Mathematical Methods in Economics 2005

23rd International Conference

14 - 16 September 2005



Czech Society for Operations Research

Czech Econometric Society

Faculty of Informatics and Management
University of Hradec Králové

CZECH SOCIETY FOR OPERATIONS RESEARCH
CZECH ECONOMETRIC SOCIETY
FACULTY OF INFORMATICS AND MANAGEMENT UNIVERSITY OF HRADEC KRÁLOVÉ

Proceedings of the 23rd International Conference

Mathematical Methods in Economics

2005

14th – 16th September 2005

Hradec Králové
Czech Republic

Editor: Hana Skalská
Publisher: Gaudeamus, University of Hradec Kralove
Printing: Libor Dvořák, tiskárna, E. Beneše 1562, Hradec Králové

Issue: 130 copies

ISBN 80-7041-535-5

Programme Committee

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

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

Doc. RNDr. Josef Hynek, Ph.D., MBA
Faculty of Informatics and Management, University of Hradec Králové.

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

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

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

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

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

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

Organisation Committee

Helena Holubičková
Faculty of Informatics, University of Hradec Kralove, Hradec Kralove

Michaela Nováková
Faculty of Informatics, University of Hradec Kralove, Hradec Kralove

RNDr. Petra Poulová
Faculty of Informatics, University of Hradec Kralove, Hradec Kralove

Doc. RNDr. Hana Skalská, CSc.
Faculty of Informatics, University of Hradec Kralove, Hradec Kralove

Ing. Hana Šrámková
Faculty of Informatics, University of Hradec Kralove, Hradec Kralove

Table of Contents

An Aggregated Semi-Markov Decision Model for Traffic Control with Lane Restriction

Ondrej Bartl
Department of Software Technologies, Faculty of Management Science and Informatics,
University of Žilina, Univerzitná 1, 010 26 Žilina, Slovakia,
bartl@kst.fri.utc.sk

Abstract

Feedback control can be employed for solving traffic situations on a two-lane road with one lane closed in some sector. To obtain cost-effective control rules, a semi-Markov decision model with state or state-and-action aggregation is given.

Keywords

semi-Markov decision processes, fixed-weight aggregation procedure, feedback traffic control rules, lane restriction

1 Traffic control problem with lane restriction

Non-standard situations on a road, where only two lanes exist and one of them is closed in some sector, require determination of cost-effective traffic control rules. A feedback control mechanism can be used to control traffic on the road in the closure area. If one lane is closed, e.g. because of a maintenance reason, then all the traffic in both directions in the closure sector must be led along the other, not closed lane. The directions for transit of vehicles from the one end of the closure sector to the other alternate with one another. The direction currently used for transit is an active direction and the opposite one is a passive direction. The feedback control of traffic is represented by decision making on how many vehicles to release for transit in the active direction with respect to the current state of queues of vehicles waiting at both ends of the closure sector. When transit of a group of vehicles in the active direction is finished, a decision epoch occurs. The directions interchange their modes of operation. The direction that has been passive becomes active and vice versa. A decision on the number of vehicles to be despatched to move through the closure sector in the new current active direction is made. As vehicles are supposed to arrive at both ends of the closure sector according to independent Poisson processes of arrivals, decision making is made in an environment of risk. Decisions on actions should minimise (the cost of) the consumption of time wasted by the waiting of vehicles in the queues. A semi-Markov decision process can model controlled evolution of this stochastic discrete event dynamic system. If the duration of closure is relatively long, then an infinite planning horizon may be considered for controlling the system. Under the assumption of stationarity of conditions in the control process, a stationary pure Markov policy is justified to prescribe proper control actions. Due to computational tractability, the lengths of queues are limited in the decision model. The greater the queue limits the higher the fidelity of the model. The reason to employ an aggregation procedure in such a model with a large-scale semi-Markov decision process is obvious.

