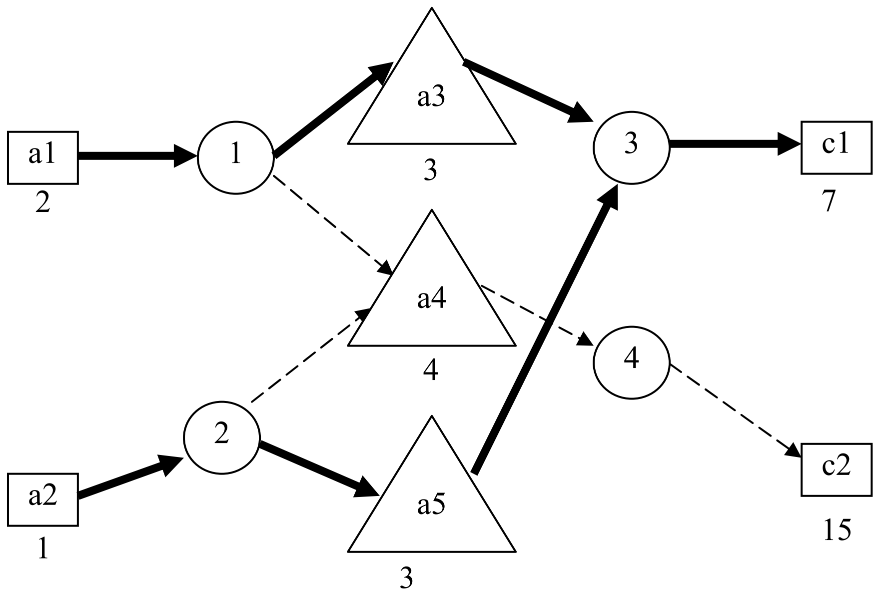


Fig. 2. Sub-optimal solution

Value of allocation is: 7 – 2 – 1 – 3 – 3 = – 2.

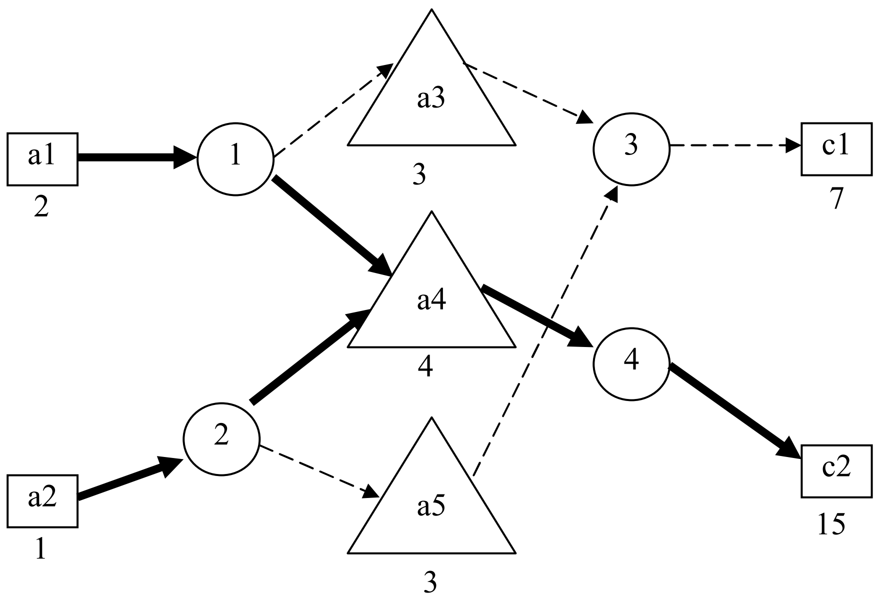


Fig. 3. Efficient solution

Value of allocation is: 15 – 2 – 1 – 4 = 8.

4. COMBINATORIAL AUCTION MECHANISM

Auctions are important market mechanisms for the allocation of goods and services. Design of auctions is a multidisciplinary effort made of contributions from economics, operations research, informatics, and other disciplines. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items, so called bundles. The advantage of combinatorial auctions is that the bidder can more fully express his preferences. This is particular important when items are complements. The auction designer also derives value from combinatorial auctions. Allowing bidders more fully to express preferences often leads to improved economic efficiency and greater auction revenues. However, alongside their advantages, combinatorial auctions raise a host of questions and challenges (Cramton et al., 2006).

Using of combinatorial auctions is promising for solving the supply chain formation problem (Walsh, Wellman, Ygge, 2000, Walsh, Wellman, 2003). Combinatorial auction mechanism is one-shot mechanism. Agents submit bids reporting costs and values, and then the auction computes an allocation that maximizes the reported value and informs the agents of results. An agent pays the price it bid for the allocation it receives. If the auction receives more money than it pays out, the proceeds are distributed evenly among all consumers.

Bidding Language

Agent *a* places a bid b_a of the form

$$<r_a ,<g_1 ,q_a^1>, \ldots , <g_n ,q_a^n>>$$

where

q_a^i is the integer quantity that agent *a* demands (positive for input demands and negative for output demands) for good g_i, and

r_a is its reported willingness to pay (or be paid, in the case of negative numbers) for the demanded bundle of goods.

Given a set of bids B, the auction computes the winning allocation from:

$$f(B) = \max_x \sum_{b_a \in B} r_a x_a$$

subject to

$$\sum_{b_a \in B} q_a^i x_a = 0 \text{ , } i = 1 \ . \ . \ . \ n,$$

where x_a = 1 if agent a wins the bid, and x_a = 0 otherwise.

Combinatorial Bidding Policies

If agents behave non-strategically (i.e., bid their true valuations) in the combinatorial auction mechanism, then the result will always be an efficient allocation.

Analysis for the case when agents behave strategically

- Assumption that it is common knowledge that consumer bid their true values.
- Producers play Bayes-Nash equilibrium strategies.
- N buyers one seller case with producer's cost and consumer's values drawn from uniform probability distribution [0,1], and buyer i has value v_i for the good, then a Bayes-Nash equilibrium bidding policy is

$$r_\pi = -k_\pi - \frac{1}{N}(1 - k_\pi) \text{ .}$$

Plausible bidding policy for complicated networks:

- A producer bids to obtain a fraction of the expected available surplus scaled by the expected proportion of its contribution to the global value.
- Let $\Pi^* \subseteq A$ be the producers participating in the efficient allocation. The contribution of these producers to the value of the allocation is Δ^* , where

$$\Delta^* = f^*(A) - f^*(A - \Pi^*).$$

- The contribution Δ_π of a producer to the value of an allocation is the difference between the efficient global value, and the global value with π excluded from the allocation,

$$\Delta_\pi = f^*(A) - f^*(A - \{\pi\}).$$

- A producer's relative contribution can then be defined in terms of its expected proportional contribution, conditional on its being part of the efficient allocation.

5. CONCLUSIONS

The task dependency network model provides a basis for understanding the automation of supply chain formation. Using of combinatorial auctions is promising for solving the supply chain formation problem. There are some ways to extend the bidding policies to accommodate more general production capabilities and consumer preferences. With these extensions it can be modeled capabilities and preferences on multi-attribute goods (e.g., goods with multiple features such as quality and delivery time, in addition to price and quantity) by simply representing each configuration as a distinct good in the network.

ACKNOWLEDGEMENTS
The research project was supported by Grant No. 402/07/0166 from the Grant Agency of Czech Republic „Combinatorial Auctions - Modeling and Analysis“.

REFERENCES

[1] Babaioff, M., Walsh, W. E. (2003): Incentive-Compatible, Budget-Balanced, yet Highly Efficient Auctions for Supply Chain Formation. In *Fourth ACM Conference on Electronic Commerce*, 64–75.

[2] P. Cramton, Y. Shoham and R. Steinberg (eds.) (2006): *Combinatorial Auctions*. MIT Press, Cambridge.

[3] Karimi, R., Lucas, C., Moshiri, B. (2007): New Multi Attributes Procurement Auction for Agent-Based Supply Chain Formation. *International Journal of Computer Science and Network Security* 7, 255-261.

[4] Simchi-Levi, D., Kaminsky, P., Simchi-Levi, E. (1999): *Designing and Managing the Supply Chain: Concepts, Strategies and Case studies.* Boston: Irwin/ Mc Graw-Hill.

[5] Tayur, S., Ganeshan, R., & Magazine, M. (1999): *Quantitative models for supply chain management*. Boston: Kluwer.

[6] Walsh, W. E., Wellman, M. P. (2003): Decentralized Supply Chain Formation: A Market Protocol and Competitive Equilibrium Analysis. *Journal of Artificial Intelligence Research* 19, 513-567.

[7] Walsh, W. E., Wellman, M. P., Ygge, F. (2000): Combinatorial Auctions for Supply Chain Formation. In *Second ACM Conference on Electronic Commerce*, 260–269

Prof.RNDr.Ing. Petr Fiala, CSc., MBA, Department of Econometrics, University of Economics, W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

CHINA'S CITIES: A DATA ENVELOPMENT ANALYSIS OF INVESTMENT CLIMATE EFFICIENCY

GERALD GROSHEK

Abstract: Since opening its economy, China has experienced an almost unbroken string of impressive GDP growth rates. However, equally important in China's recent development has been the impressive growth in internal inequality. This study measures the investment climate across 120 Chinese cities using data envelopment analysis as an alternative ranking that is compared with the World Bank's approach. The overall investment effectiveness results are then presented using a geographical information systems (GIS) approach to illustrate the competitiveness among China's cities.
Keywords: Data envelopment analysis, Chinese growth and inequality, regional investment climate.

1. Introduction

China's economic development continues to be the focus of interest and uncertainty in academic studies with much attention paid to macro measures of its strong economic growth rates. A closer look reveals a rather high degree of regional unevenness in China's development. For example GDP per capita in southeast China is 50% greater than in northeast China and 150% greater than in central and southwest China. The UNDP estimates China's Gini coefficient at 44.7 with the share of income of the lowest 20% at 4.7% and that of the richest 20% at 50% (UNDP, 2006). Consequently, the study of China's economic development has increasingly turned to the geographic disparities inherent in the gradualist approach to economic reform. Policy reforms initiated in the 1970s in the special economic zones have subsequently been spread beyond the coastal regions. Several studies (Bai, Du, Tao, and Tong forthcoming; Kaufmann, Kraay, and Zoido-Lobaton 1999, 2002; Tenev, Zhang, and Brefort 2002; and Wang and Yao 2001) explore a variety of possible causal factors behind China's uneven economic development. Standard economic growth models attribute per capita income disparities to differences in capital-labor ratios, which suggests that the unevenness in investment across China's regions follows from uneven investment within China's cities.

Recently, the World Bank (2006) provided an examination of these regional differences based on an extensive dataset of the institutional, market, infrastructure, and local policy characteristics of 120 Chinese cities. The report developed previous—and more limited—studies (Dollar et. al. 2003a, 2003b, 2003c) in an attempt to characterize how the efforts of local governments in China to improve public sector efficiency and market structures have affected economic development, factor productivity, and investment. The World Bank study employs multiple criteria to correlate investment climate and government effectiveness with a series of location characteristics. The result is a ranking that places cities into rating quintiles that is then aggregated according to geographic location. Although this is similar to previous studies, there is no attempt to disaggregate the results below six rather large regional clusters.

While offering an useful data set, the World Bank study provides an inconsistent and often confusing analysis of the factors contributing to the variance in investment climates among the 120 cities. The methodological approach is limited by an inconsistent nomenclature and a lack of clarity as data results do not follow the exposition. Several factors cited as relevant (expectation of informal payments for loans; average number of days need to clear customs; output losses due to inadequate power or transport infrastructure; industrial waste disposal; average per capita green space; infant mortality; per capita education expenditures; and unemployment rate) are not supported with data. Several factors not identified as relevant to local variation in investment climate (number of employees; miles of graded roads; electricity prices) appear in the data table, but are not covered as an independent factor in the exposition of relevant factors. While including labor costs as a factor in the city characteristics, it is unclear if the wage data are meant to reflect this indicator.

This study extends the World Bank analysis by applying data envelopment analysis to 20 city level characteristics across 120 cities. The application of data envelopment analysis[1] (DEA) to the World Bank dataset might provide greater insight into the character of China's development. DEA enables a consistent evaluation of the importance or weight of each input characteristic for each city location. In addition to providing an output or efficiency ranking of Chinese cities that is compared to previous rankings, the DEA efficiency results are mapped in a geographical information systems (GIS) framework to illustrate the pattern of regional efficiency and development in China. The GIS framework plots each city's relative success as a decision making unit (DMU) to determine if spatial considerations play a role. Here, the GIS framework will assist in determining if efficient locations are clustered according to the limited regional groupings established in the World Bank study or dispersed throughout China in some alternative pattern.

[1] See Cooper, Seiford, Tone (2006)

2. Data – The Decision Making Units, Inputs, and Outputs

Methodologically, this study constructs efficiency measures for each city location based on the data characteristics from the WB study. These characteristics form a set of inputs that contribute to each city's efficiency in maximizing output objectives as measured in terms of local shares of multinational investment and total factor productivity.

Table 1. Input Measures Leading to Effective Investment Climate

1. City characteristics
    1. per capita GDP
    2. economic growth
    3. transport costs
2. Government effectiveness
    1. taxes and fees
    2. firm expenditures on entertainment and travel
    3. time spent on bureaucratic interactions
    4. customs clearance performance
    5. balance between private sector firms and state-owned enterprises
    6. labor flexibility
    7. access to bank loans
    8. skills and technology
    9. protection of property and contract rights
    10. adequacy of power and transport
3. Harmonious society
    1. environment
        1. air quality
        2. industrial waste disposals
        3. per capita green space
    2. health
        1. health insurance coverage for permanent workers
        2. infant mortality
    3. education
        1. female share in total school enrollment
        2. per capita expenditures on education

A decision making unit (DMU) is regarded as an entity responsible for converting inputs into outputs and whose performance is to be evaluated. In this study, the DMUs are the individual 120 Chinese cities included in the World Bank study. The input data includes 20 efficiency factors identified in the World Bank study that contribute to the local investment climate. As noted in Table 1, these factors are categorized according to a series of city characteristics, government effectiveness measures, and indicators of progress towards a harmonious society. The output data include each city's share of foreign direct investment. For each city DMU, the numerical data available for each input and output need not be congruent and are assumed to be non-negative with smaller amounts preferable for inputs and larger amounts preferable for outputs so that the resulting scores reflect the DEA efficiency principles described below.

3. The Basic CCR-I Model and the Dual Problem

The study proceeds using a standard variant of data envelopment analysis based on the Charnes, Coopers, Rhodes approach. Since we are interested in minimizing inputs in Table 1 while generating at least the given levels of output (inward FDI flows), the methodology uses the Charnes-Coopers-Rhodes input-oriented model (CCR-I). As noted in Cooper, Seiford, and Tone (2005, p43), the CCR-I model used here is based on the matrix (X,Y), and is formulated as an LP problem with row vector *v* for city characteristic input multipliers and row vector *u* as FDI output multipliers. The multipliers form the variables in the LP model given in the following vector-matrix notation:

$$(LP_0) \quad \max \quad uy_0 \tag{1}$$

subject to:

$$vx_0 = 1 \tag{2}$$

$$-vX + uY \le 0 \tag{3}$$

$$v \ge 0, u \ge 0 \tag{4}$$

For a typical primal linear programming problem, this gives:

the Matrix $A = \begin{pmatrix} x_0 & 0 \\ -X & Y \end{pmatrix}$;

vector $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$;

vector $c^T = \begin{pmatrix} 0 & y_0 \end{pmatrix}$;

and vector of primal variables that comprises the input and output weights: $\begin{pmatrix} v \\ u \end{pmatrix}$

To express the dual problem of (LP_0), we employ a standard real variable θ and a nonnegative vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_n)^T$ of variables as follows:

$$(DLP_0) \quad \min \quad \theta \tag{5}$$

subject to:

$$\theta x_0 - X\lambda \geq 0 \tag{6}$$

$$Y\lambda \geq y_0 \tag{7}$$

$$\lambda \geq 0 \tag{8}$$

where for typical dual linear programming problem gives:

the Matrix $A^T = \begin{pmatrix} x_0 & -X \\ 0 & Y \end{pmatrix}$;

vector $b^T = \begin{pmatrix} 1 & 0 \end{pmatrix}$;

vector $c = \begin{pmatrix} 0 \\ y_0 \end{pmatrix}$;

and vector of dual variables including a scalar efficiency measure for each city DMU (θ) and values of the reference set: $\begin{pmatrix} \theta \\ \lambda \end{pmatrix}$.

The feasible solution for the (DLP$_0$) is $\theta = 1, \lambda_0 = 1, \lambda_j = 0 \ (j \neq 0)$ where the optimal θ (denoted by θ^*) is less than or equal to 1. The constraint in (7)—when combined with the nonzero (i.e. semi-positive) assumption for the data— constrains λ to be nonzero because $Y \geq 0$ and $y_0 \neq 0$. Also, because θ must be greater than zero (from (6)), its optimal value is bounded as follows:

$$0 < \theta^* \leq 1. \tag{9}$$

4. DEA Efficiency Results and Discussion

Each of the 20 characteristics were evaluated as inputs against the FDI inflows in each of the 120 Chinese cities (DMUs). In order to focus on the ranking of each city relative to the factors in Table 1 and to compare with the results of the World Bank (2006) study, the results examined here are limited to measures of Pareto-Koopmans efficiency that result from ratio efficiency (θ=1) and zero input slack. Table 2 presents the efficiency scores for each city ranked according to the average score of each DMU over all 20 criteria. As noted above, the feasible solution in the dual equation (DLP$_0$) includes $\theta = 1, \lambda_0 = 1, \lambda_j = 0 \ (j \neq 0)$. When all conditions within this outcome occur, the DMU is considered to be CCR or Pareto-Koopmans efficient. When an optimal efficiency score (θ^*=1) is obtained, the member state is ratio efficient and positioned along the efficient frontier.

Additionally, if no other DMU serves as an efficiency reference, there is zero slack and the city is its own reference.[2]

Table 2. DEA Top 40 Efficiency Results

Rank	DMU	Efficiency Score	Rank	DMU	Efficiency Score
1	*Zhuhai*	1.0000	21	*Dalian*	0.9434
1	Yuxi	1.0000	22	*Fuzhou*	0.9131
1	**Xiamen**	1.0000	23	Nantong	0.8762
1	Wenzhou	1.0000	24	Jiaxing	0.8720
1	**Suzhou**	1.0000	25	**Shaoxing**	0.8336
1	*Shenzhen*	1.0000	26	*Weihai*	0.8284
1	Shantou	1.0000	27	Ganzhou	0.8165
1	*Shanghai*	1.0000	28	Wuxi	0.8124
1	Nanjing	1.0000	29	**Qingdao**	0.7950
1	Lianyungang	1.0000	30	Jinhua	0.7940
1	*Jiangmen*	1.0000	31	*Foshan*	0.7902
1	Huizhou	1.0000	32	Changzhou	0.7850
1	**Hangzhou**	1.0000	33	**Yantai**	0.7825
1	Haikou	1.0000	34	Huzhou	0.7717
1	*Guangzhou*	1.0000	35	Linyi	0.7626
1	Chongqing	1.0000	36	Yancheng	0.7118
1	*Dongguan*	1.0000	37	Langfang	0.6900
18	Zhangzhou	0.9989	38	*Beijing*	0.6808
19	*Ningbo*	0.9910	39	Maoming	0.6614
20	Quanzhou	0.9748	40	Weifang	0.6553

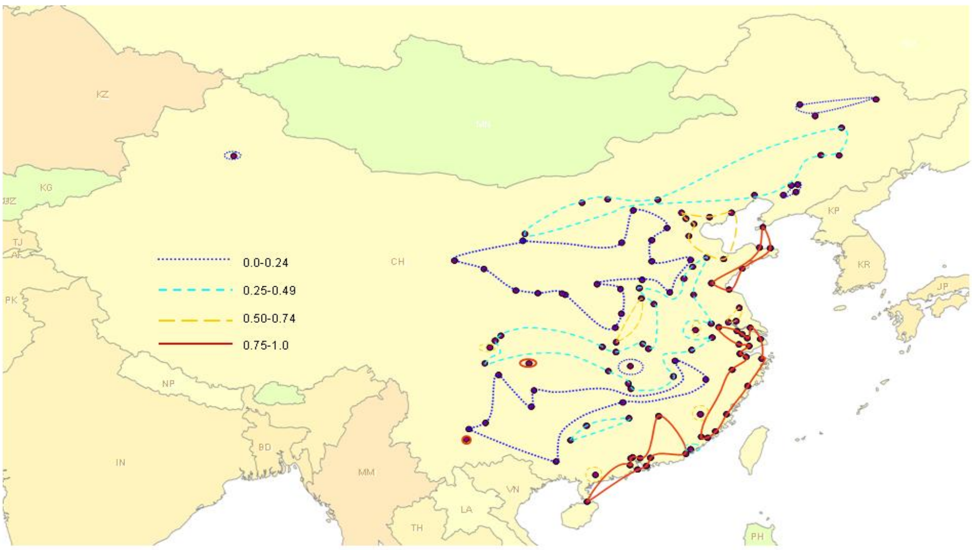
World Bank Golden Cities in **Bold;** Silver Cities in *Italics*

Of the 120 cities considered, only 17obtain CCR or Pareto-Koopmans efficiency. Three of these cities (Xiamen, Suzhou, and Hangzhou) also appear on the World Bank's list of Golden cities as denoted by bold typeface. A further six cities that achieve Pareto-Koopmans efficiency also appear as Silver cities (Shenzhen and Shanghai for example) according to the World Bank. It is evident that the World Bank study excludes several cities found to be efficient through DEA analysis. Additionally, several cities deemed by the World Bank to be Golden (Shaoxing, Qingdao, and Yantai) or Sliver (Beijing for example) rank lower in the DEA evaluation. The difference might result from the more consistent use of the input criteria and relative input weightings of the DEA evaluation that are not found in the World Bank ranking.

[2] Weak efficiency occurs when only ratio efficiency is present.

Taken together, the 120 cities can be evaluated graphically using a geographical information systems approach to illustrate any patterns in the location significance of each DMU. By dividing the city DMUs into quartiles based on their DEA efficiency score, it is possible to illustrate any patterns that might indicate spillover effects from the presence or absence of efficiency among Chinese cities and whether these characteristics are based on distinctions between large regional groupings as in the World Bank study. Inefficient cities (0,00 to 0,24 and 0,25 to 0,49) are represented by the finely hashed lines. The more CCR efficient cities are represented by the dashed lines (0,50 to 0,74) and the solid lines (0,75 t0 1,00). It is not surprising to find that the most efficient cities are located along the southeast and northern coast. However, a couple of pockets of efficiency appear in south central China where income per capita is generally lower as noted above. This result indicates that improving investment conditions in this region might result in greater income potential if extended on a scale found in costal China. In terms of inefficient cities, there are several zones of lower efficiency that exist side by side with their more efficient counterparts. Although the general view finds that efficiency is more well-established along the coast with lower efficient cities appearing as one moves north and west, there appear to be more than six regional groupings with variation in efficiency that crosses standard geographical boundaries.

Figure 1. Location of Zones of Efficiency Among Chinese Cities



5. Conclusion

This paper provided a alternate view of the efficiency of China's cities based on and application of DEA to the 2006 World Bank dataset. The results indicate that the World Bank's approach to analyzing the data might not incorporate the relevant factors in a consistent manner when identifying efficient cities. The mapping of these results demonstrates that efficiency differences across China's cities are more refined than that suggested in reference to large regional groupings. It would be productive to add a time series aspect to the DEA approach to provide insight into the significance of each input factor to each city, as well as how each city might improve its investment climate efficiency. This exercise must wait the repeated collection of annual data. Further work could expand on the efficiency results by incorporating the DEA results for input weights into a spatial framework that will enable a determination of whether Chinese cities with similar input weights are clustered or dispersed.

References

1. Bai, Chong-En, Yingjuan Du, Zhigang Tao and Sarah Y. Tong. , forthcoming. “Local Protectionism and Regional Specialization: Evidence from China’s Industries.” Journal of International Economics.
2. Cooper, William W., Seiford, Lawrence M., and Tone, Kaoru. 2006. Introduction to Data Envelopment Analysis and Its Uses: With DEA-Solver Software and References. Springer
3. Dollar, David R., Anqing Shi, Wang, Shuilin, and Lixin Xu. 2003a. “International Comparison on Chinese Investment Climate.” Comparative Economic and Social Systems. Vol. 107 (3)24-34.
4. David Dollar, Anqing Shi, Shuilin Wang, Lixin Colin Xu. 2003b. Improving City Competitiveness through the Investment Climate: Ranking 23 Chinese Cities. Washington D.C.: World Bank.
5. David Dollar, Mary Hallward-Driemeier, Anqing Shi, Scott Wallsten, Shuilin Wang, and Lixin Colin Xu. 2003c. Improving the Investment Climate in China. Washington D.C.: World Bank.
6. Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 1999. "Governance Matters." World Bank Policy Research Working Papers: Washington.
7. Kaufmann, Daniel, Aart Kraay, and Pablo Zoido-Lobaton. 2002. "Governance Matters II: Updated Indicators for 2000/01." World Bank Policy Research Working Papers: Washington.

8. Tenev, Stoyan, Chunlin Zhang, and Loup Brefort. 2002. Corporate Governance and Enterprise Reforms in China. Washington D.C.: World Bank.
9. UNDP. 2006. Human Development Report 2006 - Beyond scarcity: Power, poverty and the global water crisis. http://hdr.undp.org/hdr2006/report.cfm
10. Wang, Yan and Yudong Yao. 2001. "Sources of China's Economic Growth, 1952-1999: Incorporating Human Capital Accumulation." Policy Research Working Paper Series. No. 2650. Washington, DC. World Bank.
11. World Bank, 2006, Governance, Investment Climate, and Harmonious Society: Competitiveness Enhancements for 120 Cities in China. Poverty Reduction and Economic Management, Financial and Private Sector Development Unit, East Asia and Pacific Region. Report No. 37759-CN

prof. Gerald Groshek, Ph.D.
University of Redlands
School of Business
Redlands, California, USA

COMPARATIVE PROPERTIES OF STRUCTURAL MACROECONOMIC MODELLING

Jana Hančlová
Faculty of Economics, VŠB-Technical University of Ostrava

Abstract
This paper reviews the general and Czech literature in macroeconomic modelling from more than half a century of model-building. The first part of the paper deals with the historic approaches when constructing the core of these macroeconomic models and then with the elementary tendencies of econometric macroeconomic modelling. The second part summarises the latest development of structural macroeconomic modelling for the Czech economy. The third part is aimed at issue of macroeconomic data for evaluation of proposed models. In the summary of the paper it is optimisticly said that at present it is possible to specify the segments of the macroeconomic models of the Czech market economy and verify this model for Reuters EcoWin database quarterly data since 1993 resp. 1995.

Keywords
Macroeconomic modelling, model specification, model evaluation, econometric methodology.

1 Introduction

The overview of 60-year development of macroeconomic modelling brings a wide range of existing models which can be classified according to many criteria – time, model specification, methodology, model range, model complexity, dynamics, pre-requisitions of the modelling, modelling of a certain type of the economy, using the model etc.
The objective of the paper is to introduce an evaluation for present specification of the macroeconomic model of the Czech market economy in the context of the EU accession. The basic question is: "Is it possible to create, adequately estimate and use for prediction or simulation a smaller standard market economy model and at the same time to respect certain specific problems related to the institutional changes or shocks?"

2 History and problems of macroeconometric modelling

A minimum 50-year history of macroeconomic modelling is a basic source for understanding in the process of the creation of a new macroeconomic model and at the same time of an adequate approach for estimation, simulation and further analysis of the economic systems behaviour, investigations of the segments stability and possibilities how to use the proper economic policy tools. The background monitoring articles dealing with the history and basic approaches towards macroeconomic modelling are A. Valadkhani (2004) and K. Šmídková (1995). Valdkhani follows the history of macroeconometric modelling in the period from the late 1940s to the 1960s and show that this field has contributed to the expanding knowledge of both economists and econometrics. There were several issues invalidated the use of macroeconometric models from the early 1970s:

- theoretical contrasts with rational expectations theory,
- structural instability,
- the arbitrary division of endogenous-exogenous variables of the model,
- the possible existence of the problem of unit roots (spurious regressions) and insufficient amount of econometric „know-how".

Macroeconometric modelling in developing countries has been subject to criticisms on a greater scale because of the presence of an additional adverse factor of data unreliability. In criticising the Cowles Commission approach, which is based on the structural multi-equation modelling, motivated econometricians to devise three famous methodologies: Sims's a theoretical VAR (data-driven approach – there are no exogenous variables), Leamer's methodology (redefining the concept of exogeneity, the implementation of the Bayesian approach), Hendry methodology is the London School of Economics methodology, which accepts structural models and adopts general dynamic autoregressive distributed lag (ADL) model.
Śmídková (1995) deals with the basic directions of macroeconomic modelling in the period from the late 1950s to the 1980s. Historic cores of the models generated during 1950s- 1960s were based on Keynes' effective demand theory and were focused on behaviour analysis of final aggregate demand (sectors of private consumption, government consumption, investment and export) and Keynes'multiplicators. It usually regards the flow quantities and monetary and financial area are often exogenous inputs. In 1970s – 1980s the original macroeconomic models were extended as follows: modelling of aggregate supply behaviour (supply-side modelling), modelling of financial markets (flow of funds analysis), modelling of long-term (dis)equilibrium ((dis)equilibrium modelling) and modelling of problems for small open economy with focus on different behaviour of the domestic and foreign sector. The basic models are derived from American, Scandinavian, British and Canadian schools. The critics of using macroeconomic models in the following areas are reminded:

- an application of estimated models for the simulations of economic policies,
- a possibility to formulate and estimate in long-term periods (shocks change model parameters),
- a differentiation in advance of exogenous and endogenous variables,
- a constant model changing in a way that R^2 and t-statistics would fit the best.

The paper specifies four selected macroeconomic models:

- Klein-Goldberger model dated back to 1952 (Klein´s model),
- Rhomberg model for Canada dated back to 1959 (aimed at exchange rate issue of a small open economy),
- a macroeconomic model of London Business School of Economics – 70-ies (a big model),
- a Swedish central bank model coming from 1988 (a small highly aggregated model).

3 Selected existing macroeconomic models of the Czech economy

The insight in macroeconomic modelling of the Czech economy can be divided into two groups. The first group deals with the specification of the macroeconomic model core only with no econometric verification and the second group deals with the estimation and use of these models.

In the first group we can find two essential papers:

- Hanousek a Tůma (1994) who tried to propose a simple macroeconomic model (11 equations) for an application in the Czech transition economy. There are discussion on the specification of a model, the availability of data, the number of observations and the stability of parameters. The proposed framework is based on an IS-LM model (Keynes' scheme) and an inclusion of production functions, a Phillips curve and labour market is considered.
- Komárek (1997) who dealt with a proposal of possible segments of a future macroeconomic model of Czech economy as a small open transition economy. A question is discussed if it is possible to use a model for centrally planned economy or it is better to construct a new transformation model based on modifications of market economy models. The model structure is divided into three blocks – real, monetary and foreign. The segment proposal is based on Keynes' approach in log linear form and the issue of a Phillips curve, foreign and external debt, reactive position of the central bank and adoption processes are presented in a greater detail.

The second group of structural macroeconomic modelling of Czech economy with an evaluation includes following publications:

- Pelikán (2001) who created a Czech economy model for the prognoses and teaching. It is based on the IS-LM model describing the real and monetary market (an equation of consumption function, investment function, money demand equation, money supply, equation of import and export, money expenditures and personal incomes).
- Further publication Hušek, Pelikán (2003) compares estimations of the alternatively specified macroeconomic models for the Czech Republic.
- Monograph Tuleja (2004) attends to describe the theoretical and methodological fundaments for building of macroeconomic models, to estimate and verify this model using 11 behavioural equations for the periods of Q1_1995-Q4_2003. The macroeconomic model described both the material and financial flows in the Czech economy. The method of OLS was applied for estimation of the each behavioural equation. Paper of Hančlová (2007) continues and deals with formulating and estimating modified Tuleja's macroeconomic model of the Czech economy using simultaneous-equation relationships.

4 Macroeconomic data and their problems

Macroeconomic data of the Czech Republic for macroeconomic modelling are mostly non-consistent and they include a lot of quantitative and qualitative problems (see Rublíková, Ivaničová (2005), Komárek (1997), Hanousek, Tůma(1994)) - an economic system change, a whole statistics system change, permanent shock influence, time-series duration.

The Czech economy was a centrally planned economy for more than fifty years and the transformation process affected macroeconomic data statistics. The splitting of the Czech and Slovak Republics caused the troubles with historic macroeconomic data that have not been followed at federal level, e.g. M1 and M2 monetary aggregates. This situation can be solved by an introduction of dummy variables and revised macroeconomic statistics.

To achieve the objectives set by the Treaty on European Union, and more specifically Economic and Monetary Union, the new European System of National and Regional Accounts (ESA 95) is introduced from 1979. The new system will provide its users with a more comprehensive and more detailed methodology and in the precision and accuracy of the concepts, definitions, classifications and accounting rules which have to be applied in order to arrive at a consistent, reliable and comparable quantitative description of the economies of the Member States. The implementation of ESA95 for macroeconomic modelling of the Czech economy implies the problems of time-series methodology changes and approximation of non-followed quarterly macroeconomic data e.g. private consumption.

The influence of shocks in a transitional economy is monitored by the internal changes (institutional changes) and by foreign markets (e.g. oil price shocks, stock exchange trading). It is presumed to use time series of the Reuters EcoWin database including quarterly real data since 1993.

To sum up this part it is observed that time series for macroeconomic modelling of the Czech economy are constantly non-consistent but the situation is getting better for the Czech transitional economy modelling.

5 Conclusions

In the first part the paper deals with the historic development of structural macroeconomic modelling from the point of a model core proposal with regard to the methodology. The above mentioned approaches are separated in general and later, with focus on the Czech economy. Further on, the present issues of gaining macroeconomic data are presented and problem solution is suggested.

To sum up I would like to declare optimistically that nowadays after 15years of market economy introduction and after the EU accession we can specify the segments of the macroeconomic model of the Czech market economy and verify it for the Reuters EcoWin database quarterly data since 1993, resp.1995. This estimated model should be investigated further on towards short-term predictions and simulations.

Acknowledgement

The research work on this paper was supported by the Czech Science Foundation, grant number 402/08/1015 - Macroeconomic Models of the Czech Economy and Economies of the other EU Countries.

References

[1] Hančlová, J. (2007). Macroeconomic simultaneous-equation model of the Czech Republic. In: *AIESA 2007 - participation on building of society based on knowledge.* Bratislava May 17-18, 2007 [CD-ROM]. Bratislava: University of Economics, Faculty of Economic Informatics,pp. 1-6. ISBN 978-80-225-2334-9.

[2] Hanousek, J., Tůma, Z. (1994). Makroekonomický model české ekonomiky: návrh možného přístupu. *Finance a úvěr*, 44 (1), 18 – 29.

[3] Komárek, L. (1997). Návrh segmentů strukturálního makroekonomického modelu České republiky. *Politická ekonomie*, 4, 548 – 562.

[4] Pelikán, J. (2001). Komplexní model české ekonomiky, *Acta Oeconomica Pragensia.* 9, ss. 27-40.

[5] Hušek, R., Pelikán, J. (2003). *Aplikovaná ekonometrie. Teorie a praxe.* 1. vyd. Praha: Professional Publishing, 263 s. ISBN 80-86419-29-0.

[6] Rublíková, E., Ivaničová, Z. (2005). Econometric Approach of the Estimation of the Single Mundell-Fleming Equations. In: *Prace Naukowe Akademii Ekonomicznej we Wroclawiu: Inwesycje finansowe i ubezpeczenia – tendencje swiatowe a polski rynek.* Tom 2, č. 1088, (2005), s. 201-208. ISSN 0324-8445.

[7] Šmídková, K. (1995). Vývoj přístupů k makroekonometrickému modelování. *Politická ekonomie*, 1, 113 – 124.

[8] Tuleja, P. (2004). *Makroekonomický model České republiky.* Vědecká monografie. Studia Oeconomica. Karviná: OPF SU Karviná, 2004. 196 s. ISBN 80-7248-256-3

[9] Valadkhani, A. (2004). History of macroeconometric modelling- lessons from past experience. *Journal of Policy Modeling,* 26, 265-281.

Jana Hančlová, Department of Mathematical Methods in Economics, Faculty of Economics,
VŠB-Technical University of Ostrava, Sokolská tř. 33, 701 21 Ostrava, Czech Republic,
e-mail: jana.hanclova@vsb.cz

COALITION-PROOFNESS VERSUS EQUILIBRIUM IN INFINITE HORIZON GAMES

Milan Horniaček, Institute of Economics, Faculty of Social and Economic Sciences, Comenius University

Abstract: The concept of subgame perfect equilibrium (to which folk theorems for infinite horizon games apply) requires only immunity to unilateral deviations in all proper subgames. Strong perfect equilibrium takes into account also deviations by coalitions of players. Nevertheless, it fails to exist in many games of economic interest. The immunity to deviations by the grand coalition is often in conflict with the immunity to deviations by smaller coalitions. In the present paper we discuss some possibilities of resolution of this conflict.

1. INTRODUCTION

A subgame perfect equilibrium is the most widely used solution concept for infinite horizon discrete time non-cooperative games with discounting of future payoffs. (We restrict attention to infinite horizon discrete time games with discounting of future payoffs. Nevertheless, some ideas and conclusions in this paper apply also to infinite horizon discrete time games with other evaluation relations and to continuous time games.) Folk theorems for these games state that each strictly individually rational vector of discounted average payoffs is the equilibrium vector of discounted average payoffs in some subgame perfect equilibrium if the discount factor is close enough to one.

It is plausible that long term rivalry can lead to cooperation between (some or all) players. Therefore, it is desirable to develop and study solution concepts for infinite horizon games that take into account also deviations by (non-singleton) coalitions of players. Rubinstein's (1980) concept of strong perfect equilibrium requires immunity to deviations by all coalitions of players in each proper subgame. Unfortunately, it fails to exist in many infinite horizon discrete time games with discounting of future payoffs. In particular, in infinitely repeated games it fails to exist when the Pareto efficient frontier of the set of individually rational stage game payoff vectors has no sufficiently large flat portion (i.e., a sufficiently large portion that is a subset of a hyperplane of the vector space with the dimension equal to the number of players). Maskin and Tirole (1988), Bernheim and Ray (1989) and Farrell and Maskin (1989) consider only two-player games and pay attention to deviations by the grand coalition. (Bernheim et al. (1987)

formulate their concept called Perfectly Coalition-Proof Nash equilibrium only for games with a finite number of stages.)

Infinite horizon discrete time oligopoly games with discounting of future payoffs belong to the best known and most widely studied infinite horizon discrete time games of economic interest. In these games deviations of coalitions other than then grand one (which contains all players) do not cause a problem. Under usual assumptions (e.g. a Cournot oligopoly with a homogeneous product and the price falls to zero for some sum of outputs), each firm can (when it has or can build a sufficient production capacity) prevent any competitor from earning a single period profit exceeding its continuation average discounted profit in a subgame in which it is punished. Thus, each firm can punish a deviation by the coalition containing all other firms. Also, a deviation by any coalition that does not contain all firms can be punished by the remaining firms. At the same time, it is worth noting that – for the same reason – in each deviating coalition (with at least two members) any deviation by its strict subcoalition can be punished by the remaining coalition's members. Thus, (unlike in the case of finite horizon games studied by Bernheim et al. (1987)) we cannot disregard deviations by coalitions on the basis of their vulnerability to deviations by strict subcoalitions.

From all this it follows that deviations by the grand coalition represent the main problem in construction of equilibria immune to deviations by all coalitions in infinite horizon discrete time oligopoly games with discounting of future payoffs. In order to avoid these deviations continuation equilibrium payoff vectors during punishments of deviations by smaller coalitions, including unilateral deviations, would have to be (at least weakly) Pareto efficient. This is often impossible to achieve. Thus, there is a conflict between coalition-proofness (immunity to deviations by non-singleton coalitions and especially deviations by the grand coalition) and standard concept of (subgame perfect) equilibrium (i.e., immunity to unilateral deviations). In the present paper we discuss some possibilities of resolution of this conflict. In the next section we deal with solution concepts that impose restrictions on deviations by the grand coalition taken into account. In Section 3 we discuss possibilities based on restrictions on the type of strategies that the players use.

2. Restrictions on the grand coalition's deviations taken into account

Restrictions on the grand coalition's deviations reduce the set of strategy profiles to which the players are allowed to (collectively) switch in any (proper) subgame. The simplest requirement is the immunity to renegotiations from one continuation equilibrium to another continuation equilibrium of the same (subgame perfect) equilibrium in an infinitely countably repeated game with observable actions. This approach is justified by the fact that (in such game) all subgames are (proper and) identical. Thus, at the beginning of each period, irrespective of history, players should have the option of collectively renouncing their prescribed strategies, and adopting the strategies prescribed for any other subgame. This implies that an equilibrium should not have two continuation equilibria such that the continuation equilibrium payoff vector in one of them strictly Pareto dominates the continuation equilibrium payoff vector in the other one. Farrell's and Maskin's (1989) concept of weakly renegotiation-proof equilibrium (WRPE; it coincides with Bernheim's and Ray's (1989) concept of internally consistent equilibrium) expresses the latter requirement. Nevertheless, it is formulated only for two-player infinitely repeated games.

Of course, an infinite repetition of a stage game Nash equilibrium is a WRPE for any discount factor. Nevertheless, we are interested in WRPEs in which the equilibrium vector of average discounted payoffs is (strictly) Pareto efficient with respect to the set of individually rational stage game payoff vectors. Usually, in such equilibria any player can increase his single period stage game payoff by deviating from the stage game strategy profile prescribed for the first period of the repeated game.

Consider a continuation equilibrium in which a punishment of a previous deviation by player $i \in \{1,2\}$ takes place. Player i's payoff in this continuation equilibrium has to be lower than his payoff in the continuation equilibrium that would have been played if he had not deviated (otherwise, i would not be punished). Therefore, the principle of weak renegotiation-proofness requires that payoff of player $j \in \{1,2\} \setminus \{i\}$ in the former continuation equilibrium is no lower than in the latter.

There are approaches in the literature that extend the principle of weak renegotiation-proofness to games with more than two players (and combine it with other requirements that are not relevant for two-player games). Horniaček (1996) considered a strictly weakly renegotiation-proof equilibrium (SWRPE) in an infinitely countably repeated game with observable actions. It is a semi-strict quasi-strong perfect equilibrium (SQSPE) that has no two continuation equilibria such that payoff vector in one of them weakly Pareto dominates payoff vector in the other. (The attribute 'strictly' reflects the fact that, unlike in WRPE of Farrell and Maskin (1989), also weak Pareto domination between continuation equilibrium payoff vectors is ruled out.) A SQSPE is immune in each subgame to all deviations by all coalitions other than the grand one (including, of course, singleton coalitions) that would increase a continuation payoff of at least one member of the coalition without decreasing continuation payoff of any other member. For each coalition C other than the grand one, a SWRPE has to have a continuation equilibrium in which at least one member of C is punished for a previous deviation by C, his continuation equilibrium payoff is lower than his payoff in the continuation equilibrium that would have been played if C had not deviated, and at least one of the other players has a continuation equilibrium payoff higher than his payoff in the continuation equilibrium that would have been played if player C had not deviated.

Farrell (1993) defines quasi-symmetric WRPE as WRPE in which all innocent players (i.e., non-deviators) block renegotiation from a continuation equilibrium triggered by a deviation back to the initial continuation equilibrium. (That is, during a punishment of player i all other players have continuation equilibrium payoff no lower than their continuation equilibrium payoff when no player is punished.) This requirement implies that some payoff vectors that can be sustained (for the discount factor close enough to one) as an equilibrium vector of average discounted payoffs in a SWRPE cannot be sustained as an equilibrium vector of average discounted payoffs in a quasi-symmetric WRPE. Moreover, Farrell does not pay attention to deviations by non-singleton coalitions other than the grand one.

Example. Let stage game be the symmetric Cournot oligopoly producing a homogeneous product, with $n \geq 2$ firms, (normalized) inverse demand function $p:[0,\infty)\to[0,1]$ in the form $p(Q)=\max\{1-Q,0\}$ (where Q is the sum of outputs of the oligopolists) and cost function $c_i:[0,\infty)\to[0,\infty)$ in the form $c_i(y_i)=\gamma y_i$, $\gamma\in(0,1)$, (where y_i is the output of firm i) for each firm $i\in\{1,...,n\}$. This stage game is infinitely countably repeated with the same discount factor $\delta\in(0,1)$ for all firms.

The interior of the (weak and also strict) Pareto efficient frontier of the set of individually rational payoff vectors in the stage game, which we denote by $\wp$, is the set of all vectors with n positive components that sum up to $\left(\frac{1-\gamma}{2}\right)^2$ (which is the monopoly profit in this setting). Suppose that we want to sustain a vector $v^* \in \wp$ as the vector of average discounted payoffs in a SWRPE in the repeated game for values of discount factor close enough to one. The easiest way to satisfy the requirements of SWRPE is to punish each deviation by each coalition $C \neq \{1,...,n\}$ by producing (for sufficiently large but finite number of periods) vector of outputs y (that, of course, depends on C) with $y_i = 0$ for each firm $i \in \{1,...,n\} \setminus \{j\}$, where $j \notin C$, and $y_j > 0$. In order to punish deviations (by firms other than j or coalitions of them) during the punishment by restarting it, the best response against outputs of other firms in y has to give each firm $i \neq j$ single period profit lower than v_i^*. This requires

$$y_j > 1 - \gamma - 2\sqrt{v_i^*}\ . \tag{1}$$

At the same time, we require that output vector y gives firm j single period profit exceeding v_j^*. (If this requirement is satisfied then firm j's continuation equilibrium profit when it is a punisher exceeds its continuation equilibrium profit when it is not a punisher.) Thus, we want that $y_j \in \left(\frac{1-\gamma}{2} - \frac{1}{2}\sqrt{(1-\gamma)^2 - 4v_j^*}, \frac{1-\gamma}{2} + \frac{1}{2}\sqrt{(1-\gamma)^2 - 4v_j^*} \right)$. This leads to the requirement

$$-\frac{1-\gamma}{2} + \frac{1}{2}\sqrt{(1-\gamma)^2 - 4v_j^*} + 2\sqrt{v_i^*} > 0\,. \tag{2}$$

Condition (2) is satisfied for the symmetric distribution of the monopoly profit (i.e., $v_i^* = \frac{(1-\gamma)^2}{4n}$ for each $i \in \{1,...,n\}$) for any $n \geq 2$. Thus, for each $n \geq 2$ there is an interval of values of the discount factor close to one for which the symmetric distribution of the monopoly profit is sustainable as the equilibrium vector of average discounted profits in a SWRPE in the repeated game.

In order to sustain (for discount factor close enough to one) the symmetric distribution of monopoly profit in Farrell's quasi-symmetric WRPE, during the punishment of a single player deviation the $n-1$ punishers have to produce together output (using (1)) $Q > 1 - \gamma - 2\sqrt{\frac{(1-\gamma)^2}{4n}}$. The distribution of Q among them has to be such that each of them earns profit no lower than $\frac{(1-\gamma)^2}{4n}$. This requires that $Q \in \left[\frac{1-\gamma}{2} - \frac{1-\gamma}{2}\sqrt{\frac{1}{n}}, \frac{1-\gamma}{2} + \frac{1-\gamma}{2}\sqrt{\frac{1}{n}} \right]$.

This leads to the condition

$$\frac{1-\gamma}{2}+\frac{1-\gamma}{2}\sqrt{\frac{1}{n}}>1-\gamma-2\sqrt{\frac{(1-\gamma)^2}{4n}}\,. \tag{3}$$

Condition (3) is satisfied only for $n<9$. Thus, for $n\geq 9$ (irrespective of the value of the discount factor) the symmetric distribution of the monopoly profit is not sustainable as the equilibrium vector of average discounted profits in a quasi-symmetric WRPE in the repeated game.

Farrell and Maskin (1989) consider also a strong renegotiation-proof equilibrium. (This concept coincides with a strongly consistent equilibrium of Bernheim and Ray (1989).) It is a WRPE with the property that no one of its continuation equilibrium payoff vectors is strictly Pareto dominated by a continuation equilibrium payoff vector of another WRPE. So far there are no generalizations of this solution concept to games with more than two players.

3. RESTRICTIONS ON THE PLAYERS' STRATEGIES

So far one type of restriction has been imposed in the literature on the players' strategies in connection with immunity of an equilibrium to deviations by coalitions. Maskin and Tirole (1988) applied Markov strategies (and Markov perfect equilibrium) to an infinitely countably repeated symmetric Bertrand duopoly with discrete sets of possible prices, constant marginal costs and no fixed costs. In odd-numbered (even-numbered) periods t firm 1 (firm 2) chooses its price, which remains unchanged until period $t+2$. Markov strategies prescribe the same behavior in all subgames that are identical. - in Maskin's and Tirole's (1989) model in all subgames in the first period of which the same firm cannot change its price and its price is fixed at the same level.

Maskin's and Tirole's (1988) concept of a renegotiation-proof Markov perfect equilibrium requires for each subgame immunity to all deviations by the grand coalition (that increase a continuation payoff of both players) to another profile of Markov strategies. For discount factors close to one the unique renegotiation-proof Markov perfect equilibrium is the simple monopoly kinked demand curve equilibrium (SMKE). In this equilibrium each firm (i) cuts price to the monopoly price p^m when the market price (the other firm's price fixed for the current period) is above p^m, (ii) cuts its price immediately to $\underline{p}$ when the market price is between $\underline{p}$ and p^m; and (iii) raises its price to p^m when the market price is below $\underline{p}$. The fact that in each period only one firm can change its price and the choice of $\underline{\underline{p}}$ ensure that no firm can gain by abandoning the punishment of a previous deviation by the other firm. At the same time, the choice of $\underline{p}$ has the consequence that SMKE is not strictly Pareto dominated in terms of payoffs by another Markov perfect equilibrium (that is not its continuation equilibrium).

REFERENCES

Bernheim, Douglas, Peleg, Bezalel, and Whinston, Michael D.: Coalition-Proof Nash Equilibria I. Concepts. Journal of Economic Theory, 42 (1987), p. 1-12.
Bernheim, Douglas and Ray, Debray: Collective Dynamic Consistency in Repeated Games. Games and Economic Behavior, 1 (1989), p. 295-326.
Farrell, Joseph: Renegotiation in repeated oligopoly games. Mimeo, University of California at Berkeley, 1993.
Farrell, Joseph and Maskin, Eric: Renegotiation in Repeated Games. Games and Economic Behavior, 1 (1989), p. 327-360.
Horniaček, Milan: The Approximation of a Strong Perfect Equilibrium in a Discounted Supergame. Journal of Mathematical Economics, 25 (1996), p. 85-107.
Maskin, Eric and Tirole, Jean: A Theory of Dynamic Oligopoly II: Price Competition, Kinked Demand Curves, and Edgeworth cycles. Econometrica, 56 (1988), p. 571-599.
Rubinstein, Ariel: Strong Perfect Equilibrium in Supergames. International Journal of Game Theory, 9 (1980), p. 1-12.

Milan Horniaček
Institute of Economics
Faculty of Social and Economic Sciences
Comenius University in Bratislava
Odbojárov 10/A
P.O.Box 129
SK-82005 Bratislava
Slovakia

DYNAMIC DIFFERENCE EQUATION MODEL FOR GDP PROJECTION

Prof. Ing. Mgr. ek. Jaroslav Husár, CSc.

KOVE, FHI Bratislava

Introduction

We have developed a difference equation model for the GDP projections.[1] It is a macroeconomic model based on three principal macroeconomic variables – GDP, or gross domestic product (*Y*), *C*, or consumption expenditures and *I*, or investments expenditures. In this model we assume that the time period is cut up into periods, and the variables move in jumps one period to the next. The reason was that the economic data are presented in the form of quantities relevant to a period rather than to a point of time.

1. The basic structure of the interaction model

In this section we want to describe a model based on second-order difference equation. The principal macroeconomic variables in this model are GDP (*Y*), consumption expenditures (*C*) and investment expenditures (*I*). The developed model is an interaction model as it can be seen from the relationship between investment and consumption expenditures. The model consits of these equations:

$$Y(t) = C(t) + I(t) \tag{1}$$

$$C(t) = \alpha Y(t - 1) \tag{2}$$

$$I(t) = \beta[C(t) - C(t-1)] \tag{3}$$

$$Y(0) = Y_0\,,\ \alpha > 0\,,\quad Y(1) = Y_1 \qquad \beta > 0$$

To solve this model we are obliged to use two steps. First of all, from the model we can derive the appropriate difference equation. This may be done by substituting the second and third equation into the first. We will get:

$$Y(t) - \alpha(1 - \beta)Y(\mathrm{t} - 1) + \alpha\beta Y(t-2) = 0 \tag{4}$$

[1] The model was developed in VEGA research project at FHI, Economic University Bratislava.

Having this difference equation we can in the second step to solve it. As it can be seen it is homogenous second order difference equation. We want to get a path of the GDP so we can substitute

$$Y(t) = ax^t \tag{5}$$

And getting

$$x^t - \alpha(1 - \beta)x^{t-1} + \alpha\beta x^{t-2} = 0 \tag{6}$$

respectively after small manipulation

$$x^{t-2}\,[(x^2 - \alpha(1 - \beta)x + \alpha\beta] = 0 \tag{7}$$

Thus $Y(t) = ax^t$ is a solution if[2]

$$x^2 - \alpha(1 - \beta)x + \alpha\beta = 0 \tag{8}$$

i. e, if

$$x = \frac{\alpha(1+\beta) \pm \sqrt{\alpha^2 + (1+\beta^2) - 4\alpha\beta}}{2} \tag{9}$$

As we know, there are two possible values for x:

$$x_1 = \frac{\alpha(1+\beta) + \sqrt{\alpha^2 + (1+\beta^2) - 4\alpha\beta}}{2} \tag{9a}$$

and

$$x_2 = \frac{\alpha(1+\beta) - \sqrt{\alpha^2 + (1+\beta^2) - 4\alpha\beta}}{2} \tag{9b}$$

both of which will satisfy the characteristic equation. It means that we have two solutions[3]:

$$Y(t) = a_1 x_1^t \quad \text{a} \quad Y(t) = a_2 x_2^{\mathrm{t}} \tag{10}$$

[2] R. Shone: Economic Dynamics. CUP, Cambridge, 2003.
[3] G. Gandolfo: Economic Dynamics. Springer, London 1996.

When there are two such possible solutions, a more general solution would bethe sum of the two:

$$Y(t) = a_1 x_1^t + a_2 x_2^{t.} \tag{11}$$

where a_1 and a_2 are arbitrary constants.

This general solution will now be handled differently, depending upon the nature of the roots x_1 and x_2 of this quadratic equation. There are three possible cases:

- The roots are real and unequal,
- The roots are real and equal
- The roots are complex, i.e., they involves imaginary numbers.

In the first case we can proceed directly to step III. We need to introduce two initial conditions so as to specify the arbitrary constants a_1 and a_2:

$$Y_0 = Y(0) = a_1 x_1^0 + a_2 x_2^0 = a_1 + a_2 \tag{12}$$

$$Y_1 = Y(1) = a_1 x_1^1 + a_2 x_2^1 \tag{13}$$

Solving these two equations for a_1 and a_2, we get

$$a_1 = \frac{Y_1 - Y_0 x_2}{x_1 - x_2} \tag{14}$$

$$a_2 = \frac{Y_1 - Y_0 x_1}{x_2 - x_1} \tag{15}$$

So the general solution thus becomes:

$$Y(t) = \left(\frac{Y_1 - Y_0 x_2}{x_1 - x_2} \right) x_1^t + \left(\frac{Y_1 - Y_0 x_1}{x_2 - x_1} \right) x_2^t \tag{16}$$

where x_1 and x_2 are given above in terms of α and β. We are in a very good situation. The various possible time path of the GDP can bee explored.

The time path may be oscillating or not, and damped or explosives, depending on the value of x_1 and x_2. It may be noted that the larger numerical root will eventually predominate, though the smaller numerical root may predominate early, if its *a* value is large.

We know that there exists another possibility. In the second case, when roots of the auxiliary equation are *real and equal*, the form of the solution is a little different:

$$Y(t) = a_1x^t + a_2x^{t.} \quad (17)$$

where

$$x = \frac{\alpha}{2}(1+\beta) \quad (18)$$

Now we can proceed to step four.

Step IV. Introduce the initial conditions:

$$Y_0 = a_1 \quad (19)$$

a $Y_1 = a_1x + a_2x$ (20)

So the coefficients are:

$$a_1 = Y_0$$

$$a_2 - \frac{Y_1 - Y_0x}{x} \quad (21)$$

and the solution can be waritten like this:

$$Y(t) = Y_0x^t + \frac{Y_1 - Y_0x}{x}tx^t \quad (22)$$

But we know that this solution may be re-written as

$$Y(t) = \left[Y_0(1-t) + Y_1t\right]x^{t-1} \quad (23)$$

As we mentioned at the begininig, there exist the third possibility. In the case the roots are complex. That is, they involve the imaginary number $\sqrt{-1}$. This is generaly designated as *i*. It means that $i = \sqrt{-1}$. The general solution

$$Y(t) = a_1 x_1^t + a_2 x_2^{t.} \tag{24}$$

May now be re-written. First, let the complex numbsers bse

$$x_1 = c + di \quad \text{where} \quad c = ½(1 + \beta) \tag{25}$$

$$x_2 = c - di \quad \text{where} \quad d = ½\left[4\alpha\beta - \alpha^2(1+\beta)^2\right]^{1/2} \tag{26}$$

This can be, as it is known, expressed in trigonometric form.[4] After some mathematicla manipulation we finaly com to thi solution:

$$Y(t) = a_1 x_1^t + a_2 x_2^{t.} =$$

$$= (c^2 + d^2)^{t/2}[(a_1 + a_2)\, cos\, tB + (a_1 - a_2)\, i\, sin\, tB] \tag{27}$$

It may be noted that com=plex numbers always appear in pairs that are called conjugate, as $c + di$ and $c – di$, so that this mmethod of procedure is general. We do not develop more this idea. We see that the explanation of the model requires to solve lot of problems, so the author has to discuss all the principal problems. Modelling deals not *how an economy works in reality* but it deals with what the model can say as far as the economic system functioning is concerned. Our models showed this very clearly.