2 A glance at the non-aggregated semi-Markov decision model

The corresponding non-aggregated semi-Markov decision model is described in [1]. The time set of the closure system is the set $T = [0, \infty)$. Decision epochs are denoted by random variables T_n and review intervals or stages by random variables $\tau_n = T_{n+1} - T_n$, $n \in E$, where the set $E = \{0,1,...\}$ is the set of ordinal numbers of decision epochs as well as stages. The state of the system reflects the situation in the queues of both directions together with the indication of the current active direction. The 0–1

notation is used to denote particular directions, where the one direction is a 0-direction and the other is a 1-direction. A random variable $\boldsymbol{X}_n \equiv \boldsymbol{X}(T_n) = (X_n^0, X_n^1, X_n^2)$ represents the system state at the *n*th decision epoch, $n \in E$. The component X_n^l, $l = 0,1$, reveals the amount of vehicles in the queue of the *l*-direction at epoch T_n and the component X_n^2 indicates the number of the active direction in the *n*th review interval $[T_n, T_{n+1})$. The lengths of the queues are limited by the numbers $L^l < \infty, l = 0,1,$ in the model. New vehicles arriving in the direction with a saturated queue are not registered in the related component of the state variable being regarded as rejected. The state space *S* is represented by the set $S = \{\boldsymbol{i} = (i^0, i^1, i^2) : i^0 \in S^0, i^1 \in S^1, i^2 \in S^2\}$, where $S^l = \{0,1,...,L^l\}, l = 0,1,$ and $S^2 = \{0,1\}$. After having reviewed the state of the system at a decision epoch, an action must be selected by a decision. Action $A_n \equiv A(T_n)$ is the transit of A_n vehicles through the closure sector in the active direction. If the queue in the active direction is empty then no vehicle is obviously dispatched and after a short switching time interval a decision on the transit in the opposite direction is made. The convention that at least one vehicle must be dispatched, if the queue in the active direction is not empty, is employed for decision making to avoid too long stays of vehicles in the queues. The set $U(\boldsymbol{i})$ of feasible actions in state $\boldsymbol{i} = (i^0, i^1, i) \in S$ is given by

$$U(\boldsymbol{i}) = U(i^0, i^1, i) \equiv U_{\boldsymbol{i}} = U_{i^0, i^1, i} = \begin{cases} \{0\}, & \text{if } (i^0 = 0, i^1 \in S^1, i = 0) \text{ or } (i^0 \in S^0, i^1 = 0, i = 1), \\ \{1, 2, \ldots, i^0\}, & \text{if } i^0 \in S^0 \setminus \{0\}, i^1 \in S^1, i = 0, \\ \{1, 2, \ldots, i^1\}, & \text{if } i^0 \in S^0, i^1 \in S^1 \setminus \{0\}, i = 1. \end{cases} \quad (1)$$