3. Experimental results

Our developed model was not only used to estimate the parameter α and β but also to show his value as of the tool for making projections. Scientific thoroughness requires to say that this is only the first illumination of the problem. Why? We discovered that the model is very sensitive to longer time period. It is an explosive model. In the Table 1 we can se the estimated values of tha parameters and variables that are rquired by the model.

Table 1: Values of model parameters an variables

---

[4] Gandolfo, D.: Economic Dynamics, Springer, London, 1996.

Multiplier-accelerator model			
		values	
	α	0,63	
	β	5,7	
	α (1+β)	4,221	
	$\alpha^2(1+\beta^2)$	17,81684	
	4$\alpha\beta$=	14,364	
	sqr t=	1,858182	
	x1	3,039591	
	x2	1,181409	
	Y(0)	1471	
	Y(1)	1636	
	a_1	-54,813	
	a_2	1525,813	

From the Table 2 it is seen that as a zero yea was considered the year 2005. The next year was the year 2006. After substitution of values from Table 1 into the model (16) we got these GDP projected values:

Table 2: GDP values

Results	
t	**Y**
0	1471,00
1	1636,00
2	1623,20

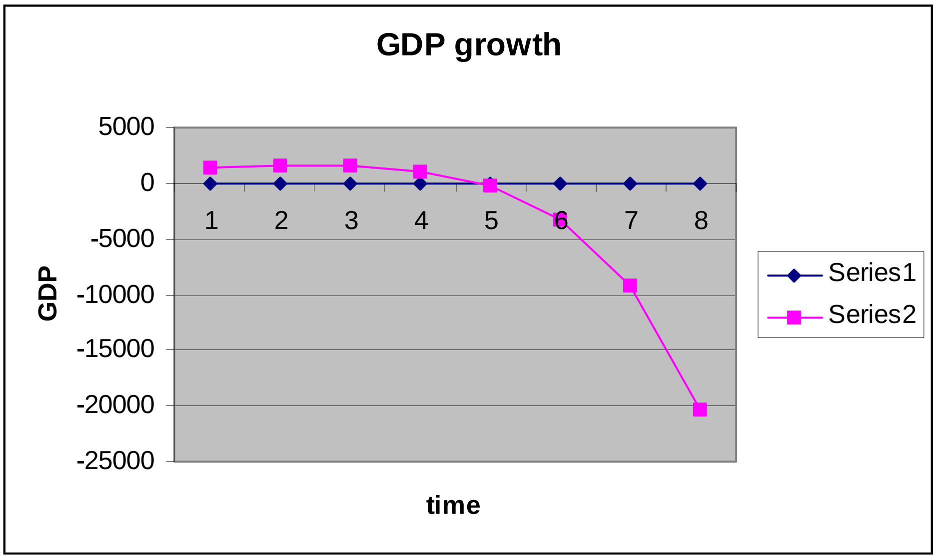
We see that the values we got from the model coincide very good with the original values of the Slovak economy. The projected value of the GDP for the year 2007 given in last table is 1 623,2 mld. Sk. We see that the path of GDP has started the oscillation. A more detailed analysis should be done if we want to get a deep understanding of path development.

Using the equation (23) we got these results:

Table 3: GDP growth, roots are real and equal

alfa=	0,75	
beta=	3	
	Y(0)=	1471
	Y(1)=	1636
	t	Y(t)
	0	1471,0
	1	1636,0
	2	1598,3
	3	1113,8
	4	-254,8
	5	-3270,4
	6	-9237,8
	7	-20355,0

As we mentioned earlier, the time path may be explosive or damped. In this case on the graph we see how the GDP behaves based on the data from table.



So, modelling the GDP we are obliged to get deep understanding of the model. To understand the *properties of the model* is the first prerequisite for macroeconomic analysis and using it for prediction.

Key words: dynamic model, gross domestic product, consumption expenditures, investment expenditures, solution of the model.

LITERATURE

1. FRANK, K.- MORVAY K.: Hodnotenie fiškálnej a monetárnej politiky v roku 2005 na Slovensku. Ekonomický časopis 54/2006, č. 5
2. GABRIELOVÁ, H.: Súčasné trendy vo vývoji produkcie na Slovensku. Ekonomický časopis 54/2006, č. 5
3. HICKS, J. R.: A Contribution to the Theory of Trade Cycle. Lodon:Oxford University Press, 1950.
4. HUSÁR, J.: Makroekonomická analýza. Bratislava: Vydavateľstvo Ekonóm, 2007.
5. OKÁLI, I.: Hlavné trendy vývoja hospodárstva Slovenska v roku 2005 a odhad jeho hospodárskeho rastu v roku 2006. Ekonomický časopis 54/2006, č. 5
6. SAMUELSON P. A.: Interactions between the Multiplier Analysis and the Principle of Acceleration. Review of Economic Statistics, May 1939.
7. SHONE, R.: Economic Dynamics. CUP, Cambridge, 2003.

Monary union without
fiscal coordination may
discipline policymekers
(Beetsma,Bovenberg 1999)

MONETARY MODEL AND FISCAL POLICY - METHODOLOGICAL APPROACH

Ivaničová Zlatica, Rublíková Eva

Summary:
In a currency union a policy mix problem can arise since monetary policy is centralized and fiscal policy is decentralized. As the Slovak economy enters the monetary union, an important question arises. How to manage the interaction between a centralized monetary policy focused on Euro area price stabilization and a decentralized fiscal policy aimed at the fiscal instruments to stabilize Slovak national output.

Keywords:
Monetary policy, fiscal policy, aggregate supply, aggregate demand (output), aggregate inflation, demand shocks, supply shocks, aggregate shocks, idiosyncratic shocks,

Introduction:
Coordination problems between monetary and fiscal policy arise through two types of interdependencies. First, a change in fiscal policy in a single country may, via its effects on inflation, cause a monetary policy response that affects all member countries of the currency union. This monetary policy externality may tend to imply that non-cooperative fiscal policy is too expansionary, since single countries of the currency union do not take the monetary policy fully into account. Second, there is an interdependence via trade links, where a demand spillover effect, caused by a since a domestic fiscal expansion, could benefit trading partners through an increase in import demand. A cost spillover effect also occurs since higher domestic activity leads to higher domestic wage/price increases.

This paper uses recent literature in macroeconomic rules (De Grauwe, 2003; Dixit and Lambertini, 2000; Svensson, 1997) to analyze the interaction between monetary policy and trade externalities for various shocks (demand or supply) as presented in Anderson (2005).

1. Structure of the model

The model structure is built on an aggregate supply curve (Philips curve relation) and the aggregate demand relation (output). It is important to specify the interactions between member countries of the monetary union in respect to inflation and output. The currency union is considered to be a closed area to allow a focus on the direct interdependencies between member countries of the currency area. Monetary policy is determined an independent central bank aiming at targeting region-wide inflation. Fiscal authorities are fully informed about monetary policy objectives, but decide on individual fiscal policy before the monetary common policy is fixed.

The currency area has n member counties indexed by i, which are all symmetric in relation to exogenous shocks.

1.1. Aggregate supply

Nominal wage increase are assumed to be determined by an inflation augmented Philips curve in which the wage increase in time $t+1$ in country i $(\Delta w_{i,t+1})$ depends on current consumer prices $(\pi_{i,t}^C)$, of output (demand) $(y_{i,t})$, fiscal policy captured by fiscalindicator (g_{it}) and a variable defined innovation (u_{it})

$$\Delta w_{i,t+1} = \pi_{i,t}^C + \alpha_y y_{i,t} + \alpha_g g_{i,t} + \alpha_u u_{i,t}; \qquad \alpha_y > 0 \tag{1}$$

The wage increase depends on the current consumer price inflation and increase of demand (output). Fiscal policy may have separate effects on wages (price inflation). Public expansion financed by value added and excise taxes (consumption taxes) add to inflationary pressures in the economy $(\alpha_g > 0)$, whereas income tax increases may lead to wage moderation $(\alpha_g < 0)$. Additionally, fiscal policy has an indirect effect on inflation via its effect on aggregate activity. The interpretation of α_u is stated in section 1.3 below.

Consequentlly increases in domestic producer prices are driven by domestic wage increases $\left(\pi_{i,t} = \Delta w_{it}\right)$ and the consumer price index in monetary union is defined over prices of domestic and foreign produced commodities i.e. $\left(\pi_{i,t}^{C} = \pi_{t}\right)$. The domestic producer price inflation can be written

$$\pi_{i,t+1} = \pi_t + \alpha_y y_{i,t} + \alpha_g g_{i,t} + \alpha_u u_{i,t} \tag{2}$$

1.2 Aggregate demand (output)

Aggregate demand (output) in country i is determined by the interest rate in the monetary union (fixed monetary policy), inflation rate (difference between inflation rate in country i and monetary union, increase of the demand (output) in monetary union, fiscal policy realized by fiscal indicator in country i, and variable defined effect representing a contemporaneous innovation or on shock in country i, such that

$$y_{i,t} = -\beta_r\left(r_t^N - E_t\pi_{t+1}\right) - \beta_\pi\left(\pi_{i,t} - \pi_t\right) + \beta_y y_t + \beta_g g_{i.t} + \beta_u u_{i,t}\,, \tag{3}$$

where E_t denotes the expectations operator in time t. In this model, demand is negativly affected by increases in expected real interest rates. The second term of equation (3) represents the role of competitiveness in domestic and foreign producer prices, or the. terms of trade effect. The third term captures how aggregate income in the monetary union affects demand for the products of country i. The effect of fiscal policy on demand is defined by fourth term. The last term represents the contemporaneous innovation on shock. It is assumed that only unanticipated (unexpected) shocks to the inflationary process have a real effect.

1.3 Shocks

Shock variables are expected to have a value of zero ($E_t u_{i,t+j}$ for all *i,j > 0)*, and variance of σ_u^2. Shocks are uncorrelated over time and aggregate value of shocks across the monetary union is $u_t \equiv \frac{1}{n}\sum_i u_{i,t}$.

In alternative interpretation of the shocks is possible.:

1. demand shock – if a shock at the same time increase demand $(\beta_u > 0)$ and the inflation pressure $(\alpha_u > 0)$ in the economy and vice versa
2. supply shock – if a shock at the same time increase demand $(\beta_u > 0)$ and reduces inflation $(\alpha_u < 0)$ in the economy and vice versa.

Both shocks can be aggregate or idiosyncratic shocks.

An aggregate shock is understood to be a shock that is common for all countries, i.e. $u_{i,t} = u_{jt}$ for all i, j in which case $u_{i,t} = u_{j,t}$ for all i. An idiosyncratic shock is one that does not have aggregate affects, i.e. $u_t = 0$, implying that a positive effect in some countries must be accompanied by a negative effect for others. It is a pure redistributive shock with no aggregate consequences.

1.4. Aggregate inflation and aggregate demand (output)

Inflation aggregate over countries can be written (from the formulation (1))

$$\pi_{t+1} = \pi_t + \alpha_y y_t + \alpha_g g_t + \alpha_u u_t, \tag{5}$$

that means, there is no long-run trade-off between activity and inflation. On the basis of aggregation in (3) $\left(\pi_t = \frac{1}{n}\sum_i \pi_{i,t};\quad y_t = \frac{1}{n}\sum_i y_{i,t};\quad g_t = \frac{1}{n}\sum_i g_{i,t};\quad u_t = \frac{1}{n}\sum_i u_{i,t}\right)$ is possible to defined the increase of aggregate demand (output) as

$$y_t = \frac{1}{1-\beta_y}\left[-\beta_r\left(r_t^N - E_t\pi_{t+1}\right) + \beta_g g_t + \beta_u u_t\right]. \tag{6}$$

Following stated considerations is possible to observe that

$$y_{i,t} - y_t = -\beta_\pi\left(\pi_{i,t} - \pi_t\right) + \beta_g(g_{i.t} - g_t) + \beta_u(u_{i,t} - u_t) \tag{7}$$

That is, demand (output) in a member country differs from the mean demand in the monetary union if it either

1. has a different producer price inflation (wage increase),
2. pursues a different fiscal policy,

3. is affected by idiosyncratic shocks.

2. Monetary model – monetary policy

The monetary authority aims at strict inflation targeting in the sense that expected inflation is kept at its target value $(E_t\pi_{t+1} = 0)$ and at controlling the nominal interest rate (r_t^N), hence real rate of return is effectively determined $(r_t \equiv r_t^N - E_t\pi_{t+1})$. Equation (5) therefore gives:

$$y_t = \frac{1}{\alpha_y}\left[E_t\pi_{i+1} - \pi_t - \alpha_g g_t - \alpha_u u_t\right]. \tag{8}$$

A combination of the formulas (8) and (6) and the condition of inflation targeting gives the following relation

$$E_t\pi_{t+1} = \pi_t + \frac{\alpha_y}{1-\beta_y}\left[-\beta_r\left(r_t^N - E_t\pi_{t+1}\right) + \beta_g g_t + \beta_u u_t\right] + \alpha_g g_t + \alpha_u u_t = 0. \tag{9}$$

Mathematical modification of this relation allows a definition of the real interest rate

$$r_t^R = r_t^N - E_t\pi_{t+1} = \frac{1-\beta_y}{\alpha_y\beta_r}\left[\pi_t + \left(\frac{\alpha_y}{1-\beta_y}\beta_g + \alpha_g\right)g_t + \left(\frac{\alpha_y}{1-\beta_y}\beta_u + \alpha_u\right)u_t\right] \tag{10}$$

where the relation $(r_t^N = E_t\pi_{t+1})$ expresses the monetary policy instrument. Equation (10) is the monetary reaction function. This function results in a tightening of monetary policy in response to an increase in inflation or a more expansionary fiscal policy. Results from the equation (10) mean that the higher value of the fiscal policy indicator (g_t) leads to a higher value of the monetary policy instrument.

Reworking the monetary restriction in equation (5), it is possible to define the following relationship

$$y_t = -\frac{1}{\alpha_y}\left(\alpha_g g_t + \alpha_u u_t\right).$$

These condition highlight that, with strict inflation targeting, the problem of stabilizing demand becomes a question of choosing the aggregate fiscal stance such that it corrects for the direct inflation consequences of the aggregate shock.

References:

1. Andersen T., M.: (2005) Fiscal Stabilization Policy in a Monetary Union with Inflation Targeting, In: Journal of Macroeconomics 27, pp 1-29
2. Beetsman, R.,M.,W.,J., Bovenberg, A.,L.: (1998) Monetary union without fiscal coordination may discipline policymakers, In: Journal of International Economics 45, pp. 239-258
3. Chocholatá, M.: (2004) Analysis of SKK/CZK Exchange Rate. In: Ekonomické rozhľady, ročník XXXIII, číslo 3, EU Bratislava
4. De Grauwe, P.: (2003) Economics of Monetary Union, fifth ed. Oxford University Press
5. Dixit, A., Lambertini, L.: (2000) Fiscal discretion destroys monetary commitment; Monetary-ficsal policy interactions and commitment versus discretion in a monetary union, Princeton University and UCLA
6. Husár, J.(2003) Aplikovaná makroekonómia, Sprint vfra, Bratislava
7. Lukáčik M., Lukáčiková A., Szomolánzi, K.: (2007) The long-run determinants of money dmend in Slovakia, Zborník z 13. medzinárodnej vedeckej konferencie Quantitative Methods in Economics, Multiple Criteria Decision Making, EU Bratislava
8. Surmanová K.: (2006), Prístup k analýze Phillipsovej krivky pomocou panelových dát, In: Ekonomika a informatica, 2/2006, FHI EU, Bratislava

The paper was wrought in the framework of the research project VEGA 1/4652/07 “Analysis of the current problems of the Slovak economy before the entrance into the European Monetary Union -econometric approach”

ADVANCE IN OPTIMISATION SOFTWARE FOR LP AND MIP PROBLEMS[1]

Josef Jablonský, Department of Econometrics, University of Economics Prague

1 INTRODUCTION

The aim of the paper is to discuss computational performance of current optimization packages for solving large scale LP optimization problems and mixed integer LP problems (MIP). I our paper [1] in 2002 we have tested several solvers on examples from the MIPLIB 3.0 library which contains several tens of test MIP problems submitted by researchers or practitioners over the world. In [1] we have compared several available solvers each other – LINGO 7.0, XPRESS 13.0, CPLEX 7.5 and XA 10.0. The primary aim of this paper is to test different releases of CPLEX solver which is one of the most powerful optimization solvers at all. The latest version of CPLEX solver is 11.0 issued in fall 2007. The capability and performance of this version will be compared with the previous two versions 9.0 and 10.1 and partially with the results of CPLEX 7.5 presented in [1]. The results given by CPLEX solvers will be compared to the latest release of XPRESS-MP solver which is the most important competitor of CPLEX solver.

CPLEX is the product of ILOG company that belongs to the worldwide leaders in optimisation software. CPLEX is the optimisation solver without any user-friendly environment. It is very often used as a solver in modelling and optimisation systems as GAMS, MPL for Windows, AIMMS, AMPL, etc. The users work with this system by means of DOS commands window and enter all the commands and the parameters of the solver directly from the prompt row. There are available five solvers for different class problems – primal and dual simplex method, network solver, barrier solver and MIP solver. The last mentioned one was used in our experiments. The more information about the CPLEX solver and other products of ILOG can be found on the web page www.ilog.com.

XPRESS-MP system is product of Dash Optimization, Ltd. and its latest release 18.0 was introduced in fall 2007. XPRESS-MP is a set of powerful optimisation tools that containing solvers for linear, quadratic and MIP problems. More information about XPRESS-MP can be found on www pages www.dashoptimization.com.

[1] The paper is supported by the Grant Agency of Czech Republic – grant no. 402/06/0150

2 BENCHMARKING OF MIP PROBLEMS - MIPLIB 3.0 LIBRARY

MIPLIB 3.0 is an electronic library of real-world pure and mixed integer programs. This library was created in 1996 and it has become a standard test set used to compare the performance of MIP solvers. The library is available on several mirror servers – one of them is http://miplib.zib.de. Information for each problem include size of the problem, number of integer and/or binary variables, optimum value of objective function for integer and appropriate continuous problem (if they are known), information describing the structure of the constraint matrix, etc. The library is continuously modified – the "older" problems that are currently easily solvable are removed and they are replaced by "newer" harder problems. MIPLIB library contains currently 60 test problem that are split into three classes:

Table 1 – List of easily solvable MIPLIB problems

Problem name	Opt. sol.	Rows	Cols	GIN	BIN
10teams	924.00	230	2025	0	1800
aflow30a	1158.00	479	842	0	421
air04	56137.00	823	8904	0	8924
air05	26374.00	426	7195	0	7195
cap6000	-2451377.00	2176	6000	0	6000
Disktom	-5000.00	399	10000	0	10000
fast0507	174.00	507	63009	0	63009
Fiber	405935.00	363	1298	0	1254
fixnet6	3983.00	478	878	0	378
gesa2	25779856.37	1392	1224	168	240
gesa2-o	25779856.37	1248	1224	336	384
harp2	-73899798.00	112	2993	0	2993
manna81	-13164.00	6480	3321	3321	0
mas74	11801.19	13	151	0	150
mas76	40005.05	12	151	0	150
misc07	2810.00	212	260	0	*259*
mod011	-54558535.01	4480	10958	0	96
modglob	20740508.09	291	422	0	98
mzzv11	-21718.00	9499	10240	251	9989
mzzv42z	-20540.00	10460	11717	235	11482
noswot	-41.00	182	128	25	75
nw04	16862.00	36	87482	0	87482
opt1217	-16.00	64	769	0	768
p2756	3124.00	755	2756	0	2756
pk1	11.00	45	86	0	55
pp08aCUTS	7350.00	246	240	0	64

pp08a	7350.00	136	240	0	64
qiu	-132.87	1192	840	0	48
rout	1077.56	291	556	15	300
set1ch	54537.75	492	712	0	240
vpm2	13.75	234	378	0	168

- the problems that can be solved relatively easily (within one hour) by commercial solvers like CPLEX or XPRESS-MP (31 problems),
- the problems that have been solved and their integer optimal solution is known (21 problems),
- the class of problems with unknown optimal solution (8 problems).

Table 2 – Time of solving of MIPLIB problems (sec)

Problem name	CPLEX 7.5	CPLEX 9.0	CPLEX 10.1	CPLEX 11.0	XPRESS
10teams	121.00	30.80	24.69	7.71	5.10
aflow30a	NEW	64.80	33.91	19.76	41.20
air04	535.00	29.03	33.64	17.32	43.10
air05	344.00	26.65	28.51	20.23	34.00
cap6000	40.00	14.82	1.01	1.40	8.70
disktom	NEW	472.77	93.76	11.83	3.40
fast0507	xxx	4675.76	3136.84	4319.22	1727.70
fiber	3.00	0.38	0.51	0.37	0.60
fixnet6	2.00	0.43	1.11	1.54	1.10
gesa2	5.00	1.19	0.72	0.81	0.50
gesa2-o	14.00	1.95	5.53	2.54	0.60
harp2	xxx	xxx	xxx	259.48	239.50
manna81	NEW	xxx	5477.01	0.34	0.70
mas74	NEW	1467.68	1088.17	558.84	4398.20
mas76	NEW	65.24	123.18	94.74	227.70
misc07	657.00	78.34	12.94	45.68	64.20
mod011	49532.00	124.55	128.21	91.03	68.50
modglob	2.00	0.31	0.22	0.34	0.40
mzzv11	NEW	949.71	640.17	400.95	382.20
mzzv42z	NEW	127.21	130.12	124.73	36.30
noswot	xxx	2736.53	2790.12	9826.47	2236.70
nw04	60.00	59.09	56.41	67.35	4.60
opt1217	NEW	xx	1119.19	0.16	0.20
p2756	4.00	0.56	0.54	0.70	0.80
pk1	441.00	89.39	83.29	198.57	218.10
pp08aCUTS	18.00	2.43	3.04	3.10	1.20
pp08a	6.00	1.41	0.95	2.12	0.90

qiu	3653.00	214.71	58.36	60.60	173.70
rout	xxx	1755.40	32.80	46.58	120.70
set1ch	4.00	0.75	0.68	0.61	0.40
vpm2	21.00	1.21	1.29	0.88	0.80

In our experiments we will solve just the first problem class. The reason is that for the last class it is hard to get even any integer feasible solution. For the problems of the second class integer feasible solutions can be found easily but the solvers usually do not stop due to increasing size of tree sizes. The problems of the first group are listed with their basic characteristics in Table 1. The table contains the optimal integer solution, the total number of rows/columns and the number of integer (GIN) and binary (BIN) variables for each problem.

NETLIB library is the set of LP test problems available on http://www-fp.mcs.anl.gov/OTC/ /Guide/TestProblems/LPtest/index.html. The library is split into three groups – feasible, infeasible and large problems. Most of the problems (but not all of them) are available in MPS format and they are LP problems without any integer variables.

Table 2 contains some results of our experiments – computational times of solving the first class of MIPLIB problems in seconds. The results for CPLEX 7.5 are taken from our study in 2002 and that is why not all the problems included in the MIPLIB library in 2002 are not included in the current library and vice versa. The symbol "NEW" in the first column of Table 2 means that the appropriate problem is the new one in the current MIPLIB and was not included in the library in 2002. The symbol "xxx" means that the appropriate problem was unsolvable by CPLEX 7.5 in 2002 and that is why the time is not available.

All the problems from the current MIPLIB were solved successfully by latest versions of CPLEX and XPRESS solvers as it is stated in Table 2. In the previous versions of CPLEX at least one problem was not solved successfully. The fastest times for each problem are shaded in Table 2. XPRESS is the fastest solver in the most cases and it has the lowest sum of solving times (little more than 10000 sec.). It is surprising that the latest version of CPLEX is slower that the previous versions in several cases. The total solving time is more than 16000 sec. but this sum contains almost 10000 sec. for *noswot* problem. Without taking into account this problem the CPLEX 11.0 has the lowest total tine at all. The progress in solvers is clear by comparing the current results to results given by CPLEX 7.5 in 2002 – all the problems are solved and the solving time is significantly lower in the current versions in almost all testing cases.

3 SOLVING LARGE SCALE DEA MODELS

Solving LP problems without any integer constraints is quite easy and the problems with several thousands or tens of thousands rows and/or variables can be solved usually within several seconds or minutes. The problem *stp3d* from MIPLIF library contains 159488 constraints and 204880 variables (all of them are binary). The LP relaxation of this problem was solved with almost one million simplex iterations in less than three hours by the latest release of XPRESS primal simplex solver and in approx. 30 minutes by CPLEX 11.0 primal simplex solver. The problems of lower sizes (less than ten thousands of constraints and variables) are solved in several seconds. The NETLIB library mentioned in the previous section contains mostly much lower problems and that is why it is without any sense to test them. By high quality solver as CPLEX and XPRESS all of NETLIB problems are solved within seconds or even fractions of seconds.

Some of the operational research models suppose solving of high number of LP problems of a smaller or average size. Typical examples of such models are data envelopment analysis models – they are used for evaluation of relative efficiency and performance of the set of homogenous decision making units. The standard DEA radial model with input orientation (often called CCR-I model) can be formulated as follows:

minimise $$z = \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{i=1}^{r} s_i^+\right),$$

subject to $$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq}, \qquad i = 1, 2, ..., m, \qquad (1)$$

$$\sum_{j=1}^{n} \lambda_j y_{ij} - s_i^+ = y_{iq}, \qquad i = 1, 2, ..., r,$$

$$\lambda_j \geq 0,\ s_i^+ \geq 0,\ s_i^- \geq 0.$$

where m/r is the number of input/output characteristics of n decision making units, $X = (x_{ij}, i=1,2,\ldots,m, j=1,2,\ldots,n)$ is the matrix of inputs and $Y = (y_{ij}, i=1,2,\ldots,r, j=1,2,\ldots,n)$ is the matrix of outputs, $\lambda = (\lambda_1, \lambda_2,\ldots,\lambda_n)$, $\lambda \geq 0$, is vector of weights assigned to evaluated units, $\mathbf{s}^+$ a $\mathbf{s}^-$ are vectors of slack variables in input and output space. In order to evaluate the efficiency of one unit of the set, i.e. unit q in our formulation, it is necessary to solve the above formulated LP problem of the size $(m+r)$ rows and $(n+m+r+1)$ variables. That is why the evaluation of efficiency of all DMUs of the decision set supposes solving of n linear programming optimisation problems (1) or its modifications. Another possibility is 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:

$$\begin{aligned}
&\text{minimise} && \sum_{q=1}^{n}\left(\theta_q - \varepsilon\left(\sum_{i=1}^{r} s_{iq}^{+} + \sum_{i=1}^{m} s_{iq}^{-}\right)\right) && \\
&\text{subject to} && \sum_{j=1}^{n} y_{ij}\lambda_{jq} - s_{iq}^{+} = y_{iq}, \qquad i = 1,2,...,r,\ q = 1,2,...,n, && (2)\\
& && \sum_{j=1}^{n} x_{ij}\lambda_{jq} - s_{iq}^{-} = \theta_q x_{iq}, \qquad i = 1,2,...,m,\ q = 1,2,...,n, && \\
& && \lambda_{jq} \ge 0, s_{iq}^{+} \ge 0, s_{iq}^{-} \ge 0, \theta_q \ge 0. &&
\end{aligned}$$

Optimisation problem (2) contains $n(n+m+r+1)$ variables and $n(m+r)$ constraints. Let us consider the problem with 2000 DMUs, 3 inputs and 2 outputs. We have written a code in Mosel language (Mosel is a modelling language used within XPRESS-MP modelling environment) that prepares and solves the optimisation problem with 5 constraints and 2006 variables 2000 times. The total time of the optimisation run was 33.1 sec. on the PC with Intel 2 GHz processor, 2 GB RAM. The same problem formulated as the model (2) has 10000 rows and 4120000 variables. Of course this problem is of a special structure and the total number of non-zero elements is low. This problem is solved on he same system and PC within 56.1 sec. (a significant part of this time is composing of the problem). The DEA problem with 1000 DMUs and 10 inputs and 10 outputs was solved within 12.3 sec. by repeating the model (1) and within 19.8 sec. by solving the model (2) – 1021000 variables and 20000 constraints. The Mosel code for DEA model (1) is presented below:

```
model DEA
   uses "mmodbc", "mmxprs"
parameters
   eps = 0.0000001
end-parameters
declarations
   DMU: set of string
   INP = 1..3
   OUT = 1..2
   X: array(DMU, INP) of real
   Y: array(DMU, OUT) of real
   EffScore: array(DMU) of real
   CSTR: string
```

```
end-declarations
   CSTR := 'mmodbc.excel:firmy.xls'
initializations from CSTR
   DMU as 'noindex;dmu'
   X as 'noindex;vstupy'
   Y as 'noindex;vystupy'
end-initializations
declarations
   lambda: array(DMU) of mpvar
   splus: array(INP) of mpvar
   smin: array(OUT) of mpvar
   theta: mpvar
end-declarations
forall(q in DMU) do
  eff := theta - eps * sum(j in INP) splus(j) - eps * sum(k
in OUT) smin(k)
  forall(j in INP) input(j):=(sum(i in DMU) X(i,j)*lambda(i))
+ splus(j) = X(q,j)*theta
  forall(k    in    OUT)    output(k):=(sum(i    in    DMU)
Y(i,k)*lambda(i)) - smin(k) = Y(q,k)
  minimize(eff)
  EffScore(q) := getobjval
  forall(j in INP) sethidden(input(j), true)
  forall(k in OUT) sethidden(output(k), true)
end-do
end-model
```

The experiments show that it is no problem to analyse efficiency by DEA models for very big number of units and/or inputs and outputs. Both the models (1) and (2) can be used but computing time seems to be lower when one solves *n* repeated smaller optimisation problems than one big problem (2). Both the results are in seconds and the models are solvable quite easily. The problems of the same or lower size were unsolvable still several years ago only.

4 CONCLUSIONS

Progress in optimization software was very significant in the last five years as it is shown by the presented computational experiments. Many problems originally unsolvable can be solved and its optimal solution is found. Optimal solution of harder problems from MIPLIB library (second class) can often be found or the current solution is very close to the optimal one. Problem is that the branching trees in this case increase very fast and the virtual memory allocated to the trees is not sufficient. Further progress will lead to higher performance of software and hardware and the problems currently unsolvable will be solved more or less easily in the time horizon of next five years.

REFERENCES

[1]Jablonský, J., Bartoň, J.: Benchmarks for mixed intger optimisation packages. In: Mathematical Methods in Economics, Ostrava 2002, pp. 145-150.
[2]Jablonský, J.: Models for efficiency evaluation of decision making units. Politická ekonomie, 2004 (2), pp. 206–220.
[3]MIPLIB3 library: http://miplib.zid.de
[4]Mittelman,H.D., Spellucci,P.: Benchmarks for optimisation software. http://plato.asu.edu/bench.htm
[5]XPRESS-MP essentials – an introduction to modelling and optimisation. Dash Associates, 2007.
[6]www.ilog.com
[7]www.dashoptimization.com

Prof. Ing. Josef Jablonský, CSc., Department of Econometrics, Faculty of Informatics and Statistics, University of Economics Prague, W. Churchill Sq. 4, 130 67 Praha 3
e-mail: jablon@vse.cz, http://nb.vse.cz/~jablon

SPECIFIC PROPERTIES OF LOCATION PROBLEMS

Marta Janáčková, Žilinská univerzita v Žiline
Alžbeta Szendreyová, Žilinská univerzita v Žiline

Abstract: This paper deals with the importance of prohibitive constants and with their impact on the solution of the location problem with maximal service distance.

INTRODUCTION

The design and the processing of the distribution systems solve various problems in the prax. One of the more frequent problems is the location problem, in which central warehouses, industrial complexes and other facilities can be located with the minimal costs of the complex realization of the distribution system. To design the uncapacitated location problem, we dispose of a method, which is able to processes real road or railway networks costs of Slovak republic. The road network consists of just about 3000 customers and hundreds of possible centers. We solve this problem in a admissible time.

To locate the rescue system centers it is necessary to fulfil an additional constraint. It is necessary to reach to the customers from the center in a short time. The same problem appears in the location of the fire brigade station, chemical or military facilities etc. This problem can be solved as uncapacitated location problem with maximal service distance.

Some properties of uncapacitated location problem

In the basic uncapacitated location problem *(1)-(5)*, we assume that a customer can be serviced by an arbitrary center.

Minimize $$\sum_{i\in I} f_i\, y_i + \sum_{i\in I}\sum_{j\in J} e_0\, d_{ij}\, b_j\, z_{ij} \qquad (1)$$

Subject to $$\sum_{i\in I} z_{ij} = 1 \qquad \text{for } j\in J \qquad (2)$$

$$y_i - z_{ij} \geq 0 \qquad \text{for } i\in I,\ j\in J \qquad (3)$$

$$y_i \in \{0, 1\} \qquad \text{for } i\in I \qquad (4)$$

$$z_{ij} \in \{0, 1\} \qquad \text{for } i\in I,\ j\in J \qquad (5)$$

where the used coefficients have the following meaning:

f_i ... fixed charge for considered period for holding center *i*,

e_0 ... prime cost for one unit transport along one distance unit on the link between a center and customer,

b_j ... the total demand of the customer j in the considered period,

d_{ij} ... the distance between a center and a customer,

I ... the set of possible center locations (e.g. warehouses),

J ... the set of the customers.

If we add a constraint to model *(1)-(5)* that the distance from a customer to the assigned center should be less than a given constant and we obtain the uncapacitated location problem with maximal service distance. The associated constraint has the form:

$$d_{ij}\, z_{ij} \le D_{max} \qquad pre\ i \in I,\ j \in J \qquad (6)$$

The lower bound of D_{max} in testing problems will be chosen so that each customer can be serviced at least from one center. The upper bound of D_{max} is constituted by the maximal distance between a customer and a center. To be able to solve this problem by an exact algorithm for the basic uncapacitated location problem, we introduce penalty constant D_{proh} and new distance coefficients $\underline{d}_{ij}$, which are obtained from real-network coefficient by the following way:

- $\underline{d}_{ij} = d_{ij}$ pre $d_{ij} \le D_{max}$ a pre $i \in I$, *(7)*
- $\underline{d}_{ij} = D_{proh}$ pre $d_{ij} > D_{max}$ a pre $i \in I$, *(8)*

The objective of a penalty constant determination is to prevent a customer from to be assigned to a center, which violates constraints *(6).*

If we substitute d_{ij} in the relation *(1)* with $\underline{d}_{ij}$ we will take the relation

$$Minimize \quad \sum_{i \in I} f_i\, y_i + \sum_{i \in I} \sum_{j \in J} e_0\, b_j\, \underline{d}_{ij}\, z_{ij} \qquad (9)$$

After necessary reformulation we obtain a new model. The expression *(9)* with the constraints *(2)- (5)* is the new model of uncapacitated location problem with maximal service distance. The distance coefficients are deformed by penalty constant in comparison to real network.

This model with rearranged objective function can be solved by exact algorithms, which were designed for the classical (original) uncapacitated location problem, in which no distance restriction is considered.
Such way we solved a lot of instances of the uncapacitated location problem with maximal service distance. The instances had the cost matrix of size 3000x100. Solving process of some instances did not finish correctly. The incorrect end was mostly caused by a memory overflow.
The selection of the applicable size of penalty constant is very important for the used method. The penalty constant must be large enough to prevent from assigning a customer to a center, which do not comply with the specified distance. It must not be either very large. During the computation the penalty constant is used very often and its accumulation in same variable causes the mentioned overflow. In such instances it is possible to scale down the value of some input data, e.g. the cost matrix or the demands of customers. We do it by application of the coefficient of the proportionality. However, it induces the rounding of numbers and distortion of the results.
The goal of this paper was to find the interval for the prohibitive constant for the instances of different dimensions. We would to find how the coefficient of the proportionality influences the accuracy of the final result.

CONCLUSIONS

Experiment showed that not only the dimension of the instances but also the geographical location of the candidates to the center locations have influence on the overflow of the memory. The analysis of the other studied parameter is a large-scaled problem and it needs much more time to evaluate the experiments. So, we will get it in the next article.

BIBLIOGRAPHY

[1] ERLENKOTTER, D.: A Dual-Based Procedure for Uncapacitated Facility Location. Operations Research, Vol. 26, No 6, 1978, 992-1009
[2] JANÁČEK, J., BUZNA Ľ.: A Comparison Continuous Approximation with Mathematical Programming Approach to Location Problems. Central European Journal of Operations Research, Vol. 12, No 3, Sept. 2004, 295-305

ACKNOWLEDGEMENT: This work has been supported by research project VEGA 1/3775/06.

RNDr. Marta Janáčková, CSc., Žilinská univerzita Žilina, Strojnícka fakulta, Katedra aplikovanej matematiky, Hurbanova 15, 010 26 Žilina, Slovenská republika, tel.: +421 41 513 4989, E-mail: marta.janackova@fstroj.utc.sk

RNDr. Alžbeta Szendreyová, PhD., Žilinská univerzita Žilina, Fakulta riadenia a informatiky, Katedra dopravných sietí, Veľký Diel, 010 26 Žilina, Slovenská republika, tel.: +421 41 513 4205, E-mail: alzbeta.szendreyova@fri.utc.sk

APPROXIMATIVE COVERING MODELS FOR EMERGENCY SYSTEM DESIGN

Jaroslav Janáček, Faculty of Management Science and Informatics, University of Žilina

1. INTRODUCTION

The medical emergency system design represents a crucial task for each responsible designer due to the interaction of two opposite demands on the system performance. On the one hand the designer is forced not to exceed a given number of located ambulance vehicles and to solve a large facility location problem. On the other hand, he must ensure a given quality of the service for potential patients. This quality is usually given by a fixed time limit, in which some ambulance vehicle should reach an arbitrary located potential patient [1], [8], [9]. This last demand is hard to meet in practise, because of random travel times on a real road network and furthermore, even if we have modelled the problem using estimated travel times; the formalized problem need not have any feasible solution. Due to this fact, there are used various surrogate criteria which try to express a size of affliction of potential patients, which are situated outside the fixed time limit from the nearest ambulance location [4], [5]. The following two criteria may represent the main approaches to quantification of patient affliction size.

$$\sum_{\substack{j\in J \\ t_{i(j),j} \succ T^{\max}}} b_j \tag{1}$$

$$\sum_{\substack{j\in J \\ t_{i(j),j} \succ T^{\max}}} b_j (t_{i(j),j} - T^{\max}) \,. \tag{2}$$

In the above-formulated criteria, J denotes set of dwelling places (customers),
b_j denotes number of inhabitants at the dwelling place j, $i(j)$ denotes the located centre, which is the time-nearest one to j and t_{ij} denotes the time of traversing from the place i to j. The first criterion expresses the size of the part of population, which is out of the time limit T^{max} considering the time nearest service centre and the second criterion is the sum of positive differences between the shortest access time and the time limit T^{max}.
To describe the problem of emergency system design minimizing potential patient affliction, mathematical programming models can be formed and the convenient means of problem solution can be sought. As both allocation and covering models can be used to describe a problem with the first criterion, the second criterion enabled only allocation model construction.

Unfortunately, the real-sized instances of the associated problems must be solved for enormous numbers of possible ambulance locations (thousands), what disqualifies the allocation models due the fact that they include approximately one million of allocation variables. In the frame of this paper, we present a new modelling approach to the emergency system design problem with the criterion (2). Our approach is based on a specific reformulation of allocation model to the form of generalized covering model, which comprises in order much less variables than the allocation one and preserves the properties, which enable its quick solvability using a standard IP-solver.

2. THE ALOCATION AND COVERING MODELS OF EMERGENCY SYSTEM

The medical emergency system design problem belongs to the family of location problems [6], in which it must be decided on ambulance vehicle locations. The served area consists of dwelling places situated in nodes of a road network. These dwelling places form a finite set J. The number of inhabitants of dwelling place $j \in J$ will be denoted as b_j. The emergency service system design problem can be formulated as a decision about location of at most p ambulance vehicles at some places from a larger set I of possible locations so that the value of chosen criterion is minimal.

Let j be dwelling place location and i be a possible location of ambulance vehicle, then t_{ij} denotes the accessibility time of dwelling place j from ambulance vehicle location i. Let us express price c_{ij} of the j-th dwelling place assignment to possible location i accordingly to [4], then we obtain $c_{ij} = b_j$, if $t_{ij} > T^{max}$ and $c_{ij} = 0$ for the criterion (1) and $c_{ij} = b_j(t_{ij} - T^{max})$, if $t_{ij} > T^{max}$ and $c_{ij} = 0$ for the criterion (2). The allocation model of the problem can be formulated as follows:

Minimize $$\sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij} \qquad (3)$$

Subject to $$\sum_{i \in I} z_{ij} = 1 \qquad \text{for } j \in J \qquad (4)$$

$$z_{ij} \le y_i \qquad \text{for } i \in I \text{ and } j \in J \qquad (5)$$

$$\sum_{i \in I} y_i \le p \qquad (6)$$

$$z_{ij} \in \{0,1\} \qquad \text{for } i \in I \text{ and } j \in J \qquad (7)$$

$$y_i \in \{0,1\} \qquad \text{for } i \in I. \qquad (8)$$

In this model, a decision on ambulance location at place $i \in I$ is modelled by the zero-one variable $y_i \in \{0,1\}$, which takes the value 1 if an ambulance should be located at i and it takes the value 0 otherwise. Furthermore, the

allocation variables $z_{ij}\in\{0,1\}$ for each $i\in I$ and $j\in J$ are introduced to assign the dwelling place j to the possible location i (z_{ij}=1).
The associated problem can be solved by the approach reported in [3] or [7], where the Lagrangean multiplier is introduced for constraint (6), and the constrain is relaxed. Then the problem takes form of the uncapacitated location problem. To solve this problem, the procedure *BBDual* [6] was designed and implemented based on principle presented in [2]. This procedure was embedded into the dichotomy algorithm, which was used to find propper value of the Lagrangean multiplier.
The medical emergency system design problem connected with criterion (1) can be also modelled using a set of the auxiliary zero-one variables x_j, which express by the values 0 or 1, whether the demand of dwelling place $j \in J$ is or is not satisfied. To be able to recognize, whether dwelling place j is or is not accessible from location i, we introduce zero-one constant a_{ij} for each pair $<i, j> \in I\times J$. The constant a_{ij} is equal to 1 if and only if dwelling place j can be reached from location i in the access time T^{max}, i.e. $t_{ij} \le T^{max}$. Otherwise, the constant a_{ij} is equal to 0. Then we can formulate the problem as so called covering model:

Minimize $$\sum_{j\in J} b_j x_j \qquad (9)$$

Subject to $$x_j + \sum_{i\in I} a_{ij} y_i \ge 1 \qquad \text{for } j\in J \qquad (10)$$

$$\sum_{i\in I} y_i \le p \qquad (11)$$

$$x_j\in\{0,1\} \qquad \text{for } j\in J \qquad (12)$$

$$y_i\in\{0,1\} \qquad \text{for } i\in I. \qquad (13)$$

The objective function (9) gives the volume of uncovered demands. The constraints (10) ensure that the variables x_j are allowed to take the value 0, if there is at least one ambulance vehicle located in the access time T^{max} from the dwelling place location j. The constraint (11) puts the limit p on the number of located vehicles [9].
As concerns the solving technique for the problem described by the second model, it can be noted that all of them belongs to the family of integer programming problems, more precisely zero-one integer programming problems and can be theoretically solved by any commercial solver, which contains some general integer programming algorithm, e.g. the branch and bound method or the branch and cut method. These general algorithms are able to solve to optimality real-sized covering problems, but only small instants of the allocation problem due to the huge cardinality of set of allocation variables z_{ij}.

3. THE APPROXIMATE COVERING MODEL

The excellent performance of general IP-sover on real-sized instances of covering problem inspired us with an idea to reformulate the medical emergency system design problem with criterion (2) to some form of covering problem. We are willing to pay for the goal by loss of preciseness as concerns the objective function value.

Taking into account these presumptions, we have divided the interval of possible access times into zones. The considered time interval starts at zero and ends at $max\{t_{ij}: i \in I, j \in J\}$ and is divided by increasing values of T^1, T^2 T^3, …, T^r into r+1 zones. The zone 0 corresponds to the interval <0, T^1>, the zone 2 corresponds to the interval <T^1, T^2>, etc. up to the zone r, which corresponds to the interval <T^r, $max\{t_{ij}: i \in I, j \in J\}$ >. We may assume that $T^1 = T^{max}$ in the studied problem.

In addition to the zero-one variable $y_i \in \{0,1\}$, which takes the value 1 if an ambulance should be located at location i, and which takes the value 0 otherwise, we introduce auxiliary zero-one variables x_{jk}, which express by the values 1 or 0, whether the access time of the dwelling place $j \in J$ from the nearest located ambulance is or is not bigger than T^k for k=1, … , r.

Let us denote by e_k for $k=1, .., r$ a width of k-1 th zone. I.e. $e_1=T^2- T^1$, $e_2=T^3- T^2$ and so on. Then the expression $e_1x_{j1}+ e_2x_{j2}+ e_3x_{j3}+\ldots+ e_rx_{jr}$ is an approximation of $t_{ij}-T^{max}$, preciseness of which is given by the zone width. Similarly to the covering model, we introduce zero-one constant a_{ij}^k for each triple $<i, j, k> \in I \times J \times \{1, \ldots, r\}$. The constant a_{ij}^k is equal to 1 if and only if dwelling place j can be reached from location i in the access time T^k, i.e. $t_{ij} \le T^k$. Otherwise, the constant a_{ij}^k is equal to 0. Then the associated covering-type model can be formulated as follows:

$$\text{Minimize} \quad \sum_{j \in J} \sum_{k=1}^{r} b_j e_k x_{jk} \tag{14}$$

$$\text{Subject to} \quad x_{jk} + \sum_{i \in I} a_{ij}^k y_i \ge 1 \qquad \text{for } j \in J \text{ and } k=1, \ldots, r \tag{15}$$

$$\sum_{i \in I} y_i \le p \tag{16}$$

$$x_{jk} \ge 0 \qquad j \in J \text{ and } k=1, \ldots, r \tag{17}$$

$$y_i \in \{0,1\} \qquad \text{for } i \in I. \tag{18}$$

In this model, the objective function (14) gives an estimation of the volume of uncovered demands. The constraints (15) ensure that the variables x_{jk} are allowed to take the value 0, if there is at least one ambulance vehicle located in the access time T^k from the dwelling place location j. The constraint (16) puts the limit p on the number of located vehicles.

4. NUMERICAL EXPERIMENTS

We have followed one goal only in this experimental part of work. We wanted to find how fast is a standard IP-solver when used on huge instances of problem described by the model (14)-(18). For these experiments, we used the medical emergency system design problem with the data originating at the Slovak road network with 2916 dwelling places, which represent aggregations of potential patients. The number of 223 points (locations) was taken into consideration as the value p in the solved instances. The access times t_{ij} were computed from real distances coming out from vehicle speeds connected with the individual link classes. The considered speed scenario was $\mathbf{v}$ = <105, 95, 75, 60, 50>, which are assumed average speeds in kilometer per hour on highways, roads of first, second and third class and on the local roads respectively. The set of candidate locations was formed from all towns and villages with more than 300 inhabitants and present positive ambulance locations. This way, the set of 2284 candidate locations was obtained. The zone widths took values from 1 to 2 minutes and the nine solved instances differ only in the number of zones. The instances for r = 1, 2, …, 9 were solved, where the instance with r=1 corresponds to a problem with criterion (1). The experiments were performed using the general optimisation software Xpress-IVE. The associated code were run on a personal computer equipped with the Intel Core 2 6700 processor with parameters: 2.66 GHz and 3 GB RAM. The results of numerical experiments are reported in Table 1, where each column corresponds to one instance of the problem and is specified by the value of r.

r	1	2	3	4	5	6	7	8	9
Objective	4.98e+06	6.45e+06	7.66e+06	8.63e+06	9.38e+06	9.93e+06	1.05e+07	1.07e+07	1.08e+07
Time [s]	1.687	3.219	11.859	8.531	10.688	14.313	18.485	21.938	18.968
Loc	223	223	223	223	223	223	223	223	223

Table 1

The rows contain the objective function value (Objective) of optimal solution, the computation time in seconds (Time [s]) and the number of located ambulances (Loc).

5. CONCLUSIONS

The presented results of numerical experiments show that the approximative models constitute a promising approach to huge location problems, which have been resisted to recent attempts to their solving. Even when the critical size of solved problems, i.e. the number of possible locations, exceeds considerably two thousands, the computational times are moderate in comparison with the location-allocation approaches.

BIBLIOGRAPHY

[1] CURRENT, J.; DASKIN, M.; SCHILLING, D.: Discrete network location models. In: Drezner, Zvi (ed.) et al. Facility location. Applications and theory. Berlin: Springer, 2002, pp 81-118

[2] ERLENKOTTER, D.: A Dual-Based Procedure for Uncapacitated Facility Location. Operations Research, Vol. 26, No 6, 1978, pp 992-1009

[3] JANÁČEK, J.: The Medical Emergency Service System Design. In: Advances in Transport Systems Telematics, ed. Mikulski, J., Katowice, 2006, pp 443-449

[4] JANÁČEK, J.: Time Accessibility and Public Service Systems. In: Proceedings of the conference "Quantitative Methods in Economics", Bratislava, December 6. - 8. 2006, pp 57-63

[5] JANÁČEK, J.: Safety on Roads from View of Emergency System Design. In: Journal of Information, Control and Management Systems, Vol. 4, No 2/1, 2006, pp 91-101

[6] JANÁČEK, J.; BUZNA Ľ.: Facility location in distribution systems. EDIS Žilina University Publisher, Žilina, 2007, 142 p.

[7] JANÁČKOVÁ, M., SZENDREYOVÁ, A.: An impact of set of centres of distribution system design problem on computational time of solving algorithm. In: Advances in Transport Systems Telematics, ed. Mikulski, J., Katowice, 2006, pp 109-112

[8] JÁNOŠÍKOVÁ, L.: Emergency Medical Service Planning. Communications –Scientific Letters of the University of Žilina, Vol. 9, No 2, 2007, pp 64-68

[9] MARIANOV, V.; SERRA, D.: Location problems in the public sector. In: Drezner, Zvi (ed.) et al. Facility location. Applications and theory. Berlin: Springer, 2002, pp 119-150

ACKNOWLEDGMENT: This work was supported by grant VEGA 1/3775/06.

Prof. RNDr. Jaroslav Janáček, CSc., Faculty of Management Science and Informatics, University of Žilina, 01026 Žilina, Slovak Republic, jaroslav.janacek@ fri.uniza.sk

THE MINIMAL NETWORK OF HOSPITALS AS A BICRITERION LOCATION PROBLEM

Ľudmila Jánošíková
University of Žilina, Faculty of Management Science and Informatics,
Department of Transportation Networks

Abstract

In the paper, the design of the minimal network of hospitals using mathematical programming methods is described. The problem is formulated as a bicriterion location problem with two objectives. The former objective respects the transportation accessibility of hospitals by regular patients and the latter represents the accessibility by emergency vehicles. A set of solutions can be generated by varying weights of the particular criteria. Our computational experiments performed for the area of the Slovak Republic compare several solutions with the present location of hospitals and the proposal of the Ministry of Health Care of the Slovak Republic from July 2007.

1. Introduction

In July 2007 the Ministry of Health Care of the Slovak Republic submitted to the Government a proposal defining which hospitals are necessary to provide health care in the Slovak Republic. The list of hospitals was denoted as the minimal network of hospitals. As a consequence, health insurance companies will be obliged to make a contract with the hospitals included in the minimal network while the funding of the other hospitals will be not guaranteed by the state. Among other reasons, in the argumentative report the Ministry said that the listed hospitals would provide urgent health care to population. The network of 45 hospitals was designed in such a way that an emergency ambulance could reach the nearest hospital in 30 minutes from any place of accident. Moreover, the network respected regional geographical accessibility without the emergency medical service (EMS). After a broad discussion the Government adopted the Order No 504/2007 Coll. of the National Council of the SR (effective from 15th November 2007), which lists only 12 general hospitals. This number is absolutely insufficient and it cannot guarantee the accessibility defined in the argumentative report.

The goal of this paper is to reveal the possibilities of using mathematical methods for the design of a network of hospitals with special attention paid to the accessibility criteria mentioned above. The network design problem can be formulated as a location problem, which aims to locate a predefined number of hospitals. More precisely, we face a discrete network location problem [1, 2], because facilities to be located (hospitals), as well as customers (cities and villages, where potential patients live) are located at nodes of a transportation network (the road network of the Slovak Republic in our case). There exist several types of network location models depending on the optimality criterion. The main topic of our paper is to define a proper model respecting both:

a) transportation accessibility by EMS vehicles,

b) regional transportation accessibility for people who need an examination or treatment in a hospital and use individual or public transport.

2. Mathematical model

Let us describe the problem more precisely. Hospitals to be included into the minimal network may be chosen from the set of present 68 general hospitals. Children and professional hospitals are not considered. To formalise the problem, let us denote the set of possible hospital locations by I. Hospitals provide health care service to people living in cities and villages. All 2916 municipalities of the Slovak Republic are regarded as „customers" of the service. Let us denote the set of all municipalities by J. An individual municipality j has b_j inhabitants.

The question to be answered first is: "How to express transportation accessibility of hospital i by customer j?" Transportation accessibility is defined as resistance, which must be overcome on the way from location i to location j. This resistance is usually the travel time. Both the locations i and j are nodes of the road network. Based on the road quality, each road segment belongs to a class from a finite classification system. Further, an average speed of a vehicle movement can be associated to each road class. Then an estimation of the necessary traversing time for each road segment can be obtained from the segment length and the average speed corresponding to the segment class. Using segment times, the accessibility time t_{ij} is enumerated as the time length of the shortest path connecting i and j. Because the average speeds are not constant but they depend on the vehicle used, the accessibility time $t_{ij}(\boldsymbol{v})$ is a function of vector $\boldsymbol{v} = \langle v_1, v_2, \ldots, v_r \rangle$ of the speeds, at which the vehicle runs on the road of class 1, 2, ..., r (notation taken from Janáček [3]). It follows that the assignment of a municipality to a hospital also depends on a given speed scenario $\boldsymbol{v}$. In the further presentation symbol $i(\boldsymbol{v}, j)$ stands for the hospital included in the minimal network, which is the time-nearest one to municipality j considering the speeds given by $\boldsymbol{v}$.

First we formulate the criteria to be respected by the design. Let us begin with the criterion b) representing regional transportation accessibility using individual or public transport. The criterion (C1) can be expressed as the total travel time potential patients spend on their way to the nearest hospital:

$$\sum_{j \in J} b_j t_{i(\boldsymbol{v},j),j}(\boldsymbol{v}) \tag{1}$$

where the speed vector $\boldsymbol{v}$ corresponds to individual or public transport. Clearly, the best design is where people spend the less time by travelling.

The second requirement stated by the Ministry says that the maximal ambulance travel time from a municipality to a hospital should not exceed 30 minutes, i.e. $t_{i(u,j),j}(\boldsymbol{u}) \le T^{max}$ (= 30 minutes) should be true for every $j \in J$. (Now, patients are transported by emergency vehicles therefore another speed vector $\boldsymbol{u}$ is used.) In terms of location theory this means that every municipality should be covered by the service. However, this is an unrealistic requirement especially in the geographic conditions of the Slovak Republic where a part of population lives in small villages far away from hospitals. So there may be that some municipalities left uncovered in a design, but they still must be serviced. To maximise transportation accessibility by the EMS for all population the total travel time for uncovered municipalities should be minimised. The criterion (C2) becomes:

$$\sum_{\substack{j \in J \\ t_{i(\boldsymbol{u},j),j}(\boldsymbol{u}) > T^{max}}} b_j t_{i(\boldsymbol{u},j),j}(\boldsymbol{u}) \tag{2}$$

The resulting objective function is a linear combination of criteria C1 and C2. The criteria are in conflict. Criterion C1 attracts hospitals to economic and administrative centres, where a lot of potential patients live. On the other hand, criterion C2 pulls hospitals to the edges of the region. A set of compromise (noninferior) solutions can be generated by varying weights of the particular criteria. The decision maker can then evaluate alternative solutions possibly considering other quantitative criteria such as maximum and average travel distances, as well as qualitative criteria such as the specialization of hospitals.

In the mathematical model, two types of bivalent variables are used. Variables $y_i \in \{0, 1\}$ model the decision whether hospital i will be included in the network ($y_i = 1$) or not ($y_i = 0$). Because the individual contribution of a municipality to the criterion value depends on the distance to the nearest opened hospital, we need other bivalent variables z_{ij} for each pair $<i, j>$. Variable z_{ij} takes the value 1 if hospital i is the nearest one to municipality j. Otherwise, it takes the value 0.

For the sake of simplicity, two cost coefficients are used in the objective function:

1. $c_{ij} = b_j t_{ij}(\boldsymbol{v})$,
2. $d_{ij} = \begin{cases} b_j t_{ij}(\boldsymbol{u}) & \text{if } t_{ij}(\boldsymbol{u}) > T^{max} \\ 0 & \text{otherwise} \end{cases}$.

Now the formulation of the model can be written as:

$$\text{Minimise} \quad \alpha \sum_{i \in I} \sum_{j \in J} c_{ij} z_{ij} + \beta \sum_{i \in I} \sum_{j \in J} d_{ij} z_{ij} \tag{3}$$

$$\text{Subject to} \quad \sum_{i \in I} z_{ij} = 1 \qquad \text{for } j \in J \tag{4}$$

$$z_{ij} \leq y_i \qquad \text{for } i \in I, j \in J \tag{5}$$

$$\sum_{i \in I} y_i \leq p \tag{6}$$

$$y_i \in \{0, 1\} \qquad \text{for } i \in I \tag{7}$$

$$z_{ij} \in \{0, 1\} \qquad \text{for } i \in I, j \in J \tag{8}$$

The objective function (3) is a linear combination of criteria C1 and C2 with weights α and β, where $\alpha + \beta = 1$. Constraints (4) ensure that each municipality is assigned to exactly one hospital. Constraints (5) are so called binding constraints, which force the variable y_i take the value 1, whenever a municipality is assigned to location i. Constraint (6) puts the limit p on the number of opened hospitals. It is apparent that the objective will decrease with the increasing p, so (6) may be written as a strict equality. Consequently, the model (3)-(8) is identical to the p-median location problem. A major advantage of this formulation is that several efficient solution procedures exist for the p-median problem allowing large-sized problems solving.