The action space *U* is represented by the set $U = \{0,1,\ldots,L\}$, where $L = \max\{L^0, L^1\}$. The following notation is quite useful in the further text: $U^l(i^l) = U(i^0, i^1, l), i^l \in S^l, i^{1-l} \in S^{1-l}, l = 0,1,$ as well as $U^l = \bigcup_{i^l \in S^l} U^l(i^l) = \{0,1,\ldots,L^l\}, l = 0,1.$ Then $U = U^0 \cup U^1$. In state and action variables, a convention is employed that state variables X_n^0, X_n^1 represent the numbers of registered vehicles in the queues at decision epoch T_n just before releasing A_n vehicles to move through the closure sector in the active direction and state variable X_n^2 is the indication of the active direction during review interval $[T_n, T_{n+1})$ from the decision epoch T_n onwards up to the next decision epoch. Note that action A_n is an action chosen by the decision at epoch T_n and remains valid within the *n*th stage, $n \in E$, i.e. $A(t) = A_n \equiv A(T_n)$ for all $t \in [T_n, T_{n+1})$. Denoting by $\tau(\boldsymbol{i}, a)$ the stage length function, the duration of the *n*th stage $\tau_n = T_{n+1} - T_n$ is given by $\tau_n \equiv \tau(\boldsymbol{X}_n, A_n)$, $n \in E$. The expected stage length function $\overline{\tau}(\boldsymbol{i}, a) = E[\tau(\boldsymbol{X}_n, A_n) | \boldsymbol{X}_n = \boldsymbol{i}, A_n = a]$ is $\overline{\tau}(\boldsymbol{i}, a) = \overline{\tau}(i^0, i^1, i, a) = T(i, a)$, $\boldsymbol{i} \in S$, $a \in U(\boldsymbol{i})$, where the deterministic function $T(i, a)$ represents either a transit time of a group of $a > 0$ vehicles through the closure sector in the *i*-direction or a switching time in the *i*th direction if $a = 0$. If $C(\boldsymbol{i}, a)$ denotes the cost function then the cost accompanying the delay of vehicles in the queues at both ends of the closure sector during the *n*th stage is random variable $C_n \equiv C(\boldsymbol{X}_n, A_n), n \in E$. The expected cost function $\overline{C}(\boldsymbol{i}, a) = E[C(\boldsymbol{X}_n, A_n) | \boldsymbol{X}_n = \boldsymbol{i}, A_n = a]$ is given by $\overline{C}(\boldsymbol{i}, a) = \overline{C}(i^0, i^1, i, a) = \overline{W}(\boldsymbol{i}, a) + \overline{R}(\boldsymbol{i}, a)$, $\boldsymbol{i} \in S$, $a \in U(\boldsymbol{i})$. It reflects the expected cost $\overline{W}(\boldsymbol{i}, a)$ associated with the waiting of registered vehicles in the queues within a stage and the expected cost $\overline{R}(\boldsymbol{i}, a)$ connected with the rejections of new vehicles during a review interval (see [1] for details). The transition probabilities $p_{i,j}(a) \equiv p(\boldsymbol{i}, \boldsymbol{j}, a)$ such that $p(\boldsymbol{i}, \boldsymbol{j}, a) = p(i^0, i^1, i, j^0, j^1, j, a) = P\{\boldsymbol{X}_{n+1} = \boldsymbol{j} \mid \boldsymbol{X}_n = \boldsymbol{i}, A_n = a\}$, $\boldsymbol{i}, \boldsymbol{j} \in S$, $a \in U(\boldsymbol{i})$ are stated by the

product $p(\boldsymbol{i},\boldsymbol{j},a)=p^0\left(i^0 \xrightarrow{i^1,i,a} j^0\right)p^1\left(i^1 \xrightarrow{i^0,i,a} j^1\right)p\left(i \xrightarrow{i^0,i^1,a} j\right)$, where $p^l\left(i^l \xrightarrow{i^{1-l},i,a} j^l\right)$ is the transition probability for the change in the amount of registered vehicles in the queue of the *l*-direction, $l=0,1$, and $p\left(i \xrightarrow{i^0,i^1,a} j\right)$ is the transition probability for the change of the number of the active direction. Recall that amounts $Z^l(t)$ of vehicles arriving to the closure sector in the *l*-direction during a *t*-time-unit interval are Poisson distributed with mean $\lambda_l t$. The probability of *k* arrivals of vehicles in the *l*-direction within a time period of *t* time units is then given by the following formula $p_k^l(t)\equiv P\left\{Z^l(t)=k\right\}=e^{-\lambda_l t}\left(\lambda_l t\right)^k/k!,\quad k=0,1,\ldots,\quad t\geq 0,\quad l=0,1$. Using the equality function $e(l,i)=1-l+(2l-1)i=1-\operatorname{abs}(l-i), l,i\in\{0,1\}$, returning 1 if *l* equals *i*, and 0 otherwise, we can state that, for $\boldsymbol{i},\boldsymbol{j}\in S$, $a\in U(\boldsymbol{i})$,

$$p^l\left(i^l \xrightarrow{i^{1-l},i,a} j^l\right)\equiv P\left\{X_{n+1}^l=j^l \,\middle|\, X_n^l=i^l, X_n^{1-l}=i^{1-l}, X_n^2=i, A_n=a\right\}$$
$$=\begin{cases}\sum_{k=L^l-i^l+e(l,i)a}^{\infty} p_k^l\left(T(i,a)\right), & j^l=L^l,\\ p_{j^l-i^l+e(l,i)a}^l\left(T(i,a)\right), & i^l-e(l,i)a\leq j^l<L^l,\\ 0, & 0\leq j^l<i^l-e(l,i)a,\end{cases} \tag{2}$$
$$0\leq i^l\leq L^l, l=0,1, i\in S^2=\{0,1,\}, a\in U(\boldsymbol{i})\subset U, n\in E,$$