Varying the values of α and β, alternative configurations of hospital locations are generated. A proper value of p is an open issue. It cannot be defined in advance. Again, the decision maker can experiment with different values decreasing from the present number of hospitals to a minimal value. The minimal number of needed hospitals can be estimated by solving the Location Set Covering Problem (LSCP), which locates the minimal number of hospitals in such a way as to cover all municipalities covered by present-day hospitals.

3. Computational results

In our computational experiments we propose several alternatives of the weights α and β and evaluate their impact on the final deployment of hospitals. As the Ministry of Health Care wants to reduce the number of hospitals, we evaluate the location of the minimal number of hospitals, which was set by the LSCP to 46.

Our experiments were performed with $\boldsymbol{v} = \langle 85, 75, 55, 40, 30 \rangle$, which are average speeds for individual transport in kilometres per hour on highways, roads of first, second and third class and on the local roads respectively (see [4]). The components of vector $\boldsymbol{u}$ for emergency vehicles were higher by 20 kilometres per hour.

We compared the following solutions:

a) the present deployment of hospitals,

b) the proposal of the Ministry of Health Care of the Slovak Republic from July 2007,

c) the solution of the Maximal Covering Location Problem (MCLP)[1],

d) the solutions of the bicriterion model (3)-(8) for different weights α and β ranging from 0 to 1.

Table 1 compares the solutions in terms of the number of located hospitals, the total travel time (criterion C1), the total travel time by EMS for uncovered population (criterion C2), the average travel time by individual transport (criterion C1 divided by the population of the Slovak Republic), and the number of people living behind the 30-minutes limit.

To see the impact of the hospital reduction, the first row of the table evaluates particular criteria for the present location of hospitals. Comparison with mathematical solutions reveals that mathematical models (precisely the MCLP model and the bicriterion model with weights α = 0, β

[1] The objective of the MCLP is to maximise the population covered with the limited number of hospitals.

= 1) can locate a considerably reduced number of hospitals in such a way that the emergency accessibility decrease is negligible. However, increase in the average travel time is significant. On the other hand, the Ministry proposal emphasizes the individual accessibility regardless of the emergency accessibility. In this proposal, the number of people living behind the 30-minutes limit is the highest among of all solutions. The same is true for the total travel time for uncovered population. The bicriterion model may be used to generate compromise solutions. Every solution for α above 0 and β below 1 overcomes the Ministry proposal in the individual as well as emergency accessibility with almost the same number of located hospitals. For higher values of α, the first criterion outweighs the second one and the solution will preferably locate hospitals to big cities at the expense of the emergency accessibility of small villages located far away from regional centres.

Table 1

Comparison of proposals

Design		No. of hospitals	Total travel time [minutes]	Total travel time for uncovered population [minutes]	Average travel time [minutes]	No. of uncovered population
Present hospitals		68	45 856 194	1 331 050	8.53	39 593
Ministry Proposal		45	59 960 609	3 438 529	11.15	95 084
MCLP		46	64 052 935	1 335 111	11.91	39 593
Bicriterion model						
α	β					
0	1	46	64 283 124	1 335 111	11.95	39 593
0.05	0.95	46	59 693 458	1 386 041	11.10	41 200
0.1	0.9	46	58 239 355	1 507 191	10.83	44 805
0.3	0.7	46	57 936 484	1 547 633	10.77	45 782
0.4	0.6	46	57 607 419	1 741 374	10.71	51 141

0.5	0.5	46	57 607 419	1 741 374	10.71	51 141
1	0	46	57 607 419	1 741 374	10.71	51 141

Mathematical models were implemented in the general optimisation software *Xpress-MP*. Their solution takes about 30 seconds on a personal computer equipped with the Intel Core 2 6700 processor with 2.66 GHz and 3 GB of RAM.

4. Conclusion

In the paper, the design of the network of hospitals using mathematical programming methods is described. The resulting bicriterion location model respects both the transportation accessibility of hospitals by regular patients and the accessibility by emergency vehicles. By setting various weights for the particular criteria, several alternative solutions can be generated. Our computational experiments reveal the impact of particular criteria on the hospital deployment and on the accessibility in consequence.

References

1. CHURCH, R.L., CURRENT, J.R., STORBECK, J.E.: A bicriterion maximal covering location problem which considers the satisfaction of uncovered demand. In *Decision Sciences*, Volume 22, January 1991, p. 38-52.
2. CURRENT, J., DASKIN, M., SCHILLING, D. Discrete network location models. In DREZNER, Z., HAMACHER, H.W. (editors) *Facility Location* : Applications and Theory. 1st ed. Heidelberg : Springer-Verlag Berlin, 2004. ISBN 3-540-21345-7.
3. JANÁČEK, J. The medical emergency service system design. In MIKULSKI, J. (editor) *Advances in Transport Systems Telematics*. Katowice : Jacek Skalmierski Computer Studio, 2006. ISBN 83-917156-4-7.
4. PEŇÁZ, T., HORÁK, J. Určování dopravní dostupnosti pro dojížďku do zaměstnání při individuální neveřejné dopravě. In *Sborník konference GIS Ostrava 2004*. Ostrava : VŠB-TU Ostrava, 2004. ISSN 1213-239X.

Acknowledgement

This paper was supported by grant VEGA 1/3775/06.

doc. Ing. Ľudmila Jánošíková, CSc., University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, e-mail: Ludmila.Janosikova@fri.uniza.sk

A REMARK ON NONLINEAR FUNCTIONALS AND EMPIRICAL ESTIMATES

Vlasta Kaňková, Department of Econometrics Institute of Information Theory and Automation Prague, Czech Republic

Abstract. Optimization problems depending on a probability measure correspond very often to economic situations. Since the probability measure is there very often completely unknown, statistical estimates (based on date) have to replace mostly the unknown probability measure to obtain at least an approximate solution and an approximate optimal value. Properties of such statistical estimates have been investigated many times in the case of linear dependence of an objective function on the probability measure. However, this assumption is not fulfilled just in many economic models, see e.g. Markowitz model or some risk measures. We try to cover some of these complicated cases.

Keywords: Stochastic optimization problems, nonlinear functionals, empirical estimates, convergence rate, portfolio selection, Markowitz model, deviations risk measures.

1. INTODUCTION

Let (Ω, S, P) be a probability space; $\xi\,(:= \xi(\omega) = [\xi_1(\omega), \ldots, \xi_s(\omega)])$ an $s-$dimensional random vector defined on (Ω, S, P); $F\,(:= F(z), z \in R^s)$ the distribution function of ξ; $F_i, i = 1, \ldots, s$ one–dimensional marginal distribution functions corresponding to F; $P_F, Z(:= Z_F)$ the probability measure and support corresponding to F. Let, moreover, $g_0(:= g_0(x, z))$ be a real–valued (say continuous) function defined on $R^n \times R^s$; $X \subset R^n$ be a nonempty "deterministic" set. If the symbol E_F denotes the operator of mathematical expectation corresponding to F, then many economic applications correspond to a class of stochastic optimization problems that can be introduced in the form:

Find

$$\phi(F) = \inf\{\mathrm{E}_F g_0(x, \xi) \mid x \in X\}. \qquad (1)$$

In applications very often the "underlying" probability measure P_F has to be replaced by empirical one; evidently, then the solution is sought w.r.t. the "empirical problem":

Find

$$\phi(F^N) = \inf\{\mathrm{E}_{F^N} g_0(x, \xi) \mid x \in X\}, \tag{2}$$

where F^N denotes an empirical distribution function determined by (mostly) an independent random sample $\{\xi^i\}_{i=1}^N$ corresponding to the distribution function F. If $X(F), X(F^N)$ denote the optimal solution sets of the problems (1) and (2), then under rather general assumptions $\phi(F^N), X(F^N)$ are "good" stochastic estimates of $\phi(F), X(F)$ (see e.g. [2], [5], [6], [8], [12], [15], [16], [17], [18], [19]). There were introduced assumptions guaranteeing the consistency, asymptotic normality and convergence rate. Especially, it means in the last case that there were introduced assumptions under which

$$P\{\omega : N^{\beta} \mid \phi(F) - \phi(F^N) \mid > t\} \xrightarrow[(N \longrightarrow \infty)]{} 0 \quad \text{for} \quad t > 0, \beta \in (0, \frac{1}{2}). \tag{3}$$

To obtain this relation the Hoeffding inequality (see e.g. [4], [6]), large deviation (see e.g. [5]), Talagrand approach (see e.g. [13]) and the stability results (see e.g. [10]) have been employed.

Furthermore, let us consider a simple "underlying" classical portfolio problem:

Find

$$\max \sum_{k=1}^{n} \xi_k x_k \quad \text{s.t.} \quad \sum_{k=1}^{n} x_k \le 1, \quad x_k \ge 0, \quad k = 1, \ldots, n, \quad s = n, \tag{4}$$

where x_k is a fraction of the unit wealth invested in the asset k, ξ_k denotes the return of the asset $k \in \{1, 2, \ldots n\}$. If $\xi_k, k = 1, \ldots, n$ are known, then (4) is a linear programming problem. However, $\xi_k, k = 1, \ldots, n$ are mostly random variables with unknown realizations in a time decision. If we denote

$$\mu_k = \mathrm{E}_F \xi_k, \quad c_{k,j} = \mathrm{E}_F (\xi_k - \mu_k)(\xi_j - \mu_j), \quad k, j = 1, \ldots n, \tag{5}$$

then it is reasonable to set to the portfolio selection two–objective optimization problem:

Find

$$\max \sum_{k=1}^{n} \mu_k x_k, \quad \min \sum_{k=1}^{n} \sum_{j=1}^{n} x_k c_{k,j} x_j \quad \text{s. t.} \quad \sum_{k=1}^{n} x_k \le 1, \quad x_k \ge 0, \quad k = 1, \ldots, n. \tag{6}$$

Evidently, there exists only rarely a possibility to find an optimal solution simultaneously with respect to the both criteria. Markowitz suggested (see e.g. [3]) to replace the problem (6) by one–criterion optimization problem of the form:

Find

$$\phi^M(F) = \max\{\sum_{k=1}^{n} \mu_k x_k - K\sum_{k=1}^{n}\sum_{j=1}^{n} x_k c_{k,j} x_j\} \quad \text{s. t.} \quad \sum_{k=1}^{n} x_k \le 1, \quad x_k \ge 0, \quad k = 1,\ldots,n, \tag{7}$$

where $K \ge 0$ is a constant. Obviously, for every $K \ge 0$ there exists $\lambda \in \langle 0,1\rangle$ such that the problem (7) is equivalent to the following one:

Find

$$\phi^\lambda(F) = \max\{\lambda\sum_{k=1}^{n} \mu_k x_k - (1-\lambda)\sum_{k=1}^{n}\sum_{j=1}^{n} x_k c_{k,j} x_j\} \quad \text{s. t.} \quad \sum_{k=1}^{n} x_k \le 1, \quad x_k \ge 0,\ k = 1,\ldots,n. \tag{8}$$

Evidently,

$$\sigma^2(x) = \sum_{k=1}^{n}\sum_{j=1}^{n} x_k c_{k,j} x_j = \mathrm{E}_F\{\sum_{j=1}^{n} \xi_j x_j - \mathrm{E}_F[\sum_{j=1}^{n} \xi_j x_j]\}^2, \quad x = (x_1,\ldots,x_n)$$

can be considered as a risk measure, that can be replaced by

$$\sigma(x) = \sqrt{\mathrm{E}_F\{\sum_{j=1}^{n} \xi_j x_j - \mathrm{E}_F[\sum_{j=1}^{n} \xi_j x_j]\}^2}, \quad x = (x_1,\ldots,x_n). \tag{9}$$

For more details, see e.g. [1], where an analysis of the corresponding relationship (according to multiobjective optimization theory) is introduced. Konno and Yamazaki introduced in [11] another risk measure $w(x)$ by

$$w(x) = \mathrm{E}_F\,|\sum_{k=1}^{n} \xi_k x_k - \mathrm{E}_F[\sum_{k=1}^{n} \xi_k x_k]\,|. \tag{10}$$

Moreover, they have recalled that $w(x) = \sqrt{\tfrac{2}{\pi}}\sigma(x)$ in the case of mutually normally distributed random vector $\xi = (\xi_1,\ldots,\xi_n)$.

Evidently, $\mathrm{E}_F\,|\sum_{k=1}^{n} \xi_k x_k - y|$ is a Lipschitz function of y, as well as the objective function $[\lambda\sum_{k=1}^{n} \mu_k x_k + (1-\lambda)w(x)], \lambda \in \langle 0,1\rangle$ is a Lipschitz function of $y := \mathrm{E}_F[\sum_{k=1}^{n} \xi_k x_k]$. Some others risk measures fulfilling the Lipschitz property can be found e.g. in [14].

To introduce more general problems covering (10), let $h(x, z) = (h_1(x, z), \ldots, h_{m_1}(x, z))$ be m_1–dimensional vector function defined on $R^n \times R^s$, $g_0^1(x, z, y)$ be a function defined on $R^n \times R^s \times R^{m_1}$. According to this new situation we replace the problem (1) by a stochastic programming problem in the form:

Find

$$\phi(F) := \phi^1(F) = \inf\{\mathrm{E}_F g_0^1(x, \xi, \mathrm{E}_F h(x, \xi)) \,|\, x \in X\}. \tag{11}$$

2. PROBLEM ANALYSIS

To investigate the empirical estimates of the problem (11), evidently, we can employ the assertion introduced in [10].

Lemma 1. [10] Let G be an arbitrary s dimensional distribution function. If

1. $g_0^1(x, z, y)$ is for every $x \in X, z \in R^s$ a Lipschitz function of $y \in Y$ with a Lipschitz constant $L^y(x, z)$; where
   $Y = \{y \in R^{m_1} : y = h(x, z) \quad \text{for some} \quad x \in X, z \in R^s\}$,
2. for every $x \in X, y \in Y$ there exist finite mathematical expectations
   $\mathrm{E}_F L^y(x, \xi), \quad \mathrm{E}_F h(x, \xi), \quad \mathrm{E}_G h(x, \xi),$
   $\mathrm{E}_F g_0^1(x, \xi, \mathrm{E}_F h(x, \xi)), \quad \mathrm{E}_F g_0^1(x, \xi, \mathrm{E}_G h(x, \xi)), \quad \mathrm{E}_G g_0^1(x, \xi, \mathrm{E}_G h(x, \xi)),$

then for every $x \in X$ it holds that

$$\begin{aligned} &|\, \mathrm{E}_F g_0^1(x, \xi, \mathrm{E}_F h(x, \xi)) - \mathrm{E}_G g_0^1(x, \xi, \mathrm{E}_G h(x, \xi)) \,| \leq \\ &\qquad \mathrm{E}_F L^y(x, \xi) \left\| \mathrm{E}_F h(x, \xi) - \mathrm{E}_G h(x, \xi) \right\|_{m_1}^2 + \\ &|\, \mathrm{E}_F g_0^1(x, \xi, \mathrm{E}_G h(x, \xi)) - \mathrm{E}_G g_0^1(x, \xi, \mathrm{E}_G h(x, \xi)) \,|. \end{aligned} \tag{12}$$

($\|\cdot\|_{m_1}^2$ denotes the Euclidean norm in R^{m_1}.)

Consequently, the assumptions guaranteeing the relation (3) can be employed in this case to obtain new results (for the problem (11)). However, the classical Markowitz problem (7) as well as some deviations risk measures mentioned in [14] is not covered by the problem (11). A special case (dealing with the Markowitz model) has been considered in [9]. To recall this assertion we denote by the symbol $C(:= C(n \times n))$ the matrix with elements $c_{k,j}, k, j = 1, \ldots, n$ defined by the relation (5). Furthermore, we denote by the symbols $x(F), x(F^N)$ the solutions of the problem (7) and the corresponding empirical problem.

Proposition 1. Let Z_F, X be compact sets, $t > 0, \{\xi^i\}_{i=1}^N$ independent random sample, $N = 1, 2, \ldots, \beta \in (0, \frac{1}{2}),$ then

$$P\{\omega : N^\beta \,|\, \phi^M(F) - \phi^M(F^N) \,|> t\} \xrightarrow[(N \longrightarrow \infty)]{} 0.$$

If, moreover $M > 0$ and the matrix C is positive definite, then also

$$P\{\omega : N^\beta \left\| x(F^N) - x(F) \right\|_n^2 > t\} \xrightarrow[(N \longrightarrow \infty)]{} 0.$$

Proof. The assertion of Proposition 1 follows immediately from Theorem 3 [9].

3. SOME AUXILIARY ASSERTIONS

In this section we prove some auxiliary assertions. To this end we consider $s = 2$ and set $\xi_1 := \bar{\xi}, \xi_2 := \bar{\eta},$ where $\bar{\xi} := \bar{\xi}(\omega), \bar{\eta} := \bar{\eta}(\omega)$ are random valuables defined on (Ω, S, P) with finite second moments. If we denote by the symbols $F(:= F_{(\bar{\xi}, \bar{\eta})}), F_{\bar{\xi}}, F_{\bar{\eta}}$ the distribution functions of the random vector $(\bar{\xi}, \bar{\eta})$ and marginal distribution functions of the random valuables $\bar{\xi}$ and $\bar{\eta},$ then

$$
\begin{gathered}
|\,\mathrm{E}_F[(\bar{\xi} - \mathrm{E}_F\bar{\xi})(\bar{\eta} - \mathrm{E}_F\bar{\eta})] - \mathrm{E}_{F^N}[(\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi})(\bar{\eta} - \mathrm{E}_{F^N}\bar{\eta})]\,| \le \\
|\,\mathrm{E}_F\bar{\xi}\bar{\eta} - \mathrm{E}_{F^N}\bar{\xi}\bar{\eta}\,| + |\,\mathrm{E}_F\bar{\xi}\,\mathrm{E}_F\bar{\eta} - \mathrm{E}_F\bar{\xi}\,\mathrm{E}_{F^N}\bar{\eta}\,| + |\,\mathrm{E}_F\bar{\xi}\,\mathrm{E}_{F^N}\bar{\eta} - \mathrm{E}_{F^N}\bar{\xi}\,\mathrm{E}_{F_N}\bar{\eta}\,| \le \\
|\,\mathrm{E}_F\bar{\xi}\bar{\eta} - \mathrm{E}_{F^N}\bar{\xi}\bar{\eta}\,| + |\,\mathrm{E}_F\bar{\xi}\,|\,|\,\mathrm{E}_F\bar{\eta} - \mathrm{E}_{F^N}\bar{\eta}\,| + |\,\mathrm{E}_{F^N}\bar{\eta}\,|\,|\,\mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi}\,|.
\end{gathered}
\tag{13}
$$

Consequently for $t > 0$

$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_F[(\bar{\xi} - \mathrm{E}_F\bar{\xi})(\bar{\eta} - \mathrm{E}_F\bar{\eta})] - \mathrm{E}_{F^N}[(\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi})(\bar{\eta} - \mathrm{E}_{F^N}\bar{\eta})] \,|\geq t\} \leq$$
$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_F\bar{\xi}\bar{\eta} - \mathrm{E}_{F^N}\bar{\xi}\bar{\eta} \,|> \frac{t}{3}\} + P\{\omega : N^{\beta} \,|\, \mathrm{E}_F\bar{\xi} \,|\,|\, \mathrm{E}_F\bar{\eta} - \mathrm{E}_{F^N}\bar{\eta} \,|> \frac{t}{3}\} +$$
$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_{F^N}\bar{\eta} \,||\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3}\}. \tag{14}$$

If we set $\Omega_1^N(t) = \{\omega : |\, \mathrm{E}_{F^N}\bar{\eta} - \mathrm{F}_F\bar{\eta} \,| < \sqrt{t}\}, \quad \Omega_1^{N,c}(t) = \Omega - \Omega_1^N(t)$ and assume (without loss of generality) that $\mathrm{E}_{F_{\bar{\eta}}}\bar{\eta} > 0$, then

$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_{F^N}\bar{\eta} \,|\,|\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3}\} = P\{\omega : N^{\beta} \,|\, \mathrm{E}_{F^N}\bar{\eta} \,|\,|\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3} \bigcap \Omega_1^N(t)\} +$$
$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_{F^N}\bar{\eta} \,|\,|\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3} \bigcap \Omega_1^{N,c}(t)\} \leq$$
$$P\{\omega : N^{\beta}(\sqrt{t} + \mathrm{E}_F\bar{\eta}) \,|\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3} \bigcap \Omega_1^N(t)\} + P\{\omega : \Omega_1^{N,c}(t)\} \leq$$
$$P\{\omega : N^{\beta} \,|\, \mathrm{E}_F\bar{\xi} - \mathrm{E}_{F^N}\bar{\xi} \,|> \frac{t}{3(\sqrt{t} + \mathrm{E}_F\bar{\eta})}\} + P\{\omega : N^{\beta} \,|\, \mathrm{E}_{F^N}\bar{\eta} - \mathrm{F}_F\bar{\eta} \,|> \sqrt{t}\}. \tag{15}$$

Lemma 2. Let $\bar{\zeta} = \bar{\xi}\bar{\eta} \,(:= \bar{\xi}(\omega)\bar{\eta}(\omega))$. Let moreover $F_{\bar{\zeta}}$ denote the distribution function of $\bar{\zeta}$. If

1. $P_{F_{\bar{\xi}}}, P_{F_{\bar{\eta}}}$ are absolutely continuous with respect to the Lebesgue measure on R^1 (we denote by $f_{\bar{\xi}}, f_{\bar{\eta}}$ the probability densities corresponding to $F_{\bar{\xi}}, F_{\bar{\eta}}$),
2. there exist constants $C_1^{\bar{\xi}}, C_2^{\bar{\xi}}, C_1^{\bar{\eta}}, C_2^{\bar{\eta}} > 0$ and $T > 0$ such that

$$f_{\bar{\xi}}(z) \leq C_1^{\bar{\xi}} \exp\{-C_2^{\bar{\xi}} \,|\, z \,|\} \quad \text{for} \quad z \in (-\infty, -T) \bigcup (T, \infty),$$
$$f_{\bar{\eta}}(z) \leq C_1^{\bar{\eta}} \exp\{-C_2^{\bar{\eta}} \,|\, z \,|\} \quad \text{for} \quad z \in (-\infty, -T) \bigcup (T, \infty),$$

then, there exist constants $C_1^{\bar{\zeta}}, C_2^{\bar{\zeta}} > 0, \bar{T} > 1$ such that for $z > \bar{T}$

$$F_{\bar{\zeta}}(-z) \le \frac{C_1^{\bar{\zeta}}}{C_2^{\bar{\zeta}}} \exp\{-C_2^{\bar{\zeta}} \sqrt{z}\}, \quad (1-F_z) \le \frac{C_1^{\bar{\zeta}}}{C_2^{\bar{\zeta}}} \exp\{-C_2^{\bar{\zeta}} \sqrt{z}\}.$$

Proof. First, evidently, for $z > T$

$$F_{\bar{\xi}}(-z) \le \frac{C_1^{\bar{\xi}}}{C_2^{\bar{\xi}}} \exp\{-C_2^{\bar{\xi}} z\}, \quad 1-F_{\bar{\xi}}(z) \le \frac{C_1^{\bar{\xi}}}{C_2^{\bar{\xi}}} \exp\{-C_2^{\bar{\xi}} z\},$$

$$F_{\bar{\eta}}(-z) \le \frac{C_1^{\bar{\xi}}}{C_2^{\bar{\eta}}} \exp\{-C_2^{\bar{\eta}} z\}, \quad 1-F_{\bar{\eta}}(z) \le \frac{C_1^{\bar{\eta}}}{C_2^{\bar{\eta}}} \exp\{-C_2^{\bar{\eta}} z\}.$$

Consequently, if $\bar{\xi}(\omega) = \bar{\eta}(\omega)$ a.s., then

$$P\{\omega : \bar{\zeta}(\omega) < -z\} = 0, \quad P\{\omega : \bar{\zeta}(\omega) > z\} = P\{\omega : |\bar{\xi}(\omega)| > \sqrt{z}\} = 2\frac{C_1^{\bar{\xi}}}{C_2^{\bar{\xi}}} \exp\{-C_2 \sqrt{z}\},$$

if $\bar{\xi}(\omega) \ne \bar{\eta}(\omega)$, then evidently for $z > 1$

$$P\{\omega : \bar{\zeta}(\omega) > z\} = P\{\omega : \bar{\xi}(\omega)\bar{\eta}(\omega) > z\} \le$$
$$P\{\omega : \bar{\xi}(\omega)\bar{\eta}(\omega) > z; \quad |\bar{\xi}(\omega)| > \sqrt{z}\} + P\{\omega : \omega : \bar{\xi}(\omega)\bar{\eta}(\omega) > z; \quad |\bar{\eta}(\omega)| > \sqrt{z}\} \le$$
$$P\{\omega : |\bar{\xi}(\omega)| > \sqrt{z}\} + P\{\omega : |\bar{\eta}(\omega)| > \sqrt{z}\} \le 2\frac{C_1^{\bar{\xi}}}{C_2^{\bar{\xi}}} \exp\{-C_2^{\bar{\xi}} \sqrt{z}\} + 2\frac{C_1^{\bar{\eta}}}{C_2^{\bar{\eta}}} \exp\{-C_2^{\bar{\eta}} \sqrt{z}\}.$$

Evidently, the assertion of Lemma 2 is valid.

Furthermore, it follows from Lemma 2 that for $z > \bar{T}$ there exist D_1, D_2, D_3 such that

$$\int_z^{+\infty} (1 - F_{\bar{\zeta}}(u))du \le D_1 \sqrt{z} e^{-D_2 \sqrt{z}} + D_3 e^{-D_2 \sqrt{z}}. \tag{16}$$

4. CONVERGENCE RATE

To introduce more general assertion dealing with the Markowitz problem we employ an approach employing in [10].

Theorem. Let X be a compact set. Let, moreover, $\left\{\xi^i\right\}_{i=1}^N$ be an independent random sample corresponding to the distribution function F, $N = 1, 2, \ldots$. If

1. $P_{F_i}, i = 1, \ldots, n$ are absolutely continuous with respect to the Lebesgue measure on R^1 (we denote by f_i probability densities corresponding to P_{F_i}),
2. there exist constants $C_1^i, C_2^i > 0$ and $T > 0$ such that
$$f_i(z) \le C_1^i \exp\{-C_2^i z\} \quad \text{for} \quad z \in (-\infty, -T) \bigcup (T, \infty), \quad i = 1, \ldots, n,$$

then for $t > 0, \beta \in (0, \frac{1}{2})$
$$P\{\omega : N^\beta \,|\, \phi^M(F) - \phi^M(F^N) \,|> t\} \xrightarrow[N \longrightarrow \infty)]{} 0.$$

If, moreover $M > 0$ and the matrix C is positive definite, then also
$$P\{\omega : N^\beta \left\| x(F^N) - x(F) \right\|^2 > t\} \xrightarrow[N \longrightarrow \infty)]{} 0.$$

Proof. Employing the relations (14), (15), (16), the assertion of Lemma 2 and the technique employed in [10] we obtain the assertion of Theorem.

CONCLUSION

The paper deals with empirical estimates in the case of nonlinear functional. First, the functional, fulfilling the Lipschitz property are analyzed. Furthermore, the Markowitz case is studied separately. Evidently, the second moment employed in the Markowitz case is a special case of more general class of the deviations risk measures.

Of course, results of mathematical statistics can be employed to extend the class of probability measures guaranteeing the results of Theorem. However, the "statistical" approach did not covered the above mentioned deviations risk measure that can be covered by approach employed in this contribution. Moreover, a similar approach can be employed for stability investigation.

Acknowledgment The research was supported by the Czech Science Foundation under Grants 402/07/1113, 402/08/0107 and 402/06/0990.

REFERENCES

1. CABALLERO, R., CERDÁ, R., MIÑOZ, M. M., REY, L. and

STANCU–MINASIAN, I.M.: Efficient Solution Concepts and Their Relations in Stochastic Multiobjective Programming. Journal of Optim. Theory and Applications (2001), 110, 1, 53–74.
2. DUPAČOVÁ, J., WETS, R. J.-B.: Asymptotic Behaviour of Statistical Estimates and Optimal Solutions of Stochastic Optimization Problems. Ann. Statist. (1984), 16, 1517–1549.
3. DUPAČOVÁ, J., HURT, J. and ŠTÉPÁN, J.: Stochastic Modelling in Economics and Finance. Kluwer, London 2002.
4. HOEFFDING, W.: Probability Inequalities for Sums of Bounded Random Variables. Journal of Americ. Statist. Assoc. (1963), 58, 301, 13–30.
5. KANIOVSKI, Y. M., KING, A. J. and WETS, R. J.-B.: Probabilistic Bounds (via Large Deviations) for the Solutions of Stochastic Programming Problems. Annals of Oper. Res. (1995), 56, 189–208.
6. KAŇKOVÁ, V.: An Approximative Solution of Stochastic Optimization Problem. In: Trans. 8th Prague Conf. 1974. Academia, Prague 1978, 349–353.
7. KAŇKOVÁ, V., LACHOUT, P.: Convergence Rate of Empirical Estimates in Stochastic Programming. Informatica (1992), 3, 497–523.
8. KAŇKOVÁ, V.: A Note on Estimates in Stochastic Programming. J. Comput. Math. (1994), 56, 97–112.
9. KAŇKOVÁ, V.: Multiobjective Programs and Markowitz Model. In: Proceedings of International Conference "Quantitative Methods in Economics (Multiple Criteria Decision Making XI)", Slovak Society for Operations Research&University of Economics in Bratislava, Bratislava 2004, 109–117.
10. KAŇKOVÁ, V.: Empirical Estimates via Stability in Stochastic Programming. Research Report UTIA 2007, No. 2192.
11. KONNO, H. and YAMAZAKI, H.: Mean–Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Markt. Magement Science (1991), 37, 5, 519–531.
12. DE MELLO, TITO HOMEN : On Rates of Convergence for Stochastic Optimization Problems under Non–i.i.d. Sampling. Speps 2006.
13. PFLUG,G.Ch.: Stochastic Optimization and Statistical Inference. In: Stochastic Programming (A. Ruszczynski and A.A. Shapiro, eds.) Handbooks in Operations Research and Managemennt Science, Vol 10, Elsevier, Amsterodam 2003, 427-480.
14. PFLUG, G. CH., RÖMISCH, W.: Modeling, Measuring and Managing Risk. World Scientific Publishing Co.Pte. Ltd., Singapore 2007.
15. RÖMISCH, W., WAKOLBINGER, A.: Obtaining Convergence Rates for Approximations in Stochastic Programming. In: Parametric

Optimization and Related Topics (J. Guddat, ed.). Akademie Verlag, Berlin 1987, 327–343.
16. SHAPIRO, A.: Quantitative Stability in Stochastic Programming. Math. Program. (1994), 67, 99–108.
17. VOGEL, S.: Stability Results for Stochastic Programming Problems. Optimization 1988, 19, 269–288.
18. WANG, L., WANG, J.: Limit Distribution of Statistical Estimators for Stochastic Programs with Dependent Samples. Z. Angew. Math. Mech. (1999), 79, 257–266.
19. WETS, R. J. - B.: A Statistical Approach to the Solution of Stochastic Programs with (Convex) Simple Recourse. Research Report, University of Kentucky, USA 1974.

VLASTA KAŇKOVÁ
Department of Econometrics
Institute of Information Theory and Automation
Academy of Science of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic
e-mail: kankova@utia.cas.cz

HEURISTIC METHODS FOR „MANY-TO-MANY" DISTRIBUTION SYSTEM DESIGN PROBLEM

Michal Koháni, University of Žilina, Slovakia

Abstract: Many-to-many distribution system is a special case of transportation system, where flows of goods from primary sources to customers are concentrated in terminals to create a bigger flow between terminals. This model belongs to discrete quadratic programmes and the optimal solution can't be found because of time purposes. First method for solving the problem reformulates the model to the form of linear programming problem to be able to use algorithms for integer programming. Second method is to use a metaheuristic method to solve the problem. Genetic algorithm or SUPRA method seems to be good for this purpose.
Keywords: Many-to-many distribution system, linearization, transportation system

1 INTRODUCTION

A transportation system, which has approximately the same number of primary sources as number of customers, seems to be a marginal case of a distribution system. This case includes such instances as National Postal Network or cargo railway system [5], which provides transport of carriages between railway stations. In these cases, demands of customers form a matrix of yearly flows from sources to places of destination. We denote this matrix as $\boldsymbol{B} = \{b_{sj}\}$, for $s \in S$ and $j \in J$, where S is a set of sources and J denotes set of customers. The fact that unit cost of transportation is smaller when bigger bulks of items are transported, approves concentration of flows between different pairs of source and customer to stronger flows at least on a part of their way. This flow concentration needs terminals, in which transhipment of transported items is performed and bigger bulks are formed or, on the other side, where bulks are split into smaller groups designated to different customers.

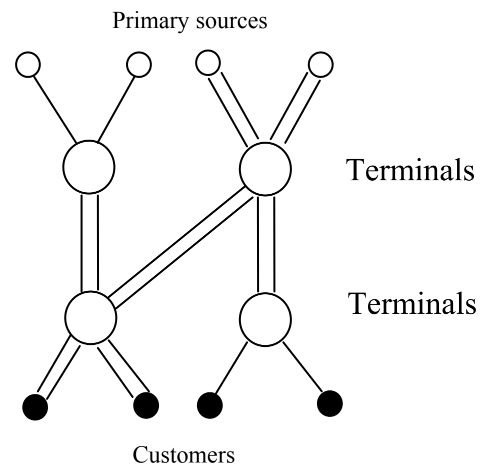


Figure 1

On the contraries to the classical distribution systems, in which big bulks leave primary source, another situation emerges in the many-to-many distribution systems. Primary sources send relative small bulks of items and it is useful to concentrate them to bigger bulks in the terminals located near the sources and then to send these bigger bulks to remote terminals and to split them there (see Figure 1).

We restrict ourselves here to the distribution system, in which a customer/source is assigned to only one terminal and an exchange of the consignments between the customer/source and other primary sources or customers is done via this assigned terminal, as it is shown in Fig. 1. Furthermore, we consider the general case, in which any source is also a customer, what is the case of post offices, railway stations and so on. We do not make any difference between a primary source and a customer hereafter and we introduce the set $J' =\{1, ..., n\}$, of customer-sources, for which matrix $\boldsymbol{B}$ gives by coefficients b_{sj} the yearly volume of the consignments, which are sent from the object s to the object j and it gives by coefficients b_{js} the total yearly volume sent from the object j to the object s.

2 Model of many-to-many distribution system design problem

Let us consider a case with a linear cost estimation function with unit cost e_0 for one item transport along unit distance on the way from a primary source to a terminal or from a terminal to a customer. Next, let us consider unit cost e_1 for one item transport along unit distance on the way between terminals. Furthermore we denote possible terminal locations by symbol $i=1, ..., m$, where each place i is associated with yearly fixed charges f_i for building and performance of the terminal i and with unit cost g_i for transhipment of one unit in the terminal. In accordance to the previous definition, we denote by $s \in J'$ object which sends consignments with the yearly total amount b_{sj} from s to $j=1, ..., n$. Symbol d_{pq} denotes the distance between objects $p \in J'$ and $q \in J'$. Our goal is to assign each sending or receipting object to exactly one terminal so that the total yearly cost of the designed system be minimal. If we denote by $y_i \in \{0, 1\}$ for $i= 1, ..., m$ the bivalent variable, which corresponds to the decision if a terminal will $(y_i =1)$ or will not $(y_i =0)$ be built at place i and if we introduce the variable $z_{ij} \in \{0, 1\}$ for $i =1, ..., m$ a $j =1, ..., n$, which says if object j will $(z_{ij} =1)$ or will not $(z_{ij} =0)$ be assigned to place i, then we can establish following mathematical programming model of the problem.

$$Minimise \sum_{i=1}^{m} f_i y_i + \sum_{i=1}^{m} \sum_{j=1}^{n} (e_0 d_{ij} + g_i)(\sum_{s=1}^{n} b_{js} + \sum_{s=1}^{n} b_{sj}) z_{ij} +$$

$$+ \sum_{i=1}^{m} \sum_{k=1}^{m} e_1 d_{ik} \sum_{j=1}^{n} \sum_{s=1}^{n} b_{sj} z_{ij} z_{ks} \ . \qquad (1)$$

Subject to $\sum_{i=1}^{m} z_{ij} = 1$ *for* $j=1, \ldots, n,$ *(2)*

$z_{ij} \le y_i$ *for* $i=1, \ldots, m, j=1, \ldots, n,$ *(3)*

$y_i \in \{0, 1\}$ *for* $i=1, \ldots, m,$ *(4)*

$z_{ij} \in \{0, 1\}$ *for* $i=1, \ldots, m, j=1, \ldots, n.$ *(5)*

The model belongs to discrete quadratic programmes due the third term of *(1)*.

3 SOLVING TECHNIQUES

The model belongs to discrete quadratic programmes due the second term of *(1)*. The exact solution of this type of problems is hard due the time reasons. For approximate solving of this problem some of approximate or heuristic methods can be used, for example SUPRA algorithm or genetic algorithm.
These methods don't give us an exact solution, so we need to know, how „far" from optimal solution we are. For knowing this, we have to estimate this optimal value and we can use lower bound solution of the problem. Solution techniques for lower bound solution of the problem were mentioned in [7].

4 SUPRA ALGORITHM

The SUPRA algorithm is a heuristic method based on the gradient method. This algorithm gives us a good combination for the set of parameters $\mathbf{p} \in P$, which optimize the objective function of given problem. This combination of parameters is the outcome of the solving and optimizing process. Solving process can be written as:
Minimize $F(\mathbf{p})$
Subject to $\mathbf{p} \in P$
Input to this algorithm is the solved problem $\boldsymbol{a}$ (written in the form of mathematical model) and the set of initial parameters $\boldsymbol{p}^{\boldsymbol{c}}$.
The basic step of this algorithm tries to search the best solution in the surrounding of actual solution by the method of the best proofs combinated with a local search in the direction of the statistical gradient. This new solution, a set of parameters $\boldsymbol{p}^{\boldsymbol{n}}$ with appropriate objective value $F(\boldsymbol{p}^{\boldsymbol{n}})$ is given to the algorithm for next processing.

The basic step of this algorithm can be described:

4.1 Generating of proofs:
A new set of parameters can be achieved:
$\boldsymbol{p}^j = \boldsymbol{p}^c + \boldsymbol{r}^j$, where $\boldsymbol{r}^j$ is randomly generated and $\boldsymbol{p}^j \in P$.
$\boldsymbol{r}^j = \boldsymbol{w} + \boldsymbol{x}$, where $\boldsymbol{w}$ is a **parameter of learning** and $\boldsymbol{x}$ is a randomly generated vector with balanced probability distribution with maximal value A, where A is a size of difference.

Actualization of w^i depends on the parameter of memory $B \in (0, 2)$ and the intensity of learning $C \in (0, 1)$:

$$w_i = B * w_i + C * ((F(\mathbf{p^c}) - F(\mathbf{p^j})) * (p^j_i - p^c_i)) \qquad (6)$$

The statistical gradient $\boldsymbol{r}$ is actualized by:

$$\boldsymbol{r} := \boldsymbol{r} + ((F(\mathbf{p^c}) - F(\mathbf{p^j})) * (p^j_i - p^c_i)) \qquad (7)$$

4.2 Local search in the direction of statistical gradient:
$\boldsymbol{p} = \boldsymbol{p}^c + \alpha \boldsymbol{r}$, where α is size of step for searching in the direction of statistical gradient.

The SUPRA algorithm finish the solving process after achieving a given number of steps or given number of steps after the last improving the objective value.
In the case of "many-to-many" distribution system design problem we can use the SUPRA algorithm. The objective value of the solved problem will be *(1)* in the form of a linear uncapacitated location problem. In this case the BBDual procedure can be used for obtaining of the value of solution. Initial parameters $\boldsymbol{p}^c$ will be parameters β_i for example, mentioned in [4].

5 GENETIC ALGORITHM
A genetic algorithm is a search technique used to find exact or approximate solutions to optimization and search problems. Genetic algorithm is categorized as metaheuristic.
Genetic algorithms are implemented as a computer simulation in which a population of abstract representations (called chromosomes) of candidate solutions (called individuals) to an optimization problem evolves toward better solutions. Traditionally, solutions are represented in binary as strings

of 0s and 1s. The evolution usually starts from a population of randomly generated individuals and happens in generations. In each generation, the fitness ratio of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness ratio), and modified (recombined and possibly randomly mutated) to form a new population. The new population is then used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached.

When we want to solve the many to many distribution system design problem using the genetic algorithm, we must describe chromosome and the fitness ratio. The chromosome **y** = $(y_1, y_2, \ldots, y_m)$ can be described as the vector of bivalent variables y_i **.** The fitness ratio can be the objective function of the original problem *(1)*.

The genetic algorithm in general can be described by these steps:

0. Initialization of initial population
1. Choose **parents**
2. **Crossover**
3. **Mutation**
4. If not enough members of the new , go to the step 1, else go to the step 5
5. **Selection**
6. If the point of finish is achieved, **finish,** elsewhere go to the step 1

6 NUMERICAL EXPERIMENTS

The set of test problems was created from real network and from real many-to-many problem, which was formulated in the frame of the project [5] in which railway infrastructure and yearly carriage flows were analyzed. This original problem, in which 53 possible locations and almost five hundred railway stations-customers were considered, was cut into sequence of smaller subproblems, in which 5, 9 or 15 possible locations were considered and where the number of considered railway stations ranged from ten to eighty by step of ten. The original fixed charges were adjusted for a particular subproblem proportionally to its total flow size to obtain non-trivial results. This way, three instances came into being for each problem size. The first one has the original fixed charges and the next two instances with modified fixed charges, what led to different solutions.

A goal of this investigation was to compare, how precise are these linearization methods in comparision wiht exact solutions and which of this method gives more precise upper bound of solution in this problem. Experiments were made in Delphi 7 enviroment.

In the Table 1 are described solutions for smaller subproblem consists of 5 possible locations of terminal and number of custumers 10 to 90 (step by 10). In the column "Optimal" are solutions of exact linearisation of the problem [3]. Tha value of gap is the difference between optimal solution value and value of appropriate method.

Table 1 *- 5 possible locations*

	Optimal	**SUPRA**		**Genetic alg.**	
Customers	*Object. value*	*Object. value*	***Gap (%)***	*Object. value*	***Gap (%)***
10	89363	89363	***0***	89363	***0***
20	233059	233059	***0***	233059	***0***
30	323475	323546	***0,021949***	323475	***0***
40	373653	373799	***0,039074***	373653	***0***
50	403270	403545	***0,068193***	403270	***0***
60	473211	473252	***0,008664***	473211	***0***
70	497436	497477	***0,008242***	497436	***0***
80	509545	509595	***0,009813***	509545	***0***
90	513898	513947	***0,009535***	513898	***0***

7 CONCLUSION

From these above mentioned experiments we can see, that genetic algorithm gives us solutions, which are equal to exact solution objective value. This equality of values can be reffected by the size of the problem and by the number of the possible terminals. In our case the number of possible locations is small and the genetic algorithm can pass over all possible solutions and choose the best one. When the number of possible locations of terminals will increase, the results will not be so exact and will depend on the set of parameters of the genetic algorithm (number of individuals in population and the number of population changes from the last change of the best solution).

Solution of the SUPRA method is near to the solution of exact linearization method (mentioned in [4]).

REFERENCES

[1] Buzna, Ľ.: *Návrh štruktúry distribučného systému pomocou spojitej aproximácie a diskrétneho programovania*, dizertačná práca, FRI ŽU, Žilina, 2003

[2] Gábrišová, L.: *Limits of the multiplier adjustment approach to capacitated location problem*, 2005 In: Komunikácie - vedecké listy Žilinskej univerzity - Roč. 7, č. 4 (2005), pp. 52-55

[3] Janáček, J., Koháni, M: *Exact Solving of the Many-To-Many Distribution System Design Problem.* Journal of Information, Control and Management Systems, FRI ŽU, Žilina, 2005

[4] Janáček, J., Koháni, M.: *The Algorithm Supra for Marshalling Yard System Optimization*, Eurnex-Žel 2007 conference, Žilina, June 2007

[5] Janáček a kol.: *Správa o riešení pilotného projektu úlohy 5.3 „Programové riešenie pre optimalizáciu riadenia prepravnej práce"*, Fakulta riadenia a informatiky ŽU, Žilina, 2001

[6] Koháni, M.: *"Using a genetic algorithm for „Many to Many" Distribution System Design Problem"*, zborník konferencie INFOTRANS 2007, Dopravní fakulta Jana Pernera, Univerzita Pardubice, Pardubice, september 2007, str. 163-174 (ISBN 978-80-7194-989-3)

[7] Koháni, M.: *Lower Bound for the "Many-to-Many" Distribution System Desing Problem*", Journal of Information, Control and Management Systems, Volume 4, No.2/1, Fakulta riadenia a informatiky, Žilinská univerzita, Žilina, 2006, p. 135-142

ACKNOWLEDGMENT

This work was supported by grant VEGA 1/3775/06.

Ing. Michal Koháni, PhD.
Department of Transportation Networks, Faculty of Management and Informatics,
University of Žilina, Univerzitná 1, 010 26 Žilina
kohani@frdsa.fri.utc.sk

MODELLING OF UNEMPLOYMENT FOR OKEČ (SUR MODEL)

Peter Komáromy, Department of Operations Research and Econometrics, University of Economics, Dolnozemská 1/b, 852 35 Bratislava, Slovakia

Abstract: This article deals with the problem of estimation of a model it consists from more than one equation. There are problems such an autocorrelation, non-significant parameters; random units do not have normal distribution, etc. We can solve all these problems with the particular tests, statistics and methods. I have applied this macroeconomic model and all mentioned tests to particular sectors of OKEČ and I have created multi-equation model for all sectors of OKEČ. This multi-equation model is named as a SUR model. SUR model is defined in the next passages of the article. For the estimation and all the tests and statistics I have used programme Eviews5.
Keywords: econometric, macroeconomy, OKEČ, SUR-model, estimation, statistics, hypothesis, autocorrelation, normal distribution, significance

1. INTRODUCTION

Models with seemingly unrelated regressions are models, where the structural matrix of endogenous variables is diagonal. This means, every equation contains only one endogenous variable – dependent variable. We have m linear equations, they can be written in this form:

$$y_1 = X_1\beta_1 + u_1, \qquad u_1 \sim N(0, \sigma^2_{u1}I_n)$$
$$y_2 = X_2\beta_2 + u_2, \qquad u_2 \sim N(0, \sigma^2_{u2}I_n)$$
$$\vdots$$
$$y_m = X_m\beta_m + u_m, \qquad u_m \sim N(0, \sigma^2_{um}I_n)$$

In every equation there is n observations. There is $k_i + 1$ of explanatory variables and this number can be different in every other equation. Variables y_i and u_i are n-dimensional vectors, X_i is $(n \times k_i + 1)$–dimensional matrix and β_i is $k_i + 1$-dimensional vector. It seems, this is not a system of equations, but only an ensemble of equations. Covariance between random units in two separate equations in the same time represents the only one connection between these two equations. Covariance between random units in two separate equations in different time equals 0. If there is not a correlation between various equations, we can use least square method to estimate the equation. If there is a correlation between various equations, we can not use a least square method, because estimation by this method would not be effective and that is why we use generalized least square method to estimate the equations. In this method residuals from least square method are used for estimation of a variance - covariance matrix Ω. Estimation in this method is more effective than estimation in least square method, because t-statistics shows a better values.

2. Modelling of unemployment for particular sectors of OKEČ

I have modelled the unemployment for 14 sectors of OKEČ. I have used some explanatory variables (real minimal wage, employee bonuses, labour productivity and number of hours worked). Variables in various sectors are marked with various indexes, for example unemployment in sector A. + B. is specified as NEZ1 or labour productivity in sector D. is specified as PP3. Variable RMM (real minimal wage) reaches the same values in every sector of OKEČ and that is why I did not use indexes by this variable. My main goal is modelling unemployment with explanatory variable RMM, so I tried to get variable RMM in every equation. This variable (also other 4 variables) can be used in normal form (in present time) or in delayed form (1, 2, 4, 8 observations delay). I have used also variables with 4 and 8 observations delayed and these variables could be important part of estimation, because I have worked with quarterly dates. The main importance of other explanatory variables is to improve equations and significance of RMM variable. In the next part of my work I have estimated all 14 equations by the least square method, I have tested parameters of every equation by t-statistics and p-value, I have tested autocorrelation by Breusch-Godfrey test, I have tested by Jarque-Bera test if random units has a normal distribution and I have also estimated parameters of equations by two-stage least square. In this method I have used all predetermined variables of system as the instruments. I have noticed p-value by all these tests and by the help of p-value I have rejected or accepted null hypothesis. I have applied these tests to all 14 equations. I did not test heteroskedasticity, because this problem is not by time series actual. In this article I will present only the statistics of (for example) 3.equation, because statistics of all 14 equations would take a place of many pages.

Sector D (industrial manufacture)

Dependent Variable: NEZ3
Method: Least Squares

Date: 01/04/08 Time: 12:23				
Sample (adjusted): 1996Q3 2006Q4				
Included observations: 42 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.470442	5.641072	-1.678837	0.1014
NEZ3(-1)	1.046148	0.042840	24.41998	0.0000
RMM(-1)	-0.009636	0.002170	-4.440357	0.0001
ODM3(-1)	0.001858	0.000483	3.842600	0.0004
R-squared	0.945610	Mean dependent var		80.24762
Adjusted R-squared	0.941316	S.D. dependent var		19.78052
S.E. of regression	4.791806	Akaike info criterion		6.062085
Sum squared resid	872.5335	Schwarz criterion		6.227577
Log likelihood	-123.3038	F-statistic		220.2172
Durbin-Watson stat	1.555107	Prob(F-statistic)		0.000000

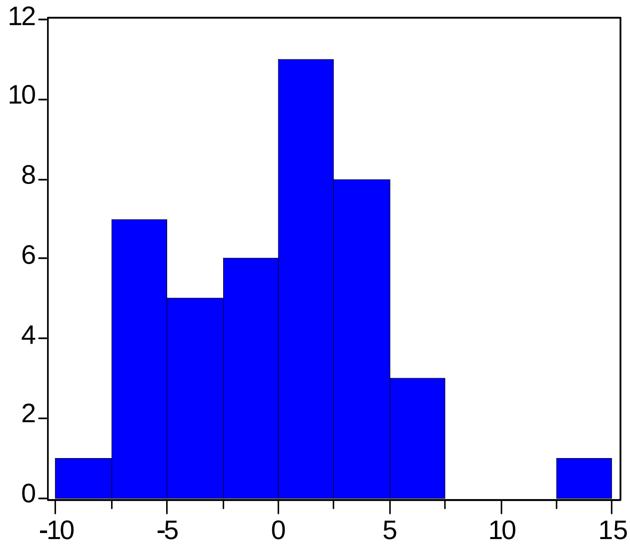
We can see unemployment as a dependent variable and unemployment, real minimal wage and employee bonuses as explanatory variables, but all explanatory variables are delayed on 1 observation. Probability by all these variables is lesser than value 0,05 and it means, I can reject the null hypothesis and all parameters by explanatory variables are significant. Only parameter β_0 is not significant, but this fact is not important. Coefficient R-squared reaches very good value (94,56%) and it means, mentioned variables represent 94,56% variability of 3. equation. By the help of F-statistics (probability by F-statistics) I can see, equation is significant, because probability is lesser than value 0,05. By the help of Durbin-Watson statistics I could test autocorrelation, but in this equation there are some delayed variables and that is the reason, why I can not use Durbin-Watson statistics to test this equation.

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	2.006441	Prob. F(4,34)	0.115721
Obs*R-squared	8.020844	Prob. Chi-Square(4)	0.090818
Test Equation:			
Dependent Variable: RESID			

Method: Least Squares
Date: 01/10/08 Time: 20:38
Sample: 1996Q3 2006Q4
Included observations: 42
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.729625	5.528990	0.493693	0.6247
NEZ3(-1)	-0.033649	0.044681	-0.753092	0.4566
RMM(-1)	0.000362	0.002132	0.169945	0.8661
ODM3(-1)	-6.31E-05	0.000476	-0.132454	0.8954
RESID(-1)	0.258891	0.166124	1.558418	0.1284
RESID(-2)	-0.125741	0.175383	-0.716948	0.4783
RESID(-3)	0.108717	0.175190	0.620567	0.5390
RESID(-4)	0.326499	0.173128	1.885883	0.0679

I have tested autocorrelation by the help of Breusch-Godfrey test. Probability by Chí-square is higher than value 0,05 and that is the reason why I do not reject null hypothesis and autocorrelation coefficient equals to 0. It means, in 3. equation there does not exist any autocorrelation. Another argument of missing autocorrelation is, probability by variables RESID (-1), RESID (-2), RESID (-3), RESID (-4) is higher than value 0,05.



Series: Residuals	
Sample 1996Q3 2006Q4	
Observations 42	
Mean	-1.21e-15
Median	0.086130
Maximum	13.70259
Minimum	-8.525111
Std. Dev.	4.613167
Skewness	0.416664
Kurtosis	3.292595
Jarque-Bera	1.365082
Probability	0.505331

Further I have tested, if random units do have a normal distribution. First I looked on this graph and I see random units could have a normal distribution. Second I looked on the table next to graph and I see probability by Jarque-Bera statistics reaches a value 0,505331 and I can not reject the null hypothesis and that is the reason why random units do have a normal distribution.

Dependent Variable: NEZ3
Method: Two-Stage Least Squares
Date: 03/27/08 Time: 17:08
Sample (adjusted): 1998Q2 2006Q4
Included observations: 35 after adjustments
Instrument list: NEZ1(-1) NEZ1(-4) RMM(-4) ODM1(-1) NEZ2(-1) RMM(-1) ODPHOD2(-2) NEZ3(-1) ODM3(-1) NEZ4(-1) PP4 NEZ5(-1) PP5 ODPHOD5 ODPHOD5(-4) NEZ6(-1) ODM6(-1) RMM PP7 ODPHOD7(-1) PP8(-4) PP9(-1) ODM9(-1) ODM9(-4) PP10(-1) PP10(-8) ODPHOD10(-2) ODPHOD10(-8) NEZ11(-4) PP11 ODM11 ODM11(-4) NEZ12(-1) ODM12(-1) PP12(-4) NEZ13(-1) NEZ13(-2) NEZ13(-8) ODM13(-1) ODPHOD13 ODM14(-1) ODM14(-2) ODPHOD14(-4)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.784578	7.550102	0.103916	0.9179
NEZ3(-1)	0.977562	0.053357	18.32109	0.0000
RMM(-1)	-0.009807	0.002223	-4.412560	0.0001
ODM3(-1)	0.001745	0.000504	3.463604	0.0016

R-squared	0.923165	Mean dependent var	86.01429
Adjusted R-squared	0.915729	S.D. dependent var	16.28612
S.E. of regression	4.727772	Sum squared resid	692.9067
Durbin-Watson stat	1.647102	Second-stage SSR	692.9067

Further I have estimated this equation by the help of 2-stage least square and by the estimation I have used all predetermined variables of a model as the instruments. We can see, all explanatory variables are significant also in this time and coefficient R-squared reaches value 92, 32% and that is well result.

3. MODEL OF MORE THAN ONE EQUATION

After estimating parameters of all 14 equations I have estimated all 14 equations whole as a system. I have done it by the help of 2-stage least square, 3-stage least square and SUR-method. Next I will present specification of all 14 equations:

nez1=c(10)+c(11)*nez1(-1)+c(12)*nez1(-4)+c(13)*rmm(-4)+c(14)*odm1(-1)
nez2=c(20)+c(21)*nez2(-1)+c(22)*rmm(-1)+c(23)*odphod2(-2)
nez3=c(30)+c(31)*nez3(-1)+c(32)*rmm(-1)+c(33)*odm3(-1)
nez4=c(40)+c(41)*nez4(-1)+c(42)*rmm+c(43)*rmm(-4)+c(44)*pp4
nez5=c(50)+c(51)*nez5(-1)+c(52)*rmm(-1)+c(53)*pp5+c(54)*odphod5+c(55)*odphod5(-4)
nez6=c(60)+c(61)*nez6(-1)+c(62)*rmm(-1)+c(63)*odm6(-1)
nez7=c(70)+c(71)*rmm+c(72)*rmm(-4)+c(73)*pp7+c(74)*odphod7(-1)
nez8=c(80)+c(81)*rmm+c(82)*rmm(-4)+c(83)*pp8(-4)
nez9=c(90)+c(91)*rmm(-1)+c(92)*pp9(-1)+c(93)*odm9(-1)+c(94)*odm9(-4)
nez10=c(100)+c(101)*rmm(-4)+c(102)*pp10(-1)+c(103)*pp10(-8)+c(104)*odphod10(- 2)+c(105)*odphod10(-8)
nez11=c(110)+c(111)*nez11(-4)+c(112)*rmm(-1)+c(113)*pp11+c(114)*odm11+c(115)*odm11 (-4)
nez12=c(120)+c(121)*nez12(-1)+c(122)*rmm(-1)+c(123)*odm12(-1)+c(124)*pp12(-4)
nez13=c(130)+c(131)*nez13(-1)+c(132)*nez13(-2)+c(133)*nez13(-8)+c(134)*rmm(-1)+c(135) *odm13(-1)+c(136)*odphod13
nez14=c(140)+c(141)*rmm(-4)+c(142)*odm14(-1)+c(143)*odm14(-2)+c(144)*odphod14(-4)

PP – labour productivity, NEZ – unemployment, ODM – employee bonuses, RMM – real minimal wage, ODPHOD – number of hours worked

All parameters in these 14 equations are significant, in all 14 equations does not exist autocorrelation and random units in all 14 equations do have a normal distribution. Further I will present estimation by 3-stage least square:

System: MAKRO
Estimation Method: Three-Stage Least Squares
Date: 01/10/08 Time: 21:00
Sample: 1998Q2 2006Q4
Included observations: 35
Total system (balanced) observations 490
Linear estimation after one-step weighting matrix

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	-9.911437	5.619579	-1.763733	0.0785
C(11)	0.377326	0.083093	4.541035	0.0000

C(12)	0.566527	0.077962	7.266744	0.0000
C(13)	-0.001457	0.000417	-3.495434	0.0005
C(14)	0.004267	0.000878	4.861998	0.0000
C(20)	6.485064	1.989547	3.259569	0.0012
C(21)	0.677823	0.073448	9.228549	0.0000
C(22)	-0.000619	0.000204	-3.040369	0.0025
C(23)	-0.000495	0.000180	-2.751540	0.0062
C(30)	9.241318	5.727068	1.613621	0.1074
C(31)	0.894486	0.039704	22.52898	0.0000
C(32)	-0.009773	0.001596	-6.123104	0.0000
C(33)	0.001690	0.000352	4.798483	0.0000
C(40)	1.043183	0.379313	2.750194	0.0062
C(41)	0.748216	0.076526	9.777234	0.0000
C(42)	0.000816	0.000312	2.617200	0.0092
C(43)	-0.001215	0.000327	-3.712476	0.0002
C(44)	4.50E-06	8.85E-07	5.088521	0.0000
C(50)	88.75839	7.216722	12.29899	0.0000
C(51)	0.394334	0.055109	7.155564	0.0000
C(52)	-0.002980	0.000499	-5.978120	0.0000
C(53)	-8.07E-05	2.07E-05	-3.897753	0.0001
C(54)	-0.000411	8.89E-05	-4.619254	0.0000
C(55)	-0.000272	9.05E-05	-3.005941	0.0028
C(60)	6.502857	2.558266	2.541901	0.0114
C(61)	0.913845	0.046525	19.64192	0.0000
C(62)	-0.003004	0.000762	-3.942095	0.0001
C(63)	0.000723	0.000245	2.943602	0.0034
C(70)	35.56388	4.337165	8.199799	0.0000
C(71)	0.003708	0.001195	3.103689	0.0020
C(72)	-0.004135	0.001234	-3.351932	0.0009
C(73)	-0.000141	2.59E-05	-5.425331	0.0000
C(74)	-0.000490	0.000150	-3.256983	0.0012
C(80)	16.26305	2.008743	8.096133	0.0000
C(81)	0.004679	0.001027	4.554316	0.0000
C(82)	-0.004755	0.001057	-4.496623	0.0000
C(83)	-4.98E-05	1.36E-05	-3.652417	0.0003
C(90)	1.960997	0.347607	5.641418	0.0000
C(91)	0.001006	0.000167	6.023628	0.0000
C(92)	-5.43E-07	1.16E-06	-0.468578	0.6396
C(93)	-0.000715	0.000155	-4.601702	0.0000
C(94)	-0.000489	0.000133	-3.667107	0.0003
C(100)	1.534691	3.779240	0.406085	0.6849

C(101)	-0.000898	0.000214	-4.193854	0.0000
C(102)	-5.00E-05	7.79E-06	-6.411943	0.0000
C(103)	7.08E-05	6.39E-06	11.08124	0.0000
C(104)	-0.000177	2.93E-05	-6.041643	0.0000
C(105)	0.000270	4.40E-05	6.137487	0.0000
C(110)	18.83897	2.129077	8.848419	0.0000
C(111)	0.319871	0.093277	3.429263	0.0007
C(112)	-0.002538	0.000398	-6.371102	0.0000
C(113)	-5.08E-05	8.43E-06	-6.023728	0.0000
C(114)	-0.001554	0.000274	-5.662941	0.0000
C(115)	0.002423	0.000347	6.992542	0.0000
C(120)	7.936789	1.450027	5.473546	0.0000
C(121)	0.580658	0.086029	6.749533	0.0000
C(122)	0.001364	0.000287	4.748919	0.0000
C(123)	-0.000779	0.000181	-4.296871	0.0000
C(124)	-0.000124	2.85E-05	-4.361604	0.0000
C(130)	-7.849685	3.721342	-2.109369	0.0355
C(131)	0.965519	0.090158	10.70924	0.0000
C(132)	-0.437114	0.094882	-4.606925	0.0000
C(133)	-0.226064	0.086984	-2.598911	0.0097
C(134)	0.000802	0.000347	2.312037	0.0213
C(135)	0.000690	0.000206	3.354530	0.0009
C(136)	0.000120	4.98E-05	2.406773	0.0165
C(140)	-93.00881	13.03593	-7.134803	0.0000
C(141)	-0.005964	0.001956	-3.049451	0.0024
C(142)	0.014445	0.002279	6.338684	0.0000
C(143)	0.010003	0.002152	4.649155	0.0000
C(144)	0.001774	0.000226	7.835571	0.0000

We can see, only the parameter C (92) is not significant (except the β_0 constant).

4. Conclusion

Econometrical research is very difficult, because Slovak economy do not have such as parameters as do have economies in the Western Europe and that is the reason, why many econometrical tests and statistics applied to the data of Slovak economy, do not have very good results. I tried to model sectors of OKEČ and I spent many hours, as long as I find equations, they are suitable for the tests I applied. But there are many other tests and statistics I did not apply, for example test of stationarity, multi-equation tests as the test of serial correlation, test of bypassing important predetermined variable and so on.

REFERENCES

1. Husár, J. – Lukáčik, M.: Aplikovaná ekonometria. Bratislava: Ekonóm, 2004

Ing. Peter Komáromy
Department of Operations Research and Econometrics
University of Economics Bratislava
Dolnozemská 1/b, 852 35 Bratislava, Slovakia
Tel.: +421 904 311 135
Email: p.komaromy@gmail.com

UNEMPLOYMENT AND ITS DEPENDENCE WITH INFLATION

Brian König

Abstract
The contribution deals with unemployment and its dependence with inflation rate through the Phillips Curve. The first part of article is devoted to a short description of unemployment and inflation development. There is also mentioned the expression NAIRU. The second part is aimed at initial Phillips Curve which expresses an inverse and non-linear relation between inflation and unemployment. The last part of the article is focused on Expectations-augmented Phillips Curve which takes into account the expectations of particular agents.

Keywords
Unemployment, inflation, NAIRU, Philipse Curve, Expectations-augmented Phillips Curve.

1. Introduction

If we want to measure the efficiency of the economy as a whole, it is necessary to know the indicators which mainly influence the economy. Among the crucial indicators which describe the development of market economy belongs also the unemployment and the inflation rate. High dependency between both of this indicators is the main subject of this contribution. One of the most famous theories which searches the relation between unemployment and inflation rate is the Phillips Curve.

2. Uneployment and its dependence with inflation rate

After the Second World War the employment increase was one of the main target of the economic policy for most nations. At that time, it was found that some actions of the government focused on decreasing unemployment were followed by an increase of the price level. It was discovered that in the market economy full employment isn´t maintainable without the increase of the inflation rate in a long term. At first, full employment had to fulfill two conditions:

- everybody who wants to work must have a job (there is no space for involuntary unemployment, there are possible frictional and structural unemployment only),

- inflation rate is regulated, there is no accelerating or deccelerating process.

Scholar often refer to the term NAIRU. The acronym comes from the english expression „*non accelerating inflation rate of unemployment* “. This NAIRU expresses the rate of unemployment at which inflationary pressures are stable. Therefore NAIRU is connected with natural rate of unemployment but not necessarrily the same. We can describe it by the relation:

$$\pi = \pi_e - \varepsilon \,.\, (u - u^*)$$

where: u^* - natural rate of unemplyoment (or NAIRU in this case),
u - actual rate of unemployment,
π_e - expected inflation,
π - actual inflation,
ε - paramater of responsiveness between inflation and unemployment rate.

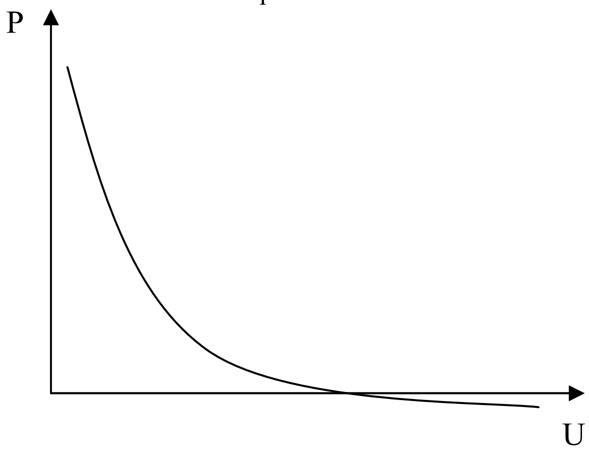
- If $u < u^*$ for a longer period, inflationary expectations rise and inflation tends to accelerate,
- if $u > u^*$ for a longer period, inflationary expectations fall and inflation tends to decelerate,
- if $u = u^*$, inflation tends to stay stable, unless there is an exogenous shock.

Relation described above is explained by Phillips Curve.

3. The Phillips Curve

The first theoretical scope was associated with A. W. Phillips. The natural form of Phillips Curve shown a cross dependency between unemployment and changes in the wage level. Afterward economist P. A. Samuelson and R. M. Solow replaced changes in the wage level by the inflation rate and Phillips Curve was understood as a relationship between the inflation and the unemployment. However this assumption was conditioned by close dependency between the inflation and changes in the wage level. Phillips noticed that there was inverse and non-linear relationship beetwen this two indicators. Inflation and unemployment represent a phenomena between which we can choose: either low unemployment near by higher inflation or lower price level at interest high uneployment. But you can´t lower both.

Picture nr.1 – Phillips Curve



where: P – inflation rate,
U– unemployment.

On the Picture nr. 1 we can see that the curve is sloped down from left to right and its left side is steeper. We can give a simple explanation. An increase in labour demand leads to a faster growth in nominal wages. This leads to an increase price level. On the contrary the right side of the curve is flater bacause an increase in unemployment leads to a lower change in nominal wages. The conjuction with the abscissa axis is equivalent to the natural rate of unemplyoment.

Evolution approved that the mentioned relation between unemployment and inflation stands only in limited conditions. Uncertainty about the stability of this relation was occurred in 70th years of the last century when most of the market economies experienced the high inflation together with the high unemployment. There was shown that inflation and unemployment can exist next to each other. This fenomen called as the stagflation Phillips Curve couldn´t express.

Stagflation is the situation in which the economy stagnates and the inflation grows. Stagflation happens if the national product decreases near by increasing inflation. It usually happens when the economy can´t get from the stage after the recesion, unemployment is on high level and even though inflation increases.

Existence of high unemployment and inflation has made serious problems in almost every market economy. Famous economist made effort in order to be able to express this unacceptable situation. One of the most convincing explanations was given by a member of the monetarism Milton

Friedman. He created an analogical Phillips Curve which is called Expectations-augmented Phillips Curve.

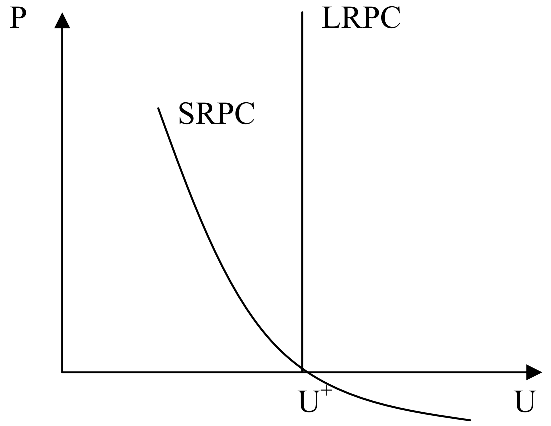
4. Expectations-augmented Phillips Curve

Friedman emphasized the aspect of amazement due to the difference between real and anticipated (expected) values. He demonstrated the existence of a natural rate of unemployment which can be achieved only in case when the expectations of economic subjetcs (particularly expectations regarding development of inflantion) will be realized. He said that unanticipated changes in inflation cause systematic errors in the employees and employer´s behaviour. It might cause temporal increasing of unemployment under the natural level of unemployment. However Friedman considered this deviation as temporal until it is eliminated by adaptation to expectations. He took natural unemployment as a rate generated by forces of the market without any interventions of fiscal or monetary policy (it means without any shocks which could cause unexpected inflation).

Friedman claimed that the market is in equilibrium by the natural level of unemployment, i.e. there is no involuntary unemployment. If unemployment diverts from its natural level, forces of the market adjust unemployment at its previous level.

Monetarists especially emphasized the understanding of the unexpected changes. Unanticipated inflation is different in its consequences. Most of the economic subjects include their expectation regarding future development of prices into their economic decisions (for example wage decisions) in order to keep the previous real income. If there is an unexpected change in inflation, for example caused by a change in the economic policy of government, existing long-term agreements and liabilities don't allow particular agents instantly adjust to an unexpected change in inflation. Adaptation to the mentioned change will be realized with some time delay.

Picture nr.2 – Expectations-augmented Phillips Curve



where: LRPC – long-run Phillips Curve,
SRPC – short-run Phillips Curve,
U^+ - natural unemployment,
P – inflation rate,
U – unemployment.

The main point in Friedman´s theory of Phillips Curve is the segmentation of money demand on expected and unexpected. If there is an unexpected increase in money demand then entrepreneurs who follow the expectation of price growth in their production are able to pay higher nominal wages than before in order to attract other employees and increase their own production. On the other side, employees are not able to recognize the difference between growth in nominal and real wages and they will increase labour supply. It will cause temporal decreasing of unemployment. However that is only a short-run state. In the next period employees will recognize that growth in nominal wages doesn´t mean increase of real wages, wrong expectations will be eliminated and unemployment returns to the previous level of unemployment.

Friedman said that inadequate expectations are the root cause of the deviation of the economic system from equilibrium. However economic subjects can be wrong only in short-run, after that they will revise their expectations and equilibrium will retrieve to the previous level.

Friedman distinguish short-run and long-run Phillips Curve. The initial Phillips Curve is actual only in short-run period when inflationary expectations deviate from real higher inflation and are gradually adjusting to the real level of inflation.

In a long-run period Phillips Curve has a vertical shape and equal to the line on the natural level of unemployment (see picture nr.2).

In the monetarists view of the inflation has a marked role money supply. They suppose that in a short-run period changes in the money supply can influence real macroeconomic indicators (such as employment, production, investment...) but this influence is just temporary. In a long-run period these indicators return to their previous levels, but inflation rises. In a long-run term period the money supply is considered as neutral.

References

[1] Husár, J.: Aplikovaná makroekonómia. Bratislava: Sprint, 2003

[2] Husár J.- Szomolányi K.: Modelovanie ekonomických procesov 1. Bratislava: Ekonóm, 2005

[3] Martincová M.: Nezamestnanosť ako makroekonomický problém. Bratislava: IURA EDITION, 2002

[4] Samuelson P. A. – Nordhaus W. D.:Ekonómia. Bratislava: Elita, 2000

[5] Táncošová, J.: Inflácia. Bratislava: IURA EDITION, 2000

[6] http://www.bized.co.uk/virtual/bank/economics/mpol/inflation/

THE SHORTEST PATH PROBLEM (MULTIPLE EDGES BETWEEN TWO ADJACENT VERTICES)

Miroslav Krumplík, Tomáš Domonkos

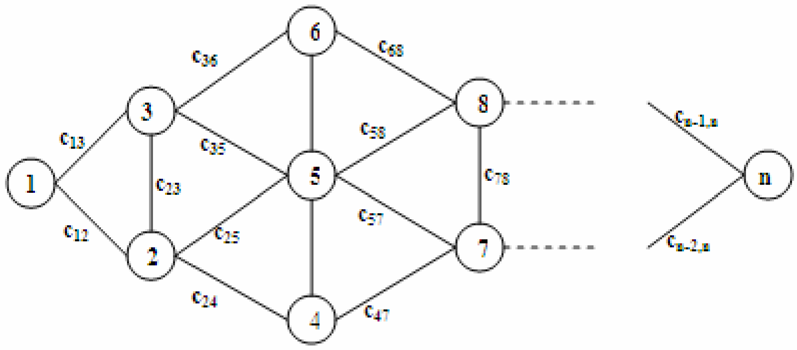
Abstract: The shortest path problem is often solved by companies which would like to transport their goods from one place to another (mainly export/import companies). The companies order transport service from cargo companies (specialized on transport). For export/import companies is very important to find the shortest way between two places but they need to know which cargo company to use on the specific way as there are a lot of companies ordering their services on one way. This page describes alternate approaches of determining the shortest path between two vertices in case there are multiple edges between two adjacent vertices.

1. INTRODUCTION

This page describes alternate approaches of determining the shortest path between two vertices in the planar graph with positive edge weights. It deals with condition that between two vertices exists more than one way (more than one edges). In real life we can find such situation when we would like to transport the commodity between two places and we can choose Cargo Company which will transport our commodity. In this case, the result will be not only the best (shortest) way between two places but also which company will be used for specific way (edge).

In this paper we will work with following conditions:

- We have the non-directed planar graph, where vertices represent cities (places) and edges represent possible ways between these cities. We have *n* vertices.
- Weights of the edges represent transportation costs between two cities for one unit of goods.
- More than one Cargo Company offers its transport service between two places.
- Set c_{ij}^k as costs for transport service between vertex *i* and vertex *j* offering by company *k*. If there is no way between vertex *i* and *j* for company *k* set c_{ij}^k = M (enough high number).



Picture No. 1

- For finding the shortest path between two vertices in the graph we will use following mathematical model (let's say we would like to find shortest path between vertex *one* and vertex *n*):

$$\textbf{MIN} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} c_{ij} \tag{1.1}$$

$$\sum_{j=1}^{n} x_{1j} = 1 \tag{1.2}$$

$$\sum_{j=1}^{n} x_{jn} = 1 \tag{1.3}$$

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} x_{ji} \quad \textit{for } i = 2,3,..,n-1 \,, \tag{1.4}$$

$$x_{ij} \geq 0 \textit{ for } i,j = 1,2,..,n$$

where x_{ij} represent flow between vertices i and j. Restriction (**1.2**) says vertex *1* is the source of flow, flow is one unit. Restriction **(1.3)** says vertex n is the final vertex. Restriction **(1.4)** means that what flows in into the vertex has to flow out from the vertex.

2. POSSIBILITY OF SWITCHING CARGO COMPANY IN EACH VERTEX

In this caption we would need to include one more restriction to the restrictions above:

- In each city (vertex) we can choose which company will transport our goods to the next city (vertex). We have m companies offering their transport service.

If there is no restriction for switching Cargo Company in each vertex, we would need to find the lowest cost ways between each vertices and use only these ways for finding the shortest path in the graph. For example, if *company A* offers transport for $5 between vertex *1* and *2*, *company B* offers transport for $8, it is logical that we will not use *company B* in case that the shortest path lies on the edge *(1,2)*.

In order to find minimal cost for transport of our goods between city *1* and city *n*, we would need to modify our model. Set c_{ij} as the best price for transport from city *i* to city *j* for one unit of goods:

$$c_{ij} = \min_k c_{ij}^k \quad \text{for } i,j = 1,2,..,n \text{ and } k = 1, 2, ..,m,$$

then our mathematical model for finding shortest path will be as follows:

$$\textbf{MIN} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} c_{ij} \qquad \textbf{(2.1)}$$

$$\sum_{j=1}^{n} x_{1j} = 1 \qquad \textbf{(2.1)}$$

$$\sum_{j=1}^{n} x_{jn} = 1 \qquad \textbf{(2.3)}$$

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} x_{ji} \quad \text{for } i = 2,3,..,n-1, \qquad \textbf{(2.4)}$$

$$c_{ij} = \min_k c_{ij}^k \quad \text{for } i,j = 1,2,..,n \text{ and } k = 1, 2, ..,m, \qquad \textbf{(2.5)}$$

$$x_{ij} \geq 0 \text{ for } i,j = 1,2,..,n$$

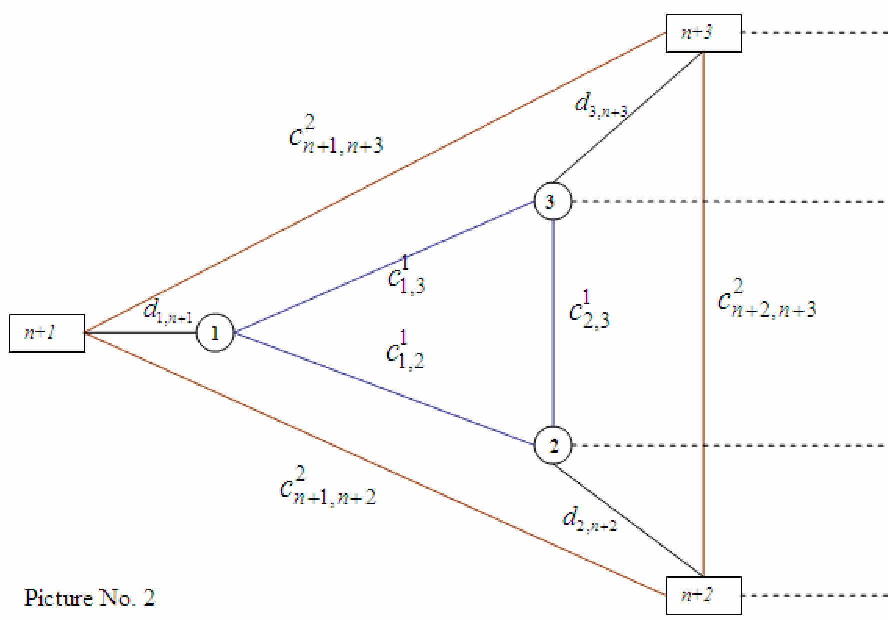
When we find the solution, we need to remember (for interpretation) which company we choose to be used on each way.

3. POSSIBILITY OF SWITCHING CARGO COMPANY IN EACH VERTEX WITH COSTS OF TRANSLOADING

We still work with condition that in each city we can choose which company we will use for transport our goods to the next city. When we change company in the city, we have to pay costs of transloading our goods from one company to another.

- Set d_{ii} as costs for transloading the goods from one company to another in the city *i*. Let's say that it doesn't matter (costs are the same) if we transload our goods from company A to company B or from company B to company A.
- From now we will have just <u>two companies</u> offering their transport service.

- Each city (vertex) will be divided by two, so we will have (n + n) cities instead of n.



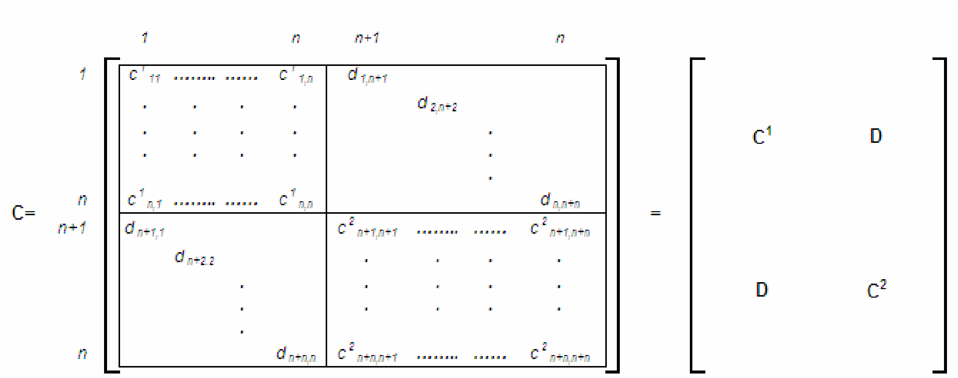
Picture No. 2

We have to bear in mind that city *one* and city (n + *1*) is the same city, weight of edge between these "two" cities represent costs of transloading goods from one company to another. Blue color edge's weight represents costs for transport service between vertex i and vertex j offering by first company (c^1_{ij}). Red color edge's weight represents costs for transport service between vertex i and vertex j offering by second company (c^2_{ij}).

When we divide one city into two, we would need to modify indexes for transloading costs.

- Set $d_{i,i+n}$ as costs for transloading the goods from one company to another in the city i. Let's say that it doesn't matter (costs are the same) if we transload our goods from company A to company B or from company B to company A, so $d_{i,i+n}= d_{i+n,i}$.

After these modifications matrix **C**=$\{c_{ij}\}$ consists of these individual matrixes:



where $\mathbf{C^1} = \{c^1_{ij}\}$ - costs for transport service between vertex i and vertex j offering by first company.
$\mathbf{C^2} = \{c^2_{ij}\}$ - costs for transport service between vertex i and vertex j offering by second company.
$\mathbf{D} = \{d_{ii}\}$ - costs for transloading the goods from one company to another in the city i. $d_{ij} = M$ for $i \neq j$.
Now we can start creating the Model. We will use Matrix **C** which consists of matrixes $\mathbf{C^1}$, $\mathbf{C^2}$ and **D.** Objective function will be the same as in the model mentioned above:

$$\textbf{MIN} = \sum_{i=1}^{n+n} \sum_{j=1}^{n+n} x_{ij} c_{ij} \qquad \textbf{(3.1)}$$

From the city 1 we can use first or second company to the next city. Let's set that vertex 1 is the source of the flow and let's set $d_{1,j} = 0$ *for* $j = 1,2,..,n$ (same as $c_{1j} = 0$ for $j = n+1, n+2,..n+n$) which allow us to use also second company from vertex $n+1$ in the beginning of transport:

$$\sum_{j=1}^{n+n} x_{1j} = 1 \qquad \textbf{(3.2)}$$

Now we have to work with situation that one of the city n or city $n+n$ is the final destination of the flow in the model (vertices n and $n+n$ are the same place, where vertex n is the final destination when we use first company for transport on the last way and vertex $n+n$ is the final destination when we use second company for transport on the last way. The destination of the flow has to be one of these two vertices, so we can write:

$$\sum_{j=1}^{n+n} x_{j,n} + \sum_{j=1}^{n+n} x_{j,n+n} = 1 \qquad \textbf{(3.3)}$$

What flows in into the vertex has to flow out from the vertex:

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} x_{ji} \quad \textit{for } i = 2,3,..,n\text{-}1,n+1,\ldots,n+n\text{-}1 \qquad \textbf{(3.4)}$$

$$x_{ij} \geq 0 \textit{ for } i,j = 1,2,..,n$$

When we have solution, we need for better interpretation deploy matrix **X** to the parts in the same way how we deployed Matrix **C**.

4. CONCLUSION

In this page we have shown alternate approaches for modeling the shortest path problem with multiple edges between two vertices. This situation is closed to reality, especially when between two places more than one company offers its transport service. Then the problem is not only to find the shortest path between two vertices but also which company will transport our goods on specific way. We have to take into consideration costs of transloading when switching the companies.

Literature:

[1] IVANIČOVÁ, Z. - BREZINA, I. - PEKÁR, J.: *Operačný výskum*, Bratislava: Edícia ekonómia, 2002

[2] PELIKÁN J.: *Diskrétní modely,* Vysoká škola ekonomická v Praze, Praha, 1999

Ing Miroslav Krumplik
Department of Operations Research and Econometrics
University of Economics Bratislava
Dolnozemská 1, 852 35, Bratislava, Slovakia
e-mail: miroslav.krumplik@email.cz

Ing Tomáš Domonkoš
Department of Operations Research and Econometrics
University of Economics Bratislava
Dolnozemská 1, 852 35, Bratislava, Slovakia
e-mail: tdomonkos@gmail.com

SENSITIVITY OF RESULTS TO STRUCTURAL PARAMETERS FOR FCNFP PROBLEMS

Peter Lošonský, University of Žilina, Faculty of Management Science and Informatics

Abstract
The fixed charge network flow problem (FCNFP), which is known to be NP-hard, is an extension of the classical network flow problem (NFP) in which a fixed cost is incurred, independent of the amount transported, along with a variable cost that is proportional to the amount shipped. FCNFP can be solved by enhanced dynamic slope scaling procedure (EDSSP) with tabu scheme which was developed in the past. This heuristic will be evaluated according to their efficiency for solving FCNFP problems with different combination of the structural parameters.

Keywords: fixed-charge, tabu search, heuristic algorithm, adaptive parameter tuning

1 Introduction

The fixed charged network flow problem (FCNFP) is one of the combinatorial problems in the minimum cost network flow problems. Besides the variable cost on each arc in the network, a fixed setup cost occurs whenever there is a non-zero flow on an arc. The fixed setup costs are very common in practical applications and may represent handling fees, changeover times, charter rental and docking fees, etc. The basic decision “to go or not to go” can be modeled by imposing fixed charges on the arcs of a network. FCNFP is known as NP-hard and usually formulated as a mixed-integer programming (MIP) model.
An enhanced dynamic slope scaling procedure (EDSSP) is used for solving FCNFP problems. This heuristic will be evaluated according to its efficiency for solving FCNFP problems with different combination of structural parameters estimated by SUPRA adaptive method.

2 The fixed charge network flow problem

The fixed charge network flow problem can be formulated as a minimum-cost network flow problem with additional constraints on the choice of the arcs. Let $D=(V,H)$ represent a directed network, where H is the set of all edges and V is the set of all nodes (suppliers, customers, transshipment points) and b_i represent their demands. Let f_{ij} represent the fixed cost of activating the route between nodes i and j (represented by decision variable y_{ij}), and c_{ij} represent the cost per unit flow over the edge $(i,j) \in H$. Flow variables are represented by x_{ij}. Let u_{ij} represent upper bound (capacity) for edge $(i,j) \in H$. The objective is to choose edges which are to be opened and the size of the shipment on those edges, so that the total cost of meeting demand, given the supply constraints, is minimized.

$$\min \sum_{(i,j)\in H} (c_{ij}.x_{ij} + f_{ij}.y_{ij}) \tag{1}$$

st

$$\sum_{j\in V_i^-} x_{ji} - \sum_{j\in V_i^+} x_{ij} = b_i \qquad \forall i \in V \tag{2}$$

$$x_{ij} \le u_{ij}.y_{ij} \qquad \forall (i,j) \in H \tag{3}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in H \tag{4}$$

$$x_{ij} \ge 0 \qquad \forall (i,j) \in H \tag{5}$$

As additional notation is used $V_S = \{i \in V : b_i < 0\}$ to denote the set of supply nodes, $V_D = \{i \in V : b_i > 0\}$ to denote the set of demand nodes (customers), $V_0 = \{i \in V : b_i = 0\}$ to denote the set of transshipment nodes, $V_i^+ = \{j \in V : (i,j) \in H\}$ and $V_i^- = \{j \in V : (j,i) \in H\}$. Another Constraints (2) stand for classical flow conservation at nodes, while constraints (3) guarantee that y_{ij} takes value 1 whenever x_{ij} is positive.

Despite its similarity to a standard NFP problem, FCNFP is significantly harder to solve because of the discontinuity in the objective function Z introduced by the fixed costs.

3 Enhanced dynamic slope scaling procedure

The dynamic slope scaling procedure (DSSP) is developed by Kim and Pardalos [2]. The computational merit of this approach is to approximate a solution for FCNFP by solving successive LP problems with recursively updated coefficients of the objective function. At each iteration, the linear factor is adjusted to reflect both the current variable cost and the fixed cost. Initially, when there is no flow through the arcs, a non-linear cost of an arc is transformed into a linear factor given as:

$$\bar{c}_{ij}^0 = c_{ij} + \frac{f_{ij}}{u_{ij}} \qquad \forall i,j \tag{6}$$

The objective function of FCNFP is then approximated as a linearized NFP problem:

$$\min \sum_{(i,j)\in H} \bar{c}_{ij}.x_{ij} \tag{7}$$

st

$$\sum_{j\in V_i^-} x_{ji} - \sum_{j\in V_i^+} x_{ij} = b_i \qquad \forall i \in V \tag{8}$$

$$x_{ij} \le u_{ij} \qquad \forall (i,j) \in H \tag{9}$$

$$y_{ij} \in \{0,1\} \qquad \forall (i,j) \in H \tag{10}$$

$$x_{ij} \ge 0 \qquad \forall (i,j) \in H \tag{11}$$

Note that the constraints of the FCNFP are not changed in this linear approximation. At each iteration, a new linear approximation is solved with adjusted linear factors. Unfortunately DSSP stops at a local optimum which can be still far away from the exact solution. In order to improve the performance of DSSP, was proposed a metaheuristic that combines DSSP with the tabu scheme to force DSSP to continue the additional search when DSSP fails to make any progress. This metaheuristic is called Enhanced dynamic slope scaling procedure (EDSSP) and its detailed description of all iteration steps and used variables can be found in [1].

We may focus only on those variables where SUPRA tries to find good setting of their values:

- *divno* – size of set for selecting unvisited arcs during diversification process
- *Inten_Iter* – maximum number of iterations of the intensification process (IP)
- *Diver_Iter* – maximum number of iterations of the diversification process (DP)
- *tb_max_sh* – maximum number of iterations for an arc that must be tabu in the IP
- *tm_max_sh* - maximum number of iterations for an arc that must not be tabu in the IP
- *tb_max_lm* - maximum number of iterations for an arc that must be tabu in the DP
- *tm_max_lm* - maximum number of iterations for an arc that must not be tabu in the DP

4 SUPRA heuristic algorithm

The SUPRA algorithm is a heuristic method based on the gradient method. This algorithm gives us a good combination for the set of parameters $\mathbf{p} \in P$, which optimize the objective function of given problem. This combination of parameters is the outcome of the solving and optimizing process. Solving process can be written as:

Minimize $F(\mathbf{p})$

Subject to $\mathbf{p} \in P$

Input to this algorithm is the solved problem $\boldsymbol{a}$ (written in the form of mathematical model) and the set of initial parameters $\boldsymbol{p}^{\boldsymbol{c}}$.

The basic step of this algorithm tries to search the best solution in the neighbourhood of actual solution by combination of exploratory steps and a local search in the direction of the statistical gradient. This new solution, a set of parameters $\boldsymbol{p}^{\boldsymbol{n}}$ with appropriate objective value $F(\boldsymbol{p}^{\boldsymbol{n}})$ is given to the algorithm for next processing.

During the first phase – generating of exploratory steps (Figure 1) are randomly generated vectors $\boldsymbol{r}^j$ $(j = 1..s)$ to create s new points $\boldsymbol{p}^{\boldsymbol{kj}}$ $(j = 1..s)$.

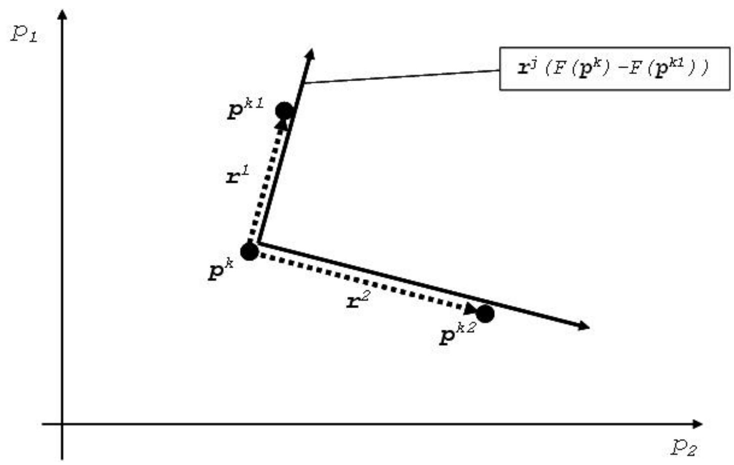


Figure 1 – SUPRA – Phase 1

1. Generate random vector $\boldsymbol{r}^j = \boldsymbol{w} + \boldsymbol{x}$ and create point $\boldsymbol{p}^{kj} = R(\boldsymbol{p}^k + \boldsymbol{r}^j)$.
2. If $F(\boldsymbol{p}^{kj}) < F(\boldsymbol{p}_{max})$ then update best found solution $\boldsymbol{p}_{max} := \boldsymbol{p}^{kj}$.
3. Update values of $\boldsymbol{r}$ (statistical gradient) and $\boldsymbol{w}$ (parameter of learning).

Local search is used during the second phase to explore the neighbourhood of current set of parameters $\boldsymbol{p}^{\boldsymbol{k}}$ in the direction of statistical gradient modified in the first phase. Statistical gradient $\boldsymbol{r}$ is normalized for the basic length of the step α. Next computation is performed with decreasing distance between the step α and starting point until maximum amount of attempts for improving the solution is achieved (Figure 2).

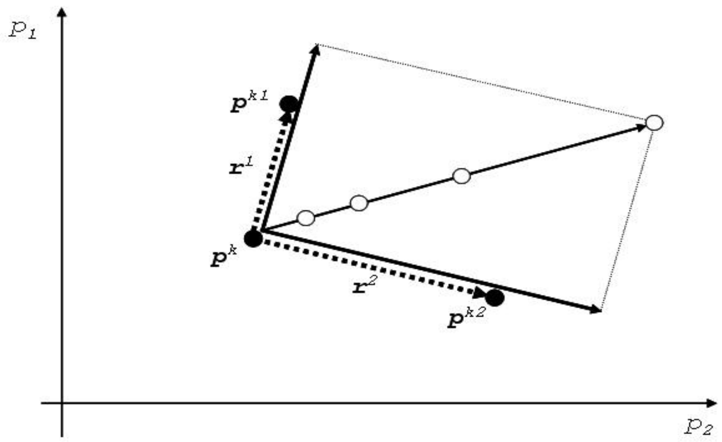


Figure 2 – SUPRA – Phase 2

The SUPRA algorithm finishes the search process after achieving a given number of steps or given number of steps after the last improving the objective value.

Detailed description of all iteration steps and used variables of this adaptive metaheuristic method can be found in [3].

Computational experiments

Five fixed charge network flow problems were generated for computation experiments with 10 suppliers, 40 transshipment points and 13 customers. The variable costs were chosen from interval of discrete values ranging from 3 to 8, fixed costs from 800 to 3200 and upper limits for edges from 400 - 600. Total supply is 3 000 units.

Different combinations of EDSSP parameters were used in computation in order to obtain better results. Parameters were allowed to vary in the following ranges: *divno* ∈ <20,40>, *Inten_Iter* ∈ <1,10>, *Diver_Iter* ∈ <20,30>, *tb_max_sh* ∈ <1,9>, *tm_max_sh* ∈ <1,9>, *tb_max_lm* ∈ <1,9>, *tm_max_lm* ∈ <1,9>.

The hardware platform is Intel Pentium CPU 3.00 GHz (2 CPUs), 1024 MB RAM running on Windows XP operating system.

Exact objective values of generated test problems were obtained by XPRESS-Mosel optimization software to compute the gaps by using following formulas

$$Gap = Objective\_Value - Objective\_Value\_Exact \quad (6)$$

$$Gap = \frac{Objective\_Value - Objective\_Value\_Exact}{Objective\_Value\_Exact} * 100\% \quad (7)$$

Table 1 shows best combination of EDSSP parameters estimated by SUPRA.

	Exact value	Objective value	Gap	Parameters
FCNFP_1	64828	69267	4439 (6,85 %)	Divno=27, Inten_Iter=5, Diver_Iter=20, tb_max_sh=1, tm_max_sh=1, tb_max_lm=1, tm_max_lm=1
FCNFP_2	55795	55886	91 (0,16 %)	Divno=22, Inten_Iter=1, Diver_Iter=23, tb_max_sh=6, tm_max_sh=9, tb_max_lm=1, tm_max_lm=9
FCNFP_3	68568	71363	2795 (4,08 %)	Divno=38, Inten_Iter=10, Diver_Iter=30, tb_max_sh=1, tm_max_sh=5, tb_max_lm=1, tm_max_lm=6
FCNFP_4	66109	68675	2566 (3,88 %)	Divno=27, Inten_Iter=1, Diver_Iter=20, tb_max_sh=1, tm_max_sh=9, tb_max_lm=9, tm_max_lm=3
FCNFP_5	56292	58892	2600 (4,62 %)	Divno=36, Inten_Iter=2, Diver_Iter=30, tb_max_sh=1, tm_max_sh=6, tb_max_lm=3, tm_max_lm=9
		Avg.	***2498 (3,92 %)***	

Table 1: Best combination of EDSSP parameters estimated by SUPRA

Some of the promising values of EDSSP parameters were tested for all test problems according to obtained results from Table 1. Results can be seen in Tables 2 – 5.

	Exact value	**Objective value**	**Gap**
FCNFP_1	64828	69655	4827 (7,45 %)
FCNFP_2	55795	57595	1800 (3,23 %)
FCNFP_3	68568	73039	4471 (6,52 %)
FCNFP_4	66109	69613	3504 (5,30 %)
FCNFP_5	56292	64010	7718 (13,71 %)

	Exact value	**Objective value**	**Gap**
FCNFP_1	64828	69655	4827 (7,45 %)
FCNFP_2	55795	56892	1097 (1,97 %)
FCNFP_3	68568	74585	6017 (8,78 %)
FCNFP_4	66109	71665	5556 (8,40 %)
FCNFP_5	56292	61219	4927 (8,75 %)
		Avg.	***4484,8 (7,07***

			Avg.	***4464 (7,24 %)***

Table 2: EDSSP results for Divno=20, Inten_Iter=1, Diver_Iter=30, tb_max_sh=1, tm_max_sh=1, tb_max_lm=1, tm_max_lm=1

				%)

Table 3: EDSSP results for Divno=20, Inten_Iter=1, Diver_Iter=30, tb_max_sh=9, tm_max_sh=9, tb_max_lm=9, tm_max_lm=9

	Exact value	**Objective value**	**Gap**
FCNFP_1	64828	69655	4827 (7,45 %)
FCNFP_2	55795	56674	879 (1,58 %)
FCNFP_3	68568	74585	6017 (8,78 %)
FCNFP_4	66109	73485	7376 (11,16 %)
FCNFP_5	56292	62754	6462 (11,48 %)
		Avg.	***5112,2 (8,09 %)***

Table 4: EDSSP results for Divno=20, Inten_Iter=1, Diver_Iter=30, tb_max_sh=1, tm_max_sh=9, tb_max_lm=1, tm_max_lm=9

	Exact value	**Objective value**	**Gap**
FCNFP_1	64828	69655	4827 (7,45 %)
FCNFP_2	55795	56588	793 (1,42 %)
FCNFP_3	68568	74775	6207 (9,05 %)
FCNFP_4	66109	69613	3504 (5,30 %)
FCNFP_5	56292	63510	7218 (12,82 %)
		Avg.	***4509,8 (7,21 %)***

Table 5: EDSSP results for Divno=20, Inten_Iter=1, Diver_Iter=30, tb_max_sh=9, tm_max_sh=1, tb_max_lm=9, tm_max_lm=1

Conclusion

As we can see to see from Tables 2 – 5, the best combination of EDSSP parameters is *Divno* = 20, *Inten_Iter* = 1, *Diver_Iter* = 30, *tb_max_sh* = 9, *tm_max_sh* = 9, *tb_max_lm* = 9, *tm_max_lm* = 9. However these values are the highest values from used ranges for parameters, so it might be interesting to find out in the future, if higher values will bring better solutions.

REFERENCES

[1] KIM D., PAN X., PARDALOS P: „*An Enhanced Dynamic Slope Scaling Procedure with Tabu Scheme for Fixed Charge Network Flow Problems*“, Computational Economics 27, 2006, pp. 273 – 293

[2] KIM D., PAN X., PARDALOS P: „*A solution approach to the fixed charge network flow problem using a dynamic slope scaling procedure*“, Operations Research Letters 24, 1999, pp. 195 – 203

[3] KOHÁNI, M.: *Optimalizácia návrhu štruktúry komunikačného systému „od mnohých k mnohým“*, dizertačná práca, 2007, pp. 73-76

Acknowledgement

This work has been supported by research VEGA 1/3775/06.

Kontaktná adresa

Žilinská Univerzita v Žiline, Fakulta riadenia a informatiky, Katedra dopravných sietí, Univerzitná 8215/1, 010 26 Žilina

IS SLOVAK MONETARY POLICY TIME CONSISTENT? [1]

Martin Lukáčik, Karol Szomolányi

Abstract

By entrance to European Monetary Union, Slovak economy will give up its own monetary policy. How credible this policy has been? Is it necessary to remove it by common European monetary policy?

Kydland and Prescott (1977) introduced a time-consistency problem of the economic policy. By verifying of the time-inconsistency problem we can evaluate the monetary policy. In flavour of quadratic preferences, Ireland (1999) found a way how to verify time-inconsistency problem of the monetary policy. Using this approach we will show that there is no time-consistency problem of the Slovak monetary policy for short run.

Keywords: *time-inconsistency problem of the monetary policy, National Bank of Slovakia, European Monetary Union, vector autoregression*

JEL Classification: E5, C1

1 Introduction

Nobel Prize laureates in 2005, Kydland and Prescott (1977), offered an explanation of inflation behaviour by introducing a model of the time-inconsistency of the monetary policy problem. Society's preferences to low inflation and output higher than potential product, expressed by a quadratic lost function, makes an incentive of monetary authority, who cannot commit its policy, to deviate from original plans. By assumption that the money growth has effect on inflation (see Romer; 2006 for example) and under rational expectations, time inconsistent monetary policy produces an inflation bias.

Economic environments and monetary policies differ by timing of the decision making by agents. We refer their as "with commitment" and "without commitment". In environment without commitment agents behave sequentially rational (Chari, Kehoe and Prescott; 1989). Central bank announces target inflation, before it chooses the monetary policy, but, as it

[1] This paper is supported by the Grant Agency of Slovak Republic – VEGA, grant no. 1/4652/07 "The Analysis of Actual Problems of Slovak Economy Development before the Entrance to European Monetary Union – Econometrical Approach".

cannot commit its policy, it can only choose the actual inflation level after agents choose their expectations. The marginal cost of slightly higher inflation rate is zero and marginal cost of slightly higher output level is positive. Since there is no uncertainty and agents understand the time-inconsistency problem of monetary authority, the expected level of inflation is higher and monetary authority has to deviate from the announced policy. This is time-inconsistency problem of the monetary policy and it implies high inflation even though output does not rise.

Even if, potentially, there is a recourse of the time-inconsistency problem of the monetary policy in every time and in all economies, there have been economies, both in the past and in the presence, with low inflation (USA in 50's of the last century, Germany in the most of the post-war period, significant inflation problem neither have been at Slovakia since its origin). These phenomena economists have explained by reputation (see Barro; 1986), delegation (see Rogoff; 1985), punishment equilibrium (see Barro and Gordon; 1983b) and incentive contracts (see Persson and Tabellini; 1993).

There are econometric techniques helping us to verify and measure time-inconsistency problem of the monetary policy in the economy. Ireland (1999) answered to the question whether the time-consistency problem explained the behaviour of inflation in the United States economy in 1960-1997. He used Barro and Gordon's (1983a) theory, where preferences of agents are expressed by quadratic square function. Surico (2008) removed quadratic preferences by asymmetric and offered a way how to explain inflation in economy with no commitment and how to measure inflation bias.

There is no place in the paper to present more detailed overview of the time-inconsistency problem of the monetary policy. To learn more about the problem we refer reader to study papers mentioned above. More comfortable Slovak or Czech reader can study the problem from papers of authors: Szomolányi (2005, 2006) and Szomolányi, Lukáčiková and Lukáčik (2007, 2008).

The aim of the paper is to verify the time-inconsistency problem of Slovak economy. There is no place to use both mentioned techniques. In the paper we will follow Ireland's conception of explanation of the inflation. As well as U.S. data used by Ireland, Slovak data imply that there is time consistency problem of the Slovak monetary policy for long run perspective, but not for short run perspective.

The paper can be divided into sections: the short introduction and review of papers dealing with the time-inconsistency problem of the monetary policy has been in introduction; in the second section is Ireland's modification of

Barro and Gordon's theory; in the third section are empirical results using Slovak data and the paper is closed by conclusion.

2 Inflation and Unemployment

Barro and Gordon (1983a) modified Kydland and Prescott's model. Let us assume that the relation between inflation and unemployment is given by expectation augmented Phillips Curve:

$$U_t = U_t^n - \omega\left(\pi_t - \pi_t^e\right), \quad \omega > 0 \tag{2.1}$$

where U_t denotes unemployment rate and $U^n{}_t$ denotes natural unemployment rate, π_t is inflation rate and $\pi^e{}_t$ is expected inflation rate in time t and ω is a Phillips Curve parameter. We assume that agents have perfect access to information and expectations are rational.
Preferences of agents are expressed by a quadratic lost function of inflation and unemployment rates:

$$L^{BG} = \frac{1}{2}\left(U_t - kU_t^n\right)^2 + \frac{\beta}{2}\pi_t^2, \quad 0 < k < 1, \quad \beta > 0 \tag{2.2}$$

where the parameter k express wishing of the monetary authority to reduce actual unemployment rate below the natural unemployment rate. For simplicity we assume that by society desired inflation rate is zero.
Ireland (1999) modified Barro and Gordon's model by two assumptions. First, the natural unemployment rate fluctuates over time in response to a ε_t real shock according to the autoregressive process:

$$U_t^n - U_{t-1}^n = \lambda\left(U_{t-1}^n - U_{t-2}^n\right) + \varepsilon_t, \quad -1 < \lambda < 1 \tag{2.3}$$

where ε_t is serially uncorrelated and normally distributed with mean zero and standard deviation σ_ε.
The second Ireland's assumption is that the monetary authority perpetrates control error. The actual inflation rate then differs from the target (planned) $\pi^c{}_t$ inflation rate by the term:

$$\pi_t = \pi_t^c + \eta_t \tag{2.4}$$

where control error η_t is serially uncorrelated and normally distributed with mean zero, standard deviation σ_η, and covariance $\sigma_{\varepsilon\eta}$ with ε_t.
By the commitment assumption we assume that central bank can commit its policy – it set inflation before agents make expectations. Since expectations are rational, expected inflation rate equals to real inflation rate and so we can express Phillips Curve by the form: $U_t = U^n{}_t$. With commitment central bank minimizes the lost function (2.1) subject to Phillips Curve $U_t = U^n{}_t$. It follows that if can monetary authority commit its policy he will choose zero target inflation; $\pi^c{}_t = 0$.

Monetary authority without commitment, chooses target inflation $\pi^c{}_t$ after setting of inflation expectations $\pi^e{}_t$ by private sector but before the realization of the real shock ε_t..
The central bank set $\pi^c{}_t$, to minimize the lost function (2.2) subject to the Phillips Curve (2.1) and subject to the control error (2.4). We solve the problem of monetary authority by substituting Phillips Curve (2.1) into the lost function:

$$\min_{\pi_t^c} E_{t-1}\left\{\frac{1}{2}\left[\left(1-k\right)U_t^n-\omega\left(\pi_t^c+\eta_t-\pi_t^e\right)\right]^2+\frac{\beta}{2}\left(\pi_t^c+\eta_t\right)^2\right\}$$

where E_{t-1} denotes expectations in the beginning of the period t-1, or in the end of the period t, respectively. The first order condition of the problem we can write as:

$$\omega E_{t-1}\left[\left(1-k\right)U_t^n-\omega\left(\pi_t^c+\eta_t-\pi_t^e\right)\right]=bE_{t-1}\left(\pi_t^c+\eta_t\right) \qquad (2.5)$$

Since agents know the structure the economy and they understand time-inconsistency problem of monetary authority, in equilibrium they correctly predict the monetary policy: $\pi^c{}_t = \pi^e{}_t$. By combining this equilibrium condition, the first order condition (2.5) all together with the fact that $E_{t-1}(\eta_t) = 0$, we will gain an inflation bias:

$$\pi_t^c=\pi_t^e=\omega A E_{t-1}U_t^n \qquad (2.6)$$

where A denotes the term $(1\text{-}k)/\beta>0$. The inflation bias follows from the impossibility of monetary authority to commit its own policy and is positively correlated with the expected natural rate of unemployment $E_{t-1}U^n{}_t$. Here, as in Barro and Gordon's original model, the equilibrium inflation rate moves together with the natural rate of unemployment.
By combining of the Phillips Curve (2.1) with the control error (2.4) and with the inflation bias (2.6) we will get:

$$U_t=U_t^n-\omega\eta_t \qquad (2.7)$$

The term (2.7) express that a fluctuation of the unemployment rate around the natural unemployment rate is caused by the monetary policy control error. By substituting of the (2.3) the (2.7) becomes:

$$U_t=U_{t-1}^n+\lambda\Delta U_{t-1}^n+\varepsilon_t-\omega\eta_t \qquad (2.8)$$

where $\Delta U^n{}_{t-1} = U^n{}_{t-1} - U^n{}_{t-2}$ denotes the change of the natural unemployment rate in time t-1. By combining of equations (2.3), (2.4) and (2.6) we will get:

$$\pi_t=\omega A U_{t-1}^n+\omega A\lambda\Delta U_{t-1}^n+\eta_t \qquad (2.9)$$

Terms (2.8) and (2.9) separately indicate that inflation and unemployment are both nonstationary, inheriting unit roots from the underlying process for the natural rate. Together, however, they imply that a linear combination of inflation and unemployment rates is stationary:

$$\pi_t - \omega A U_t = -\omega A \varepsilon_t + \left(1 + \omega^2 A\right) \eta_t \tag{2.10}$$

From the term (2.10) it follows that Barro and Gordon's theory imposes conditions for long-run behaviour of both inflation and unemployment: according to the model these variables should be nonstationary but cointegrated. The statistical test of the cointegration will assign whether Barro and Gordon's model can successfully explain the inflation in the economy. Ireland showed that both inflation and unemployment in the U.S. economy have been really nonstationary and cointegrated in years 1960-1997, and so, for long-run perspective, there has been the time-inconsistency problem in the economy.

Taking differences of the (2.7), solving for $\Delta U^n_t = U^n_t - U^n_{t\text{-}1}$ and by substituting the solution into (2.1) we will gain:

$$\Delta U_t = \lambda \Delta U_{t-1} + \varepsilon_t - \omega \eta_t + \omega\left(1 + \lambda\right) \eta_{t-1} - \omega \lambda \eta_{t-2} \tag{2.11}$$

where $\Delta U_t = U_t - U_{t\text{-}1}$ denotes change of the unemployment rate in time . Equations (2.10) and (2.11) together form a vector ARMA(1,2) for stationary linear combination of inflation and unemployment and stationary change of the unemployment:

$$\begin{aligned} \begin{pmatrix} \pi_t - \omega A U_t \\ \Delta U_t \end{pmatrix} &= \begin{pmatrix} 0 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \pi_{t-1} - \omega A U_{t-1} \\ \Delta U_{t-1} \end{pmatrix} + \begin{pmatrix} -\omega A & 1 + \omega^2 A \\ 1 & -\omega \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} + \\ &+ \begin{pmatrix} 0 & 0 \\ 0 & \omega\left(1 + \lambda\right) \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1} \\ \eta_{t-1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & -\omega\lambda \end{pmatrix} \begin{pmatrix} \varepsilon_{t-2} \\ \eta_{t-2} \end{pmatrix} \end{aligned} \tag{2.12}$$

From restrictions of equations and between equations of the (2.12) scheme it follows that Barro and Gordon's theory imposes conditions for short-run behaviour of both inflation and unemployment. The statistical test of these conditions will assign how well Barro and Gordon's model explains a dynamic relation of these two variables. Using U.S. economy data in years 1960-1997, Ireland showed that in comparison with restricted autoregression (2.12), unrestricted autoregression has statistically better corresponded to the dynamic structure of unemployment and inflation in U.S. economy. It follows therefore that in short run perspective there has been no time-inconsistency problem of U.S. monetary economy.

3 Slovak Monetary Policy

There are used seasonally adjusted quarterly time series of Slovak economy before the revision of the end 2007, so we used data from 1995Q1 till 2007Q2.

The inflation is measured by three methods; the first is equivalent to Ireland's way – it means the percent change of the GDP deflator between two consecutive quarters; the second is the percent change of the CPI and the third way the percent change of the HCPI.

The unemployment rate is derived from the Labour Force Survey using the ILO approach. All the statistics are taken over from the database SLOVSTAT published by The Statistical Office of the Slovak Republic and EUROSTAT database.

The stationary analysis is performed with Augmented Dickey-Fuller test (ADF) and Phillips-Perron test (PP). The second unit root test is preferred in case of a high number of the autoregression terms in test equation solving the autocorrelation.

Test of co-integration is realized by Johansen's procedure. Engle-Granger's two step method is refused because of the problem with autocorrelation of the residuals of the long-term relationship.

Parameters of unrestricted VARMA are directly estimable, but the estimation of restricted VARMA process, which is also needed for the Ireland's test, is computable only by the state space procedure.

The same order of integration necessary for the co-integration of two variables is confirmed only for the seasonally adjusted time series of GDP deflator and unemployment rate, which is the same result as Ireland's one.

Johansen's procedure also verifies his conclusion, because we find the co-integration relationship between these variables, in contrast of Engle-Granger's two step method. This conclusion implies for long-run perspective, there has been the time-inconsistency problem in the Slovak monetary policy.

The directly estimated unrestricted VARMA process for stationary linear combination of inflation and unemployment and stationary change of the unemployment however hasn't desired statistical properties, so that in short run perspective we can't do any time-inconsistency conclusions about Slovak monetary economy.

4 Conclusion

Presence of the time-inconsistency problem of the monetary policy controlled by The National Bank of Slovakia for long-run perspective denotes the non-existence of reputation of actual maker of monetary policy in Slovakia. This conclusion implies the entrance to European Monetary Union in case of reputation European Central Bank will have a contribution from the long-run perspective, despite of losses of possibilities to positive influences on Slovak economy. Question arises whether the date of the entrance is appropriate or this moment should come later.