$$p\left(i \xrightarrow{i^0,i^1,a} j\right)\equiv P\left\{X_{n+1}^2=j \,\middle|\, X_n^0=i^0, X_n^1=i^1, X_n^2=i, A_n=a\right\}=\begin{cases}1, & j=1-i,\\ 0, & j=i,\end{cases} \tag{3}$$
$$i\in S^2=\{0,1\}, i^0\in S^0, i^1\in S^1, a\in U(\boldsymbol{i})\subset U, n\in E.$$

Cost-effective feedback control of traffic in the closure sector is prescribed by an optimal stationary pure Markov policy $\left(d^*\right)^\infty$ minimising the long-run expected average cost criterion $a^{d^\infty}(\boldsymbol{i})$ for each initial state $\boldsymbol{i}\in S$. The decision function $d^*: S\to U$ reveals that $d^*(\boldsymbol{i})=d^*\left(i^0,i^1,i\right)$ vehicles must be released to move through the closure sector in the active direction when the state of the system at a current decision epoch is $\boldsymbol{i}=\left(i^0,i^1,i\right)\in S$.

3 The corresponding aggregated semi-Markov decision model

To build a computationally tractable decision model that is congruent with reality on the road, a fixed-weight aggregation procedure [5] can be employed. It requires to aggregate properly particular values in the sets of feasible values in the state space (and sometimes also in the action space) of the non-aggregated decision model. Then the transition, cost and time structure of the aggregated model must be determined from the corresponding structure of the non-aggregated model. Consequently an optimal control policy is computed in the aggregated model and finally a suboptimal control policy for the non-aggregated model is constructed by a disaggregation procedure from the optimal policy obtained for the aggregated model.

3.1 Partitioning of sets in state and action spaces to aggregate their elements

3.1.1 State aggregation

The fixed-weight aggregation procedure for the aforementioned decision model starts with aggregating values of the first two components of the vector state variable $\boldsymbol{X}(t)$. The sets $S^l, l=0,1,$ are partitioned into their subsets $S_0^l, S_1^l, \ldots, S_{M^l-1}^l, S_{M^l}^l$ by partitioning points $n_0^l<n_1^l<\ldots<n_{M^l-1}^l$, where

$n_0^l = 0$ and $n_{M^l-1}^l < L^l$. The subset $S_{k^l}^l = \{n_{k^l-1}^l + 1, n_{k^l-1}^l + 2, \ldots, n_{k^l}^l\}$ is associated with the value k^l of the *l*th component, $l = 0,1$, of the aggregate state variable $\tilde{\boldsymbol{X}}(t) = (\tilde{X}^0(t), \tilde{X}^1(t), \tilde{X}^2(t))$. We define that $n_{-1}^l = -1$ and $n_{M^l}^l = L^l$, $l = 0,1$, to make clear the specification of subsets $S_{k^l}^l$. As the values of the last component of the state variable $\boldsymbol{X}(t)$ are not aggregated, we get that the aggregate state space is $\tilde{S} = \{\boldsymbol{k} = (k^0, k^1, k^2) : k^0 \in \tilde{S}^0, k^1 \in \tilde{S}^1, k^2 \in \tilde{S}^2\}$, where $\tilde{S}^l = \{0,1,\ldots,M^l\}, l = 0,1$, and $\tilde{S}^2 = S^2 = \{0,1\}$. A function $f^l : \tilde{S}^l \to S^l$ such that $f^l(k^l) = \max\{i^l : i^l \in S_{k^l}^l\} \equiv n_{k^l}^l$, $k^l \in \tilde{S}^l, l = 0,1$, refers to the points $n_{k^l}^l$. We can partition the sets S^l quasi-uniformly with the same number *w* of elements in each subset $S_{k^l}^l$, $k^l \in \tilde{S}^l$, except for the subsets S_0^l and maybe $S_{M^l}^l$, $l = 0,1$. Then $M^l = \lceil L^l / w \rceil$, i.e. M^l is the upper integer part of quotient L^l/w, which means that $M^l = \min\{n \in I : n \geq L^l/w\}$, where *I* is the set of integers. The partitioning points are $n_{k^l}^l = wk^l$ for $k^l \in \{0,1,...,M^l - 1\}$, $l = 0,1$. As $n_{M^l}^l = L^l$, we have $n_{M^l}^l = \min\{L^l, wM^l\}$, because $L^l \leq wM^l$, $l = 0,1$.