References

[1] Barro, R. J.: Reputation in a Model of Monetary Policy with Incomplete Information. Journal of Monetary Economics 17, pp. 1-20, 1986

[2] Barro, R. J. – Gordon, D.: A positive Theory of Monetary Policy in a Natural Rate Model. Journal of Political Economy 91, pp. 589-610, 1983a

[3] Barro, R. J. – Gordon, G. B.: Rules, Discretion, and Reputation in a Model of Monetary Policy. Journal of Monetary Economics 12, pp. 101-121, 1983b

[4] Ireland, P. N.: Does the Time-Consistency Problem Explain the Behavior of US Inflation? Journal of Monetary Economics 38, pp. 215-220, 1999

[5] Ištvániková, A.: Optimalizačný prístup k formovaniu hospodárskej politiky na základe ekonometrického modelu. 7. medzinárodná vedecká konferencia Kvantitatívne metódy v ekonómii a podnikaní, Bratislava 2001

[6] Kydland, F. – Prescott, E.: Rules Rather than Discretion: The Inconsistency of Optimal Plans. Journal of Political Economy, vol. 85, 1977

[7] Persson, T. – Tabellini, G.: Designing Institutions for Monetary Stability. Carnegie-Rochester Conference Series on Public Policy 39, 1993

[8] Rogoff, K.: The Optimal Degree of Commitment to an Intermediate Monetary Target. Quarterly Journal of Economics 100, pp. 1169-1190, 1985

[9] Romer, D.: Advanced Macroeconomics. Third edition McGraw-Hill, Berkeley, 2006

[10] Surico, Paolo, "Measuring the Time Inconsistency of US Monetary Policy". Economica, Vol. 75, Issue 297, pp. 22-38, February 2008

[11] Szomolányi, K.: Časová konzistencia politiky v monetárnej ekonomike. Dizertačná práca, 2006

[12] Szomolányi, K: Podstata a povaha problému časovej konzistencie ekonomickej politiky. Ekonomika a informatika III. ročník, 2/2005

[13] Szomolányi, K. – Lukáčiková, A. – Lukáčik, M: Problém časovej konzistencie ekonomickej politiky, zdanenie kapitálu. Ekonomika a informatika V. ročník, 1/2007

[14] Szomolányi, K. – Lukáčiková, A. – Lukáčik, M: Problém časovej konzistencie ekonomickej politiky, model dlhu. Ekonomika a informatika VI. ročník, 1/2008

Martin Lukáčik and Karol Szomolányi
Faculty of Economic Informatics
University of Economics
Dolnozemská cesta 1
852 35 Bratislava

AN INFINITE SERVERS MODEL FOR MOTOR VEHICLES DISMANTLING AND RECYCLING IN A SCARCE ENERGY AMBIENCE

FERREIRA, Manuel Alberto M. (ISCTE),
FILIPE, José António Candeias Bonito (ISCTE),
COELHO, Manuel Francisco Pacheco (ISEG),
ANDRADE, Marina (ISCTE)

Abstract. We propose the M|G|∞ queue system to analyze a situation at which conventional motor vehicles will be out of use as far as the usual energy sources will be exhausted. Two possibilities are posed: recycling or dismantling. Using the hazard rate function we conclude that when the rate of dismantling and recycling of motor vehicles is greater than the rate at which they become idle, the situation tends to get balanced. The motor vehicles that are unused with the conventional energies either turn useful with another kind of energies or are included in other useful devices.
Key words: Motor Vehicles, Recycling, Dismantling, M|G|∞, Hazard Rate Function.

Mathematics Subject Classification: Primary 60E05, 60G07; Secondary 60K25.

1 INTRODUCTION

Hardin's paper of "The tragedy of the commons" [7] is a reference for the problems that traditionally occur in the resources area. Theoretical study of Commons became extremely important in the analysis of the consequences of human behaviour when human specie exploits Earth resources [1and 6]. Scientists have long claimed for the need of changing human behaviour and we are getting close to the last chance to change things. Some apocalyptic studies point out this evidence very clearly [9].
New situations may occur very suddenly. The production of oil has already got its peak and new oil productions will occur in the future with decreasing rents until the complete depletion of this resource. The other non-renewable sources of energy will have the same end.
We may have to reorder the priorities and to reorganize structures in societies. We have now to produce the new kind of energies (clean energies) at a major scale. The big problem that remains is to know if the transition period from non-renewable sources of energy to the renewable sources is enough to overcome the big problems related with the destruction of Earth.

All the wastes that people have made for so many decades must be overcome, as well. Anyway, many kinds of new problems will occur. However, what is important now is to know how quickly changes may happen while we develop the new sources of energy in order to create a new economy and a reorganized society.
Two situations may occur. First, it is possible to reverse climatic changes and to reverse all the related problems with a fast change of direction from the old to the new sources of energy. Anyway, many types of equipment must be recycled to respond to the new situation as far as factories begin to produce new equipments responding to the new circumstances.
The other situation represents a tragic scenario where there is no enough time to perform a smooth transition and societies would have a period of big privations [7]. In our opinion, such a scenario will not be the prevailing one if institutions realize about the big problems of all kinds for our global civilization.

2 A QUEUE MODEL FOR MOTOR VEHICLES' DISMANTLING AND RECYCLING

We consider the M|G|∞ queue system [8] where customers arrive according to a Poisson process at rate λ [4]. They receive a service whose length is a positive random variable with distribution function G (.) and mean α. Each customer as soon as it arrives at the system, immediately finds an available server. Each customer service is independent from the others customers' services and from the arrival process. The traffic intensity is given by $\rho = \lambda\alpha$.
With this model we intend to analyze a situation at which motor vehicles arrive at the system getting idle and leave the system as soon as they are recycled or dismantled. Both situations are modelled with the same purpose in the model. Our interest in studying this situation is precisely to see how the system may recover to a balanced situation in which motor vehicles get operational or get dismantled (in this situation materials would become employed as components in other applications).
Let $N(t)$ be the number of busy servers (or, what is the same, the number of customers being served) in the instant t, in a M|G|∞ system. If we consider $p_{0n}(t) = P[N(t) = n | N(0) = 0]$, $n = 0,1,2,...$, we may have [2]:

$$p_{0n}(t) = \frac{\left(\lambda \int_0^t [1 - G(v)]dv\right)^n}{n!} e^{-\lambda \int_0^t [1-G(v)]dv}, \quad n = 0,1,2,... \qquad (1).$$

So if the initial instant is a moment at which the system is empty, the transient distribution is Poisson with mean $\lambda\int_0^t[1-G(v)]dv$.

The stationary distribution is the limit one:

$$\lim_{t\to\infty} p_{0n}(t)=\frac{\rho^n}{n!}e^{-\rho}, \quad n=1,2,... \tag{2}$$

This queue system, as any other, has a sequence of busy periods and idle periods. A busy period begins when a customer arrives at the system, finding it empty.

Let's see the distribution of the number of customers being served in the instant t in the M|G|∞ system, when the initial instant is the moment at which a busy period begins, that gets relevant for our purposes.

Be $p_{1'n}=P[N(t)=n|N(0)=1']$, $n=0,1,2,...,$ and $N(0)=1'$ the initial instant at which a customer arrives at the system and the number of customers being served turns from 0 to 1. This means that a busy period has just begun [3].

So, when $t\geq 0$, we may have a situation that represents [5]:

1. the customer that arrived at the system at the initial instant has left the system with a probability $G(t)$, or he remains in the system, with probability $1-G(t)$;
2. the other servers, which were empty at the beginning (initial instant), may be now empty or busy with 1, 2, ... customers, with probabilities given by $p_{0n}(t)$, $n=0,1,2,...$.

The two subsystems, the one of the initial customer and the one of servers initially empty, are independent. Consequently:

$$p_{1'0}(t)=p_{00}(t)G(t)$$

$$p_{1'n}(t)=p_{0n}(t)G(t)+p_{0n-1}(t)(1-G(t)), \quad n=1,2,... \tag{3}$$

We also have for this situation:

$$\lim_{t\to\infty} p_{1'n}(t)=\frac{\rho^n}{n!}e^{-\rho}, \quad n=0,1,2,... \tag{4}$$

For the M|M|∞ system (exponential service times), the equations 3 are applicable even when $N(0)=1$ (when the initial instant is a moment at which there is a customer in the system; that does not enforce that this is the moment at which we are turning from 0 to 1 customer to be served). This results from the lack of memory of the exponential distribution.

If $g(t)$ is the probability density function related to $G(t)$, and if we call $h(t)$ the hazard rate function, we'll have [10]:

$$h(t)=\frac{g(t)}{1-G(t)} \qquad (5).$$

The function $h(t)$ is the rate at which services end.
So,
Proposition 1:
If $G(t)<1$, $t>0$, continuous and differentiable and if

$$h(t)\geq \lambda G(t),\ t>0 \qquad (6)$$

$p_{1'0}(t)$ is a non-decreasing function.
Dem:
It's enough to observe that $\frac{d}{dt}p_{1'0}(t)=p_{00}(t)(1-G(t))\left(\frac{g(t)}{1-G(t)}-\lambda G(t)\right)$.
Besides, we may note that

$$h(t)\geq \lambda,\quad t>0 \qquad (7)$$

is a sufficient condition for the result in 6.
So, if the rate at which the services end is greater or equal than the rate of arrivals, we conclude that $p_{1'0}(t)$ does not decrease.
For the system M|M|∞, the equation 7 is equivalent to

$$\rho \leq 1 \qquad (8).$$

Considering $\mu(1',t)$ and $\mu(0,t)$ the mean values of the distributions given by 3 and 1, respectively, we'll have

$$\mu(1',t)=\sum_{n=1}^{\infty}np_{1'n}(t)=\sum_{n=1}^{\infty}nG(t)p_{00}(t)+\sum_{n=1}^{\infty}np_{0n-1}(t)(1-G(t))=$$

$$=G(t)\mu(0,t)+(1-G(t))\sum_{j=0}^{\infty}(j+1)p_{0j}(t)=\mu(0,t)+(1-G(t)).$$

So,

$$\mu(1',t)=1-G(t)+\lambda\int_0^t[1-G(v)]dv \qquad (9).$$

Proposition 2:
If $G(t)<1$, $t>0$, continuous and differentiable and if

$$h(t)\leq \lambda,\quad t>0 \qquad (10).$$

$\mu(1',t)$ is a non-decreasing function.
Dem:
It's enough to observe that, considering equation 9, $\frac{d}{dt}\mu(1',t)=(1-G(t))(\lambda-h(t))$.

Besides, if the rate at which services end is lesser or equal than the rate at which customers arrive $\mu(1',t)$ is a non-decreasing function. We can note additionally that, for the M|M|∞ system, the equation 10 is equivalent to

$$\rho \geq 1 \quad (11).$$

3 RESULTS AND COMMENTS

According to our study interests, the customers are the motor vehicles that become idle. The arrival rate is the rate at which the motor vehicles become idle. The service time for each one is the time that goes from the instant they get idle until the instant they become recycled or dismantled. The service time hazard rate function is the rate at which the motor vehicles become recycled or dismantled.

An idle period for our M|G|∞ system should be a one at which there were no motor vehicles idle. In a busy period there are always continuously idle motor vehicles.

The equation 6 shows that if the dismantling and recycling rate is greater or equal than the rate at which motor vehicles get idle, the probability that the system gets empty (that is, there are no idle motor vehicles) does not decrease with time. This means that the system has a tendency to become balanced as far as time goes on.

The equation 10 shows that if the dismantling and recycling rate is lesser or equal than the rate at which motor vehicles get idle, the mean number of motor vehicles in the system does not decrease with time. This means that the system has a tendency to become unbalanced as far as time goes on.

Consequently, we conclude that when the rate of dismantling and recycling of motor vehicles is greater than the rate at which they become idle, the system has a tendency to get balanced. In this situation, the motor vehicles that become unused with the conventional energy turn useful with another kind of energy or get included in other useful devices.

We must note that it is important the recycling or the dismantling of motor vehicles but, more than that, it is essentially relevant the cadence at which these actions are performed. Moreover, we give a reference for this cadence: λ, the rate at which motor vehicles get idle.

4 AN ECONOMIC ANALYSIS AS A COMPLEMENT TO THE MODEL

We have seen that rates λ and $h(t)$ are determinant to monitor the way that the system of motor vehicles recycling and dismantling may be managed.

We consider now additionally p as the probability for the motor vehicles arrivals destined to recycling and (1-p) as the probability for the motor

vehicles arrivals destined to dismantling. Let $h_i(t)$, $c_i(t)$ and $b_i(t)$, $i=1,2$ be the hazard rate function, the mean cost and the mean benefit, respectively for recycling when $i=1$ and dismantling when $i=2$.
With these new variables we can perform an economic analysis (beyond other considerations that may be posed) to evaluate about the interest of recycling and dismantling.
We will analyze this situation in a global approach and not in a selfish way, considering the individuals point of view.
So, we may consider the total cost per unit of time for motor vehicles recycling and dismantling as:

$$C(t) = pc_1(t)\lambda + (1-p)c_2(t)\lambda \tag{12.}$$

Furthermore, the benefit per unit of time resulting from recycling and dismantling is given by:

$$B(t) = b_1(t)h_1(t) + b_2(t)h_2(t) \tag{13.}$$

From an economic point of view, it must be B(t)>C(t).
To conclude about the advantage of recycling, we have the following:

$$b_1(t) > \max\left[\ \frac{p\lambda c_1(t) + (1-p)\lambda c_2(t) - b_2(t)h_2(t)}{h_1(t)}, 0\ \right] \tag{14.}$$

With $G_1(t)$ and $G_2(t)$ both exponential, equation 14 becomes:

$$b_1(t) > \max\left[\ (pc_1(t) + (1-p)c_2(t))\rho_1 - \frac{\alpha_1}{\alpha_2} b_2(t), 0\ \right] \tag{15.}$$

To conclude about the advantage of dismantling we have, in the same conditions:

$$b_2(t) > \max\left[\ \frac{p\lambda c_1(t) + (1-p)\lambda c_2(t) - b_1(t)h_1(t)}{h_2(t)}, 0\ \right] \tag{16}$$

and

$$b_2(t) > \max\left[\ (pc_1(t) + (1-p)c_2(t))\rho_2 - \frac{\alpha_2}{\alpha_1} b_1(t), 0\ \right] \tag{17.}$$

So, there are minimum benefits above which, from an economic point of view both, recycling and dismantling, are interesting. The most interesting is the one for which this minimum benefit is the least. By other words: in a global perspective, it is more efficient the activity that corresponds to a lower level for the minimum interesting benefit.

Recycling seems to be as much interesting as far as it is more economically profitable and our inequalities 14 to 17 are tools that may be applied to evaluate this interest.
Note that instead of the instantaneous reasoning we can make it for a period of time of length T. It's enough to perform the respective integral calculus between 0 and T. Of course, the conclusions are analogous.

5 MAIN CONCLUSIONS

Our model contributes for a better understanding of this kind of problems and it (or some modified versions of it) may be applied to study some other social and economic phenomena such as unemployment, health or projects of investment, for example, with interesting results.
An extension of our model has also permitted to get conclusions about the economic advantages of recycling and dismantling.
The application of the model to the phenomenon studied in our paper shows that our model is very useful and that its conclusions and results are quite simple to understand. Just through a theoretical analysis of the model, we have evidenced some remarkable topics for analyzing the evolution of the studied system (or another one, whichever it is, since it is according the assumptions of the model).
In practice, it is essential to estimate λ and $h(t)$ to get conclusive particular results for the available data about the system. This will give us the tools to monitor the situation and to suggest solutions. A correct estimation of λ will depend on the arrivals process to be Poisson, in real. Additionally, in general, it is correct to admit that with very large populations, such as the one we are dealing with, the estimation of $h(t)$ is usually technically complicated. So, frequently, the best to do is to estimate directly $h(t)$ instead of estimating first the service time distribution and then computing $h(t)$.
A particular situation at which the computation is easier is the exponential service time one, for which $h(t) = 1/\alpha$.

REFERENCES

[1] Agrawal, A. (2002). Common resources and institutional sustainability. In E. Ostrom, T. Dietz, N. Dolsak, P. C. Stern, S. Stonich, & E. U. Weber (Eds.), *The drama of the commons* (pp. 42-53). Washington DC: National Academy Press

[2] Carrillo, M. J. (1991), Extensions of Palm's Theorem: A Review. Management Science. Vol. 37 N.º 6, 739-744.

[3] Ferreira, M. A. M (1988), *Redes de Filas de Espera.* Master Thesis presented at IST. Lisboa.

[4] Ferreira, M. A. M (1996), Comportamento Transeunte do Sistema M|G|∞ - Aplicação em problemas de doença e de emprego. *Revista de Estatística*. Vol. 3., 3.º Quadrimestre. INE, 23-31.
[5] Ferreira, M. A. M. (1998), Aplicação da Equação de Ricatti ao Estudo do Período de Ocupação do Sistema M|G|∞. *Revista de Estatística*. Vol. 1., 1.º Quadrimestre. INE, 23-28.
[6] Ferreira, M.A.M., Filipe, J. A. and Coelho, M. (2008), A Queue Model for Motor Vehicles Dismantling and Recycling in a Scarce Energy Ambience. Aplimat- Journal of Applied Mathematics, Volume 1(2008), Number 1. Section Finantial and Actuary Mathematics. Bratislava. ISSN: 1337-6365.
[7] Filipe, J. A. (2006), *O Drama dos Recursos Comuns. Um caso de Aplicação da Teoria dos Jogos aos Comuns da Pesca. Estudo da Cooperação aplicada à pesca da sardinha nas Divisões VIIIc e IXa do ICES*. PhD Thesis presented at ISCTE. Lisboa.
[8] Hardin (1968), The Tragedy of the Commons, *Science*,162, pp.124-148.
[9] Kelly, F. P. (1979), *Reversibility and Stochastic Networks*. John Wiley and Sons. New York
[10] Kunstler, J. H. (2006), *O fim do petróleo – O grande desafio do século XXI*. Editorial Bizâncio. Lisboa.
[11] Ross, S. (1983), *Stochastic Processes*. John Wiley and Sons. New York.

Current Address

Manuel Alberto Martins Ferreira, Professor Catedrático
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel.+351 217 903 000
E-mail: manuel.ferreira@iscte.pt

José António Candeias Bonito Filipe, Professor Auxiliar
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel.+351 217 903 000
E-mail: jose.filipe@iscte.pt

Manuel Francisco Pacheco Coelho, Professor Auxiliar

ISEG – Instituto Superior de Economia e Gestão
SOCIUS – Centro de Investigação em Sociologia Económica e das Organizações
Rua do Quelhas, 6, 1200-781 Lisboa (Lisbon, Portugal), Tel.+351 213 925 800
E-mail: coelho@iseg.utl.pt

Marina Alexandra Pedro Andrade, Professora Auxiliar
ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa
UNIDE – Unidade de Investigação e Desenvolvimento Empresarial
Av. Forças Armadas 1649-026 Lisboa (Lisbon, Portugal), Tel.+351 217 903 000
E-mail: marina.andrade@iscte.pt

OPTIMIZATION OF THE SHARPE RATIO AND THE OMEGA IN MUTUAL FUNDS PORTFOLIO

Ing. Zdenka Milánová
Comenius University in Bratislava, Faculty of Social and Economic Sciences,

Abstract: The violation of the assumption of normal distributed returns of assets leads to searching for new measures of performance, which allow taking into consideration another characteristic then first and second moments. In this paper is compared performance of two portfolios. The first one is generated by maximization of Sharpe ratio and the second one by maximization of Omega function. The portfolio consists of six mutual funds that are traded on Slovak financial market.

Keywords: Sharpe ratio. Omega function.

Introduction

The popularity of alternative investment instruments leads to searching for performance measures which are able to capture different characteristics of returns unlike (in contrast with) classic approach which is based on mean - variance analysis.
There are different performance measures that maximize final wealth for a given time period[1]. They can be divided on classic performance measures (Treynor measure, Sharpe ratio, Jensen measure), performance measures wit asymmetric preferences (Sortino ratio, performance according to semi – absolute deviation (ROAS), Omega) and performance measures used in practice (Total return, Calmar ratio, Sterling ratio, Information ratio)
In this paper is compared performance of the portfolio that is generated by maximization of Sharpe ratio an Omega function.

Portfolio Selection model with Sharpe Ratio

Sharpe ratio is quantitative performance measure that has basis in financial theory and is based on exact tracking of risk and returns of the assets. Preferences of the investor are explained by analysis in mean – variance space (Markowitz, 1952) and by Capital Asset Pricing Model (CAPM) of Sharpe[2].

---

[1] Pedersen, Ch. S. - Rudholm-Alfvin, T .: Selecting a Risk-Adjusted Shareholder Performance Measure, Journal of Asset Management; Oct. 2003; 4, 3, pp. 152 –172.
[2] Sharpe, W.F.: Capital Asset Prices, Journal of Finance, September 1964, pp.425-442.

Sharpe ratio gives into relation excess return and risk of the portfolio. An excess return is measured as a difference between expected return of the portfolio $E(R_P)$ and risk free rate r_f. Risk of the portfolio is measured by standard deviation σ_P. Sharpe ratio can be write as follows:

$$\textit{Sharpe ratio} = \frac{E(R_P)-r_f}{\sigma_P}.$$

For each risk measure there is performance measure that enable to identify above – average, normal and below - average performance. Performance measure $\rho(.)$ is a ratio between expected excess return and relative risk measure that is maximized by investor. By maximizing of performance measure $\rho(.)$ investor obtains market portfolio, which represents benchmark for market.

Portfolio selection model with performance measure as a goal, investor is interested in growth of capital and he can take into consideration risk connected with growth of this capital. Sharpe ratio[3] is a classic performance measure that takes into account first and second moment of returns distribution. If returns distribution is asymmetric it can not to capture al information that contains time series of returns. Portfolio selection model with Sharpe ration can be formulated as follows:

$$\max \frac{E(R_P)-r_f}{\mathbf{w}^T.\mathbf{C}.\mathbf{w}} \qquad (1)$$

s.t.

$$\mathbf{w}^T \cdot \mathbf{e} = 1,$$
$$\mathbf{w} \geq 0,$$

where **C** is covariance matrix, **w** is vector of asset weights in the portfolio and **e** is vector of ones.

Portfolio Selection Model with Omega Function

Omega function [4] is a new performance measure that utilizes all information that time series of returns contains. It allows to evaluate and to arrange portfolios of the investor. It enables to divide interval of returns on loss interval and profit interval, where loss interval is under the threshold of

---

[3] Sharpe, W.F.: *Mutual Fund Performance*, Journal of Business, 1966, pp.119-138.

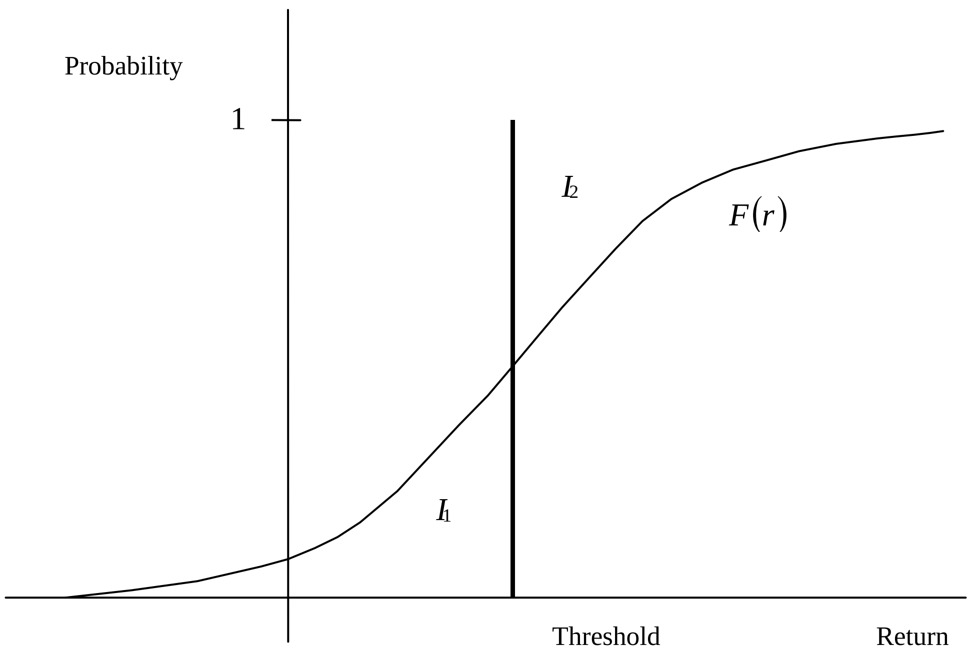
[4] Keating, C. – Shadwick, W .:*A Universal Performance Measure*, The Finance Development Centre, London, Working paper (May). 2002.

http://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2005/Keating_A_universal_performance.pdf. (29.1.2008) returns and profit is above threshold of returns (Picture 1). Threshold is labeled as *h*. Omega function sis defined as follows:

$$\Omega(r) = \frac{\int_h^b (1 - F(x))dx}{\int_h^r F(x)dx} = \frac{I_2}{I_1},$$

where (a,b) is interval of returns and F is cumulative distribution function.

Omega equals to 1 if *h* is average return. It is computed from historical data and do not need estimation. It means that exploits all information from time series and so it is equally statistically significant as time series. Omega represents combination of all moments. Only one assumption is decision rule, where investor prefers more than less.



Picture 1: Cumulative distribution function.

Usually, Omega gives different organization of funds, portfolios and assets than classic performance measures. The reason is that it uses additional information from time series. In the cases when higher moments are not significant, Omega offers same results as traditional performance measures and do not need to estimate risk and expected return. If higher moments are significant, Omega provides corrections of these approximations and it can cause changes in weights of optimal portfolio selection model.

Portfolio selection model as a maximization of Omega $\Omega(\mathbf{w})$ can be formulated as follows[5]:

$$\max \Omega(\mathbf{w}) = \frac{\sum_{t=1}^{T} \max\left[\mathbf{w}^T \mathbf{r}_t - h, 0\right]}{\sum_{t=1}^{T} \max\left[h - \mathbf{w}^T \mathbf{r}_t, 0\right]}, \qquad (2)$$

subject to

$$\mathbf{w}^T \cdot \mathbf{e} = 1,$$
$$\mathbf{w} \geq 0$$

where $\mathbf{r}_t$ is a vector of asset returns in time period *t*, for $t = 1,2,....T$.

Performance Optimization of Mutual Funds Portfolio

Portfolio selection model with Sharpe ratio and Omega function is applied on data about six mutual funds. Portfolio consists of two money market funds (KBC MULTI CASH CSOB SKK, ING International (II) Slovak Money Market Fund), two bond funds (KBC RENTA EURORENTA, ING International (II) Slovak Bond Fund) and two stock funds (KBC Equity Fund Central Europe, ING Visegrad Stock Fund). We have weekly data from January 2, 2004 to January 31, 2008.

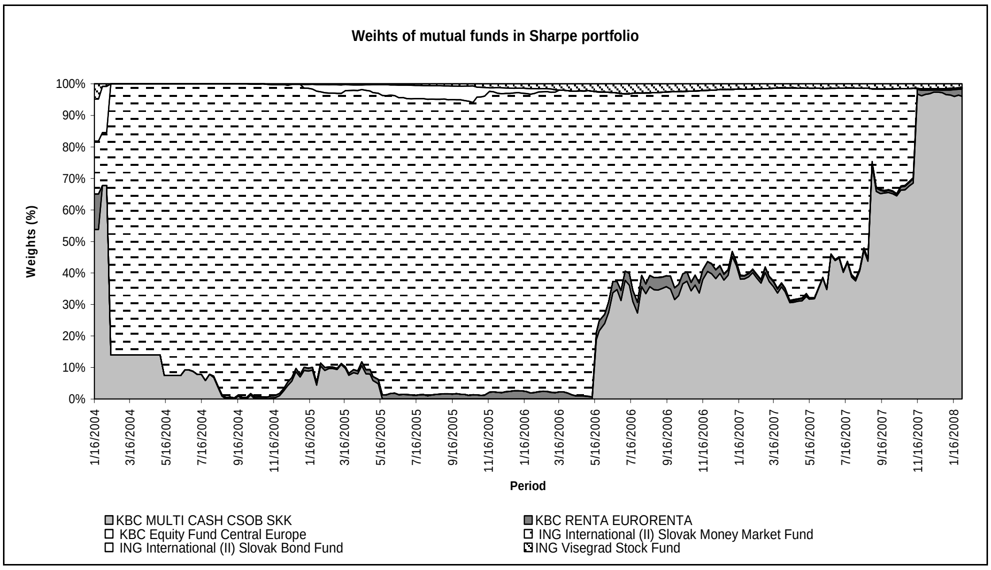
Next, we are going to track portfolio performance of the Sharpe portfolio and Omega portfolio. In each period it is needed to find optimal portfolio at the basis of previous historic returns. There are 213 periods and in each one it is needed to compute covariance matrix and vector of expected returns for Sharpe portfolio and cumulative distribution function for Omega portfolio. To automate these processes two procedures were created in Visual Basic for Application in MS Excel

On the Picture 2 is illustrated development of the composition of mutual funds portfolio generated as a maximization of Sharpe ratio. Risk free rate is given at the level of 3 %. On the Picture 3 is shown development of the composition of mutual funds portfolio, which is generated as a maximization of Omega function. Threshold equals to 3 %.

---

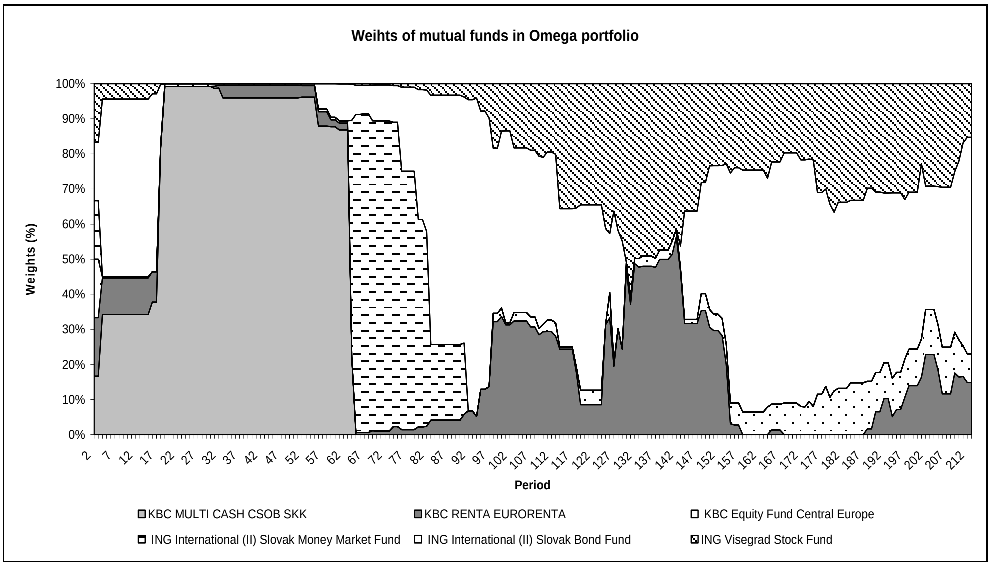
[5] Avouyi-Dovi, S., Morin, S., Neto, D.: *Optimal Assets Allocation with Omega Function*, http://www.effasebc.org/EBC%20Presentations/Last%20Two%20Years/2004%20Oct%20 Zurich/Optimal%20Allocation%20with%20Omega%20-%20HkappaEFFAS-ebc.pdf, (24.2.2008).

From Picture 2 it is apparent, that portfolio generated by maximization of Sharpe ration is very poor diversified. In majority tracked time period ING International (II) Slovak Money Market Fund has the biggest weights in Sharpe portfolio. In the case of maximization of Omega function, portfolio is by far diversified, thanks to cumulative distribution function, which allows using all information from time series.



Picture 2: Weights of individual mutual funds in portfolio generated as a optimization of Sharpe ratio.

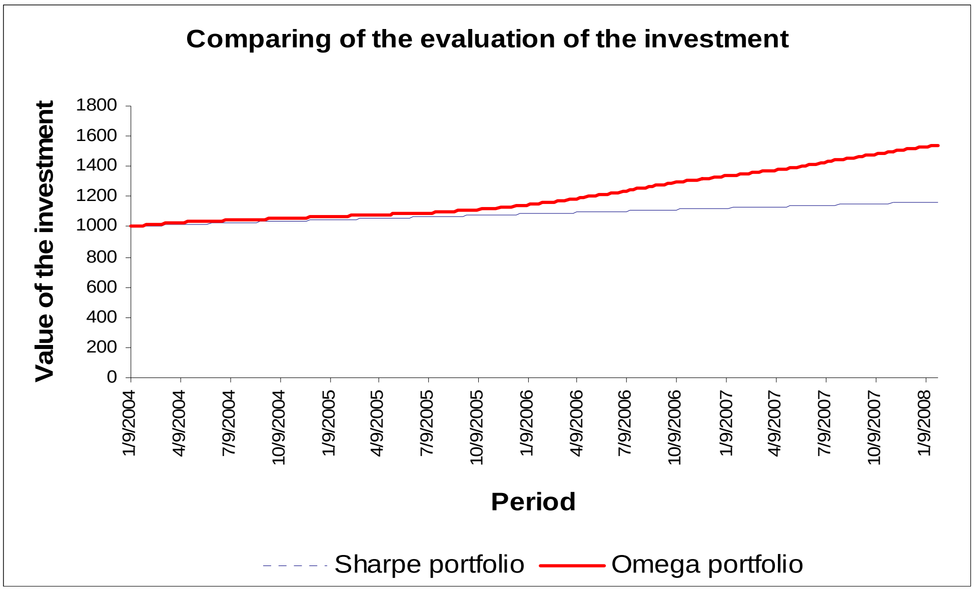
Let suppose that at the beginning investor invests 1000 EUR in two portfolios. The first one has weights generated by Sharpe ratio optimization and the second one by Omega function optimization. We are interested what the value of this capital is at the end of the period. Developments of individual investment in the portfolio are illustrated at the Picture 4. At the end of period we can obtain 1165, 75 EUR in the case of Sharpe portfolio (16, 75 % performance per four years) and 1539, 43 EUR in the case of Omega portfolio (53, 94 % performance per four years).



Picture 3: Weights of individual mutual funds in portfolio generated as a optimization of Omaga function ratio.

Conclusion

Determining of the performance and searching for a combination of assets that maximize this performance is one of investor goals. If returns not only new investment products but also stocks and bonds are not normally distributed, classic performance measures as Sharpe ratio provides poor results and new approaches as Omega function have became interested. Omega function exploits cumulative distribution function to utilize all information that contains time series of returns. In this paper were compared these two approaches at the basis of six mutual funds. In each period these two approaches were used to generate optimal portfolio that has maximal performance. Portfolio generated by maximization of Omega function offers much better evaluation of initial investment than portfolio generated by optimization of Sharpe ratio.



Picture 4: Development of the investment invested into the portfolio generated by optimization of Sharpe ratio and Omega function.

Bibliography

1. Avouyi-Dovi, S., Morin, S., Neto, D.: *Optimal Assets Allocation with Omega Function*, http://www.effasebc.org/EBC%20Presentations/Last%20Two%20Years/2004%20Oct%20Zurich/Optimal%20Allocation%20with%20Omega%20-%20HkappaEFFAS-ebc.pdf, (24.2.2008).
2. Markowitz, H.: *Portfolio Selection*, The Journal of Finance, March 1952, pp. 77-91.
3. Keating, C. – Shadwick, W .:*A Universal Performance Measure*, The Finance Development Centre, London, Working paper (May). 2002. http://faculty.fuqua.duke.edu/~charvey/Teaching/BA453_2005/Keating_A_universal_performance.pdf. (29.1.2008)
4. Pedersen, Ch. S. - Rudholm-Alfvin, T .: Selecting a Risk-Adjusted Shareholder Performance Measure, Journal of Asset Management; Oct. 2003; 4, 3, pg. 152 –172.
5. Sharpe, W.F.: Capital Asset Prices, Journal of Finance, September 1964, pp.425-442.
6. Sharpe, W.F.: *Mutual Fund Performance*, Journal of Business, 1966, pp.119-138.

Contact
Ing. Zdenka Milánová
Institute of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Odbojárov 10/A, P.O. Box 129, 820 05 Bratislava 25.
Tel: (421 2) 50 11 76 26, email: milanova@fses.uniba.sk

ANALYZING THE IMPACT OF VALUE CHANGES IN ELASTICITY OF SUBSTITUTION AND ELASTICITY OF TRANSFORMATION TO THE STABILITY OF APPLIED CGE MODEL

Veronika Miťková
Institute of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava

Abstract
The goal of this work is to analyze the impact of value changes in elasticity of substitution in Armington function of composite supply and value changes in elasticity of transformation in CET functions of output transformation to the stability of the applied general computable equilibrium model. The model construction allows analyzing the impact of changes in a selected sector of economy aspect to the aggregated sector of other sectors. The automotive sector is selected as the analyzed sector since the Slovak economy is strongly influenced by dynamics in this sector. According to the simulations and analyzes it is obvious that the model is stabile from the point of view of elasticity values changes and so it is appropriate for foreign trade (and policy) analyzes.

Keywords: Armington function, CET function, elasticity of substitution, elasticity of transformation

JEL: C68, F10

Introduction

The research presented in this paper was inspired by work of Kazybayeva and Tanyeri –Abur (2003) about trade policy adjustments in general equilibrium analysis where the parameters of trade functions were changed and the impact to the model results were observed. They conclude that the qualitative results remain the same under a wide range of different CES and CET parameters; thus the interpretations and conclusions apply to the entire spectrum of parameter variation in the model.

McDaniel's and Balistreri's (2003) research about influence of Armington function parameters to the model results begins with note that a few robust findings emerge from the econometric literature. The first one is that long-run estimates are higher than short-run estimates, the second that more disaggregate analyses find higher elasticities, and the last one that reduced-form time-series analyses find lower elasticities relative to cross-sectional studies. For the experiments they used the steady-state equilibrium as the baseline. The model was greatly simplified by aggregating up to include only three goods and four regions. The standard nested-Armington structure was used in which the lower-tier substitution elasticity (between imported varieties) is twice the substitution elasticity between imported and domestic varieties. This model is similar to the model used in our early work and the goal of this research is to test the model to the sensitivity to the impact of the changes in elasticity of substitution and elasticity of transformation.

The Model

A standard computable general equilibrium model (CGE) captures all financial or physical flows in the economy. The database of the model we are working with consists of a social accounting matrix (SAM), which synthesizes these flows on activity, commodity, household, factor, institution, saving and investment, tax and rest of the world accounts. From the mathematical point of view, the model is a set of simultaneous equations describing the agents' behavior; some of them are nonlinear. The first order nonlinear conditions capture the behavior on the production and consumption side; it means that behavior is derived by maximizing profit and utility. The equations include conditions which must be fulfilled for the system, not necessary for individual agents. The equations include factor and commodity markets and such macroeconomic aggregates as savings and investments, government and rest of the world.

We focus on the impact of the value changes in elasticity of substitution and elasticity of transformation to the stability of the applied computable general equilibrium model in this work. The model was developed for analyzing the impact of policy changes to Slovakian automotive sector which was selected for analysis since the Slovak economy is strongly influenced by dynamics in this sector in the past years (Miťková, Mlynarovič, 2007).

For more than twenty years before there was a principal question in the modeling agenda how to deal with export, import, balance of payments and balance of trade in the CGE models. After trying various approaches, a general consensus was reached by adopting the Armington assumption. This assumption regards imperfect substitutability and it is extended to the modeling of exports as well. The use of the Armington function in trade differs from the standard neoclassical trade model in which all goods are tradable and all domestically produced goods are perfectly substitutable with imports. According Zhang (2006) the standard treatment is not fully compatible with the empirical results: the domestic relative price of tradeables is fully determined by world prices. Neoclassical models result in the full transmission of world price changes and in extreme specialization in production. In the Armington framework, the economy is less responsive to world price changes, thus dampening the move toward specialization (Zhang, 2006). Also Khan (2004) quotes that the most common approach now is to specify sectoral constant elasticity of substitution (CES) import demand functions, export transformation functions that assume constant elasticity of transformation (CET) and aggregation functions based on these.

Armington elasticities specify the degrees of substitution in demand between similar products produced in different countries. They are critical parameters which, along with model structure, data and other parameters, determine the results of policy experiments.

The Composite Supply Function

The supply side is modeled as composite supply of import and domestic use of domestic output by Armington function.

$$QQ_c = \alpha q_c \cdot \left[\delta_c^q \cdot QM_c^{-\rho_c^q} + \left(1-\delta_c^q\right)\cdot QD_c^{-\rho_c^q}\right]^{\frac{-1}{\rho_c^q}}$$

Equation 1

It is supposed an imperfect substitutability between imports and domestic output sol domestically and it is captured by a CES aggregation function in which the composite commodity that is supplied domestically is "produced" by domestic and imported commodities, and enters the function as inputs. The optimal mix between imports and domestic product is given by Equation 1 $QQ_c = \alpha q_c \cdot \left[\delta_c^q \cdot QM_c^{-\rho_c^q} + \left(1-\delta_c^q\right)\cdot QD_c^{-\rho_c^q}\right]^{\frac{-1}{\rho_c^q}}$

Equation 1.

$$QQ_c = \alpha q_c \cdot \left[\delta_c^q \cdot QM_c^{-\rho_c^q} + \left(1-\delta_c^q\right)\cdot QD_c^{-\rho_c^q}\right]^{\frac{-1}{\rho_c^q}}$$

$$\frac{QM_c}{QD_c} = \left(\frac{PD_c}{PM_c}\cdot\frac{\delta_c^q}{1-\delta_c^q}\right)^{\frac{1}{1+\rho_c^q}}$$

Equation 2 The share parameter of composite supply function is derived from Equation 1and a relationship between elasticity of substitution and an exponent of composite supply $\rho_c^q = \frac{1}{\sigma_c^q - 1}$. Now it is possible to derive the composite supply function form:

$$QQ_c = \alpha q_c \cdot \left\{ \frac{1}{1 + \frac{PD_c}{PM_c} \cdot \left(\frac{QD_c}{QM_c} \right)^{\left(\frac{\sigma_c^q}{\sigma_c^q - 1} \right)}} \cdot QM_c^{-\frac{1}{\sigma_c^q - 1}} + \left[1 - \frac{1}{1 + \frac{PD_c}{PM_c} \cdot \left(\frac{QD_c}{QM_c} \right)^{\left(\frac{\sigma_c^q}{\sigma_c^q - 1} \right)}} \right] \cdot QD_c^{-\frac{1}{\sigma_c^q - 1}} \right\}^{\left(1 + \sigma_c^q\right)}$$

Equation **3**

where:

αq_c shift parameter for composite supply (Armington) function,
δ_c^q share parameter for composite supply (Armington) function,
σ_c^q the elasticity of substitution for composite supply (Armington) function,
ρ_c^q exponent $\left(-1 < \rho_c^t < \infty\right)$ for composite supply (Armington) function,
PD_c domestic price of domestic output,
PM_c import price (in domestic currency),
QD_c quantity of domestic output sold domestically,
QM_c quantity of imports,
QQ_c quantity supplied to domestic commodity demanders (composite supply).

The Output Transformation Function

Domestic output as a function of export and domestic use of domestic output is modeled by an output transformation (CET) function.

$$QX_C = \alpha t_c \cdot \left[\delta_c^t \cdot QE_c^{\rho_c^t} + \left(1 - \delta_c^t\right) \cdot QD_c^{\rho_c^t} \right]^{\frac{1}{\rho_c^t}}$$

Equation 3

It is supposed an imperfect transformability between domestic output for exports and domestic sales as a parallel to imperfect substitutability between imports and domestic output sold domestically used in Armington function. The CET function is identical to CES function except for negative elasticities of substitution. In economic terms, the difference between the Armington and CET functions is that the arguments in the first one are inputs and in the second one are outputs.[1] The optimal mix between exports and domestic sales is given by Equation 5.

$$\frac{QE_c}{QD_c} = \left(\frac{PE_c}{PD_c} \cdot \frac{1-\delta_c^t}{\delta_c^t} \right)^{\frac{1}{\rho_c^t - 1}}$$

Equation 4

The share parameter for output transformation (CET) function is derived from Equation 5 and a relationship between elasticity of transformation and an exponent of output transformation supply $\rho_c^t = \frac{1}{\sigma_c^t + 1}$. Now it is possible to derive the output transformation function form:

$$QX_c = \alpha t_c \cdot \left\{ \frac{1}{1 + \frac{PD_c}{PE_c} \cdot \left(\frac{QE_c}{QD_c} \right)^{\frac{\sigma_c^t + 2}{\sigma_c^t + 1}}} \cdot QE_c^{\frac{1}{\sigma_c^t + 1}} + \left[1 - \frac{1}{1 + \frac{PD_c}{PE_c} \cdot \left(\frac{QE_c}{QD_c} \right)^{\frac{\sigma_c^t + 2}{\sigma_c^t + 1}}} \right] \cdot QD_c^{\frac{1}{\sigma_c^t + 1}} \right\}^{\left(\sigma_c^t + 1\right)}$$

Equation 5

where:

αt_c shift parameter for output transformation (CET) function,

δ_c^t share parameter for output transformation (CET) function,

σ_c^t elasticity of transformation for output transformation (CET) function,

ρ_c^t exponent $\left(1 < \rho_c^t < \infty\right)$ for output transformation (CET) function,

QX_c quantity of domestic output,

QE_c quantity of exports,

PE_c export price (in domestic currency).

The values of elasticities σ_c^q and σ_c^t for Armington and constant elasticity transformation functions determine the intensity of the substitution between domestic supply and imports (Armington CES function), and the intensity of

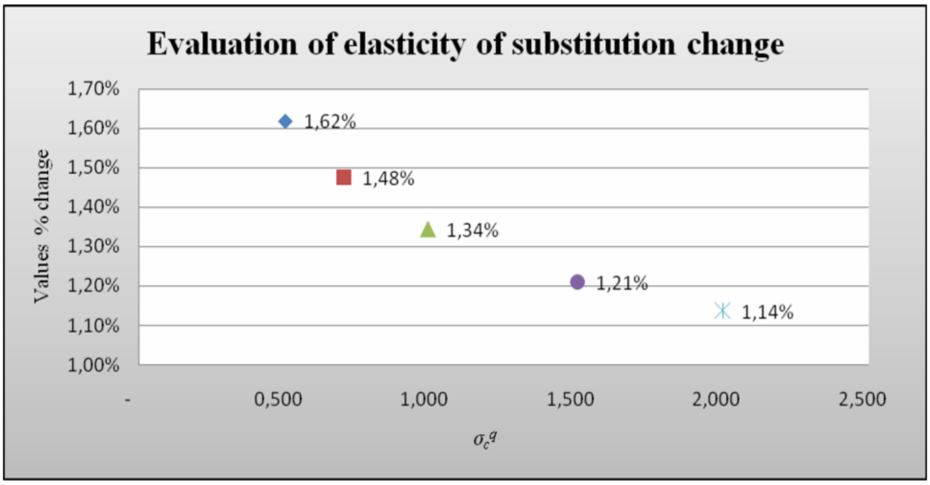
[1] There is a clear difference between equations for import demand and export supply: the quantity demanded of the imported commodity is inversely related to the import price and the quantity supplied of the exported commodity is directly related to the export price.

the transformation between domestic production and exports (CET function). To evaluate how sensitive the model is to differences in these trade elasticities an experiment with export price increase by 5% was carried out applying series of different elasticity variations as follows in the next chapter.

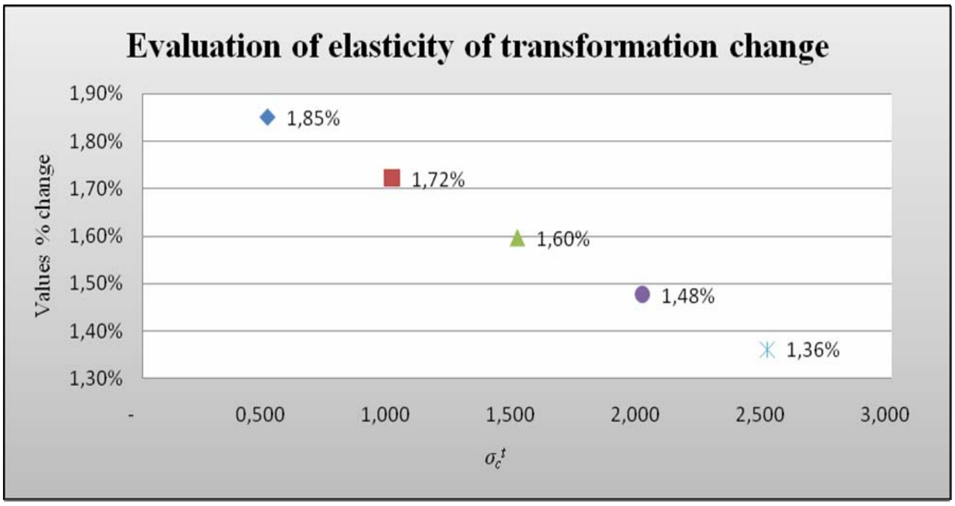
Experiments

Uniform values of elasticities of substitution and transformation were adopted as base in this applied CGE model. For value of elasticity of substitution in Armington function was set up 0.7 and for value of elasticity of transformation in CET function was set up 2.0.

In the first run the value of σ_c^q was gradually changed from 0.5 to 2.0 in the sector of automotives while the value in the sector of the aggregated rest of the world remained fixed. In the second run the value of σ_c^t was gradually changed from 0.5 to 2.5 in the sector of automotives while the value in the sector of aggregated rest of the world remains fixed again. The percentage deviations of values of aggregates as domestic consumption on automotives and on aggregated rest of commodities, GDP, government consumption on automotives, domestic and foreign price of automotives, CPI and other were obtained as shows the Picture 1, up to 1.7% for the first run and the Picture 2, up to 1.9% for the second run.



Picture 1: Graduate change of the value of the elasticity of substitution



Picture 2: Graduate change of the value of the elasticity of transformation

Conclusion

We can conclude that the model is not sensitive on changes in the Armington function elasticity of substitution value and CET function elasticity value according to this experiment. The changes in both cases were acceptable, in limit up to 2%. Notice that the higher elasticity values the lower percentage changes which can be explained by the higher flexibility of the model at these values.

References

Kazybayeva, S., Tanyeri –Abur, A. (2003) Trade Policy Adjustments in Kazakhstan: A General Equilibrium Analysis. Paper presented at the International *Conference Agricultural policy reform and the WTO: where are we heading?* Capri.

Khan, H.A. (2004) Using Macroeconomic Computable General Equilibrium Models for Assessing Poverty Impact of Structural Adjustment Policies. *ADB Institute Discussion Paper* No.12. Tokyo.

McDaniel, C., Balistreri E. (2003) A review of Armington trade substitution elasticities. *Économie internationale* 94-95, p. 301-313.

Miťková, V., Mlynarovič, V. (2007) *Modeling Policy Changes with the CGE Model for Open Economy of Slovakia*. International conference: Improforum. České Budejovice.

Zhang, X.G. (2006) Armington Elasticities and Terms of Trade Effects in Global CGE Models. *Productivity Commission Staff Working Paper*. Melbourne.

PORTFOLIO OPTIMIZATION IN TACTICAL ASSET ALLOCATION

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

Abstract

Although the strategic asset allocation is the most crucial decision required to achieve investment goals, the historical evidence suggests that tactical allocation process may offer good opportunities to enhance long – term portfolio return. The paper presents applications of portfolio selection model in a tactical asset allocation. At first by a combination of a momentum strategy and Markowitz model a dynamized investment strategy is constructed. In the second approach the momentum strategy is combined with Omega function which employs all information contained within the return series and can be used to evaluate and rank portfolio asset. Resulting portfolios are confronted with some standard approaches.

Key words: Tactical asset allocation, momentum strategy, dynamized Markowitz strategy, omega function

INTRODUCTION

Asset allocation is a crucial decision required to achieve investment goals ant its basic goal is to decide what classes of assets and in what proportions should be included in the investment portfolio. The decision on strategic asset allocation is an assumption for solving problems connected with an active asset allocation or for decisions on such changes in asset weights that exploit growing markets and good economic development for specific subcategories of the specific one. Historical data indicate that tactical asset allocation can provide opportunities for a significant increasing in portfolio returns

A basic approach to portfolio optimization from a viewpoint of basic asst classes, money market tools, bonds and equities, and its effective application based on the efficient frontier modeling and the choice of the best compromise portfolio according to the specific rule that enables to specify investor preferences on the base of his (her) risk attitude is built on the Markowitz portfolio selection model. In this case the efficient frontier modeling requires to solve a sequence of mathematical programming problems that one can write as

$$eff\ \left\{ \mathbf{E}^T \mathbf{w}; \Gamma(\mathbf{w}) \right\}$$

subject to

$$\mathbf{e}^T \mathbf{w} = 1$$
$$\mathbf{w}^l \le \mathbf{w} \le \mathbf{w}^u$$

where

$\Gamma(\mathbf{w})$ **-** scalar risk measure**,**
w **-** vector of asset weights
E - vector of expected returns,
$\mathbf{w}^l$ - vector of lower bounds on asset weights,
$\mathbf{w}^u$ - vector of upper bounds on asset weights,
e - vector of ones

and starting from known results (Zeleny (1982); Konno - Waki - Yuuki (2002)) one can write

$$\Gamma(\mathbf{w}) = \begin{cases} \Gamma(\mathbf{w}, c, \alpha, \lambda) = \Gamma(c, \alpha, \lambda) = \left[\sum_{r_k \le \lambda} \left| \mathbf{r}_k^T \mathbf{w} - c \right|^{\alpha} p_k \right]^{\frac{1}{\alpha}}, \quad \alpha > 0 \\ CVaR_{\beta}(\mathbf{w}, \beta) = \dfrac{1}{1-\beta} \mathrm{E}\left[-(\mathbf{r}^T \mathbf{w}) \middle| -(\mathbf{r}^T \mathbf{w}) \ge VaR_{\beta}(\mathbf{w}) \right] \end{cases}$$

where p_k is the probability of kth level $\mathbf{r}_k^T\mathbf{w}$ of portfolio return, c is a reference level of wealth from which deviations are measured. For example c could represent expected return of the asset, zero, the initial wealth level, the mode, the median, etc. Parameter α is the power to which deviations are raised, and thus α reflects the relative importance of large and small deviations. Parameter λ specifies what deviations are to be included in the risk measure if α > 0. Possible choices for parameter λ include ∞, c, a desired target level return, and some others. Conditional value at risk (*CVaR*) is an alternative measure of risk which maintains advantages of *VaR*, yet free from computational disadvantages of *VaR*, where β ,0 < β < 1, is the confidence level.

Let us note that there exist *VBA procedures* (Jackson – Staunton, 2001; Mlynarovic, 2005) for an effective execution of the solution process in

Excel environment which provide approximation of efficient frontier. From the view a practical application there are two main problems:

how the select the particular risk measure function,
how the estimate the expected return of assets

For $\alpha = 2$, $\lambda = \infty$ and $c = \mathbf{E}^T$w we have the model of portfolio selection in the mean – variance space that 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.

Questions on an application of different risk measures are examined by various authors (e.g. Konno-Waki-Yuuki, 2002; Miťková-Mlynarovič, 2006). A subject of interest is also question connected with a sensitivity of model solutions on inputs, e.g. expected returns. That is to say that the second restrictions on mean – variance approach consists in the fact that its recommendations for asset allocation are highly sensitive to small changes in inputs, means, in errors in estimations. It is assumed that errors in expected return estimations are 10-times more important than errors in variance estimations and 20-times more important than errors in covariance estimations. One of approaches how to solve this problem was developed by *Fisher Black* a Robert Litterman (Black – Litterman, 1992). It consists in a construction of quantitative approach for finding stabile portfolios in mean – variance space.

In this connection as a perspective approach we suggest such a modification of Markowitz model that in specified boundaries for risk looks for asset allocations with equal risk attribution of individual assets to the total portfolio risk through a goal programming application. The problem of asset allocation with the same risk contribution, or decomposition, of the total portfolio risk formally can be written in the form:

$$\min \quad \sum_{i=1}^{n} p_i + n_i$$

subject to

$$\begin{aligned} w_i(\mathbf{Cw})_i + n_i - p_i - \frac{1}{n}\mathbf{w}^T\mathbf{Cw} = 0, \quad i = 1, 2, \ldots, n \\ \mathbf{e}^T\mathbf{w} = 1 \\ \mathbf{w}^l \le \mathbf{w} \le \mathbf{w}^u \\ \sigma_l^2 \le \mathbf{w}^T\mathbf{Cw} \le \sigma_u^2 \\ p_i n_i = 0, \ \ p_i, n_i \ge 0, \quad i = 1, 2, \ldots, n \end{aligned}$$

where $\mathbf{C}$ is covariance matrix, σ_l^2 and σ_u^2 is the lower and upper bound fot the portfolio risk, n_i and p_i, i = 1, 2, ..., n, are so called negative and positive deviations, with possible nonnegative weights in the objective function, and n is the number of assets.

A MARKOWITZ STRATEGY DYNAMIZATION

As it was mentioned above Markowitz`s portfolio selection model is usually used for long term strategic asset allocation, but no for short term allocations. *Riberio* a *Loyes* (2007) just have shown that Markowitz optimization in a combination with so called momentum based tactical asset allocation can provide significant value added in the process of tactical investment strategy construction.

Most investor base tactical asset allocation decisions on consideration of value but lately also momentum, speed, in relative return across a wide set of asset classes is frequently used. In the most basic form, the strategy invests each 6 months in 5 asset classes, out of a possible 10, choosing those asset classes with the highest returns over the past 6 months. Momentum is an empirical phenomenon that contradicts market efficiency. In an efficient market, it should not be possible to build a profitable trading strategy without moving into riskier assets. But at now even well-known guardians of the efficient market hypothesis, such as *Eugen Fama* (Fama, E. and K. French, 1996), recognize the possibility that momentum profits could be due to a market inefficiency. There could be other explanation, even base on a rational framework, and there are a lot of theories that try to explain momentum. Arguments of behavioral finance are considered as the most convincing and suggest as reason for momentum underreaction and overreaction to information.

In the under reaction case, investors are unable to process available information in a timely fashion. Thus, security prices undereact to new information. In that case, prices will slowly adjust in the direction on intrinsic value, producing short-term trends. For example, new information arrives that tell investors that fundamentals are better than expected. But investors are suspicious and do not adjust their expectations fully. Prices move less than fundamentals would imply. Consequently, prices will only slowly absorb the positive past information. The overreaction story is also based on other investors` cognitive biases. Investors learn that fundamentals are better than expected. Most of the investors adjust their expectations fully, but some of the investors extrapolate this positive news into the future. Prices increase as fundamentals would predict, but they continue the upward move beyond fundamentals because of extrapolation or even trend following behavior. Both behavioral biases will make prices deviate persistently from intrinsic value.

The performance of the dynamic Markowitz strategy, which is the result of Markowitz optimization and momentum based tactical allocation combination, e.g. in comparison with a naïve strategy based on the momentum strategy, comes from the following sources:

- momentum in asset class returns – as best performing asset classes in the recent past are also more likely to outperform in the near future,
- persistence (clustering) in asset class volatility and correlation,

- stability in total risk exposure – as we can combine asset classes to maintain a reasonably constant total volatility, thus introducing a volatility timing feature in the strategy.

Table. 1: Annualized asset characteristic chosen by the momentum strategy

	R1 securities	R2 securities	R3 securities	R4 securities	O1 bonds	O2 bonds	Como-dities	Realities	Risk Free
Average return	**-2.63%**	**-12.14%**	**8.64%**	**42.81%**	**11.36%**	**10.41%**	**23.81%**	**-13.77%**	**4.62%**
St. deviation	**19.09%**	**18.41%**	**21.36%**	**24.07%**	**2.97%**	**3.85%**	**14.03%**	**20.89%**	**0.16%**
Lower semi st. deviation	**21.11%**	**20.82%**	**22.89%**	**26.94%**	**2.79%**	**4.67%**	**14.17%**	**22.44%**	**0.08%**
VaR, (0.95)	**36.90%**	**28.38%**	**39.19%**	**37.83%**	**3.83%**	**5.55%**	**21.11%**	**33.29%**	**-0.19%**
Conditional VaR	**42.10%**	**42.84%**	**46.65%**	**58.06%**	**5.31%**	**9.99%**	**30.73%**	**44.53%**	**-0.18%**
Parametric VaR	**31.40%**	**30.29%**	**35.13%**	**39.59%**	**4.89%**	**6.33%**	**23.08%**	**34.36%**	**0.27%**
Omega ratio (11%)	**0.89**	**0.81**	**0.99**	**1.17**	**1.01**	**0.94**	**1.12**	**0.83**	**0.00**

Table. 2: Investment strategies

Portfolio	R1	R2	R3	R4	O1	O2	Comm-odities	Reali-ties	Risk Free
Omega maximum	**0.00%**	**0.00%**	**0.00%**	**25.00%**	**25.00%**	**25.00%**	**25.00%**	**0.00%**	**0.00%**
Momentum_Markowitz	**0.00%**	**0.00%**	**0.00%**	**25.00%**	**25.00%**	**8.80%**	**25.00%**	**0.00%**	**16.20%**
Naivne	**11.11%**	**11.11%**	**11.11%**	**11.11%**	**11.11%**	**11.11%**	**11.11%**	**11.11%**	**11.11%**
GMV	**0.71%**	**0.00%**	**0.18%**	**0.00%**	**25.00%**	**23.81%**	**0.30%**	**0.00%**	**50.00%**
Risk decomposition	**7.69%**	**17.34%**	**7.27%**	**4.25%**	**0.00%**	**25.00%**	**12.47%**	**5.06%**	**20.92%**

Table 3: Investment strategies characteristics

	Portfolio				
	Omega	M_M	Naivne	GMV	Risk_dec
Average return	21.43%	20.38%	6.904%	7.64%	5.13%
St. deviation	8.23%	7.82%	9.869%	1.01%	8.00%
Lower semi st. dev.	10.01%	9.39%	11.129%	1.16%	9.06%
VaR (0.95)	12.69%	12.15%	17.18%	1.15%	11.60%
Conditional VaR	18.97%	17.96%	21.83%	2.05%	18.30%

Parametric VaR	13.53%	12.86%	16.23%	1.67%	13.16%
Omega ratio (11%,)	1.16	1.15	0.94	0.60	0.89

In practical application the daily data on asset returns for period from July 2, 2007 to December 31, 2007 were used and as a result momentum approach the asset presented in the Table 1 were chosen. So as the result we have four regional security indices, two regional bonds, one commodities index and one realities index. These data were completed with one tool of the money market that represents a risk free rate. This process realizes the first principle of the dynamized Markowitz strategy. The second principle then consisted in the modeling of the efficient mean-variance frontier for 6-months moving returns and risks and as the final Markowity –Momentun strategy the efficient portfolio with 8%, p.a. risk was chosen. The structure of this startegy (portfolio) and its comparison with the naïve momentum strategy, the global minimum variance portfolio, the portfolio with the same risk decomposition and with omega portfolio, its construction we will examined later, are presented in the Table 2. The Table 3 then presents corresponding characteristics of the portfolios.

OMEGA FUNCTION AND PORTFOLIO OPTIMIZATION

Many of the difficulties one can encounter in performance measurement and attribution of financial assets or their portfolios are rooted in the over – simplification that mean and variance fully describe distribution of returns. It is a generally accepted stylized fact of empirical finance that few, if any, would now challenge that returns from investment are not distributed normally. Thus in addition to mean and variance higher moments are required for a complete description.

In the last decade one can see many approaches to the portfolio and securities return analyzes and explanations that try to capture an asymmetry in investments returns. *Kearing, Shadwick* a *Cascon* (Keating, C., Shadwick, W.F., 2002a, 2002b; Cascon, A., Keating, C. and Shadwick, W. F., 2002) suggested so called *Omega* measure (function), which employs all the information contained within the return series and can be used for evaluation and ranking of portfolio assets.

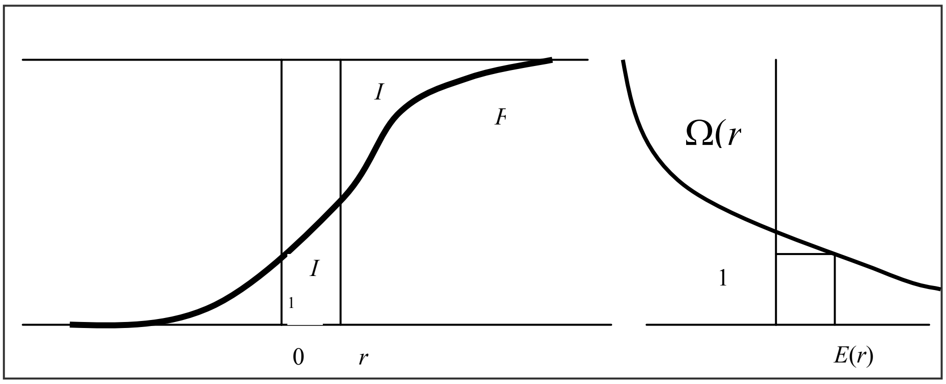
The approach based on the mathematical technique, which facilitate the analysis of returns distributions. This approach enables partitioning returns into loss and gain above and below a return threshold and then considering the probability weighted ratio of returns above and below the partitioning. The formally the measure is defined in the form

$$\Omega(r) = \frac{\int_r^b (1 - F(x))dx}{\int_a^r F(x)dx}$$

where (*a, b*) is the interval of returns and *F* is the cumulative distribution of returns. It is in other words ratio of the two areas I_2 and I_1 shown in the Picture 1a for defined loss threshold *r*, where I_2 is the area above the function *F* on the right from *r* and I_1 is the area below the function *F* on the left from *r*. If we assume this ratio for all possible return thresholds, we will have the function, illustrated in the Picture 1b, which is a characteristic one for assumed asset or portfolio.

The *Omega* function possesses many pleasing features (Cascon, A. C. Keating, W. Shadwick, 2001) that can be intuitively and directly interpreted in financial terms. As it is illustrates in the Picture 1b, *Omega* takes the value 1 when *r* is the mean return. An important feature of the measure is that it is not plagued by sampling uncertainty, unlike standard statistical estimators, as it is calculate directly from the observed distribution and requires no estimates. This function is, in a rigorous mathematical sense, equivalent to the returns distribution itself, rather than simply being an approximation to it. It therefore omits none of the information in the distribution and it is as statistically significant as the return series itself.

As a result, *Omega* is a perspective tool for financial performance measurement where what is of interest to the practitioner is the risk and reward characteristics of the return series. This is the combined effect of all its moments, rather then the individual effects of any of them, which is precisely what *Omega* provides



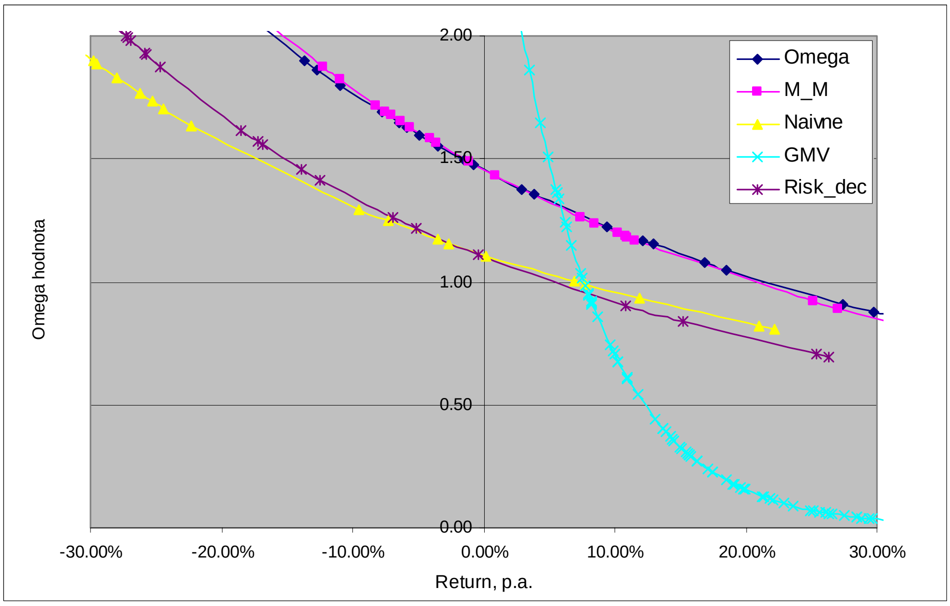
Picture 1a -b: Omega ratio and Omega function

If E_P is the required portfolio return, then as a further approach to portfolio selection can be used just the one, which for the stated performance threshold E_P looks for such portfolio that maximizes *Omega* function value. The problem can be formally written in the form

$$\max \quad \Omega(\mathbf{w}, E_P)$$

subject to

$$\mathbf{e}^T \mathbf{w} = 1$$



Picture 2: Omega functions for resulted portfolios

$$\mathbf{w}^l \leq \mathbf{w} \leq \mathbf{w}^u$$

$$\Omega(\mathbf{w}, E_P) = \frac{\int_{E_P}^{b} (1 - F(\mathbf{w}, x))dx}{\int_{a}^{E_P} F(\mathbf{w}, x)dx}$$

The resulted *Omega* portfolio for the performance threshold 11%, p.a. is compared with the other ones in the Table 2 and 3. An interesting information also provides the Picture 2, which presents *Omega* functions for assumed portfolios

REFERENCES

BLACK, F. – R. LITTERMAN (1990): Global Portfolio Optimization. Financial Analysts Journal, vol. 48, no. 5, 28 - 43

FAMA E. and K. FRENCH**:** Multifactor Explanations of asset Pricing Anomalies, Journal of Finance, January 1996

CASCON, A., C., KEATING, and W. F. SHADWICK: The Mathematics of Omega measure. The Finance Development centre, 2002

KEATING, C. and W. F. SHADWICK: A Universal Performance measure. The Finance Development centre, 2002a

KEATING, C. and W. F. SHADWICK: An Introduction to Omega. The Finance Development centre, 2002b

KONNO, H. - WAKI , H.–A. YUUKI (2002): Portfolio Optimization under Lower Partial Risk Measures. *Asia/Pacific Financial Markets* **9**, 127-140

MARKOWITZ, H.: Portfolio Selection, Journal of Finance, Vol. 7, pp.77-91, 1952

MIŤKOVÁ, V. – MLYNAROVIČ, V.: Structured Products in a Strategic Portfolio Selection. *Firma a konkurenční prostředí 2006*. Brno: MSD, International Scientific Conference, p.144 - 158

MLYNAROVIČ, V. (2001): Finančné investovanie. Teória a aplikácie. IURA Edition, Bratislava, ISBN 80-89047-16-5.

RIBERIO, R. and J. LOEYS: Markowitz in Tactical Asset Allocation. Global Market Strategy, J.P. Morgan Securities Ltd. August 7, 2007

ZELENY, M. (1982): Multiple Criteria Decision Making. McGraw Hill, New York.

VLADIMÍR MLYNAROVIČ, The Faculty of Social and Economic Sciences, Comenius University in Bratislava, Odbojárov 10/a, 820 05 Bratislava, e-mail: vladimir.mlynarovic@fses.uniba.sk

TWO APPROACHES TO VEHICLE AND CREW SCHEDULING IN URBAN AND REGIONAL BUS TRANSPORT

Stanislav Palúch
Faculty of Management Science and Informatics
University of Žilina, Slovakia

Abstract
There are two different approaches to vehicle and crew scheduling in urban and regional bus transport. In Slovak and Czech Republic simultaneous bus and crew scheduling is used. This approach supposes that drivers cannot change vehicles and hence all constraints required for drivers working time have to be imposed to running boards of vehicles. Constraints and conditions for such running boards can be formulated as nonlinear objective functions and the resulting task is a nonlinear multi criteria optimization problem. A good suboptimal algorithm was developed for this problem. In Great Britain (and many other countries) a two stage scheduling procedure is used. In the first step an optimum vehicle schedule is calculated. In the second one a large set of possible driver shifts is created and then an optimum subset is selected using set covering problem. The comparison of two mentioned approaches shows that the second approach can save dead mileage and total driver working time.

This paper was supported by VEGA as a grant No. 1/0135/08.

1. The Fundamental Vehicle Scheduling problem

Trip s is a travel from a starting point to a finishing point of a route and is considered to be an elementary amount of the work of a bus. We will say that the trip s_i *precedes* the trip s_j and we will write $s_i \prec s_j$ if the trip s_j can be linked after the trip s_i into a running board for one bus. Relation $\prec$ is irreflexive and transitive.
Running board, or *running board of a bus* is an arbitrary nonempty sequence
$T = s_1, s_2, \ldots, s_m$ of trips with property: $s_1 \prec s_2 \prec \cdots \prec s_m$.

The number m we will call *the length of running board T*. We will write $s_i \to s_j$ if both trips s_i, s_j are provided by the same bus and the trip s_i is linked immediately behind the trip s_j. Note that $s_i \to s_j$ implies $s_i \prec s_j$. For any given set $S = \{s_1, s_2, \ldots s_n\}$ of trips with precedence relation $\prec$ we can construct a bus schedule. *Bus schedule* of the set S of trips is a set of running boards $O = \{T_1, T_2, \ldots T_k\}$ of the form

$$T_1 = s_{1,1} \to s_{1,2} \to \ldots \to s_{1,n(1)-1} \to s_{1,n(1)}$$

$$T_2 = s_{2,1} \to s_{2,2} \to \ldots \to s_{2,n(2)-1} \to s_{2,n(2)}$$

....

$$T_k = s_{k,1} \to s_{k,2} \to \ldots \to s_{k,n(k)-1} \to s_{k,n(k)}$$

such that every trip of the set S occurs exactly in one running board of O. To every bus schedule $O = \{T_1, T_2, \ldots, T_k\}$ the number $C(O)$ - the cost of bus schedule O is assigned. We will say that the cost $C(O)$ is separable if it can be expressed as a sum of costs of all running boards T_1, T_2, ... T_k, i. e.

$$C(O) = \sum_{i=1}^{k} c(T_i) \tag{3}$$

where $c(T_i)$ denotes the cost of running board T_i. In most simple cases the cost of running board $T = s_1 \to s_2 \to \ldots \to s_{m-1} \to s_m$ is the sum of all linkage costs:

$$c_0(T) = \sum_{i=1}^{m-1} c(s_i, s_{i+1}) \tag{4}$$

The linkage cost denoted $c(s_i, s_j)$ represents namely dead mileage expenses, however, it may include waiting costs, line changing penalty and other penalties as well. In this case we will say that the cost $c_0(T)$ is linear. We will say that the cost $C(O)$ of bus schedule O is linear, if C is separable and the cost $c_0(T)$ of running board T is linear.

Most important objectives for vehicle scheduling are the following:

1. **O1:** Minimization of the number of running boards.
2. **O2:** Minimization of the total linkage cost.

Fundamental vehicle scheduling problem FVSP is to find a bus schedule with the minimum number of running boards and with the minimum total linear cost.

In the terms of bivalent programming the FVSP can be formulated as follows:

Let $d_{ij} = c(s_i, s_j)$ if $s_i \prec s_j$ and $d_{ij} = \infty$ otherwise. Let x_{ij} be a zero-one decision variable saying that if $x_{ij} = 1$ and $d_{ij} < \infty$, then s_i immediately follows s_j in some running board T. To solve FVSP means

$$\text{Minimize} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}.x_{ij} \tag{5}$$

$$\text{subject to} \quad \sum_{j=1}^{n} x_{ij} = 1 \quad \text{for } i = 1,2,...,n \tag{6}$$

$$\sum_{i=1}^{n} x_{ij} = 1 \quad \text{for } j = 1,2,...,n \tag{7}$$

$$x_{ij} \in \{0,1\} \tag{8}$$

We see that we got a classical assignment problem which can be solved in polynomial time. (Several related problem of fleet size minimization see in Černá [1], Černý [3] and Peško [5].)
Remark: We can change the definition of d_{ij} as follows: $d_{ij} = c(s_i,s_j)$ if $s_i \prec s_j$ and d_{ij} = the fixed costs of one bus per day. Then the optimization will give us also an optimum number *k* of vehicles.
Remark: If we set $d_{ij} = c(s_j, depot) + c(s_i, s_j) - c(depot, s_i)$ if $s_i \prec s_j$ and $d_{ij} = \infty$ otherwise where $c(depot, s_i)$ and $c(s_j, depot)$ are pull-out and pull-in expenses, the resulting bus schedule will be optimal with respect to objective

$$c_0(T) = c(depot, s_1) + \sum_{i=1}^{m-1} c(s_i, s_{i+1}) + c(s_m, depot)$$

i. e. we can get a bus schedule which minimizes total linkage cost plus total pull-out and
pull-in cost.

2. Two Step Vehicle and Crew Scheduling

The first step in this procedure is to solve the corresponding fundamental vehicle scheduling problem. Clearly, the resulting bus running boards in most cases don’t comply with conditions for driver working shift.
The second step described by Smith and Wren [12] (system IMPACS) was first used for London Transport Buses and for many other organizations later. The basic approach is to divide running boards (resulting from the first step) into so called pieces of work, then generate many thousand potential shifts by combining pieces of work, to eliminate some of the resultant shifts by intelligent heuristics, and then to use a integer linear programming process to select a subset of shifts which forms a good schedule.

Let a_{ij} is a real number, $a_{ij} = 1$ if the shift *i* contains the trip *j*, $a_{ij} = 0$ otherwise. Let x_i is a decision variable $x_i \in \{0,1\}$, $x_i = 1$ if and only if the shift *i* is selected shift, let c_i is the cost of the shift *i*. The total cost of selected shifts is $\sum_i c_i x_i$, the requirement, that every trip is covered by at least one shift can be expressed by the condition $\sum_i a_{ij} x_i \geq 1$ for every *j*. So the arising integer linear programming problem is

Minimize $\sum_i c_i x_i$ (*)

subject to $\sum_i a_{ij} x_i \geq 1$ for j=1,2,...,n (**)

$x_i \in \{0,1\}$ for every *j* (***)

Just described problem is known as the set covering problem. However it may happen that some trip is covered by more than one shift. This obstacle can be easily fixed by successive heuristic procedure which determines which shift will really contain that trip. Our computational experience with real instances shows that such situation occurs very seldom because of the total cost (*). The idea to replace the conditions (**) by $\sum_i a_{ij} x_i = 1$ leads to set partitioning problem – the resulting shifts doesn't overlay in this case, but our computational experience shows that in some cases the resulting problem hasn't a feasible solution at all or, if it has, this solution is much worse than that of the set covering problem.

2. Simultaneous Vehicle and Crew Scheduling Problem (SVCSP)

In Czech and Slovak Republic there are closer ties between drivers and vehicles. One bus schedule has to be covered by one or two crew schedules. In this case the considered bus running boards have to fulfill several conditions.

The most important and most frequent constraints for vehicle and crew scheduling are:

1. **C1:** All vehicles have to return after work to the places where they started in the morning. (As the starting points can be given one depot, several depots or this condition can be replaced by

the requirement that every bus returns after finishing the daily work to the starting point of corresponding running board.)

2. **C2:** All running boards have to fulfill the safety break condition. We will say that running board fulfills safety break condition (SB) or is feasible if in every time interval 240 minutes long there exists at least 30 minutes of safety break. This safety break can be in one continuous piece or two or three time intervals, every one of them is at least 10 minutes long.
3. **C3:** The time length of every running board must be suitable for one crew or two crews. This means that it must be in time interval $\left(t_1^1, t_2^1\right)$ if it is designed for one driver or in interval $\left(t_1^2, t_2^2\right)$ for two drivers.
4. **C4:** All vehicles have to visit the central depot once a day.
5. **C5:** All driver shifts have to contain a meal break.
6. **C6:** The driving time of every driver is limited. (This is different requirements from that of **C3**)
7. **C7:** The durations of all shifts of all drivers have to be as uniform (as equal) as possible.

Assume that $T = s_1 \to s_2 \to ... \to s_{m-1} \to s_m$ and $O = \{T_1, T_2, \dots, T_k\}$. Just mentioned constraints can be expressed by corresponding objective functions.

The constraint **C1** can be modelled by objective function

$$c_1(T) = c(s_m, s_1),$$

where $c(s_m, s_1)$ represents all expenses connected with the move of the vehicle from arrival place of the trip s_m to departure place of the trip s_1.

The modeling of constraint **CI** for $\mathbf{I} = 2,3,4,5,6$ can be achieved by the following objective function $c_i(T)$): $c_i(T) = 0$, if T complies with condition **CI**

$$c_i(T) = \infty \quad \text{otherwise}$$

The objective function for condition **C7** is

$$c_7(O) = \sum_{i=1}^{k} \left(duration(T_i)\right)^2$$

Simultaneous vehicle and crew scheduling problem SVCSP can be formulated a multicriteria assignment problem with objectives

$\sum_i c_0(T_i)$, $\sum_i c_1(T_i)$, $\sum_i c_2(T_i)$, $\sum_i c_3(T_i)$, $\sum_i c_4(T_i)$, $\sum_i c_5(T_i)$, $\sum_i c_6(T_i)$, $c_7(O)$

All mentioned objectives are separable, all except the first one are nonlinear. Their formulation by means of mathematical programming is rather complicated. We can define the complex objective function

$$C(O) = \sum_{r=0}^{6} K_r \sum_{i=1}^{k} c_r(T_i) + K_7 c_7(O),$$

where K_i for $i = 0,1,2,3,4,5,6,7$ are some suitable constants, and to formulate SVCSP as to find a bus schedule with minimum value of $C(O)$. Arising mathematical problem is probably NP-hard and therefore suboptimal methods are to be used for solving SVCSP. Several neighborhood search procedures were designed, implemented and successfully applied in practice. As the most successful heuristic method appears the one based on the following two step procedure:

1. Step. Compute optimum bus schedule O board with minimum number of vehicles and minimum total linkage cost $\sum_i c_0(T_i)$.
2. Step. Neighborhood search heuristic starting with bus schedule O trying to find a better solution with lower value of complex objective function $C(O)$ by combination of heads, bodies, and tails of running boards leading to numerous applications of matching problem.

The objective functions $c_i(T)$ for $i = 2,3,4,5,6$ appeared not to be suitable for just described heuristic algorithm, since they do not express „how far“ the actual solution O is from the ideal one. Therefore it is useful to modify the definition of the objective function $c_i(T)$ for $i = 2,3,4,5,6$ as a measure of violating the corresponding constraint **CI.** Practical experiences showed that the choice of constants K_i for $i = 0,1,2,3,4,5,6,7$ is crucial for final result and how improper values of these constants can lead to curious unexpected results.

3. Practical Experiences and Results

Simultaneous vehicle and crew scheduling was used for optimization of several tens Czech and Slovak towns (the latest are Piešťany, Trenčín, Považská Bystrica a Martin -- Vrútky) with great economical gain. The dead

mileage savings were from 6 to 20 %, the number of vehicles dropped by up to 15 %.
The comparison with two step approach used in Great Britain modified for Slovak conditions showed that in all cases we have got additional savings of dead mileage from 20% up to 70%, while the number of drivers has risen slightly. Since the research is not finished in this area we can expect even better results. However, there is a little hope that Slovak bus providers will apply this approach in short future because of poor technical state of vehicles, low professionality of drivers and long term tradition – all mentioned facts preclude changing various drivers on one bus.

References.

[1] Černá, A.: Heterogenous Bus Fleet Expliotation in Regional Bus Transport, Slovak,
(Využitie heterogenneho autobusového parku v mestskej a regionálnej doprave),
Proceedings of the 5-th International Conference on Public Transport, Dom techniky,
Bratislava, 21. 22.11.2001, pp. 93- 96.

[2] Černá, A.: Optimization of Regional Bus Transport, Czech, (Optimalizace regionální
autobusové dopravy.) Proceedings of International Conference "Transportation Science",
(Věda o dopravě), Fakulta Dopravní ČVUT Praha, 6.-7.11.2001, pp. 70-75.

[3] Černý, J., Černá, A.: Teorie řízení a rozhodování v dopravních systémech – Theory
of Management and Decision Making in Transportation Systems, Jan Perner Institute,
Prague, 2004, ISBN 80-86530-15-9

[4] Peško, Š.: Muticomodity Return Bus Scheduling Problem. Proceedings, International
Scientific Conference on mathematics, pp. 77-82, Žilina, ISBN 80-7100-578-9, (1998)

[5] Peško,Š.: The Minimum Fleet Size Problem for Flexible Bus Scheduling,
Studies of the Faculty of Management Science and Informatics,
Vol. 9, University of Žilina, 59-65, (2001)

[6] Palúch, S.: Simultaneous Bus and Crew Scheduling. International Scientific
Conference on Mathematics, June 30. – July 3., 1998

[7] Palúch,S.: Bus Scheduling as a Graph Coloring Problem, Komunikácie, Vedecké
listy Žilinskej univerzity 5. ročník,4/2003, pp.16-20, ISSN 1335-4205

Stanislav Palúch – Department of Mathematical Methods, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1 01026 Žilina, Slovakia Phone: +421 41 5134250, e-mail: paluch@fria.fri.uniza.sk

PICKUP AND DELIVERY PROBLEM

Pelikán Jan, Fábry Jan, Kořenář Václav
University of Economics Prague, W. Churchill sq. 4, 130 67 Prague 3

Abstract
Vehicle routing problem and traveling salesman problem are classical problems in operational research, this modification of those problems consists of a transport among nodes of the communication network using cyclical routes of vehicles with given capacity. A transportation demand is given by the place of pickup, the place of delivery and quantity of goods. The goal is to find cyclical routes of minimal length which ensure the transport requirements. In the paper there are two models proposed for the problem, both are demonstrated on an example. The problem is based on a case study from practice.

Keywords
Pickup and delivery problem, integer programming, heuristic methods

1 Introduction

Traveling salesman problem and vehicle routing problem and their modifications are frequently solved practical examples. In such problems the goal is to assure pick-up or/and delivery of goods in a distribution network. While in those applications it is necessary to find cyclical routes starting and ending in a depot, in the studied problem it is required to transport goods between nodes of the network. Each requirement is specified by the pickup point, the delivery point and the amount of goods which has to be transported. In [1] and [2] the problem is called pickup and delivery problem.

In the practice time windows are often given for each node in the network and the order of pickup and delivery processes has to be defined, because the goods being loaded as the last one will be unloaded as the first one. The origin and the destination nodes need not to be identical. Vehicles with different capacities can assure the transport of goods. The objective is to minimize the total transportation cost. Column generation method is used for selecting routes in the mathematical model.

In the paper, no time windows and no conditions for order of pickup and delivery are considered. All vehicles have the same capacity. Each route has to be cyclical for all included nodes. Two models are proposed in the paper;

the first one is based on the optimal flow theory, the second one uses a model of set covering problem.
Let a distribution network is given by $G = \{V, E\}$, where V is a set of n nodes and E is a set of undirected arcs. Each arc (i, j) is evaluated by minimal distance between nodes i and j. Let us denote q_{kl} the amount of goods that has to be transported from node k to node l. Vehicles with capacity V are used for pickup and delivery and they can start in any node. All routes have to be cyclical, each vehicle has to come back to the node it starts from. The objective is to minimize the length of all the routes.

Example
Let us consider the distribution network with 4 nodes; the distance matrix C and the transportation requirements matrix Q are given:

$$C = \begin{bmatrix} 0,30,40,60 \\ 30,0,60,50 \\ 40,60,0,20 \\ 60,50,20,0 \end{bmatrix} \quad Q = \begin{bmatrix} 0,0,7,5 \\ 0,0,5,7 \\ 5,6,0,0 \\ 8,0,8,0 \end{bmatrix}.$$

Vehicle capacity is V=12.

2 Optimal multi-product flow model

Let us define two additional nodes in the network: the node 0 as the source and the node n+1 as the sink. The following variables are defined in the model:

$y_{ij} \geq 0$, integer – number of vehicles going through the arc (i, j) in the direction from i to j $(i, j = 0,1,...,n+1,\ i \neq j)$,

$x_{ij}^{kl} \geq 0$, amount of goods (part of the total amount q_{kl}) transported from node i to node j $(i, j = 0,1,...,n+1,\ i \neq j; k, l = 1,2,...,n,\ k \neq l)$.

Mathematical model follows:

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} y_{ij} \rightarrow \min \tag{1}$$

$$\sum_{i=1}^{n} y_{ij} = \sum_{i=1}^{n} y_{ji},\ j = 1,2,...,n, \tag{2}$$

$$\sum_{i=0}^{n} x_{ij}^{kl} = \sum_{i=1}^{n+1} x_{ji}^{kl},\ j = 1,2,...,n; k, l = 1,2,...n, k \neq l, \tag{3}$$

$$\sum_{k,l} x_{ij}^{kl} \le V y_{ij}, i, j = 1,2,\ldots,n, i \ne j, \quad (4)$$

$$x_{0,k}^{kl} = q_{kl}, k,l = 1,2,\ldots,n, k \ne l; \quad x_{0,j}^{kl} = 0, k,l = 1,2,\ldots,n, k \ne l; j = 1,2,\ldots,n+1, j \ne k,$$
$$x_{l,n+1}^{kl} = q_{kl}, \; k,l = 1,2,\ldots,n, k \ne l; \quad x_{j,n+1}^{kl} = 0, k,l = 1,2,\ldots,n, k \ne l; j = 0,1,\ldots,n, j \ne l, \quad (5)$$

$$x_{ij}^{kl} \ge 0, i, j = 0,1,\ldots,n+1, i \ne j, \; k,l = 1,2,\ldots,n, k \ne l,$$
$$y_{kl} \ge 0, \text{ integers}, \; k,l = 1,2,\ldots,n, \; k \ne l. \quad (6)$$

The objective (1) corresponds to the sum of the evaluations of all the arcs in the solution, i.e. the total length of all routes. Equations (2) assure the vehicle will leave the location that it will visit. With respect to equations (3), amount of goods being transported from k to l entering node j leaves this node. Inequalities (4) disable exceeding the capacity of the vehicle transporting goods between nodes i and j. Equations (5) assure that both the total flow from the source 0 to node k and the total flow from node l to the sink $n+1$ are equal to the total requirement q_{kl}. All other flows from the source and to the sink are set to 0.

Example

Application of the model (1) - (6) to the example introduced above leads to the objective value equal to 260 and to the following values of variables:

$y_{12} = y_{13} = y_{24} = y_{34} = 1, \; y_{31} = y_{43} = 2$.

Hence, two routes are generated:

route A: $2 \to 4 \to 3 \to 1 \to 2$ of the length 140

and route B: $1 \to 3 \to 4 \to 1$ of the length 120.

Optimal solution:

Route A			Route B		
Node	Pickup	Delivery	Node	Pickup	Delivery
2	q_{24}=7 q_{23}=5		1	q_{14}=5 q_{13}=7	
4		q_{24}=7	3		q_{13}=7
3	q_{31}=5 q_{32}=6	q_{23}=5	4	q_{41}=8	q_{14}=5
1		q_{31}=5	1		q_{41}=8
2		q_{32}=6			

Note: Values of matrix Y provide information about the arcs that are included in the optimal routes. Generation of routes based on this information need not to be unique. In addition it is necessary to determine the depot for each route. Selection of the depot has to enable flows to be realized. If node 4 is selected as the depot for the route A, it will not be possible to realize the requirement q_{23}=5. Thus, it is possible we are not able to complete routes with depots on the base of the optimal matrix *Y*.

3 Routes generation model

The model is based on the assumption there are proposed routes satisfying all conditions of transportation, depots in all proposed routes are determined. The goal is to select routes satisfying requirements given by matrix *Q* and minimizing the total length of all routes derived from matrix *C*. A number of vehicles realizing the transport will be determined as well.
Let us assume *S* routes including arcs of distribution network, the length of each route is denoted d_s (s = 1,2,…,S). Parameter $a_{ij}^{kl}(s)$ equals 1 if goods transported on the route *s* from node *k* to node *l* uses arc (i, j), 0 otherwise.

Example

Parameters $a_{ij}^{kl}(s)$ for the route A: $2 \to 4 \to 3 \to 1 \to 2$ presented in previous chapter are defined in the following table:

(*i,j*) / (*k,l*)	2,4	2,3	4,3	4,1	3,1	3,2
(1,2)						1
(3,1)				1	1	1
(2,4)	1	1				
(4,3)		1	1	1		

In the model, the following variables are used:

$y_s \geq 0$, integer – number of vehicles on the route *s* (s = 1,2,…,S),

$x_{kl}^s \geq 0$, amount of goods (part of the amount q_{kl}) transported on the route *s* $(k, l = 1,2,...,n,\ k \neq l;\ s = 1,2,...,S)$.

Mathematical model is:

$$z = \sum_{s=1}^{S} y_s d_s \to \min \tag{7}$$

$$\sum_{s=1}^{S} x_{kl}^s = q_{kl},\ k,l = 1,2,...n, k \neq l, \tag{8}$$

$$\sum_{k,l} a_{ij}^{kl} x_{kl}^{s} \le V y_s,\ i,j = 1,2,...,n, i \ne j,\ s = 1,2,...,S, \tag{9}$$

$$x_{kl}^{s} \ge 0, k,l = 1,2,...,n,\ k \ne l,\ y_s \ge 0, \text{ integer},\ s = 1,2,...,S. \tag{10}$$

The objective function (7) corresponds to the total length of all routes. Equations (8) assure transport of required amount of goods q_{kl} from node k to node l. Inequalities (9) disable exceeding the capacity of the vehicle transporting goods on the route s between nodes i and j. The key issue in this mathematical model is generation of the routes and their number. The same routes with the different depots are considered as different routes in this model.

References

[1] Savelsbergh, M., P., M., Sol., M.: The general pickup and delivery problem. *Transportation Res.* 29 , 1995,17-29.

[2] Hang, X., Zhi-Long, Ch., Rajagopal, S., Arunapuram, S.: Solving a practical pickup and delivery problem. *Transportation science* 2003, vol.37, no.3, 347-364.

[3] Pelikán, Jan: *Diskrétní modely v operačním výzkumu*. Professional Publishing.2001. ISBN 80-86419-17-7.

Authors address:
Pelikán Jan, Fábry Jan, Kořenář Václav
University of Economics Prague, W. Churchill sq. 4, 130 67 Prague 3
pelikan@vse.cz, fabry@vse.cz, korenar@vse.cz

FLEXIBLE BUS SCHEDULING WITH OVERLAPPING TRIPS

Štefan Peško

Abstract

We investigate simplified forms of the flexible bus scheduling problem for one bus. I tis motivated by practical needs of removing overlapping trips from the real bus schedule. The goal is to minimize the total overlaps of the sequence of trips of fixed running board. The possibility of using linear programming approach with integer solutions for optimistic and pessimistic models are considered.

1 INTRODUCTION

The topics discussed in this paper are focused on important operational problems confronting the management of competitive bus transportation system.

The basic bus scheduling problem consists of assigning buses to given set of trips in running board such that:

- each trip is performed exactly once,
- each bus must start and end its work day at the same depot,
- the number of buses is as low as possible minimum,
- the operational cost is minimum.

This problem has several variations with practical restriction (on number and types of depots, meal breaks, buses types, length of the running boards) studied in Slovakia by Palúch and his colleagues [1],[2],[3],[4],[5].

We will consider the Flexible Bus Scheduling Problem (FBSP) when the set of flexible trips is given for one bus (see table 1) and no practical restrictions are presumed. This paper aims to present and discuss mathematical programming formulations and the possibility of for solving the FBSP.

2 BASIC NOTATION

In this paper we will use the terminology from Hartley [6]. A flexible trip is a journey performed by bus between two places at specific time given by the flexible time-table.

Given set of flexible trips $S = \{S_1, S_2, ..., S_n\}$, each flexible trip S_i is represented by an arbitrary ordered quintuple

Departure place	time	Arrival place	time	Trip minutes	Deadhead minutes	Overlapping minutes
A	5:30	B	5:55	25	0	
B	6:00	A	6:23	23	0	3
A	6:20	B	6:45	25	5	
C	6:50	A	7:20	30	0	
A	7:30	C	8:00	30	0	5
C	7:55	A	8:25	30	0	2
A	8:23	B	8:48	25	0	3
B	8:45	A	9:08	23	0	
:	:	:	:	:	:	:

Figure 1: Running board of bus with overlapping trips

$$S_i = (m_i^d, t_i^d, m_i^a, \tau_i, \tau_i^-, \tau_i^+) \qquad (1)$$

where

m_i^d – the departure place,
t_i^d - the scheduled departure time,
m_i^a – the arrival place,
τ_i - the trip time,
τ_i^- - the most likely departure time,
τ_i^+ - the latest likely departure time.

We will assume that $\tau_i^- \le t_i^d \le \tau_i^+$. The bus models are of the discrete-time rather that of continuous-time variety. We will suppose that t_i^d, τ_i^-, $\tau_i^+ \in \{1,2,\dots,1440\}$. Note that for flexible trip S_i the scheduled arrival time is equal $t_i^d + \tau_i$, the likely arrival time is equal $\tau_i^- + \tau_i$ and the latest likely arrival time is equal $\tau_i^+ + \tau_i$.

Let $\delta_{i,i+1}$ the times of the idle trip from place m_i^a to place m_{i+1}^d are given. The running board of the bus τ with n flexible trips is sequence

$$\tau = S_1 \prec \ldots \prec S_i \prec S_{i+1} \prec \ldots \prec S_n,$$

where we write $S_i \prec S_{i+1}$ if the bus arrives at i-th trip could start (i+1)-th trip

$$\tau_i^- + \tau_i + \delta_{i,i+1} \leq \tau_{i+1}^+. \qquad (2)$$

The overlapping time z_i (see figure 2) between trips S_i and S_{i+1} with scheduled departure times t_i^d and t_{i+1}^d is defined by value

$$z_i = \max\ \{0, t_i^d + \tau_i + \delta_{i,i+1} - t_{i+1}^d\}. \qquad (3)$$

The problem is when in given running board τ exists trips S_i with positive z_i. It is possible to find real departure times such that running board of the bus is without overlapping times?

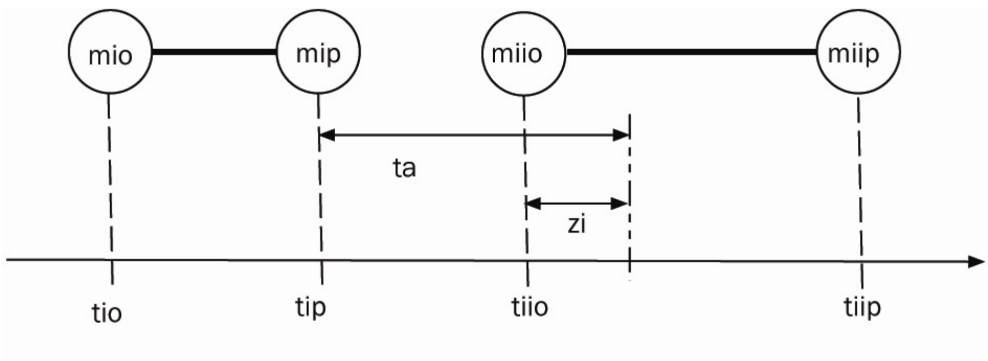


Figure 2: Overlapping time z_i between scheduled trips S_i and S_j

3 OPTIMISTIC OPTIMIZATION

The first we will optimistic assume that it is possible find real departure times

$$t_i : \tau_i^- \leq t_i \leq \tau_i^+,\ t = 1,2,\ldots,n, \qquad (4)$$

that given running board τ is without overlapping times i. e.

$$\max\{0, t_i, + \tau_i + \delta_{i,i+1} - t_{i+1}\} = 0 \text{ for } i=1,2,\ldots,n-1 \qquad (5)$$

However in practical conditions it is possible to find more solutions (4) conforming with constraints (3). One option is to find solution with minimum offsetof scheduled times and we obtain the following mathematical programming problem (OOP1):

$$\sum_{i=1}^{n} \max = |t_i - t_i^d| \;\rightarrow\; \min$$
$$t_i + \tau_i + \delta_{i,i+1} - t_{i+1} \leq 0 \quad i = 1,2,\ldots,n-1$$
$$\tau_i^- \leq t_i \leq \tau_i^+ \quad i = 1,2,\ldots,n$$

It is easy see that if we set $t_i = t_i^d - x_i^- + x_i^+$ where variable $x_i^- \geq 0$ and $x_i^+ \geq 0$ are negative, and positive ofset of the scheduled departure time than the OOP1 can be formuled as the following linear problem (OOP2):

$$\sum_{i=1}^{n} x_i^- + x_i^+ \;\rightarrow\; \min$$
$$x_i^+ + x_{i+1}^- - x_i^- - x_{i+1}^+ \leq t_{i+1}^d - t_i^d - \tau_i - \delta_{i,i+1}$$
$$0 \leq x_i^- \leq t_i^d - \tau_i^-$$
$$0 \leq x_i^+ \leq \tau_i^+ - t_i^d$$

Note that the problem OOP2 has always integer solution x_i^-, x_i^+ if exists and so the problem OOP1 has integer solution $t_i = t_i^d - x_i^- + x_i^+$.

4 PESSIMISTIC OPTIMIZATION

The models OOP1 or OOP2 don´t need to have a solution. Then we can pessimistic assume that it is possible find almost real departure times t_i only via the mathematical programming problem (POP1):

$$\sum_{i=1}^{n-1} \max\ \{0, t_i + \tau_i + \delta_{i,i+1} - t_{i+1}\} \;\rightarrow\; \min$$
$$\tau_i^- \leq t_i \leq \tau_i^+ \qquad i = 1,2,...,n$$

If we set $y_i^+ - y_i^- = t_i + \tau_i + \delta_{i,i+1} - t_{i+1}$ where variable $y_i^- \geq 0$ and $y_i^+ \geq 0$ are negative and positive ofset of the overlapping time than the POP1 can be formuled as the following linear problem (POP2):

n−1

$$
\begin{array}{rcll}
\sum_{i=1} y_i^+ & \rightarrow & \min & \\
y_i^+ - y_i^- - t_i + t_{i+1} & = & \tau_i + \delta_{i,i+1} & i = 1,2,...,n-1 \\
\tau_i^- \le t_i & \le & \tau_i^+ & i = 1,2,...,n \\
y_i^-, y_i^+ & \ge & 0 & i = 1,2,...,n-1
\end{array}
$$

Note that analogical as for the problem OOP2 the problem POP2 has always integer solution t_i, y_i^-,y_i^+ if exists and so the problem POP2 has integer solution t_i.

5 OPEN QUESTIONS

The bus companies schedule running boards of many buses. In the system KASTOR [5] exists a choice for minimum fleet size if fixed maximum overlapping of trips is feasible provisionally. Then scheduler handle with trips is necessary for design realistic running boards with computed fleet size.

It is possible to use solutions solving the optimistic and the pessimistic models presented in this paper. Open questions are following finer optimum criterion for for the running boards.

We can find a minimum of

- weighted sum of offsets for trips i.e. $\Sigma_i\ \alpha_i\ x_i^- + \beta_i\ x_i^+$,
- weighted sum of overlapping trips i.e. $\Sigma_i\ \gamma_i\ z_i$,
- longest overlapping trips i.e. $\max_i z_i$,
- longest offset of departure times i.e. $\max_i \{x_i^-, x_i^+\}$,
- irregularity of departure offsets i.e. $\Sigma_i\ (x_i^- + x_i^+)^2$,
- irregularity of overlapping trips i.e. $\Sigma_i\ z_i^2$.

The other questions deal with number of running boards in models. We hope that we can find a basic linear models – optimistic and pessimistic – for two busses too.

Acknowledgement The research of the author is supported by grant VEGA 1/0135/08

References

[1] Palúch, S.: A Graph Theory Approach to Bus Scheduling with Types of Buses, Studies
of the Faculty of Management Science and Informatics, Volume 9, pp. 69-73,
ISBN 80-7100-959-8, 2001

[2] Peško, Š.: The Minimum Fleet Size Problem for Flexible Bus Scheduling, Studies of the Faculty of Management Science and Informatics, Vol. 9, University of Žilina, pp. 59-65, 2001

[3] Palúch, S.: Bus Scheduling as a Graph Coloring Problem , Komunikácie, Vedecké listy Žilinskej univerzity, 5.ročník, 4/2003, pp.16-20, ISSN 1335-4205, 2003

[4] Czimmermann, P., Peško, Š.: Autobusové rozvrhy s dvoma typmi vozidiel, 3th International Conference APLIMAT, Bratislava, pp. 321-326, ISBN 80-227-1995-1, 2004

[5] Palúch, S., Majer, T.: K optimalizácii mestskej a prímestskej pravidelnej osobnej Dopravy, Vysoká škola jako facilitátor rozvoje společnosti a regionu, II. mezinárodní Konference 2007, sborník, 26. ledna 2007, Kunovice, Česká republika, Evropský Polytechnický institut, 245-248, ISBN 80-7314-107-8, 2007

[6] Hartley, T.: A Glossary of Terms in Bus and Crew Scheduling, In: Wren (ed.): Computer Scheduling of Public Transport , Nort-Holland Publishing Company, pp. 353-399, 1981

Current addresses

doc. RNDr. Štefan Peško, CSc.,
Dept. of Mathematical Methods,
Faculty of Management Science and Informatics,
University of Žilina, 010 26 Žilina, Slovakia,
pesko@frcatel.fri.uniza.sk

THE PORTFOLIO OPTIMIZATION MODELS FOR INSURANCE

Pinda Ľudovít – Pekár Juraj

Abstract: The aim of this article is derive the optimization model for property-liability insurance companies. We supposed the theoretical discussion sufficient to consider a simplified version of the model which includes *n*-th types of insurance and *m*-th assets alternatives. The objective of the model is to minimize the insurance company's variance of return on equity for any particular value of its expected return on equity. With the paper it is shown the application in Slovak insurance area with investment the technical reserves.

Key words: Property-liability insurance companies, variance of return, technical reserves, funds generating factor.

The non-life insurance companies have the specific structure of their portfolio. It consists of different insurance lines, risky and risk-free securities. We shall suppose that the insurance-company *n*-th different insurance products and *m*-th risky assets and risk-free assets in its portfolio. Our aim it will be to minimize the risk of return with respect to expected return of insurance-company. Let introduce the following notation

π - the net income of the company during a special time period,

P_i - premium income in the *i*-th line of insurance, $i = 1, \ldots, n$,

$\bar{r}_i$ - the rate of underwriting return in the *i*-th insurance line as a proportion of premium income, $i = 1, \ldots, n$,

$E(\bar{r}_i)$ - the expected rate of underwriting return in the *i*-th insurance line as a proportion of premium income, $i = 1, \ldots, n$,

A_j - the amount of capital invested by the company in *j*-th risky security $j = 1, \ldots m$, A_f - the amount of capital invested by the company in risk-free securities,

$\bar{r}_j$ - the rate of return on *j*-th risky asset, $j = 1, \ldots m$,

$E(\bar{r}_j)$ - the expected rate of return on *j*-th risky asset, $j = 1, \ldots m$,

r_f - the rate of return on risk-free assets,

$\bar{r}_P$ - the company's rate of return,

$E(\bar{r}_P)$ - the expected return of portfolio.

The net income of the company can be expressed as follows

$$\pi = \sum_{i=1}^{n} P_i \bar{r}_i + \sum_{j=1}^{m} A_j \bar{r}_j + A_f r_f . \tag{1}$$

The rate of return on equity is obtained by dividing both sides of equation (1) by equity at the beginning of the period K_0

$$\bar{r}_p = \sum_{i=1}^{n} \frac{P_i}{K_0} \bar{r}_i + \sum_{j=1}^{m} \frac{A_j}{K_0} \bar{r}_j + \frac{A_f}{K_0} r_f , \tag{2}$$

where $\bar{r}_P = \dfrac{\pi}{K_0}$ is the company's rate of return on equity. Equation (2) can be rewritten as follows

$$\bar{r}_p = \sum_{i=1}^{n} \bar{w}_i \bar{r}_i + \sum_{j=1}^{m} w_j \bar{r}_j + w_f r_f , \tag{3}$$

where

$\bar{w}_i$ - the premium-to-surplus ratios for insurance lines $i = 1, \dots , n$,

w_j - the asset-to-surplus ratios for risky assets $j = 1, \dots m$,

w_f - the asset-to-surplus ratios for risk-free assets $j = 1, \dots m$.

The company has two sources of investment funds: its initial equity capital K_0 and funds obtained by insurance policies. Thus, the following relationship must hold

$$\sum_{j=1}^{m} A_j + A_f = \sum_{i=1}^{n} P_i + K_0 . \tag{4}$$

This equation is known as the balance sheet constraint because it equates total assets on the left-hand side of equation with the sum of liabilities, it means funds obtained from policyholders $\left(\sum_{i=1}^{n} P_i \right)$. Dividing both sides of equation (4) by K_0, the constrain can be rewritten as

$$\sum_{j=1}^{m} w_j + w_f = \sum_{i=1}^{n} \overline{w}_i + 1 . \qquad (5)$$

Returning to equation (3), the expected value and variance can be obtained as

$$E(r_p) = \sum_{i=1}^{n} \overline{w}_i E(\overline{r}_i) + \sum_{j=1}^{m} w_j \, E(\overline{r}_j) + w_f r_f , \qquad (6)$$

and

$$\sigma^2(\overline{r}_p) = \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{w}_i \overline{w}_j \operatorname{cov}(\overline{r}_i, \overline{r}_j) + \sum_{k=1}^{m}\sum_{l=1}^{m} w_i w_j \operatorname{cov}(\overline{r}_k, \overline{r}_l) . \qquad (7)$$

The company's optimization problem is as

$$\min_{(\overline{w}_i, w_j, w_f)} \quad \sigma^2(\overline{r}_p),$$

$$E(\overline{r}_p) = E(\overline{r}_p)^* , \qquad (8)$$

$$\sum_{j=1}^{m} w_j + w_f = \sum_{i=1}^{n} \overline{w}_i + 1 . \qquad (9)$$

$$\overline{w}_i \geq 0, \quad i = 1, \ldots, n, \ w_f \geq 0, \quad w_j \geq 0, \quad j = 1, \ldots, m , \qquad (10)$$

where $E(\overline{r}_P)^*$ is a constant. By solving the minimization problem for different values $E(\overline{r}_P)^*$,
the efficient frontier is generated.

The development of invested the technical reserves the insurance companies in Slovakia are presented on Figure 1, 2, 3. In Table 1, there are given the investments the technical reserves the insurance companies on slovak insurance market to the time 1. 1. 2006. Into the risky assets are invested 0,12 % technical reserves, it means 98 123 SKK. From this we see that the insurance companies in Slovakia have not the interest to invested on risky assets and the portfolio consists mainly with risk-free securities.

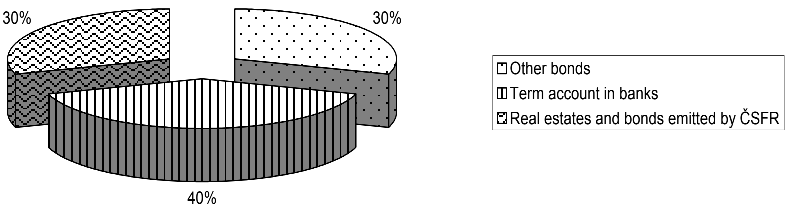


Figure 1. Technical reserves in accordance with Public notice MF SR č. 68/1991 Z.z.

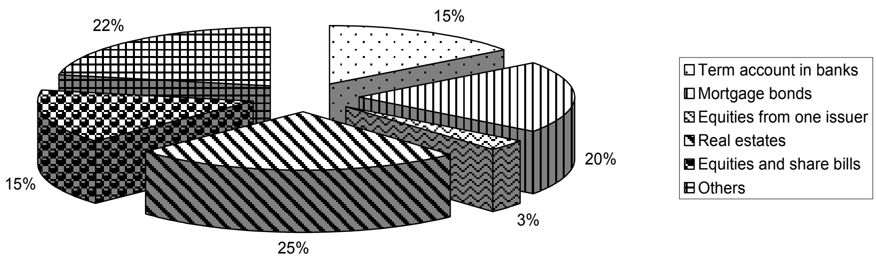


Figure 2. Technical reserves in accordance with Public notice MF SR č. 136/1996 Z.z.

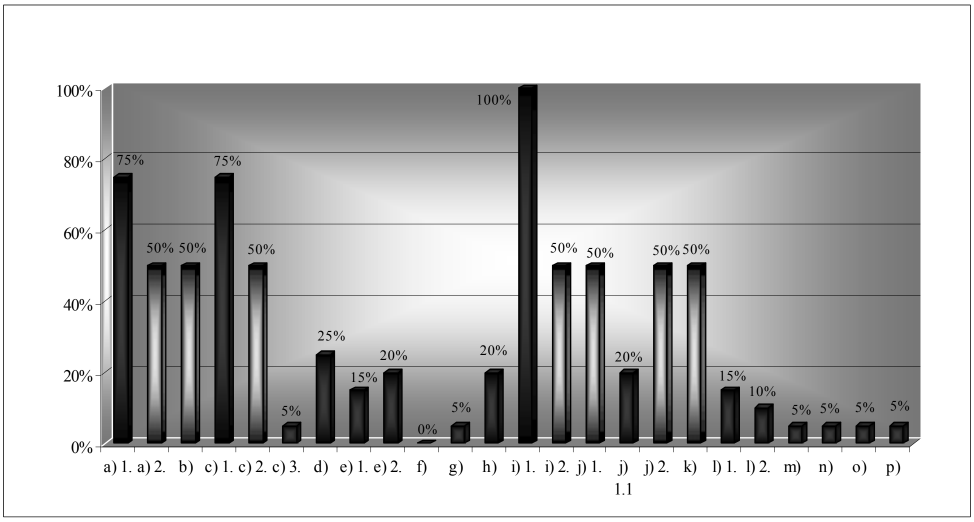


Figure 3. Limits placed technical reserves in accordance with Public notice MF SR č. 39/2005 Z.z.

		1. 1. 2006	% share from reserves
a) 1	Bonds emitted by SR or NBS, member states EÚ or their central banks	38 972 789	46,57%
i)	Term account in banks	12 140 888	14,51%
b)	Bonds emitted by banks or foreigner banks	11 448 181	13,68%
j)	Mortgage bonds	8 389 370	10,02%
a) 2	Bonds emitted by EIB, EBOR or MBOR	3 723 646	4,45%
d)	Bond accepted on market of quoted assets	3 397 955	4,06%
l)	Real estates	3 053 104	3,65%
h)	Share bills of opened share funds	1 285 783	1,54%
c) 1	Treasury bills emitted member states	749 411	0,90%
n)	Loans for insurers	288 537	0,34%
o)	Bills guaranteed by banks	147 073	0,18%
e)	Equities accepted on quoted assets market	98 123	0,12%
	Sum	**83 694 860**	**100,00%**

Tab. 1 Technical reserves in accordance with Public notice MF SR č. 39/2005 Z.z.

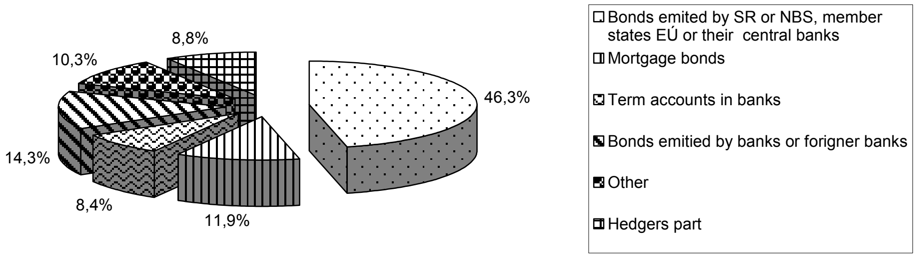


Figure 2. Technical reserves in accordance with Public notice MF SR 39/2005 Z.z. to 1.1.2007

As in the simplified model, the objective is to minimize the variance of equation (3) subject to constraints. Among the constraints is the following set, analogous to expressions (8), (9), (10)

$$E(\bar{r}_p) = E(\bar{r}_p)^*, \tag{11}$$

$$\sum_{j=1}^{m} w_j + w_f = \sum_{i=1}^{n} g_i \ \overline{w}_i + 1, \tag{12}$$

$$\overline{w}_i \geq 0, \quad i = 1,\ldots,n, \ w_f \geq 0, \quad w_j \geq 0, \quad j = 1,\ldots,m, \tag{13}$$

where

g_i - the funds generating factor for *i*-th insurance line.

These factors are introduced to relax the assumption that 1 financial unit of premiums gives rise to 1 financial unit of investment funds. Without this assumption, the balance sheet identity is written as follows

$$\sum_{j=1}^{m} A_j + A_f = \sum_{i=1}^{n} L_i + K_0, \tag{14}$$

where

L_i - the liabilities generated by writing P_i of premiums in insurance line *i*, $i = 1,...,n$,

A_j - the amount invested in the *j*-th risky security, $j = 1,...m$,

A_f - the amount invested in the risk-free security.

The relationship between premiums and liabilities is

$$L_i = P_i \cdot g_i\,, \quad i = 1,...,n\,. \tag{15}$$

where g_i is the ratio of liabilities to premiums for this line of insurance. The assets backing these liabilities constitute the investment funds obtained by writing this type of insurance. Substituting $P_i \cdot g_i$ for L_i in (14) and dividing both sides of the equation by K_0 yields the revised balance sheet constraint (12).