The values of the action variable $A(t)$ can or need not be aggregated. If they are not aggregated, then the action taken at the *n*th decision epoch is $A_n \equiv A(T_n)$, $n \in E$, and the set of feasible actions in state $\boldsymbol{k} = (k^0, k^1, k) \in \tilde{S}$ is $\tilde{U}(\boldsymbol{k}) = \tilde{U}(k^0, k^1, k) \equiv \tilde{U}_{\boldsymbol{k}} = \tilde{U}_{k^0,k^1,k}$. For all $\boldsymbol{k} = (k^0, k^1, k) \in \tilde{S}$, we have, according to the indication *k* of the current active direction, that $\tilde{U}(\boldsymbol{k}) = \{0\}$ when $k^0 = 0, k^1 \geq 0, k = 0$ or $\tilde{U}(\boldsymbol{k}) = \{1, 2, \ldots, f^0(k^0)\}$ when $k^0 > 0, k^1 \geq 0, k = 0$, and $\tilde{U}(\boldsymbol{k}) = \{0\}$ when $k^0 \geq 0, k^1 = 0, k = 1$ or $\tilde{U}(\boldsymbol{k}) = \{1, 2, \ldots, f^1(k^1)\}$ when $k^0 \geq 0, k^1 > 0, k = 1$. The corresponding action space is then the set $\tilde{U} = \{0,1,\ldots,L\} = U$, where $L = \max\{L^0, L^1\}$. Let $\tilde{U}^l$ denote the set of all admissible actions in the state-aggregated decision model when the *l*-direction is active, $l = 0,1$, and let $\tilde{U}^l(k^l) = \tilde{U}(k^0, k^1, l)$, $k^l \in \tilde{S}^l, k^{1-l} \in \tilde{S}^{1-l}, l = 0,1$, where concurrently $\tilde{U}^l(k^l) = \bigcup_{i^l \in S_{k^l}^l} U^l(i^l)$. Then $\tilde{U}^l = \bigcup_{k^l \in \tilde{S}^l} \tilde{U}^l(k^l)$ and we have that $\tilde{U}^l = \{0,1,\ldots,L^l\} = U^l$, $l = 0,1$, where U^l is the set of all possible non-aggregate actions that are available for process control in the original model when the *l*-direction is active, $\tilde{U}^0 \cup \tilde{U}^1 = \tilde{U}$.

3.1.2 Action aggregation

In case the values of the action variable $A(t)$ are aggregated, the action taken at the *n*th decision epoch is $\tilde{A}_n \equiv \tilde{A}(T_n)$, $n \in E$, and the set of feasible actions in state $\boldsymbol{k} = (k^0, k^1, k) \in \tilde{S}$ is the set $\tilde{\tilde{U}}(\boldsymbol{k}) = \tilde{\tilde{U}}(k^0, k^1, k) \equiv \tilde{\tilde{U}}_{\boldsymbol{k}} = \tilde{\tilde{U}}_{k^0,k^1,k}$. The action aggregation can be carried out by partitioning the action space *U* into its subsets $U_0, U_1, \ldots, U_B$ using the partitioning points $v_0 < v_1 < \ldots < v_{B-1}$, where $v_0 = 0$ and $v_{B-1} < L$. The subset $U_b = \{v_{b-1}+1, v_{b-1}+2, ..., v_b\}$ is then associated with the value *b* of the aggregate action variable $\tilde{A}(t)$. We define that $v_{-1} = -1$ and $v_B = L$ to complete the specification of subsets U_b for $b = 0$ and $b = B$. The aggregate action space is $\tilde{\tilde{U}} = \{0,1,\ldots,B\}$. To determine a special (non-aggregate) action as a representative of each subset U_b, we use the selection function $s : \tilde{\tilde{U}} \to U$ defined by $s(b) = \max\{a : a \in U_b\} \equiv v_b$, $b \in \tilde{\tilde{U}}$. The aggregate action $\tilde{A}_n = b$ then implies the corresponding non-aggregate action $A_n = s(b)$ to be used in the decision model. The specification of the sets $\tilde{\tilde{U}}(\boldsymbol{k}) \subset \tilde{\tilde{U}}$ of feasible aggregate actions in aggregate states $\boldsymbol{k} \in \tilde{S}$ is stated as follows. Us-