The condition (14) express the matching of assets and liabilities the insurance companies. The company's assets be chosen as far possible in such a way as to make the assets and liabilities equally responsive to the influences which affect them both. In 1952 F. M. Redington presented a theory of immunization, under which the investor is protected against small changes in the rate of interest. Redington's theory covers only small changes. There is however, another theory of immunization – full immunization, under which the investor makes a profit on any immediate changes whatever in the rate of interest.

REFERENCES

1. Dávid, A. – Pinda, Ľ.: Finančná matematika 2. EKONÓM 1998. ISBN 80-225-0996-5.
2. McCutcheon, J. J. – Scott, W. F.: An Introduction to the Mathematics of Finance. Oxford 1993. ISBN 0-7506-0388-7.
3. Mlynarovič, V.: Finančné investovanie. Teória a aplikácie. IURA Edition, Bratislava 2001, ISBN 80-89047-16-5.
4. Pinda, Ľ.: Deriváty cenných papierov. Vybrané problémy. IURA EDITION 2001. ISBN 80-88715-98-9.
5. Pinda, Ľ. – Starečková, A.: Investície a ich poistenie. EKONÓM 2006. ISBN 80-225-2152-3.
6. Škrovánková L.: Penzijné a nemocenské poistenie. Vydavaleľstvo Ekonóm , 1998
7. Škrovánková, L.: Penzijné a nemocenské poistenie. Vydavateľstvo EKONÓM 1999. ISBN 80-225-1836-0
8. Škrovánková, L.: Aktuárske metódy v dôchodkovom a nemocenskom poistení. Vydavateľstvo EKONÓM 2004. ISBN 80-225-1836-0

doc. RNDr. Ľudovít Pinda, CSc.
Department of Mathematics
University of Economics Bratislava
Dolnozemská 1, 852 35 Bratislava, Slovakia
e-mail:pinda@dec.euba.sk

Mgr. Juraj Pekár, PhD.
Department of Operations Research and Econometrics
University of Economics Bratislava
Dolnozemská 1, 852 35 Bratislava, Slovakia
e-mail:pekar@dec.euba.sk

SOLVING MULTI-CRITERIA PROBLEMS WITH FEEDBACK AND FUZZY IMPUTS

Jaroslav Ramík, Faculty of Economics, VŠB – TU Ostrava

1. Introduction.

In this paper we propose a fuzzy extension of the analytic network process (ANP), particularly an analytical hierarchical process (AHP) with feedback between criteria that uses uncertain human preferences as input information in the decision-making process. Instead of the classical Eigenvector prioritization method, employed in the prioritization stage of the ANP, a new fuzzy preference method, based on logarithmic least squares method is applied. The resulting fuzzy ANP enhances the potential of the ANP for dealing with imprecise and uncertain human comparison judgments. It allows for multiple representations of uncertain human preferences, as crisp, interval, and fuzzy judgments and can find a solution from incomplete sets of pair-wise comparisons.

When applying classical 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 normally 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 FIGURE 2. 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 AHP T.L. Saaty as early as in 1978 in [8]. Later on, Van Laarhoven and V. Pedrycz extended AHP to fuzzy pair-wise comparisons, see [11]. In his books [9] and [10], T.L. Saaty extends the AHP to a more general process with feedback called Analytic Network Process (ANP). In 2003, L. Mikhailov and M.G. Singh proposed a new method based on ANP and fuzzy data. Their approach is, however, essentially different to the approach used in this paper. Recently, Büyüközkan et al. see [2], and Mohanty et al., see [5] proposed another versions of fuzzy ANP and applied their methods in practice. Here, we propose a new and relatively simple method based on the original approaches from [1], [3] and [11], we extend it to the case of feedbacks between the decision criteria. We supply an illustrating example to demonstrate properties of 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 we have m variants (alternatives) of the decision outcomes $V_1, V_2,...,V_m$ are considered such that again $\sum_{r=1}^{m} w(V_r, C_i) = 1$, where $w(V_r,C_i)$ is a non-negative real number - an evaluation (weight) of V_r subject to the criterion C_i, $i = 1,2,...,n$. This system is characterized by the *supermatrix* (see [10])

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{21}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W_{32}} & \mathbf{I} \end{bmatrix},$$

a 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. Moreover, $\mathbf{W}$ is a *column-stochastic matrix*, i.e. the sums of columns are equal to one. Then the limit matrix $\mathbf{W}^{\infty} = \lim\limits_{k\to+\infty} \mathbf{W}^k$ (see [4]) exists and is given as follows

$$\mathbf{W}^{\infty} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_{32}\mathbf{W}_{21} & \mathbf{W}_{32} & \mathbf{I} \end{bmatrix}. \qquad (1)$$

Here $\mathbf{Z} = \mathbf{W}_{32}\mathbf{W}_{21}$ is the $m\times 1$ 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 three 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), (see [9], [10]). Consider now the dependences among the criteria, as is depicted in FIGURE 1.

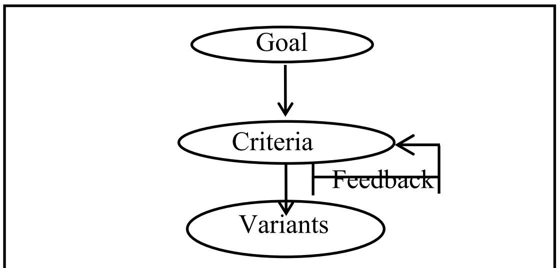


FIGURE 1. Hierarchical system with feedback

This system is given by the supermatrix

$$\mathbf{W} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W}_{21} & \mathbf{W}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{32} & \mathbf{I} \end{bmatrix}, \qquad (2)$$

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) is not column-stochastic, such that the sum of the elements in each column is equal to one, hence the limiting matrix does not exist. 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

$$\mathbf{W}^\infty = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{W_{32}}(\mathbf{I} - \mathbf{W_{22}})^{-1}\mathbf{W_{21}} & \mathbf{W_{32}}(\mathbf{I} - \mathbf{W_{22}})^{-1} & \mathbf{I} \end{bmatrix}.$$

(3)

Hence the vector

$$\mathbf{Z} = \mathbf{W}_{32}(\mathbf{I} - \mathbf{W}_{22})^{-1}\mathbf{W}_{21}$$

(4)

is used for ranking the variants i.e. for 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

$$(\mathbf{I} - \mathbf{W}_{22})^{-1} = \mathbf{I} + \mathbf{W}_{22} + \mathbf{W}_{22}^2 + \ldots$$

(5)

Then by substituting the first four terms from (5) to (4) we get

$$\mathbf{Z} = \mathbf{W}_{32}(\mathbf{I} + \mathbf{W}_{22} + \mathbf{W}_{22}^2 + \mathbf{W}_{22}^3)\mathbf{W}_{21}.$$

(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", by 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 FIGURE 2.

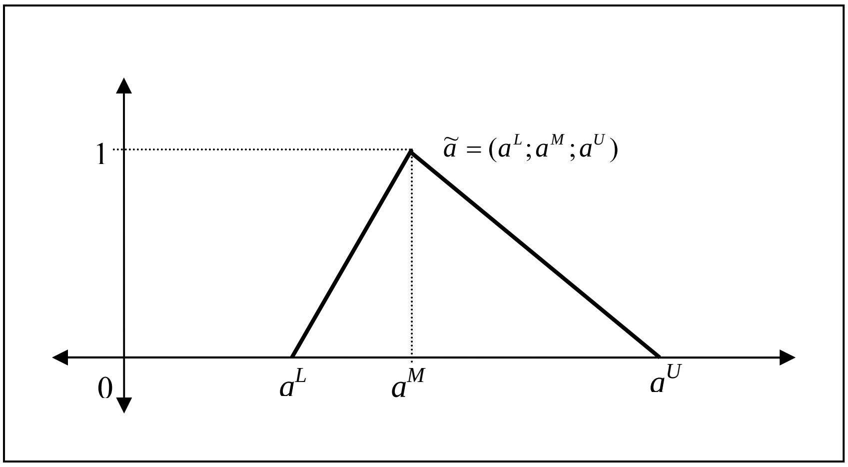


FIGURE 2. A triangular fuzzy number

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. $\widetilde{a} = (a^L; a^M; a^U)$.

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 $\widetilde{a} = (a^L; a^M; a^U)$ and $\widetilde{b} = (b^L; b^M; b^U)$, $a^L > 0$, $b^L > 0$, we obtain special formulae

$$\widetilde{a} \,\widetilde{+}\, \widetilde{b} = (a^L + b^L; a^M + b^M; a^U + b^U),$$

$$\widetilde{a} \simeq \widetilde{b} = (a^L - b^L; a^M - b^M; a^U - b^U),$$

$$\tilde{a} \,\tilde{*}\, \tilde{b} = (a^L * b^L ; a^M * b^M ; a^U * b^U),$$
$$\tilde{a} \,\tilde{/}\, \tilde{b} = (a^L / b^U ; a^M / b^M ; a^U / b^L).$$

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

$$\tilde{\mathbf{A}} = \begin{bmatrix} (1;1;1) & (a_{12}^L ; a_{12}^M ; a_{12}^U) & \cdots & (a_{1n}^L ; a_{1n}^M ; a_{1n}^U) \\ (\frac{1}{a_{12}^U} ; \frac{1}{a_{12}^M} ; \frac{1}{a_{12}^L}) & (1;1;1) & \cdots & (a_{2n}^L ; a_{2n}^M ; a_{2n}^U) \\ \vdots & \vdots & \ddots & \vdots \\ (\frac{1}{a_{1n}^U} ; \frac{1}{a_{1n}^M} ; \frac{1}{a_{1n}^L}) & (\frac{1}{a_{2n}^U} ; \frac{1}{a_{2n}^M} ; \frac{1}{a_{2n}^L}) & \cdots & (1;1;1) \end{bmatrix},$$

(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 all 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) as triangular fuzzy pair-wise comparison matrix, or
(B) as triangular fuzzy values.
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 $\widetilde{\mathbf{A}}$ defined by (7). We assume that there exists a fuzzy vectors of triangular fuzzy weights $\widetilde{w}_1, \widetilde{w}_2, ..., \widetilde{w}_r$, $\widetilde{w}_i = (w_i^L; w_i^M; w_i^U)$, i = 1,2,...,r, such that the pair-wise comparison matrix (7) is an estimation of the fuzzy matrix

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \frac{\widetilde{w}_1}{\widetilde{w}_1} & \frac{\widetilde{w}_1}{\widetilde{w}_2} & \cdots & \frac{\widetilde{w}_1}{\widetilde{w}_r} \\ \frac{\widetilde{w}_2}{\widetilde{w}_1} & \frac{\widetilde{w}_2}{\widetilde{w}_2} & \cdots & \frac{\widetilde{w}_2}{\widetilde{w}_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\widetilde{w}_r}{\widetilde{w}_1} & \frac{\widetilde{w}_r}{\widetilde{w}_2} & \cdots & \frac{\widetilde{w}_r}{\widetilde{w}_r} \end{bmatrix}.$$

Here, e.g. $r = n$ or $r = m$. We shall find the fuzzy weights $\widetilde{w}_1, \widetilde{w}_2, ..., \widetilde{w}_r$ by minimizing the fuzzy functional

$$\widetilde{H} = \sum_{i,j} \widetilde{\log}\left(\frac{\widetilde{w}_i}{\widetilde{w}_j} - \widetilde{a}_{ij} \right)^2 . \tag{8}$$

In (8), minimization of $\widetilde{H}$ 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 [3], 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 [3], 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 $\tilde{Z} = \left(\tilde{z}_1,...,\tilde{z}_n\right)^T = \left((z_1^L; z_1^M; z_1^U),...,(z_{m1}^L; z_m^M; z_m^U)\right)^T$.

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 ranking fuzzy numbers, see e.g. [7], for a comprehensive review of comparison methods see [3].

5. Example

In this section we analyze an illustrating example of decision making situation with 3 decision criteria and 4 variants. The goal of this realistic decision situation is to choose the "best" product 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, the variants according to criteria as well as the feedback between the criteria is based on the data from fuzzy pair-wise comparison matrices. Here, instead of classical Saaty´s nine-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 valuations 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 without the feedback. In this we apply the same approach, a particular case of the previous more general situation when taking crisp evaluations, i.e. $a^L = a^M = a^U$ and

Step 1: Evaluate the pair-wise comparison matrices and calculate the corresponding weights

The data for relative importance of the criteria are given by the following pair-wise comparison matrix **C**

	L	M	U	L	M	U	L	M	U	
C =	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

	L	M	U			
W21=	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**

	L	M	U	L	M	U	L	M	U	L	M	U	
A1 =	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			

	L	M	U	L	M	U	L	M	U	L	M	U	
A2 =	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. There are zeros in 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, we obtain the stochastic matrices

	L	M	U	L	M	U	L	M	U	
W32* =	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

$$\tilde{\mathbf{Z}} = \tilde{\mathbf{W}}_{32}(\mathbf{I} \tilde{+} \tilde{\mathbf{W}}_{22} \tilde{+} \tilde{\mathbf{W}}_{22}^2 \tilde{+} \tilde{\mathbf{W}}_{22}^3)\tilde{\mathbf{W}}_{21}. \qquad (6^*)$$

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 $\widetilde{Z} = (\widetilde{z}_1,...,\widetilde{z}_4)^T = ((z_1^L; z_1^M; z_1^U),...,(z_4^L; z_4^M; z_4^U))^T$ 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 $\widetilde{z}_i$, 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 FIGURE 3.

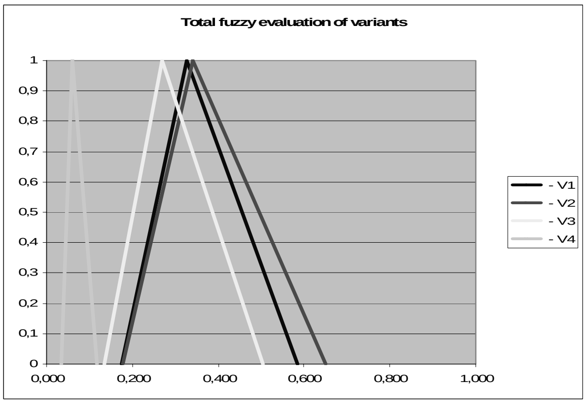


FIGURE 3. Total fuzzy evaluation of variants

Now, we solve the same problem applying classical AHP with non-fuzzy evaluations in the pair-wise comparisons and the zero feedback. We apply a particular case of the previous more general situation considering crisp evaluations, i.e. $a^L = a^M = a^U$ and $\mathbf{W}_{22} = \mathbf{0}$.

Hence, we get

W21 =

0,627	- C1
0,280	- C2
0,094	- C3

W32 =

C1	C2	C3	
0,547	0,160	0,100	- variant 1: V1
0,287	0,278	0,598	- variant 2: V2
0,121	0,467	0,266	- variant 3: V3
0,045	0,095	0,036	- variant 4: V4

By (1) we obtain the aggregating crisp evaluations of the variants:

Z =

		Rank
0,397	- V1	1
0,314	- V2	2
0,231	- V3	3
0,058	- V4	4

The situation is graphically depicted in FIGURE 4.

6. Conclusion

Considering feedback dependences between the criteria the total rank of the variants can be changed as we have demonstrated in the above example. Fuzzy (soft) evaluation of pair-wise comparisons may be more comfortable and appropriate for DM. Occurrence of dependences among criteria is realistic (and also frequent). Presence of fuzziness in evaluations may change the final rank of variants, too.

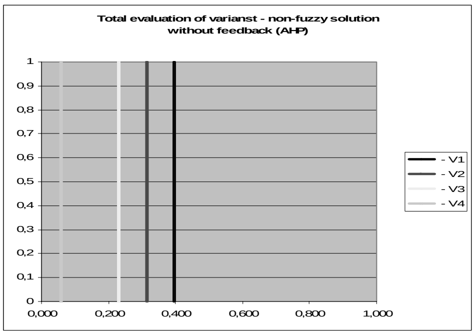


FIGURE 4. Total evaluation of variants – classical AHP

REFERENCES

[1] Buckley, J.J., Fuzzy hierarchical analysis. *Fuzzy Sets and Systems* 17, 1985, 1, pp. 233-247, ISSN 0165-0114.

[2] Büyüközkan, G., Ertay, T., Kahraman, C. and Da Ruyan, Determining the importance weights for the design requirements in the house of quality using the fuzzy analytic network approach. *Internat. J. of Intelligent Systems*, 19, 5 (2004), pp. 443 – 461, ISSN 0884-8173.

[3] Chen, S.J., Hwang, C.L. and Hwang, F.P., *Fuzzy multiple attribute decision making*. Lecture Notes in Economics and Math. Syst., Vol. 375, Springer-Verlag, Berlin – Heidelberg 1992, ISBN 3-540-54998-6.
[4] Horn, R. A., Johnson, C. R., *Matrix Analysis*. Cambridge University Press, 1990, ISBN 0521305861.
[5] Mikhailov, L., Singh, M.G., Fuzzy analytic network process and its application to the development of decision support systems. Systems, Man and Cybernetics, Part C, IEEE Transactions, 2003, 33, 1, pp. 33- 41, ISSN: 1094-6977.
[6] Mohanty, R. P., Agarwal, R., Choudhury, A. K. and Tiwari, M. K., A fuzzy ANP-based approach to R&D project selection: a case study. *Internat. J. of Production Research,* 43, 24, 2005, pp. 5199-5216.
[7] Ramik, J., Duality in fuzzy linear programming with possibility and necessity relations. *Fuzzy Sets and Systems* 157, 2006, 1, pp. 1283-1302, ISSN 0165-0114.
[8] Saaty, T.L., Exploring the interface between hierarchies, multiple objectives and fuzzy sets. *Fuzzy Sets and Systems* 1, 1978, pp. 57-68, ISSN 0165-0114.
[9] Saaty, T.L., *Multicriteria decision making - the Analytical Hierarchy Process*. Vol. I., RWS Publications, Pittsburgh, 1991, ISBN .
[10] Saaty, T.L., *Decision Making with Dependence and Feedback – The Analytic Network Process*. RWS Publications, Pittsburgh, 2001, ISBN 0-9620317-9-8.
[11] Van Laarhoven, P.J.M. and Pedrycz, W., A fuzzy extension of Saaty's priority theory. *Fuzzy Sets and Systems* 11, 1983, 4, pp. 229-241, ISSN 0165-0114.

SUPPLY CHAIN MODELING VIA ROBUST OPTIMIZATION

Marian Reiff

Abstract
In this paper, methods and ideas for r designing a supply chain network are provided. In case of the designing supply chain network, also decisions about production facilities, like allocation, material consumption, production etc. are modeled. The contribution of this paper includes the development of a methodology for supporting the decision process on a strategic level in an uncertain environment when designing the supply chain network. It also encompasses decision situations with not accurate or wrong data. The paper is dedicated to mathematical possibilities of modeling supply chain design via discrete mixed integer tasks.

Keywords
Supply chain, optimization, uncertainty

Uncertainty (risk) is general expression used in economic scientific disciplines (like finance insurance), natural sciences and technical sciences for depiction future states, conditions, for description of physical measures or simply for description of unknown. From the rank of literature that deals with uncertainty it is possible to mention the work of (Knight, 1921), that firstly deals with relationship of expression uncertainty, probability and risk. From mathematical programming point of view the expression risk presents uncertainty, witch probability distribution is possible to compute or estimate. Uncertainty or risk can be mathematically describe with various theories like probability theory (Laplace, 1812) or fuzzy logic (Zadeh, 1973)

General formulation of uncertainty in supply chain is defined by (Zimmermann, 2000) as lack of information, incompleteness information, conflict proofs, amphibolous or mistakes in measures.

Cause of uncertainty in supply chain can be assign according to (Davis, 1993) (Vidal a Goetschalckx, 2001) to four main sources:

- Uncertainty resulted from side of suppliers, it occurs for example in precisely scheduled operations, this evidence can be described by average lateness.

- Production uncertainty spring up in production process for example machine damage.
- Uncertainty in customer demand caused by mistakes in its estimation, no regular orders, etc.
- Issue of regional economic alliances like European union supported globalization trends in supply chain and by that caused new factors of uncertainty like exchange rate or reliability of distribution channels etc. (Vidal a Goetschalckx, 2001).

Majority of uncertainty factors influenced supply chain operations can be classify as short time fluctuation or longtime trends.

Short time fluctuation are implicitly accompanied in mathematical models via use of mean value from sufficiently long time period or with the use of method risk pooling - data aggregation ((Nahmias, 1997).

On the other side long time trends require different approach. For example use of stochastic or robust models. The target of robust optimization is to find solution that would be good in different situation of realization of random parameters. The definition of good solution is different from model to model and the choice of appropriate criteria of good solution is part of modeling process. For example (Bai, Carpenter a Mulvey, 1997) introduce the goal of robust optimization as finding such solution that would be close to optimal solution and would be not too sensitive to changes in realization of uncertain in input parameters. (Mulvey, Vanderbei a Zenios, 1995) describe robust optimization as process of finding such solution that value of purpose function would be close enough to every optimal solution of all possible scenarios of realization of uncertain input data. The final solution doesn’t have to be optimal to any realization of uncertain data.

Robust optimization models for designing the supply chain according (Butler, Ammons a Sokol, 2004), is possible to classify considering the “criteria of good solution” subsequently: models minimizing maximal divergence (regret models) and models minimizing variability. For example minmax models are (Kouvelis a Yu, 1997), (Gutierrez, Kouvelis a A.Kurawarwala, 1996) and variability models are (Bai, Carpenter a Mulvey, 1997), (Ahmed a Sahinidis, 1998), (Mulvey, Vanderbei a Zenios, 1995).

In rest of paper computed results are demonstrated to illustrate the difference between optimal solutions of deterministic models of supply chain design, and “good solution of robust variability model and robust minmax model.
Solution of these models is illustrated in fictive case study. The goal of study is to design distribution network of fourteen different goods from tree production facilities to twenty five customer locations. Future demand of customers is describe by ten scenarios.

In case that future demand of customer is know for sure, the decision about supply chain design is possible to make by deterministic model. Optimal solution for scenario 1, 2, 3 … 10 is in table 1:

Scenario	Optimal solution
S_1	1 135 156,74
S_2	1 092 290,86
S_3	1 214 937,46
S_4	1 031 639,65
S_5	921 318,44
S_6	816 888,54
S_7	708 432,41
S_8	597 837,15
S_9	486 763,08
S_{10}	261 010,73

Table 1

For example in case of realization of demand by scenario 9 the optimal solution would be 486 763,08.
In case of doing decision about supply chain design in uncertain environment according to solution of deterministic model with input data from scenario 9 the expect value of solution (with condition that every scenario has equal probability of realization) would be in that case 837 361,73.

In table 2 are demonstrated results for every scenario. In these cases the physical supply chain structure is forced for each scenario 1, 2, ...,10 . The binary decision variables (decision variables about physical structure of supply chain) are exogenous.

Scenario		S_1	S_2	S_3	S_4	S_5
Value of purpose function	S_1	1 135 156,74	1 135 509,67	1 135 156,74	1 135 156,74	1 135 156,74
	S_2	1 092 398,78	1 092 290,86	1 092 398,78	1 092 398,78	1 092 398,78
	S_3	1 214 937,46	1 215 965,18	1 214 937,46	1 214 937,46	1 214 937,46
	S_4	1 031 639,65	1 033 561,95	1 031 639,65	1 031 639,65	1 031 639,65
	S_5	921 318,44	923 674,47	921 318,44	921 318,44	921 318,44
	S_6	816 888,54	819 694,66	816 888,54	816 888,54	816 888,54
	S_7	708 432,41	711 698,77	708 432,41	708 432,41	708 432,41
	S_8	600 756,80	604 487,83	600 756,80	600 756,80	600 756,80
	S_9	494 148,96	498 313,72	494 148,96	494 148,96	494 148,96
	S_{10}	280 730,58	285 805,67	280 730,58	280 730,58	280 730,58
Expected value		829 640,84	832 100,28	829 640,84	829 640,84	829 640,84

Scenario		S_6	S_7	S_8	S_9	S_{10}
Value of purpose function	S_1	1 135 156,74	1 135 156,74	1 135 156,74	1 153 735,49	1 201 989,66
	S_2	1 092 398,78	1 092 398,78	1 092 398,78	1 120 691,60	1 152 496,72
	S_3	1 214 937,46	1 214 937,46	1 214 937,46	1 237 443,72	1 272 756,86
	S_4	1 031 639,65	1 031 639,65	1 031 639,65	1 047 212,75	1 084 954,92
	S_5	921 318,44	921 318,44	921 318,44	932 289,61	959 275,01
	S_6	816 888,54	816 888,54	816 888,54	823 329,59	844 584,31
	S_7	708 432,41	708 432,41	708 432,41	710 194,72	726 454,65
	S_8	600 756,80	600 756,80	600 756,80	597 837,15	609 278,39
	S_9	494 148,96	494 148,96	494 148,96	486 763,08	493 319,19
	S_{10}	280 730,58	280 730,58	280 730,58	264 119,56	261 010,73
Expected value		829 640,84	829 640,84	829 640,84	837 361,73	860 612,04

Table 2

In table 3 (variance model) and table 4 results (mimax model) of robust supply chain design are demonstrated. In third row is calculated difference between good solution and optimal solution. It fourth row this difference is calculated in proportional.

Scenario	Value of purpose function	Absolute variance	Relative variance
S_1	1 137 572,45	2 415,71	0,00212808
S_2	1 094 134,46	1 843,60	0,00168783
S_3	1 218 283,54	3 346,08	0,00275412
S_4	1 034 707,37	3 067,72	0,00297364
S_5	923 427,77	2 109,33	0,00228947
S_6	818 004,15	1 115,61	0,00136568
S_7	708 509,68	77,27	0,00010907
S_8	599 809,83	1 972,68	0,00329969
S_9	492 243,60	5 480,52	0,01125911
S_{10}	276 793,16	15 782,43	0,06046659
Expected value	830 348,60	3 721,09	0,00883333

Table 3

Scenario	Value of purpose function	Absolute variance	Relative variance
S_1	1 157 037,98	21 881,24	0,019276
S_2	1 120 691,60	28 400,74	0,026001
S_3	1 246 470,91	31 533,45	0,025955
S_4	1 048 889,19	17 249,54	0,016721
S_5	938 389,92	17 071,48	0,018529
S_6	828 566,04	11 677,50	0,014295
S_7	722 654,55	14 222,14	0,020076
S_8	606 754,98	8 917,83	0,014917
S_9	490 758,98	3 995,90	0,008209
S_{10}	264 119,56	3 108,83	0,011911
Expected value	842 433,37	15 805,87	0,017589

The computational demandingness is illustrated in table 5 These models were computed with solver Ceplex on the computer with processor 1,60 GHz a RAM 248 MB.

Model type	Deterministic model	Robust model	Minimax model
CPU [s]	0,578	65,453	6 813,421

Table 5

On the basis of the comparison of the results of expected value of overall costs of the robust model and the robust model that minimizes maximal deviation, we can contend, that using of the robust model doesn't increase significantly the value of overall costs and simultaneously doesn't requires additional managerial attention when designing the supply chain network. The usage of the robust model that minimizes maximal deviation seams to be too conservative. That's because the value of the total costs is significantly higher in comparison with other results. Thus the time demandingness of the solution process is higher for minimax tasks than for minisum tasks.

Literature

Ahmed, S. a Sahinidis, N. 1998. Robust process planning under uncertainty. In: *Industrial Engineering and Chemical Research*, 1998, č. 37, s. 1883 - 1892.

Bai, D., Carpenter, t. a Mulvey, J. M. 1997. Making a case for robust optimization models. In: *Management Science*, 1997, č. 22, s. 1037 - 1049.

Butler, R. J., Ammons, J. C. a Sokol, J. 2004. A robust optimization model for strategic production and distribution planning for a new product. In: *working paper, University of Central Florida*, 2004.

Davis, T. 1993. Effective Supply Chain Management. In: *Sloan Management Review*, 1993, č. Summer, s. 36 - 46.

Gutierrez, G., Kouvelis, P. a A.Kurawarwala. 1996. A robustness approach to uncapacitated network design problems. In: *European Journal of Operational Research*, 1996, č. 94, s. 362 - 376.

Kouvelis, P. a Yu, G. 1997. *Robust Discrete Optimization and Its Applications,* Dordecht: Kluwer Academic Publishers, 1997.

Laplace, P. S. d. 1812. Theorie Analytique des Probabilités., 1812.

McClelland, M. a Marucheck, A. E. 1986. Strategic Issues in Make-to-Order Manufacturing. In: *Production and Inventory Management*, roč. 27, 1986, č. 3, s. 82 - 95.

Mulvey, J. M., Vanderbei, R. J. a Zenios, S. A. 1995. Robust Optimization of Large-Scale Systems. In: *Operations Research*, roč. 2, 1995, č. 43, s. 264 - 281.

Nahmias, S. 1997. *Production and Operation Analysis* Chicago IL: Irwin/McGraw-Hill, 1997.
Vidal, C. J. a Goetschalckx, M. 2001. A Global Supply Chain Model with Transfer Pricing and Transportation Cost Allocation. In: *European Journal of Operational Research*, 2001, č. 129, s. 134 - 158.
Zadeh, L. A. 1973. Outline of a New Approach to the Analysis of Complex System and Decision Processes. In: *IIE Transactions*, 1973, s. 28 - 44.
Zimmermann, J. H. 2000. An Application Orientated View of Modeling Uncertainty. In: *European Journal of Operational Research*, 2000, č. 122, s. 190 - 198.

Acknowledgement: This work has been supported by research project VEGA 1/4588/07

Author

Ing. Marian Reiff, PhD.,
Department of Operations Research and Econometrics, Faculty of Economic Informatics, University of Economics in Bratislava,
Dolnozemská cesta 1/b, 85235 Bratislava, Slovakia
reiff@euba.sk

ASYMMETRIC SIGNALS IN FINANCIAL MARKETS: THE DYNAMICS OF VOLATILITY AND THRESHOLD ADJUSTMENT MODELS

Rui Menezes, Nuno B. Ferreira and Diana Mendes

ISCTE, Av. das Forças Armadas, 1649-026 Lisboa
corresponding author: rui.menezes@iscte.pt

Abstract
Traditional modeling in applied econometrics has relied mainly on regression-like models where the behavior of positive and negative shocks on the relevant variables has been treated alike. However, this is rather unlikely in many empirical situations, in part because the behavior of economic agents differs when the innovations take different signs. This seems to be the case of volatility in the stock markets. Similarities occur in many other economic and financial contexts, which suggest that the behavior of these variables may incorporate many types of nonlinearities. One such type of nonlinearity that can be encountered in applied work is captured by the threshold autoregressive mechanism. This is a particular case of a Switching Regime Model that is generally nonlinear but linear within regimes. The purpose of this chapter is to provide an overview of some relevant aspects of the threshold mechanism applied to dynamic models (volatility) and ECM models in the context of nonstationarity of the instruments. The processes will be illustrated by real empirical data taken from the Portuguese and US stock markets.

1. Introduction

It has been widely recognized by economists that economic relationships are typically nonlinear. This is so that, for example, Granger and Teräsvirta (1993), *inter alia*, have dedicated a whole book to the subject of modeling nonlinear economic relationships. Nonlinear relationships are present in many aspects of the economic activity, and particularly so in the context of financial markets. Examples of this include the attitude of investors towards the risk and the process of generating financial variables such as stock returns, dividends, interest rates, and so on. On the other hand, the performance of an economy also presents strong signs of nonlinear behavior: *e.g.* business cycles, production functions, growth rates, unemployment, etc. Although the shape of nonlinearity in these relationships may be rather complex, there are cases where one may admit some sort of linear relationship between the relevant variables within certain regimes. This is the case when one aims to study the co-movements of stock returns volatility and some relevant macroeconomic factors. One obvious question that we may pose in this context is whether the magnitude of positive and negative responses differs for similar positive and negative variations in the predictors, in which case we can say that the underlying variables display asymmetric adjustment. Markets characterized by higher elasticity of supply are likely to show less asymmetry than their counterparts due to increased security of supply. Models of financial markets have incorporated asymmetry using GARCH-type methodologies. An alternative way to deal with these cases is to use threshold autoregressive (TAR) and momentum threshold autoregressive (M-TAR) models to address the problem of multivariate asymmetry. These methodologies are essential when the asymmetric variables are non-stationary (but not only), because of the low power of unit root and cointegration tests in such cases. In a non-stationary framework, asymmetric cointegration tests were developed by Enders and Siklos (2001) using a modified error correction model derived from the original EG testing procedure. We apply this methodology to the Portuguese and U.S. stock markets using monthly observations from January 1993 to December 2003.

2. Volatility of Stock Market Returns

There are many different ways for measuring the volatility associated with stock market returns. However, since the volatility itself is not directly observed, one needs to find a suitable estimator to measure the risk resultant from changes in stock returns. Stock returns are usually measured by

$$r_t = \ln\left(P_t / P_{t-1}\right),\ \forall t, t = 1, \cdots, T, \tag{1}$$

where P_t is the value of the underlying stock index at time t. The series should only reflect the risk of changes in market returns, so one may filter r_t by subtracting from it the series r_t' generated by a non-risky asset. In our case, we considered as r_t' the short-run Libor and treasury bill (3 months) rates obtained from the Data-Stream. The difference between r_t and r_t' is usually termed excess return (R_t).
The volatility of R_t can be estimated on the basis of the absolute deviation from the mean excess return, *i.e.*

$$w_t = f\left(\left|R_t - \overline{R}\right|\right),\ \forall t, t = 1, \cdots, T. \tag{2}$$

Alternatively, a measure of the volatility of stock returns that has been used in the empirical literature is based upon the historical standard deviations of the excess return (rolling historical volatility), and is given by

$$v_t = \left(\frac{\sum_{p=1}^{k}\left(R_{t+1-p} - \overline{R}\right)^2}{k-1}\right)^{\frac{1}{2}}. \tag{3}$$

These two measures of volatility were used to address the issue of asymmetric co-movements between stock market volatility and some relevant macroeconomic indicators: (1) dividend yield (dy), (2) earnings price ratio (epr), (3) inflation (cpi), and (4) industrial growth rate (ipi). All macroeconomic variables were found to be I(1). Inflation and industrial growth rates are expressed in terms of the natural logs of the underlying indexes.

3. Error Correction Model and Dynamic Volatility

The starting point for analyzing the relationship between stock market returns volatility and macroeconomic variables is an error correction formulation based on a standard ADL(1,1) model. As in many other circumstances, a one lag specification should be sufficient to capture the dynamic behavior of the measures of volatility used here. The model, thus, stands as

$$\Delta z_t = \lambda_0 \Delta x_t - (1-\alpha_1)(z_{t-1} - \beta_0 - \beta_1 x_{t-1}) + \varepsilon_t \,, \tag{4}$$

where $\alpha_1 < 1$. The parameter λ_0 denotes the short-run reaction of z_t to a change in x_t. The parameter $1-\alpha_1$ measures the speed of adjustment of z_t to a disequilibrium shock in the system, that is, it measures how fast the system returns to equilibrium when a shock provokes a deviation from the long-run relationship given by the regression line. The long-run equilibrium is given by $\hat{z}_{t-1} = \hat{\beta}_0 + \hat{\beta}_1 x_{t-1}$, over which the OLS residuals, $\hat{u}_{t-1}$, are zero. The model is set such that the error correction disturbances, ε_t, are white noise. Extensions of model (4) in order to allow for higher order serial correlation or multi-equation systems are straightforward to construct.

When the level variables z_t and x_t are both first-order integrated and the OLS residuals, $\hat{u}_{t-1}$, are stationary, then z_t and x_t are said to be cointegrated, CI(1,1), and there is an error correction model that represents the short and long-run behavior of the variables (Granger representation theorem). Standard estimators of the model are then super-consistent. When the level variables z_t and x_t are both stationary, the usual OLS estimation is applied.

We now turn to consider the estimation of dynamic volatility models where the variance of the error term depends on past information. These are known as autoregressive conditional heteroskedasticity (ARCH) models and were introduced by Engle (1982). In the standard model, the variance of the error term is modeled as a function of past values of the error term. This model was extended to the Generalized ARCH model (Bollerslev, 1986) by also allowing for the dependence of the variance on its own past values. Further extensions appeared in the literature allowing for asymmetric movements in the conditional variance, of which the Threshold ARCH (Zakoian, 1990; Glosten *et al.*, 1993) and the Exponential GARCH (Nelson, 1991) models are popular examples in the empirical literature. Many other extensions have been proposed by several authors in order to match specific issues for their research needs.

For our purposes, we shall focus on the specifications of the above mentioned asymmetric conditional heteroskedasticity models: TARCH and EGARCH. The mean equation in these models is the long-run equilibrium model, $z_t = \mathbf{x}_t'\beta + u_t$, where $\mathbf{x}_t'$ denotes a k-dimension vector of exogenous variables. In the TARCH specification, the conditional variance is modeled by

$$\sigma_t^2 = \omega + \delta u_{t-1}^2 + \gamma u_{t-1}^2 d_{t-1} + \theta\sigma_{t-1}^2, \qquad (5)$$

where $d_t = 1$ if $u_t > 0$, and zero otherwise. In (5), ω, δ, γ, and θ are parameters. In this model, good news ($u_t > 0$), and bad news ($u_t < 0$), have different effects on σ_t^2. The parameter γ denotes the *leverage effect* (if $\gamma > 0$), and the impact on the conditional variance is asymmetric if $\gamma \neq 0$. Estimation and testing of this model should use quasi-likelihood robust standard errors if the residuals show signs of leptokurtosis.

The EGARCH specification of the conditional variance is given by

$$\log\sigma_t^2 = \omega + \delta\left|\frac{u_{t-1}}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}}\right| + \gamma\frac{u_{t-1}}{\sigma_{t-1}} + \theta\log\sigma_{t-1}^2, \qquad (6)$$

where the interpretation of the parameters is as above. However, the leverage effect under this specification is exponential rather than quadratic as in the TARCH model. The EGARCH residuals follow a generalized error distribution.

4. Threshold Adjustment

As noted before, the long-run equilibrium relationship between two time series z_t and x_t can be estimated as a standard regression model $z_t = \beta_0 + \beta_1 x_t + \mu_t$, where β_i (i = 0, 1) are parameters, and μ_t is a disturbance term that can now be serially correlated. As above, the parameter β_1 gives the magnitude of adjustment of z_t to variations in x_t, and is the long-run elasticity of the two variables if they are measured in logs. If $\beta_1 < 1$, then shifts in x_t are not fully passed onto z_t. The estimation of the long-run equation is the first step of the threshold methodology.

The second step focuses on the OLS estimates of ρ_1 and ρ_2 in the following error correction model

$$\Delta\mu_t = I_t\rho_1\mu_{t-1} + (1 - I_t)\rho_2\mu_{t-1} + \varepsilon_t, \qquad (7)$$

where ε_t is a white noise disturbance and the residuals from the long-run equation are used to estimate $\Delta\mu_t$. I_t is the Heaviside indicator function such that

$$I_t = \begin{cases} 1 & \text{if } \xi_{t-1} \geq \tau \\ 0 & \text{if } \xi_{t-1} < \tau \end{cases}. \tag{8}$$

If in (8) $\xi_{t-1} = \mu_{t-1}$, then the model specification illustrated in (7) is called the threshold autoregressive (TAR) model. It allows for different coefficients of positive and negative variations. A sufficient condition for the stationarity of μ_t is $-2 < (\rho_1, \rho_2) < 0$. This means that the long-run equation is an attractor such that μ_t can be written as an error correction model similar to that given in (7). If $\rho_1 = \rho_2$ then the adjustment is symmetric, which is a special case of (7) and (8). Expression (7) can also contain lagged values of $\Delta\mu_t$. When μ_{t-1} is above its long-run equilibrium value, the adjustment is $\rho_1\mu_{t-1}$, and if μ_{t-1} is below its long-run equilibrium value, the adjustment is $\rho_2\mu_{t-1}$.

If in (8) $\xi_{t-1} = \Delta\mu_{t-1}$, then the model (7) is called the momentum threshold autoregressive (M-TAR) model. The M-TAR model allows the decay to depend on the previous period change in μ_{t-1}. The value of the threshold, τ, in our case, is assumed to be zero in all models.

The TAR model is designed to capture asymmetrically *deep* movements in the series of the deviations from the long-run equilibrium, while the M-TAR model is useful to capture the possibility of asymmetrically *steep* movements in the series (Enders and Granger, 1998). For example, in the TAR model if $-1 < \rho_1 < \rho_2 < 0$, then the negative phase of μ_t will tend to be more persistent than the positive phase. On the other hand, for the M-TAR model, if for example $|\rho_1| < |\rho_2|$ the model exhibits little decay for positive $\Delta\mu_{t-1}$ but substantial decay for negative $\Delta\mu_{t-1}$. This means that increases tend to persist but decreases tend to revert quickly toward the attractor.

Finally, we can perform a number of statistical tests on the estimated coefficients (and also on the residuals) in order to ascertain the validity of the error correction model outlined in (7), and subsequently if the adjustment is symmetric or not. The relevant tests on the coefficients are H_0: $\rho_1 = 0$ and H_0: $\rho_2 = 0$, for which we obtain the sample values of the *t*-statistics; and H_0: $\rho_1 = \rho_2 = 0$, for which we obtain the sample values of the *F*-statistic. The restriction that adjustment is symmetric, i.e. H_0: $\rho_1 = \rho_2$, can also be tested using the usual *F*-statistic.

If the variables in the long-run equation are stationary, the usual critical values of the *t* and *F* distributions can be used to assess the significance level of the underlying tests. However, if these variables are integrated of first order, one can use the critical values reported by Enders and Siklos (2001) to determine whether the null hypothesis of no cointegration can be rejected. If the alternative hypothesis is accepted, it is possible to test for asymmetric adjustment using the standard critical values of the *F* distribution, since ρ_1 and ρ_2 converge to a multivariate normal distribution (Enders and Granger, 1998).

5. Results

Figures 1 and 2 depict the square deviations from the mean of the excess return variables for Portugal and the US.

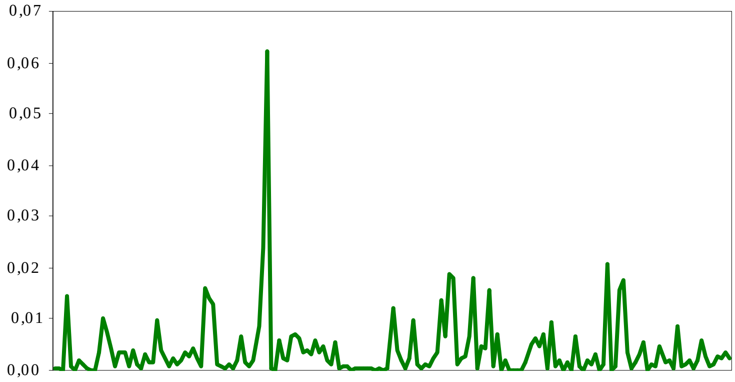


Fig. 1. Square deviations from the mean R_t for Portugal

We show the square deviations rather than the deviations themselves in order to obtain a clearer picture of the volatility issue. As may be seen, there are volatility clusters in the data for both countries, with periods of low volatility followed by periods of high volatility, which is consistent with many other findings in stock markets. However, high volatility periods also correspond to periods of negative excess return, which suggests that the impact is asymmetric. These findings trigger the need to analyze the asymmetric conditional heteroskedasticity properties of the excess return variables R_t.

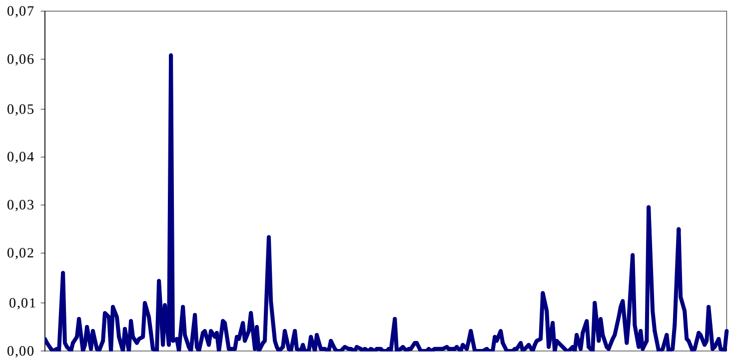


Fig. 2. Square deviations from the mean R_t for the US

The results of estimating the TARCH and EGARCH models for the Portuguese and US excess return data are presented in Table 1.

Table 1. TARCH and EGARCH results

Parameters	TARCH		EGARCH	
	Portugal	**US**	**Portugal**	**US**
ω	0.0000 ** (1.45E-07)	0.0000 ** (1.36E-07)	–0.4632 ** (0.0339)	–0.2607 ** (0.0229)
δ	0.0940 ** (0.0090)	–0.0015 (0.0071)	0.2349 ** (0.0109)	0.1171 ** (0.0112)
γ	0.0488 ** (0.0097)	0.1378 ** (0.0108)	–0.0363 ** (0.0060)	–0.1102 ** (0.0079)
θ	0.8757 ** (0.0062)	0.9230 ** (0.0059)	0.9700 ** (0.0031)	0.9817 ** (0.0021)
SIC	–6.8388	–6.5258	–6.8432	–6.5340

Notes: ** significant at the 1% level; asymptotic standard errors in parentheses.

The results show strong evidence of conditional heteroskedasticity in the data, and also that the impact of news on the conditional variance is asymmetric both for Portugal and the US. However, there is evidence that the leverage effect is quadratic but not exponential, since the EGARCH γ parameters are significantly negative in both countries.

The next question entails the possibility of asymmetric co-movements between the stock market returns volatility and macroeconomic variables. The methodology for assessing asymmetric co-movements between these variables was applied to the dataset described earlier. A total of 16 regressions were run for each country analyzed (Portugal and US). The resulting residuals were then used to perform the TAR and M-TAR tests of asymmetry. This is however only possible when the residuals of the series are convergent. The results obtained are reported in Table 2 (for Portugal) and Table 3 (for the US).

The estimators of volatility $\hat{w}_t$ and $\hat{v}_t$ were found to be stationary for Portugal on the basis of ADF tests using the Bayesian information criterion for model selection and KPSS tests. Thus, for reasons of consistency, we replaced the original non-stationary macroeconomic variables by their first-difference I(0) counterparts in these models. For the US, however, the estimator $\hat{w}_t$ is I(0) and the estimator $\hat{v}_t$ is I(1). As before, we replaced the original macroeconomic variables by their first-difference counterparts in the models that use $\hat{w}_t$. With regard to $\hat{v}_t$, the first two tests on the estimated values of ρ_1 and ρ_2 are actually cointegration tests.

For both Portugal and the US, the results obtained when $\hat{w}_t$ was used as a measure of volatility show that the null hypothesis that ρ_1 and ρ_2 are zero is rejected at the

1% level or better in all cases (separate and joint tests). The test of the null hypothesis of symmetry, however, was not rejected at significant levels in none case. An error correction model without separation of the positive and negative variations would therefore be a sufficient representation of the process under analysis. Thus, given their triviality, we shall not report the results for the series $\hat{w}_t$.

Table 2. TAR and M-TAR results for Portugal

Variable	ρ_1	ρ_2	$\rho = 0$	$\rho_1 = \rho_2 = 0$		$\rho_1 = \rho_2$		
			***t*-max**	**Φ**		***F***		***p*-value**
				TAR				
$\Delta(\text{dy}_t)$	–0.093 (0.053)	–0.156 (0.042)	–	8.301	**	0.853		0.357
$\Delta(\text{epr}_t)$	–0.080 (0.054)	–0.169 (0.043)	–	8.940	**	1.710		0.193
$\Delta(\text{lncpi}_t)$	–0.097 (0.052)	–0.135 (0.045)	–	6.281	**	0.305		0.582
$\Delta(\text{lnipi}_t)$	–0.075 (0.042)	–0.094 (0.035)	–	5.206	**	0.115		0.736
				M-TAR				
$\Delta(\text{dy}_t)$	–0.201 (0.047)	–0.023 (0.050)	–	9.240	**	6.735	*	0.011
$\Delta(\text{epr}_t)$	–0.220 (0.047)	–0.029 (0.052)	–	11.304	**	7.465	**	0.007
$\Delta(\text{lncpi}_t)$	–0.180 (0.049)	–0.066 (0.045)	–	7.711	**	2.906		0.091
$\Delta(\text{lnipi}_t)$	–0.190 (0.040)	–0.013 (0.033)	–	11.406	**	11.527	**	0.000

Notes: Dependent variable is $\hat{v}_t$ ~ I(0); * significant at the 5% level; ** significant at the 1% level; asymptotic standard errors in parentheses.

Table 3. TAR and M-TAR results for the US

Variable	ρ_1	ρ_2	$\rho = 0$		$\rho_1 = \rho_2 = 0$		$\rho_1 = \rho_2$	
			t-max		Φ		*F*	*p*-value
					TAR			
dy_t	–0.758	–0.639	–3.161	**	37.568	**	0.292	0.590
epr_t	–0.722	–0.509	–2.617	**	34.121	**	0.987	0.322
lncpi_t	–0.814	–0.599	–3.096	**	39.763	**	0.978	0.325

lnipi_t	–0.809	–0.635	–3.200	**	40.197	**	0.617	0.434
				M-TAR				
dy_t	–0.759	–0.652	–3.490	**	37.209	**	0.263	0.609
epr_t	–0.692	–0.659	–3.492	**	33.150	**	0.024	0.877
lncpi_t	–0.812	–0.592	–2.952	**	39.430	**	0.978	0.325
lnipi_t	–0.805	–0.656	–3.355	**	39.769	**	0.468	0.495

Notes: Dependent variable is $\hat{v}_t$ ~ I(1); ** significant at the 1% level; critical values for *t*-max in the TAR and M-TAR models are -2.55 and -2.47 (1%); critical values for Φ in the TAR and M-TAR models are, respectively, 9.27 and 9.14 (1%).

Turning now to the results of $\hat{v}_t$, it can be seen that the null hypothesis that ρ_1 and ρ_2 are jointly equal to zero is rejected at the 1% level or better in all cases for both countries. For the US, the *t*-max and Φ statistics are reported and were compared to the critical values computed by Enders and Siklos (2001). Rejection of the null hypothesis in this case means that the volatility is cointegrated with the macroeconomic variables used in our analysis. However, we found no signs of asymmetric adjustment in any case of the US macroeconomic variables. For Portugal, we found that only ρ_2 is significantly different from zero in the TAR specification, and ρ_1 is significantly different from zero in the M-TAR specification. Locally, however, it seems that the convergence criterion of the model is not violated, although testing the threshold could perhaps add something more on this issue. A general acceptation of the error correction model given by (7), entails testing for asymmetric adjustment using the standard *F* critical values. The test procedures carried out for Portugal leads to the conclusion that there is no asymmetry in the TAR mechanism but there is asymmetry in the M-TAR mechanism for 3 of the 4 macroeconomic variables.

The 3 cases of asymmetric co-movements in Portugal using $\hat{v}_t$ refer to dividend yields, earnings price ratios, and industrial growth rates. As can be seen, in all cases $|\rho_1| > |\rho_2|$, which means that the model exhibits little decay for negative changes but substantial decay for positive changes in volatility relative to changes in macroeconomic factors. In other words, we may conclude that in the Portuguese stock market, volatility decreases tend to persist but increases tend to revert quickly toward the attractor.

The M-TAR results for the Portuguese stock market volatility suggest that volatility is obtained via an accumulation of changes in μ_{t-1} above the threshold followed by a sharp drop to the threshold. However, a similar pattern is not observed for changes in μ_{t-1} below the threshold, thus causing asymmetry. That is, volatility departs from its equilibrium level (given by the threshold) relative to the macroeconomic factors and periodically collapses to the threshold, sharply for positive and smoothly for negative changes.

In contrast with the case of the US, the asymmetric results for Portugal may challenge market efficiency according to the efficient market hypothesis. The EMH holds that stock prices adjust rapidly and unbiasedly to new and relevant price sensitive information. Under and over price adjustments relative to its fundamental value are unpredictable so that price changes are independent and random. Given the information provided by the macroeconomic variables, one would then expect that prices adjust rapidly and symmetrically toward the equilibrium level, which is consistent with cointegration (for non-stationary variables) and symmetric adjustment. This behaviour of prices would be transmitted to variables that capture stock returns and the corresponding volatility. Volatility itself would be unpredictable under the EMH. However, predictability is not a sufficient condition for market inefficiency. It is also need to prove that predictability allows for the possibility of generating systematic abnormal gains.

Stock market asymmetry may arise because investors are risk and loss averse. Risk aversion may encourage economic agents to react quickly to bad news while reacting more reluctantly to good news. On the other hand, asymmetries may arise driven by the potential loss in an overvalued stock market. Alternative explanations based on models of structural slumps and booms are also possible (Siklos, 2002). In our context, the volatility in the Portuguese stock market drops suddenly and periodically toward the attractor when it is substantially above the equilibrium level. Higher volatility induces greater risk and potentially larger losses, so it may be seen as a sign of bad news to the investor, which prompts him to react quickly to such news, and conversely for lower volatility. For our results, however, we found no evidence of such type of behavior in the US market.

6. Conclusions

This paper employs asymmetric conditional heteroskedasticity and threshold adjustment methodologies to inquire the asymmetric nature of stock market volatility in Portugal and the US. There is clear evidence of asymmetric volatility for Portugal and the US in the univariate conditional heteroskedasticity models. However, the leverage effect seems to be quadratic rather than exponential. For the threshold adjustment models, we found no evidence of asymmetric volatility behavior in the US relative to changes in macroeconomic variables, although transmission effects are cointegrated. Standard symmetric cointegration techniques could thus be applied to the US data. For Portugal, there is evidence of transmission effects between macroeconomic variables and the stock market volatility, and also that there is “sharp” movement asymmetry of volatility in some cases of the Portuguese stock market. The results for Portugal do not seem to be consistent with the general efficient market hypothesis, although the hypothesis itself cannot be denied.

7. References

Bollerslev, T. 1986, Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, Vol. 31, pp. 307-327.
Enders, W. and Granger, C.W.J. 1998, Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business & Economic Statistics*, Vol. 16(3), pp. 304-311.
Enders, W. and Siklos, P. 2001, Cointegration and threshold adjustment, *Journal of Business & Economic Statistics*, Vol. 19(2), pp. 166-176.
Engle, R.F. 1982, Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica*, Vol. 50, pp. 987-1008.
Glosten, L., Jaganathan, R., and Runkle, D. 1993, On the relation between the expected value and the volatility of the normal excess return on stocks, *Journal of Finance*, Vol. 48, pp. 1779-1801.
Granger, C.W.J. and Teräsvirta, T. 1993, *Modelling Nonlinear Economic Relationships*, Oxford University Press, New York.
Nelson, D. 1991, Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, Vol. 59, pp. 347-370.
Siklos, P. 2002, Asymmetric adjustment from structural booms and slumps, *Economic Letters*, Vol. 77, pp. 329-333.
Zakoian, J. 1990, *Threshold Heteroskedastic Models*, manuscript, CREST, INSEE, Paris.

RISK SENSITIVE DISCRETE- AND CONTINUOUS-TIME MARKOV REWARD PROCESSES

Karek Sladký
Department of Econometrics
Institute of Information Theory and Automation
Prague, Czech Republic

Abstract: In this note we consider risk sensitive Markov reward processes with finite state space in discrete- and continuous-time setting. Explicit formulas for the growth rates and average reward optimality are obtained and connections between discrete- and continuous-time models is discussed.

Keywords: discrete-time and continuous-time Markov reward processes, exponential utility functions, growth rates, average reward optimality, certainty equivalents

1. INTRODUCTION

The usual optimization criteria examined in the literature on optimization of Markov reward processes, e.g. total discounted or mean reward, may be quite insufficient to characterize the problem from the point of the decision maker. To this end it is necessary to select more sophisticated criteria that reflect also variability-risk features of the problem. One possible approach how to attack this problem is to consider, instead of linear objective functions, exponential functions (recall that both linear and exponential functions are separable and hence appropriate for sequential decision). Introducing of exponential objective functions in decision processes proved to be appropriate in many real-life problems, for details see Howard [7]. The research of risk-sensitive optimality criteria in Markov decision processes was initiated in the seminal paper by Howard and Matheson [6] and followed by many other researchers in recent years (see e.g. [1,2,3,4,11–13]).

In this note, we consider Markov reward processes with finite state space $I = \{1, 2, \ldots, N\}$ both in discrete- and continuous-time setting. In particular, in the discrete-time case, we consider Markov chain $X^{\mathrm{d}} = \{X_n, n = 0, 1, \ldots\}$ with finite state space $I = \{1, 2, \ldots, N\}$, transition probabilities p_{ij} and one-stage immediate rewards r_{ij} accrued to the transition from state i into state j. In the continuous-time setting, the development of the considered process $X^{\mathrm{c}} = \{X(t), t \geq 0\}$ (with finite state space I) over time is governed by the transition rates $\{q(j \mid i), i, j \in I\}$. For $j \neq i$ $q(j \mid i)$ is the transition rate from state i into state j, $q(i \mid i) = \sum_{j \in I, j \neq i} q(j \mid i)$ is the transition rate out of state i. As concerns reward rates, $r(i)$ denotes the rate earned in state $i \in I$, and $r(i, j)$ is the transition rate accrued to a transition from state i to state j. Let $\xi(t)$ be the (random) reward obtained up to time t; for the discrete-time case t is an integer, say $t = n$, and we write ξ^n instead of $\xi(t)$.