ing the 'rank' function $r:U\to\tilde{\tilde{U}}$ such that $r(a)=b\Leftrightarrow a\in U_b \Leftrightarrow v_{b-1}+1\le a\le v_b$, we have obtained, for all $\boldsymbol{k}=(k^0,k^1,k)\in\tilde{S}$, that $\tilde{\tilde{U}}(\boldsymbol{k})=\{0\}$ when $k^0=0,k^1\ge 0,k=0$ or $\tilde{\tilde{U}}(\boldsymbol{k})=\left\{1,2,...,r\left[f^0\left(k^0\right)\right]\right\}$ when $k^0>0,k^1\ge 0,k=0$, and $\tilde{\tilde{U}}(\boldsymbol{k})=\{0\}$ when $k^0\ge 0,k^1=0,k=1$ or $\tilde{\tilde{U}}(\boldsymbol{k})=\left\{1,2,...,r\left[f^1\left(k^1\right)\right]\right\}$ when $k^0\ge 0,k^1>0,k=1$. Note that function *r* returns an aggregate action matching a given non-aggregate action. The number $r(a)$ can be interpreted as a rank for a non-aggregate action *a* within the set $\tilde{\tilde{U}}=\{0,1,\ldots,B\}$ of possible rank categories (which here are aggregate actions). Action aggregation can also be performed quasi-uniformly like state aggregation with the same number *w* of elements in each subset U_b, $b\in\tilde{\tilde{U}}=\{0,1,\ldots,B\}$, except for the subsets U_0 and perhaps U_B. Then $B=\lceil L/w\rceil$ and the partitioning points are $v_b=wb$ for $b\in\{0,1,\ldots,B-1\}$. Note, that $v_B=L=\min\{L,wB\}$, because $L\le wB$.

Let $\tilde{\tilde{U}}^l\left(k^l\right)=\tilde{\tilde{U}}\left(k^0,k^1,l\right)$, $k^l\in\tilde{S}^l,k^{1-l}\in\tilde{S}^{1-l},l=0,1$, be the set of feasible aggregate actions in the aggregate state component k^l when the *l*-direction is active. The set $\tilde{\tilde{U}}^l\left(k^l\right)$ is concurrently given by $\tilde{\tilde{U}}^l\left(k^l\right)=\left\{b\in\tilde{\tilde{U}}\middle|\left(\exists a\in\tilde{U}^l\left(k^l\right):a\in U_b\right)\right\}$, $k^l\in\tilde{S}^l$, $l=0,1$, which means that the set $\tilde{\tilde{U}}^l\left(k^l\right)$ is a set of all subscripts of subsets $U_b\subset U$ that contain at least one element from the set $\tilde{U}^l\left(k^l\right)$. Then the set of all feasible aggregate actions to control the closure system, when the *l*-direction is active, is the set $\tilde{\tilde{U}}^l=\bigcup_{k^l\in\tilde{S}^l}\tilde{\tilde{U}}^l\left(k^l\right)=\left\{0,1,\ldots,r\left[f^l\left(M^l\right)\right]\right\}=\left\{0,1,\ldots,r\left[L^l\right]\right\}$, $l=0,1$. Hence, $\tilde{\tilde{U}}=\tilde{\tilde{U}}^0\cup\tilde{\tilde{U}}^1$.

Now we need to specify the sets $\tilde{U}_s^l\left(k^l\right)=\left\{s(b)\middle|b\in\tilde{\tilde{U}}^l\left(k^l\right)\right\}$ of feasible special non-aggregate actions, which are given by the selection function *s*, in