We shall suppose that the obtained random reward, say ξ, is evaluated by an exponential utility function, say $u^{\gamma}(\cdot)$, i.e. utility functions with constant risk sensitivity depending on the value of the risk aversion coefficient γ. In case that $\gamma > 0$ (the risk seeking case) the utility assigned to the (random) reward ξ is given by $u^{\gamma}(\xi) := \exp(\gamma\xi)$, if $\gamma < 0$ (the risk averse case) the utility assigned to the (random) reward ξ is given by $u^{\gamma}(\xi) := -\exp(\gamma\xi)$, for $\gamma = 0$ it holds $u^{\gamma}(\xi) = \xi$ (risk neutral case).
Hence we can write for $\gamma \neq 0$

$$u^{\gamma}(\xi) = \operatorname{sign}(\gamma) \exp(\gamma\xi) \tag{1}$$

and for the expected utility we have (E is reserved for expectation)

$$\bar{U}^{(\gamma)}(\xi) := \mathrm{E}u^{\gamma}(\xi) = \operatorname{sign}(\gamma)\ \mathrm{E}[\exp(\gamma\xi)], \tag{2}$$

and for the corresponding certainty equivalent $Z^{\gamma}(\xi)$ we have

$$u^{\gamma}(Z^{\gamma}(\xi)) = \operatorname{sign}(\gamma)\ \mathrm{E}\,[\exp(\gamma\xi)] \iff Z^{\gamma}(\xi) = \gamma^{-1} \ln\{\mathrm{E}[\exp(\gamma\xi)]\}. \tag{3}$$

Finally, recall that

$$u^{\gamma}(\xi^{(1)} + \xi^{(2)}) = \operatorname{sign}(\gamma)\ u^{\gamma}(\xi^{(1)}) \cdot u^{\gamma}(\xi^{(1)}),$$

i.e. exponential utility function is separable, what is very important for sequential decision problems.

In this note we focus attention on properties of the expected utility and the corresponding certainty equivalents if the stream of obtained rewards is evaluated by exponential utility functions both for the discrete- and continuous-time models. According to our knowlege, in the existing literature risk-sensitive optimality was studied only for discrete-time models; in the present note we show how these methods work also in continuous-time Markov reward case. Attention will be focused on the connections and similarity between discrete- and continuous-time risk-sensitive Markov reward models.

2. NOTATIONS AND PRELIMINARIES

1. Discrete-Time Case

We denote by $P=[p_{ij}]$ the $N\times N$ transition matrix of the chain X^{d}. Recall that the limiting matrix $P^* = \lim_{m\to\infty} m^{-1}\sum_{n=0}^{m-1} P^n$ exists, in case that the chain is aperiodic even $P^* = \lim_{n\to\infty} P^n$. In particular, if P is *unichain* (i.e. P contains a single class of recurrent states) the rows of P^*, denoted p^*, are identical. Obviously, $r_i = \sum_{j=1}^{N} p_{ij} r_{ij}$ is the expected one-stage reward obtained in state $i\in I$ and r denotes the corresponding N-dimensional column vector of one-stage rewards. Then $P^n\cdot r$ is the (column) vector of rewards accrued after n transitions, its ith entry denotes expectation of the reward if the process X^{d} starts in state i. To simplify our analysis in what follows we assume that P is irreducible, i.e. P contains a single class of recurrent states and no transient states. Then all rows p^* of the limiting matrix P^* are positive.

Let $\xi^n_{X_0} = \sum_{k=0}^{n-1} r_{X_k,X_{k+1}}$ be the sum of transition rewards received in the n next transitions of the considered Markov chain X and similarly let $\xi^{(m,n)}_{X_m}$ be reserved for the total (random) additive reward obtained from the mth up to the nth transition. Since $\xi^n_{X_0} = r_{X_0,X_1} + \xi^{(1,n)}_{X_1}$ and $u^\gamma(\cdot)$ is separable and multiplicative we immediately conclude that

$$u^\gamma(\xi^n_{X_0}) = \mathrm{sign}(\gamma)\, u^\gamma(r_{X_0,X_1})\cdot u^\gamma(\xi^{(1,n)}_{X_1}) = u^\gamma\left(\sum_{k=0}^{n-1} r_{X_k,X_{k+1}}\right)$$

and on taking expectations and conditioning on X_1 we get

$$\bar{U}^{(\gamma)}(\xi^n_{X_0}) = \mathrm{E}\, u^\gamma(\xi^n_{X_0}) = \sum_{j=1}^{N} p_{X_0 j}\cdot\exp(\gamma\, r_{X_0 j})\cdot \mathrm{E}\, u^\gamma(\xi^{(1,n)}_j). \quad (4)$$

Supposing that the chain starts in state $X_0 = i$ then for expected utility in the n next transitions we have (E_i is reserved for expectation if the process starts in state i)

$$U_i(\gamma,0,n) := \mathrm{E}_i[\exp(\gamma \sum_{k=0}^{n-1} r_{X_k,X_{k+1}})] \quad (5)$$

Similarly, for $m<n$ if the starting state $X_m = i$ we write

$$U_i(\gamma,m,n) := \mathrm{E}_i[\exp(\gamma \sum_{k=m}^{n-1} r_{X_k,X_{k+1}})].$$

(6)

Observe that the expected utility $\bar{U}_i(\gamma,m,n) = U_i(\gamma,m,n)$ if $\gamma>0$; in case that $\gamma<0$ it holds $\bar{U}_i(\gamma,m,n) = -U_i(\gamma,m,n)$ and $U_i(\gamma,n,n) = 1$.

In what follows we shall often abbreviate $U_i(\gamma,0,n)$ by $U_i(\gamma,n)$. Similarly $U(\gamma,n)$ is reserved for the (column) vector whose ith element equals $U_i(\gamma,n)$. $\bar{P} = [\bar{p}_{ij}]$ is an $N\times N$ nonnegative matrix with elements $\bar{p}_{ij} := p_{ij}\cdot \mathrm{e}^{\gamma r_{ij}}$. Hence by (4)

$$U(\gamma,n) = \bar{P}\cdot U(\gamma,1,n). \quad (7)$$

2. Continuous-Time Case

Let $Q=[q_{ij}]$ be the $N\times N$ matrix whose ij-th element is $q_{ij} = q(j\,|\,i)$ for $i \neq j$ and for the ii-th element we set $q_{ii} = -q(i\,|\,i)$. Recall that Q defines a continuous-time Markov process $X^{\mathrm{c}} = \{X(t),\ t \ge 0\}$ with transition probability matrix $P(t)$; the ij-th element of $P(t)$, denoted $P_{ij}(t)$, is the probability that state $j\in I$ is reached at time t from initial state i. It is well known that

$$P(t) = \exp[Qt] = \sum_{k=0}^{\infty} \tfrac{1}{k!}(Q\,t)^k, \quad (8)$$

and is a solution of the Kolmogorov backward and forward equations

$$\frac{\mathrm{d}P(t)}{\mathrm{d}t} = Q\cdot P(t), \quad \frac{\mathrm{d}P(t)}{\mathrm{d}t} = P(t)\cdot Q. \quad (9)$$

Recall that $P^* = \lim_{t\to\infty} P(t)$ (with $P^*Q = QP^* = 0$) always exists, and that the sojourn time of the considered process X^{c} in state $i\in I$ is exponentially distributed with mean value $q(i\,|\,i)$.

Hence the expected value of the reward obtained in state $i \in I$ equals $\tilde{r}_i = q(i\,|\,i)\,r(i) + \sum_{j\in I,\, j\neq i} q(j\,|\,i)\,r(i,j)$ and $\tilde{r}$ is the (column) vector of reward rates at time t. Then $P(t)\cdot\tilde{r}$ is a (column) vector of expected reward rate at time t, its i-th entry is the (expected) reward rate if the process X^{d} starts in state i. To simplify our analysis in what follows we assume that $Q = [q_{ij}]$ is irreducible, i.e. Q contains a single class of recurrent states and no transient states. Then the limiting matrix P^* irreducible and positive.

Let

$$\xi_{X(0)}(t) = \int_0^t r(X(\tau))\,\mathrm{d}\tau + \sum_{k=0}^{N(t)} r(X(\tau^-), X(\tau^+))$$

be the total (random) reward obtained up to time t,

where $X(t)$ denotes the state at time t, $X(\tau^-)$, $X(\tau^+)$ is the state just prior and after the k-th jump, and $N(t)$ is the number of jumps up to time t. Similarly

$$\xi_{X(t')}(t',t) = \int_{t'}^t r(X(\tau))\,\mathrm{d}\tau + \sum_{k=N(t')}^{N(t)} r(X(\tau^-), X(\tau^+))$$

is the total (random) reward obtained in the time interval $[t',t)$; hence

$$\xi_{X(0)}(t+\Delta) = \xi_{X(0)}(\Delta) + \xi_{X(\Delta)}(\Delta, t+\Delta) \quad \text{or}$$
$$\xi_{X(0)}(t+\Delta) = \xi_{X(0)}(t) + \xi_{X(t)}(t, t+\Delta).$$

Since $u^\gamma(\cdot)$ is separable and multiplicative we have

$$u^\gamma(\xi_{X(0)}(t)) = \mathrm{sign}(\gamma)\, u^\gamma(\xi_{X(0)}(\Delta))\cdot u^\gamma(\xi_{X(\Delta)}(\Delta,t)) \quad \text{or}$$
$$u^\gamma(\xi_{X(0)}(t+\Delta)) = \mathrm{sign}(\gamma)\, u^\gamma(\xi_{X(0)}(t))\cdot u^\gamma(\xi_{X(t)}(t,t+\Delta)).$$

On taking expectations and conditioning on $X(\Delta)$ we immediately conclude that for

$$U_i^{(\gamma)}(t) := \mathrm{E}\,\{u^\gamma(\xi_i(t))\}, \qquad U_i^{(\gamma)}(t',t) := \mathrm{E}\,\{u^\gamma(\xi_i(t',t))\}$$

we have (δ_{ij} is the Kronecker symbol)

$$U_i^{(\gamma)}(t+\Delta) = \sum_{j=1}^{N} P_{ij}(\Delta)\cdot[\mathrm{e}^{\gamma r(i)\Delta}\delta_{ij} + \mathrm{e}^{\gamma r(i,j)}]\cdot U_j^{(\gamma)}(\Delta, t+\Delta).$$

Since

$$P_{ij}(\Delta) = \begin{cases} 1 + q_{ii}\Delta + o(\Delta^2) & \text{for } i = j \\ q_{ij}\Delta + o(\Delta^2) & \text{for } i \neq j \end{cases}$$

on letting $\Delta \to 0+$ we conclude that

$$U_i^{(\gamma)}(t+\Delta) = (1 + q_{ii}\Delta)\mathrm{e}^{\gamma r(i)\Delta} U_i^{(\gamma)}(\Delta, t+\Delta) + \sum_{j=1,\, j\neq i}^{N} q_{ij}\Delta \mathrm{e}^{\gamma r(ij)} \cdot U_j^{(\gamma)}(\Delta, t+\Delta) + o(\Delta^2)$$

$$\mathrm{e}^{\gamma r(i)\Delta} = 1 + \gamma r(i)\Delta + o(\Delta^2).$$

Since the process X^{c} is time homogeneous, after some manipulation we arrive at

$$\frac{\mathrm{d}U_i^{(\gamma)}(t)}{\mathrm{d}t} = (q_{ii} + \gamma r(i))\; U_i^{(\gamma)}(t) + \sum_{j=1,\, j\neq i}^{N} q_{ij}\, \mathrm{e}^{\gamma r(i)} \cdot U_j^{(\gamma)}(t) \qquad (10)$$

that can be also written in matrix form as

$$\frac{\mathrm{d}U^{(\gamma)}(t)}{\mathrm{d}t} = \bar{Q} \cdot U^{(\gamma)}(t), \qquad (11)$$

where $U^{(\gamma)}(t)$ is a (column) vector, and $\bar{Q} = [\bar{q}_{ij}]$ is an $N \times N$ matrix with nonnegative off-diagonal elements

$$\bar{q}_{ij} = \begin{cases} q_{ii} + \gamma \cdot r(i) & \text{for } i = j \\ q_{ij} \cdot \mathrm{e}^{\gamma r(i,j)} & \text{for } i \neq j \end{cases}$$

Similarly to (8) and the solution of the Kolmogorov equations (9), we can conclude that the solution of (11) takes on the following form (*e* denotes unit column vector)

$$U^{(\gamma)}(t) = \exp[\bar{Q}t]\, e = \sum_{k=0}^{\infty} \tfrac{1}{k!} (\bar{Q}t)^k e. \qquad (12)$$

Hence $U^{(\gamma)}(t)$ is a linear combination of exponential functions with the exponents being the eigenvalues of the matrix $\bar{Q} = [\bar{q}_{ij}]$ and the real part of the eigenvalues determines the growth of the elements of $U^{(\gamma)}(t)$.

3. GROWTH RATES AND AVERAGE REWARDS

In this section we analyze growth rates and average rewards both for the discrete- and continuous-time models. Attention will be focused on the similarity of the both models.

1. Discrete-Time Case

Since $\bar{P}$ is nonnegative and irreducible, in virtue of the well-known Perron–Frobenius theorem (see [5]) spectral radius of $\bar{P}$ is equal to eigenvalue of $\bar{P}$ and the corresponding eigenvector $v(\bar{P})$ can be selected strictly positive, i.e.

$$\rho(\bar{P})\cdot v(\bar{P}) = \bar{P}\cdot v(\bar{P}). \tag{13}$$

Since $v(\bar{P})$ is unique up to a multiplicative constant, we can select $v(\bar{P})$ greater or less than unit vector. Hence on iterating (7) we obtain for suitably selected numbers $\alpha_1 \le 1 \le \alpha_2$ such that, since $\alpha_1 v(\bar{P}) \le e \le \alpha_2 v(\bar{P})$, since $U(\gamma,n,n) = e$ (unit vector)

$$(\bar{P})^n \cdot [\alpha_1 v(\bar{P})] \le U(\gamma,n) \le (\bar{P})^n \cdot [\alpha_2 v(\bar{P})]. \tag{14}$$

However,

$$(\bar{P})^n \cdot v(\bar{P}) = (\rho(\bar{P}))^n v(\bar{P})$$

hence

$$\alpha_1 (\rho(\bar{P}))^n v(\bar{P}) \le U(\gamma,n) \le \alpha_2 (\rho(\bar{P}))^n v(\bar{P}) \tag{15}$$

and the growth rate of $U(\gamma,n)$ is equal to $\rho(\bar{P})$.
Since for the corresponding certainty equivalent $Z^\gamma(\xi^n)$ we have (cf. (3))

$$Z^\gamma(\xi^n) = \gamma^{-1}\ln\{\mathrm{E}[\exp(\gamma\xi^n)]\} \tag{16}$$

implying that in virtue of (13)

$$Z^\gamma(\xi^n) = \gamma^{-1}\ln\{(\rho(\bar{P}))^n \cdot const.\}. \tag{17}$$

Moreover, for the mean value of the certainty equivalent we get

$$J(\gamma) := \lim_{n\to\infty}\frac{1}{n} Z^\gamma(\xi^n) = \gamma^{-1}\rho(\bar{P}). \tag{18}$$

2. Continuous-Time Case

Since $\bar{Q}$ has nonnegative off-diagonal elements and is irreducible, there exists real eigenvalue $\sigma(\bar{Q}) > 0$ and the real part of any other eigenvalue of $\bar{Q}$ is nongreater than $\sigma(\bar{Q}) > 0$ (see e.g. [5]). Moreover, the corresponding eigenvector $v(\bar{Q})$ can be selected strictly positive, i.e.

$$\sigma(\bar{Q}) \cdot v(\bar{Q}) = \bar{Q} \cdot v(\bar{Q}). \tag{19}$$

Since $v(\bar{Q})$ is unique up to a multiplicative constant, we can select $v(\bar{Q})$ greater or less than unit vector. Observe that the above facts follow immediately from the Perron–Frobenius theorem. Since for any (real) α it holds $(\sigma(\bar{Q}) + \alpha) \cdot v(\bar{Q}) = (\bar{Q} + \alpha I) \cdot v(\bar{Q})$ (the same holds for any other eigenvalue of $\bar{Q}$) and for sufficiently large α the resulting matrix is non- negative and (13) can be applied directly.

Since the solution of the set of equations (10), (11) is given by a linear combination of exponential functions with exponents being eigenvalues of the matrix $\bar{Q}$, the growth of $U^{(\gamma)}(t)$ is governed by real parts of the eigenvalues, in particular by $\sigma(\bar{Q})$, the eigenvalue with maximum real part. Moreover, since the corresponding eigenvector $v(\bar{Q})$ is strictly positive, the growth of each $U_i^{(\gamma)}(t)$ is the same and is given by $\sigma(\bar{Q})$. Then in analogy with the discrete-time case we can conclude that for suitable numbers $\alpha_1 \le \alpha_2$ we have

$$\alpha_1 \exp[\sigma(\bar{Q})\, t]\, v(\bar{Q}) \;\le\; U^{(\gamma)}(t) \;\le\; \alpha_2 \exp[\sigma(\bar{Q})\, t]\, v(\bar{Q}) \tag{20}$$

or for $\rho(\bar{Q}) := \exp[\sigma(\bar{Q})]$

$$\alpha_1 (\rho(\bar{Q}))^t\, v(\bar{Q}) \;\le\; U^{(\gamma)}(t) \;\le\; \alpha_2 (\rho(\bar{Q}))^t\, v(\bar{Q}) \tag{21}$$

implying to the corresponding certainty equivalent

$$Z^{\gamma}(\xi(t)) = \gamma^{-1} \ln\{(\rho(\bar{Q}))^t \cdot const.\} \tag{22}$$

and for its mean value

$$J(\gamma) := \lim_{t\to\infty} \frac{1}{t} Z^{\gamma}(\xi(t)) = \gamma^{-1} \rho(\bar{Q}). \tag{23}$$

Acknowledgement. This research was supported by the Czech Science Foundation under Grants 402/08/0107 and 402/07/1113.

References

[1] Bielecki T.D., Hernández-Hernández T.D., Pliska S.R. (1999) Risk-sensitive control of finite state Markov chains in discrete time, with application to portfolio management. Math Methods Oper Res 50:167–188

[2] Cavazos-Cadena R., Montes-de-Oca R. (2003) The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. Math Oper Res 28:752–756

[3] Cavazos-Cadena R. (2003) Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space. Math Methods Oper Res 57:253–285

[4] Jaquette S.A. (1976) A utility criterion for Markov decision processes. Manag Sci 23:43–49

[5] Gantmakher F.R. (1959) The Theory of Matrices. Chelsea, London

[6] Howard R.A., Matheson J. (1972) Risk-sensitive Markov decision processes. Manag Sci 23:356–369

[7] Howard R.A. (1988) Decision analysis: practice and promise. Manag Sci 34:679–695

[8] Puterman, M.,L. (1994): Markov Decision Processes – Discrete Stochastic Dynamic Programming. Wiley, New York

[9] Ross, S.M. (1970): Applied Probability Models with Optimization Applications. Holden–Day, San Francisco, CA

[10]Sladký K. (1976) On dynamic programming recursions for multiplicative Markov decision chains. Math Programming Study 6:216–226

[11]Sladký K. (1980) Bounds on discrete dynamic programming recursions I. Kybernetika 16: 526–547

[12]Sladký, K. (2006) Some remarks on risk-sensitive optimality criteria in Markov decision processes. In: Quantitative Methods in Economics (Multiple Critera Decision Making XIII). Proceedings of the International Conference, December 6–8, 2006, Bratislava, Slovak Republic, pp. 165–173

[13]Sladký, K. (2008) Growth rates and average optimality in risk-sensitive Markov decision chains. Kybernetika 44:205–226

[14]White, D.J. (1988) Mean, variance and probability criteria in

finite Markov decision processes: A review. J Optim Theory Appl 56: 1–29

KAREL SLADKÝ
Department of Econometrics
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 18208 Praha 8
e-mail: sladky@utia.cas.cz

COMPARISON VARIOUS PORTFOLIO INSURANCE AND STATIC STRATEGIES

Branislav Tuš

Abstract:
The continuing creating and developing new insurance strategies make new questioners about their effectiveness. This paper provides and illustrative comparison of various insurance strategies with buy and hold strategy. In terms of insurance strategies it is compared the classical Constant proportion portfolio strategy with their modification using risk measures as Value at Risk and Expected Shortfall.
The results indicates that buy and hold strategy does not dominate the portfolio insurance strategies. Moreover portfolio insurance strategies can outperform buy and hold strategy measuring their risk return characteristics.

Keywords: CPPI, Value at Risk, Insurance, Expected Shortfall, Monte Carlo simulation

1. Introduction

At the beginning the portfolio strategies techniques were mainly realized trough options. The idea behind it was that a strategy which would provide protection against market losses while preserving the upward potential should have considerable appeal to a wide range of investors. This patterns is created by buying a put option on a portfolio. Recently a new insurance strategies were introduced such as constant proportion portfolio insurance (CPPI), Stop-Loss, synthetic put.
When it comes to mutual funds the option based strategy were used in closed fund where the investor could enter or exit the fund without a big penalty only on reference dates. The later strategies like CPPI are suited more for open-ended mutual funds.
The decision between choosing portfolio insurance strategy instead of buy and hold strategy is not clear and depends on investors utility function. Under the assumption that returns are log normally distributed a risk averse investor may prefer a portfolio insurance strategy to a buy and hold strategy. On the other hand PI strategies can generate excess return comparing static portfolio. So far there is no clear evidence of effectiveness of PI strategies.
In general the, investments are traditionally measured in mean – variance space. Since the PI strategies are provide with asymmetric distribution function this measures are no well suited for comparing them with classical

strategies. The difference is that the volatility in PI strategy could be created by positive returns which is good for an investor. So for this reasons measures like Value at Risk (VaR) or Expected Shortfall would present a better statistical risk measures. Although the VaR which measures the potential loss with predefined probability on a specified period, this standard measure is not coherent. Therefore Expected Shortfall (EP) has become popular.

Portfolio insurance strategies

Using stop-loss insurance means that the portfolio is fully invested in risky asset as long as its value does not reach some critical level (stop loss) , which is the minimum discounted target level. Ones the portfolio reaches the critical level the entire portfolio is switched to risk less instruments, ensuring the target is reached at the end of the period.

The synthetic put strategy is a dynamic portfolio strategy that replies the payoff of protective put (long positions in a put and in underling asset). Strike price is set to equal the desired floor. The portfolio is invested in risky and risk-free asset. The proportion is calculated using option pricing model.

Constant Proportion Portfolio Insurance – this strategy also invest in risky and risk free assets. The risky asset is determined by the cushion, which is the difference between the actual value of the portfolio and the discounted minimum value which ensure to meet the target level at the and of the reference day. The second deterministic factor is multiplier. The higher the multiplier the more risk assets are invested while the rest is invested in risk-free asset. So the investor has to choose the multiplier. The higher the multiplier the more risk assets are invested, but in case of rapid fall on the markets the sooner the investor approaches the floor and stays for the rest of the reference period in the money market. If the investor choose very conservative level of multiplier the strategy would not participate on strong markets so would not bring a big excess return over money market instruments.

So if the cushion approaches zero also exposure to the risky asset approaches zero and the portfolio is completely invested in cash. The guarantee is perfect if the portfolio value never descends below the minimum discounted floor value. This could not be meet in case of sharp drops in financial markets. But there is no ability to minimize this risk.

The typical characteristics of this strategy is that it has a right oriented returns distribution like option based strategies.

Mathematical description

V_t value of the portfolio at time t, $t>0$

V_0 initial investment, guaranteed amount

P_t	value of the floor at time t
$Ct=V_t-P_t$	„cushion“ at time t
m	multiplier
E_t	exposure , or investment in the risky asset at time t
T	maturity of the strategy
r	risk-free rate
S_t	price of the risky asset at time t

The evolution of the portfolio is described by the following stochastic deferential equation.

$$dVt = \frac{dSt}{St} \times Et + rdt \times (Vt - Et)$$

If we assume that the floor grows at fixed rate r, less than or equal to risk-free rate. The guaranteed amount being V_0.

$$P_t = P_0 e^{rt}$$

$$P_0 \geq V_0 e^{-rt}$$

We suppose the price of risky asset follows a standard model

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

With constrains on short selling

$$0<Et<p*Vt$$

Where p>0,

$$dV_t = \left\{ \begin{array}{l} V_t \times rdt; V_t \leq p_t \\ (V_t - P_t) \times dZ_t + P_t \times rdt; P_t < V_t < \frac{m}{m-p} P_t \\ V_t \times dX_t; V_t \geq \frac{m}{m-p} P_t \end{array} \right\}$$

Modifications of CPPI.

The main parameter of CPPI is the multiplier. One of the modifications is to dynamically determine the multiplier using risk characteristics.

The multiplier is determined by Expected Shortfall (or Value at Risk). Multiplier is set that portfolio value would not decease below the minimum guaranteed value on defined horizon with probability (99 %).

The other modification is related to the minimum discount value (floor). In classical model when the total portfolio raises the difference between the actual portfolio value and minimum discounted value grows which enables to invest more in risky asset. The advantage is that in rising market the proportion of risky asset grows in the bear markets is declining. In modification the floor changes when the portfolio value growths at specified amount. That means for instance if the portfolio value grows 2 %, comparing the notional value set at the beginning of the reference date, the discounted floor increases the same level.
This modification does the opposite that the classical method (in rising market the proportion of risky asset does not grows so rapidly and in declining market the proportion does not decrease so dramatically). The aim of this method is to stabilize the proportion of risky asset.

2 Performance measurement

Portfolio theory assumes that investors select portfolios that are optimal to them. Optimality is often understood as maximizing expected utility of wealth. As such, choosing between portfolios amounts to choosing between return distribution. Comparison is mostly done by comparing average excess return over risk free rate, with standard deviation with combining both and using sharpe ratio.
Unfortunately such characteristic are not suitable for PI, where there is a protection against downward and limitation against upward. Counting a number of negative occurrences is also not the right to contemplate. The asymmetric risk measures has been in recent time the put forward as an alternative to symmetric risk measures. A typical one is Value at Risk measure. The drawback of this measure is that does not coverage the extreme situation on the market. In this case a better characteristics provide Expected Shortfall which calculates average loss below the confidence level limit.
The problem of all tis measures is that if they are put together they can lead to contradictory results.

3 Objective of this study

In this paper we compare the strategies in terms of average excess return and the expected shortfall of the strategy. As an objective we also compare the values at the end of the reference period.
In this article we attempt to describe the various here above mentioned strategies. Trying to find the optimal settings of PI strategies to maximize

expected utility function. In the article we set a static strategy as an benchmark and comparing various strategies against it.
First there is presented a mathematical description of mentioned strategies.

3.1 Changing the floor value

In this section we examined the floor value. In particular the lower the floor the more risk the strategy will be, that means more risky asset will be allocated to the strategy. We do not attempt to find an optimal floor for the strategy, which is impossible without a particular distribution for risky asset returns. We try to present the impact of the change of the initial floor.

3.2 Changing the multiplier

The multiplier is the has the most biggest impact on the total performance of the strategy. We compare a strategies with fixed multiplier but set on different levels and compare it to the dynamic changing multiplier and with variable floor. .

4. Performance comparison

We employ the following terminology in discussing the various strategies discussed above:

Static strategy
Stop-Loss
CPPI
Modified CPPI

Details of simulation

As a risky asset we use a index MSCI World which is a capitalization index that comprise stock around the world.
First we use historical data for MSCI World and apply mentioned strategies.
In second we generate 100 000 of two year daily returns simulations and sort them from worst to best. We assume 250 trading days in the year.
The simulation of MSCI World were realized under this parameters.

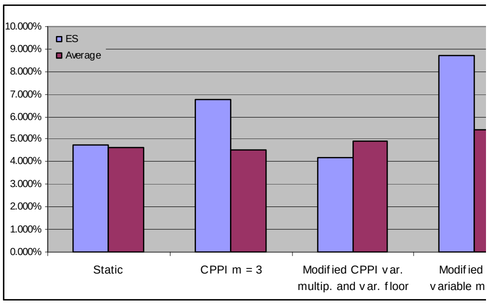
$$\mu = 7\ \%,\ \sigma = 12\ \%,\ T = 17 \text{ years},\ V_0 = 100$$

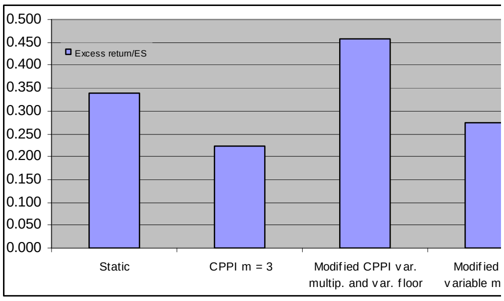
From these scenarios we chose scenario according to the quantiles. As a horizon we set a 2 year period. As a risk free rate we took a EUR 1 month money market rate.

Simulation results

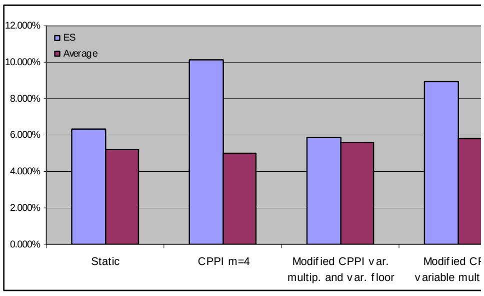
Changing the multiplier

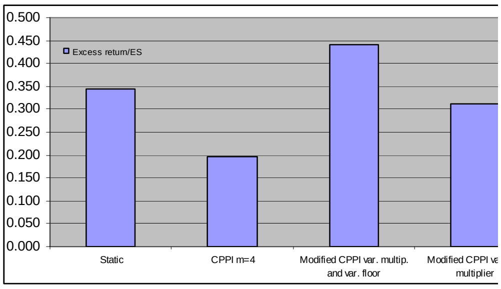
m=3


10.000%
9.000%
8.000%
7.000%
6.000%
5.000%
4.000%
3.000%
2.000%
1.000%
0.000%
ES
Average
Static
CPPI m = 3
Modified CPPI var. multip. and var. floor
Modified variable m

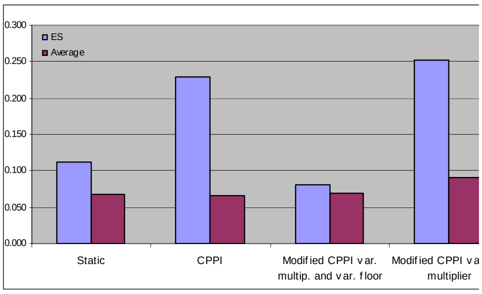

0.500
0.450
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
Excess return/ES
Static
CPPI m = 3
Modified CPPI var. multip. and var. floor
Modified variable m

m=5


12.000%
10.000%
8.000%
6.000%
4.000%
2.000%
0.000%
ES
Average
Static
CPPI m=4
Modified CPPI var. multip. and var. floor
Modified CP variable mult


0.500
0.450
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
Excess return/ES
Static
CPPI m=4
Modified CPPI var. multip. and var. floor
Modified CPPI va multiplier

m=7


0.300
0.250
0.200
0.150
0.100
0.050
0.000
ES
Average
Static
CPPI
Modified CPPI var. multip. and var. floor
Modified CPPI va multiplier

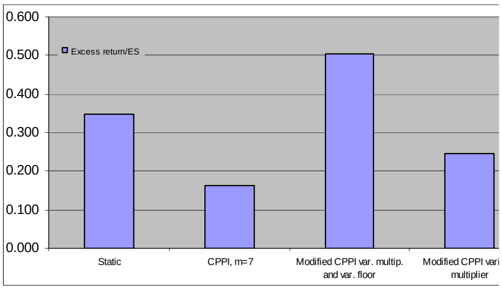
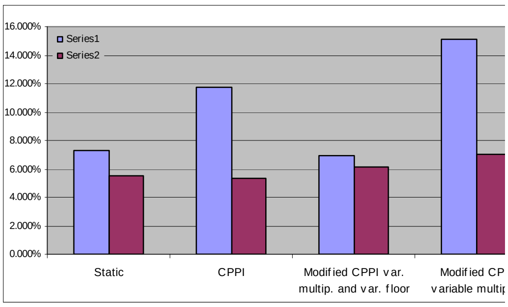

0.600
0.500
0.400
0.300
0.200
0.100
0.000
Excess return/ES
Static
CPPI, m=7
Modified CPPI var. multip. and var. floor
Modified CPPI vari multiplier

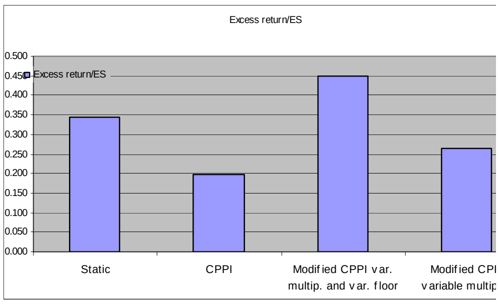
Table n.1 Performance comparison

	Average return annually	Expected Shortfall	Excess return/ES	Number of trades
Static	4.626%	4.753%	0.339211	1
CPPI m = 3	4.517%	6.756%	0.222524	17
Modified CPPI var. multip. and var. floor	4.927%	4.188%	0.456898	24
Modified CPPI variable multiplier	5.414%	8.741%	0.274536	93
	Average return annually	**Expected Shortfall**	**Excess return/ES**	**Excess return/ES**
Static	5.764%	7.946%	7.946%	1
CPPI m = 5	5.664%	13.893%	13.893%	75
Modified CPPI var. multip. and var. floor	6.144%	6.964%	6.964%	67
Modified CPPI variable multiplier	7.032%	15.136%	15.136%	75
	Average return annually	**Expected Shortfall**	**Excess return/ES**	**Excess return/ES**
Static	6.739%	11.145%	34.669%	1
CPPI m = 7	6.573%	22.909%	16.140%	132
Modified CPPI var. multip. and var. floor	6.900%	7.998%	50.314%	101
Modified CPPI variable multiplier	9.047%	25.143%	24.544%	75

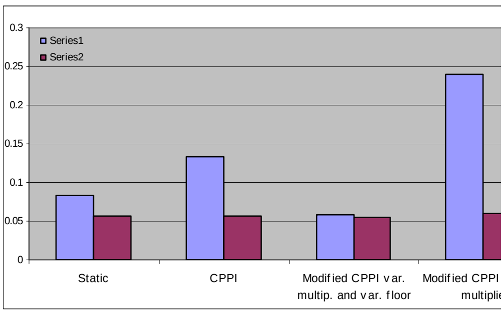
Changing the floor

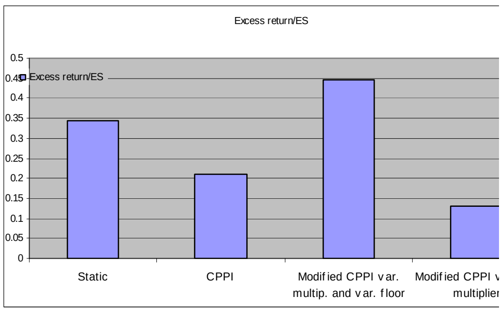
Risk budget = 1 % (Guaranteed amount = 99, V_0 = 100)



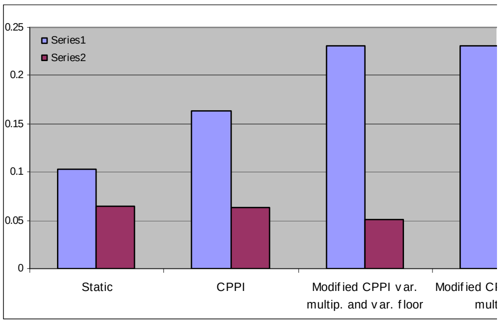


Risk budget = 2 % (Guaranteed amount = 98, V_0 = 100)





Risk budget = 4 % (Guaranteed amount = 96, V_0 = 100)



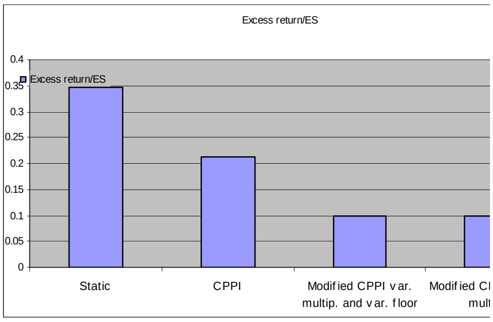


Table n.2 Performance comparison

Risk budget = 1 %	**Average return annually**	**Expected Shortfall**	**Excess return/ES**	**Number of trades**
Static	5.544%	7.330%	0.345254	1
CPPI m = 4	5.341%	11.714%	0.198694	74
Modified CPPI var. multip. and var. floor	6.144%	6.964%	0.449432	74
Modified CPPI variable multiplier	7.032%	15.136%	0.265474	104
Risk budget = 2 %	**Average return annually**	**Expected Shortfall**	**Excess return/ES**	**Excess return/ES**
Static	5.731%	8.284%	0.344656	1
CPPI m = 4	5.650%	13.275%	0.209006	108
Modified CPPI var. multip. and var. floor	5.471%	5.834%	0.444818	78
Modified CPPI variable multiplier	6.034%	23.921%	0.132021	137
Risk budget = 4 %	**Average return annually**	**Expected Shortfall**	**Excess return/ES**	**Excess return/ES**
Static	6.429%	10.266%	0.346186	1
CPPI m = 4	6.356%	16.326%	0.213167	117
Modified CPPI var. multip. and var. floor	5.516%	5.983%	0.441318	80
Modified CPPI variable multiplier	5.141%	23.048%	0.098291	164

5. Concluding remarks

The present study shows the impact on risk, return characteristic for various PI strategies comparing them to static strategy. First we demonstrated that static buy and hold strategies does not outperform a PI. Despite the PI strategies generate lower return they are associated with lower risk characteristic. Combining risk return characteristics the PI strategies presented better results.

Setting parameters in PI strategies has a crucial impact on risk return characteristics of the strategy. We demonstrated that dynamic minimal guaranteed floor with dynamic multiplier PI startegy could outperform buy and hold strategy. The dynamic strategies especially the ones with dynamic parameters performed high rebalancing frequency and trading costs.

References

1. BOULIER and KANNIGANTI, Expected performance and risk of various portfólio strategies
2. Acervi, C.Tasche, D., 2002. On the coherence of expected shortfall. Journal of Banking and Finance 26, 1487-1503
3. Benninga, S., 1990. Comparing portfolio insurance strategies. Finanzmarkt und Portfolio Management 4, 20-30
4. Black, F., Perold, A.F., 1992. Theory of constant proportion insurance. Journal of Economic Dynamic and Control 16, 403-426
5. Garcia, C.B., Gould , F.J., 1987 . An empirical study of portfolio insurance. Financial Analyst Journal 43, 44-54.

Contact

Ing. Branislav Tuš
Allianz-Slovenská dôchodková správcovská spoločnosť
Račianska 42, 831 02 Bratislava
Tel: +421 905 207 984
email: tusb@asdss.sk

WAVELET NEURAL NETWORKS PREDICTION OF CENTRAL EUROPEAN STOCK MARKETS *

Lukas Vacha[1], Jozef Barunik[1]
[1]Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic, Prague.

Abstract
In this paper we apply neural network with denoising layer method for forecasting of Central European Stock Exchanges, namely Prague, Budapest and Warsaw. Hard threshold denoising with Daubechies 6 wavelet filter and three level decomposition is used to denoise the stock index returns, and two-layer feed-forward neural network with Levenberg-Marquardt learning algorithm is used for modeling. The results show that wavelet network structure is able to approximate the underlying process of considered stock markets better that multilayered neural network architecture without using wavelets. Further on we discuss the impact of structural changes of the market on forecasting accuracy, and we find that for certain periods the one-step-ahead prediction accuracy of the direction of the stock index can reach 60% to 70%.

Keywords: neural networks, hard threshold denoising, time series prediction, wavelets.

1. Introduction
Traditional prediction methods for time series often restrict on linear regression analysis, exponential smoothing, and ARMA. These methods generally produce reasonable prediction results for stationary random time series of linear systems. In the recent decades, development in econometrics brought also methods which are capable of forecasting more complex systems, such as stock markets. These mainly include Wavelet Decomposition [10], [4], [1] and Neural Networks analysis, [5], [9]. Further on, the idea of combining both methods into single wavelets and neural networks has resulted in the formulation of wavelet networks. This research area is new, and there is tremendous potential for its development and application to various fields. This paper is one of the first attempts to fit this exciting method to Central European Stock markets, namely Prague Stock Exchange, Budapest Stock Exchange and Warsaw Stock Exchange. Our main expectation will be that this method will improve the single neural network approach.

2. Denoising with Wavelets
Classical time series denoising approaches are rooted in Fourier analysis where noise is assumed to be represented mainly as high frequency oscillations. The wavelet based denoising, assumes that analysis of time series at different resolutions might improve the separation of the true underlying signal from noise. Let us begin with description of the basics of Wavelet decomposition theory.

2.1 Wavelets

There are two types of wavelets: father wavelets φ and mother wavelets ψ. The father wavelet integrates to unity and the mother wavelet integrates to zero. The father wavelet, also called scaling function, essentially represents the smooth, trend, i.e., the low frequency part of the signal, on the other hand the mother wavelet represents the details, i.e., the high frequency part of the signal. The mother wavelet is compressed or dilated in time domain, to generate cycles to fit the actual time series. The formal definition of the father φ and mother ψ wavelet is

$$\varphi_{j,k} = 2^{-\frac{j}{2}}\varphi\left(\frac{t-2^j k}{2^j}\right), \quad \psi_{j,k} = 2^{-\frac{j}{2}}\psi\left(\frac{t-2^j k}{2^j}\right)\sqrt{a^2+b^2}$$

where j is the scale (or dilatation) and k is the translation (or shift). Commonly, many types of wavelets can be possibly used, including Haar wavelet, Mexican hat, Morlet wavelet, Daubechies wavelet etc. In the empirical part, we will use Daubechies wavelet db6.

Any time series $x(t)$ can be built up as a sequence of projections onto father and mother wavelets indexed by both j, the scale, and k, the number of translations of the wavelet for any given scale. Usually k is assumed to be dyadic. The wavelet coefficients are approximated by integrals

$$s_{J,k} \approx \int_{-\infty}^{\infty} x(t)\varphi_{J,k}(t)\,dt, \quad d_{j,k} \approx \int_{-\infty}^{\infty} x(t)\psi_{j,k}(t)\,dt$$

$j = 1,2,\ldots,J$, where J is the maximum scale. The wavelet representation of the time series $x(t)$ in $L^2(R)$ [1] can be given by

$$x(t) = \sum_k s_{J,k}\varphi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}(t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots + \sum_k d_{1,k}\psi_{1,k}(t)$$

where the basis functions $\varphi_{J,k}(t)$ and $\psi_{j,k}(t)$ are assumed to be orthogonal. When the number of observations is dyadic, the number of the wavelet coefficients of each type at the finest scale 2^1 is $N/2$, labeled $d_{1,k}$. The next scale 2^2 has $N/2^2$ coefficients, labeled $d_{2,k}$. At the coarsest scale 2^J there are $N/2^J$ coefficients $d_{J,k}$ and $s_{J,k}$.

3. Nonlinear Wavelet Denoising

The simplest method of nonlinear wavelet denoising is via thresholding. The procedure sets all wavelet coefficients that has lower value than some fixed

---

[1] Square integrable real-valued function, $\int_{-\infty}^{\infty} x^2(t)\,dt < \infty$.

constant to zero. Two thresholding rules were instrumental in the initial development of wavelet denoising, both for their simplicity and performance: hard and soft thresholding [4].

3.1 Threshold Selection

Optimal thresholding occurs when the threshold is set to the noise level, i.e., $\eta = \sigma_{\in}$. Setting $\eta < \sigma_{\in}$ will allow unwanted noise to enter the estimate while setting $\eta > \sigma_{\in}$ will destroy information that belongs to the underlying signal. Following [3] we can set a universal thresholding as

$$\eta^{U} = \widehat{\sigma}_{\in}\sqrt{2\log N},$$

where N is the sample size. In practical situations the standard deviations of noise $\sigma_{\in}$ is not known. The most commonly used estimator of $\sigma_{\in}$ is the maximum absolute deviation (MAD) standard deviation [10].

$$\widehat{\sigma}_{MAD} = \frac{median\left(\left|d_{1,1}\right|,\left|d_{1,2}\right|,\ldots,\left|d_{1,N/2-1}\right|\right)}{0.6745}.$$

The denominator is needed to rescale the numerator so that $\widehat{\sigma}_{MAD}$ is tuned to estimating the standard deviation for Gaussian white noise [4].

3.2 Hard Thresholding

In our paper we use a hard thresholding. The hard thresholding rule on the wavelet coefficients o_t is given by

$$\delta_{\eta}^{H}\left(o_t\right) = \begin{cases} o_t & \text{if } \left|o_t\right| > \eta \\ 0 & \text{otherwise} \end{cases},$$

where η is the threshold value. The operation is not a continuous mapping, it only affects input coefficients that are less or equal to the threshold η. After obtaining the thresholded wavelet coefficients using δ_{η}^{H} we compose the denoised times series via an inverse wavelet transform (IDWT) so we get $\hat{x}_{den}(t)$. For a more detailed treatment see [10].

4. Wavelet Network Structure

Wavelet Network is a network combining the ideas of the feed-forward neural networks and the wavelet decomposition. Wavelet networks use simple wavelets and wavelet network learning is performed by the standard type algorithm such as Conjugate-Gradient, or more efficient Levenberg-Marquardt [6], [8]. Neural networks can be viewed as universal approximation tools for fitting linear or nonlinear models, as [5] showed. Limiting space of this paper do not allow us to explore Neural Networks estimation methodology in detail, but reader is advised to follow i.e. [12], or [9] for very good explanation.

There are basically two main approaches to form wavelet networks. In the first approach, the wavelet decomposition is decoupled from the learning component of neural network architecture. In other words, the series are firstly decomposed / denoised using wavelets, and then fed to the neural network. In the second approach, the wavelet theory and neural networks are combined into a single method, where the inputs $x_1, \ldots, x_k$ with weights $\omega_1, \ldots, \omega_n$ are combined to estimated output in Multilayer feed-forward network (MPL):

$$\widehat{x}_{DWNN}(t) = \sum_{i=1}^{N} \omega_i f(\gamma_i x(t) + \beta_i),$$

where f is an activation function, $\gamma_i, \beta_i, \omega_i$ are network weight parameters that are optimized during learning, and N is number of hidden layers. If we feed this network with nonlinear wavelet denoised values $\widehat{x}_{den}(t)$ (see section 3) we will get the form of the wavelet neural network (DWNN) used in this paper.

5. Results

In the testing, we focus on sample of 1050 daily returns from 7.1.2004 to 3.4.2008 of value-weighted indices PX-50, BUX and WIG (Prague, Budapest and Warsaw stock exchanges respectively). The dataset was downloaded from the server www.stocktrading.cz. In the prediction task we start with denoising of the 512 data sample with db6 wavelet filter and a 3-level decomposition. Then two-layer neural network with 5 neurons in each layer and Levenberg-Marquardt learning algorithm is used to learn the sample. To avoid over-fitting of the network we use a window of 50 real out-of-sample data on which we test the estimated model on one day predictions. This algorithm is repeated 10 times with moving window of 50, so the final prediction of 500 data is obtained. Finally we compare this method to neural network approach so we can see if the wavelet denoising layer improves the forecasts. Architecture of the network is again two layers with 5 neurons and Levenberg-Marquardt optimization.

As for evaluation, we focus mainly on out-of-sample performance, as it is most important in financial time series forecasting. We consider Root Mean Square Error statistics (RMSE) to see the performance of out-of-sample prediction. Further on, we use statistics proposed by Pesaran Timmerman – SR (PT) [11], which evaluates the correctness of the signs prediction. Such statistics is often used in financial literature as the predicted positive change predicts buy signal, negative change sell signal which allows evaluating a trading strategy. Pesaran Timmerman statistics is based on the null hypothesis that a given model has no economic value in forecasting direction and is approximately normally distributed. In other words, we test the null hypothesis that the signs of the forecasts and the signs of actual variables are independent. If the prediction of signs is statistically dependent, we approached a good forecasting model with economic significance.

Finally to test the performance of the wavelet network and the neural network approach we use the statistic proposed by Clark and McCracken (CM) [2]. They compare out-of-sample accuracy for two models which are nested. The statistics is normally distributed under the null hypothesis of equal predictive ability of the two models.

Table 1: Prediction results for PX50, BUX and WIG

	PX50		BUX		WIG	
	MPL	DWNN	MPL	DWNN	MPL	DWNN
RMSE	0.8	0.175	0.08	0.11	0.37	0.41
CM	**0.78*** **		**0.25*** **		-7.6	
SR (PT)	0.52	**0.56** **	0.51	**0.53*** **	0.47	0.49

*, **, ***, 1%, 5% and 10% significance levels

To compare the results, we can see that Wavelet Neural Networks (DWNN) performed best on BUX returns with lowest RMSE, while RMSE of PX50 was little bit higher, and WIG more than double of PX50 and BUX. This would indicate that the DWNN model will forecast the PX50 and BUX returns better than WIG. This is confirmed by Pesaran Timmerman statistics which is significant on 5% level of significance for PX 50, 10% level of significance of BUX but is not significant at all for WIG. Success rate of correct forecasted direction is 0.56 for PX50, 0.53 for BUX and 0.49 for WIG. As to the Multilayer Feedforward architecture (MPL) without denoising the directional accuracy is not significant at all. But if we compare the DWNN and MPL using Clark and McCracken statistics, we can see that DWNN yields significantly better prediction accuracy for PX50 and BUX series. For WIG the results are not significantly different, which would be expected as WIG does not seem to be significantly predictable even using Pesaran Timmerman.

To sum up the achieved results we can say, that DWNN performs significantly better on the PX50 and BUX returns prediction. Although reader can notice, that the success rate is quite low. This is probably caused by quite large data sample which contains large structural changes as recent stock market crash of February 2008, etc. Even though Wavelet Neural Networks are considered as a universal approximation theorem as mentioned before, reader can see that if we feed it with data which simply cannot be approximated, its performance is poor. As the stock market structure changes in time quite quickly, testing the prediction on such a large data samples does not seem to provide reasonable predictions. Thus our last test is focused on the division of the dataset into 3 month moving windows, where we simply look at how does the Success Rate statistics of Pesaran Timmerman evolve in time.

Figure 1: Success Rate statistics of Pesaran Timmerman of PX50, BUX and WIG in time.

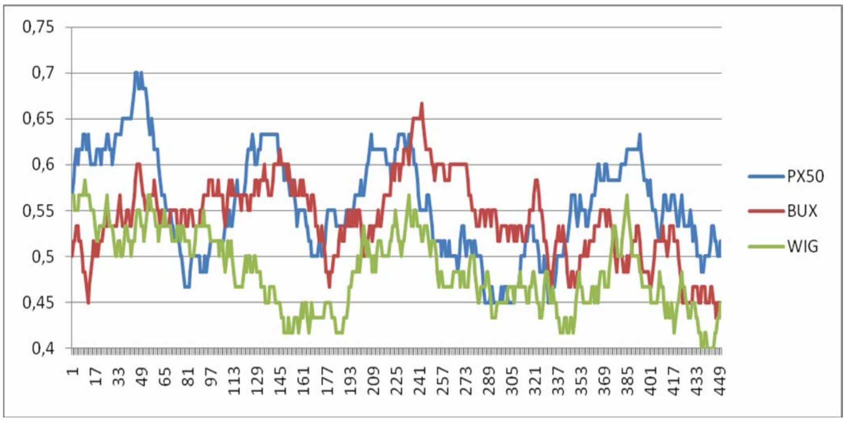


Figure 1 demonstrates that our assumption was right. For some periods, the statistics for PX50 and BUX returns is 0.6 to 0.7, which means that 60% to 70% of the one day sign change is predicted correctly. For WIG returns the predictability does not excess 55%, which can lead us to suspicion that WIG simply does not contain strong predictable patterns. As to other two tested series, we can see that the pattern strongly evolves over time, and that it would make sense to adjust appropriate method for forecasting each 3-month. This would be done using adjusting wavelets and their levels, number of hidden layer of neural network, etc. which we leave for further research.

6. Conclusion

Our results indicate that the Wavelet Neural Network might outperform simple Neural Networks while forecasting Central European stock exchanges. More concretely, PX50 and BUX returns were predicted using DWNN structures significantly better than using MPL Network structure. We also conclude that the stock market structural changes affected the final stock market direction prediction greatly. Dataset used for testing was quite large and contained structural changes such as large market crash of February 2008 which leads to significantly lower prediction accuracy of the used methods. For this reason we also used the three month moving window using which we have showed, that for some periods the prediction accuracy reached sustainable 60% to 70%, meaning 60% to 70% future directions of the stock markets were predicted correctly using PX50 and BUX data. On the other hand, the prediction accuracy using WIG data did not improve a lot. Thus we conclude that this market does not simply contain stronger predictable patterns. Further research should concentrate on the exploring of dynamically adjusted wavelet types, number of hidden layers of neural networks or other parameters specific to structural changes in the forecasted underlying series.

References:

[1] Abramovich F., Bailey T.C., Sapatinas T. (1999), Wavelet Analysis and its Statistical Applications, J. R. Statist. Soc. D, 48.

[2] Clark, Tod E., and Michael W. McCracken (2001): Tests of Forecast Accuracy and Encompassing for Nested Models, Journal of Econometrics 105: 85-110.

[3] Donoho, D. L. and Johnstone, I. M. (1994) Ideal Spatial Adaptation by Wavelet Shrinkage. Biometrica, 81, 422-455.

[4] Gençay, R., Selçuk, F., Whitcher, B. (2002), An Introduction to Wavelets and Other Filtering Methods in Finance and Economics. Academic Press.

[5] Hornik, K., (1989): Multilayer Feedforward Networks are universal Approximators, Neural Networks 2, 359-366.

[6] Levenberg, K. (1944): A Method for the Solution of Certain Problems in Least Squares. Quart. Appl. Math. 2, 164-168

[7] Mallat S. (1998), A Wavelet Tour of Signal Processing, Academic Press

[8] Marquardt, D. (1963): An Algorithm for Least-Squares Estimation of nonlinear Parameters. SIAM J.Appl.Math. 11, 431-441.

[9] McNelis, P.D. (2005): Neural Networks in Finance: Gaining predictive edge in the markets, Elsevier Academic Press, ISBN 0-12-485967-4

[10] Percival, D. B., Walden, A. T. (2000), Wavelet Methods for Time series Analysis. Cambridge University Press.

[11] Pesaran, M.H., and A.Timmermann (1992): A Simple Nonparametric test of Predictive Performance, Journal of Business and Economic Statistics 10: pp. 461-465

[12] Poggio, T. and F. Girosi (1990): Networks for Approximation and Learning. Proc. of the IEEE, vol. 78, no.9, pp. 1481-1497.

LIST OF PARTICIPANTS

Andrade Marina, SCTE - Instituto Superior de Ciências do Trabalho e da Empresa UNIDE - Unidade de Investigação e Desenvolvimento Empresarial
Bartl Ondrej, University of Žilina, Faculty of Management Science and Informatics, Slovakia,
Baruník Jozef, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic
Baxa Jaromír, ÚTIA AV ČR
Brezina Ivan, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Coelho Manuel Francisco Pacheco, Instituto Superior de Economia e Gestão SOCIUS – Centro de Investigação em Sociologia Económica e das Organizações
Čemická Katarína, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Čičková Zuzana, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Domonkos Tomáš, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Fábry Jan, Department of Econometrics, University of Economics, Prague, Czech Republic
Fendek Michal, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Ferreira Manuel Alberto M, ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa UNIDE - Unidade de Investigação e Desenvolvimento Empresarial
Ferreira Nuno B, ISCTE, Lisboa, Portugal
Fiala Petr, Department of Econometrics, University of Economics, Prague, Czech Republic
Filipe José António Candeias Bonito, ISCTE - Instituto Superior de Ciências do Trabalho e da Empresa UNIDE - Unidade de Investigação e Desenvolvimento Empresarial
Furková Andrea, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Goga Marián, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Groshek Gerald, University of Redlands School of Business Redlands, California, USA
Hančlová Jana, Faculty of Economics, VŠB-Technical University of Ostrava, Czech Republic
Horniaček Milan, Institute of Economics, Faculty of Social and Economic Sciences, Comenius University, Slovakia
Husár Jaroslav, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Chocholatá Michaela, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Ivaničová Zlatica, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Jablonský Josef, Department of Econometrics, University of Economics Prague, Czech Republic
Janáček Jaroslav, Faculty of Management Science and Informatics, University of Žilina, Slovakia
Janáčková Marta, Žilinská univerzita Žilina, Strojnícka fakulta, Katedra aplikovanej matematiky
Jánošíková Ľudmila, University of Žilina, Faculty of Management Science and Informatics, Department of Transportation Networks, Slovakia
Kaňková Vlasta, Department of Econometrics, Institute of Information Theory and Automation Prague, Czech Republic
Koháni Michal, Department of Transportation Networks, Faculty of Management and Informatics, University of Žilina, Slovakia
Komáromy Peter, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia

König Brian, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Kořenář Václav, Department of Econometrics, University of Economics, Prague, Czech Republic
Krumplík Miroslav, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Ladislav Kovaľ, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Lošonský Peter, University of Žilina, Faculty of Management Science and Informatics
Lukáčik Martin, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Lukáčiková Adriana, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Mendes Diana, ISCTE, Lisboa, Portugal
Menezes Rui, ISCTE, Lisboa, Portugal
Milánová Zdenka, Institute of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia
Miťková Veronika, Institute of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia
Palúch Stanislav, Faculty of Management Science and Informatics, University of Žilina, Slovakia
Pekár Juraj, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Pelikán Jan, Department of Econometrics, University of Economics, Prague, Czech Republic
Peško Štefan, Dept. of Mathematical Methods, Faculty of Management Science and Informatics, University of Žilina, Slovakia
Pinda Ľudovít, Department of Mathematics, University of Economics Bratislava, Slovakia
Ramík Jaroslav, Faculty of Economics, VŠB - TU Ostrava, Czech Republic
Reiff Marian, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Rublíková Eva, Department of Statistics, University of Economics Bratislava, Slovakia
Sladký Karel, Department of Econometrics, Institute of Information Theory and Automation, Prague, Czech Republic
Sojka Jozef, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Stolze Axel, Martin - Luther -Universität, Halle-Wittenberg, Germany
Surmanová Kvetoslava, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Szendreyová Alžbeta, Žilinská univerzita Žilina, Fakulta riadenia a informatiky, Katedra dopravných sietí
Szomolányi Karol, Department of Operations Research and Econometrics, University of Economics Bratislava, Slovakia
Tuš Branislav, Allianz-Slovenská dôchodková správcovská spoločnosť, Slovakia
Vacha Lukas, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague
Vladimír Mlynarovič, Institute of Economics, Faculty of Social and Economic Sciences, Comenius University in Bratislava, Slovakia
Vošvrda Miloslav, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic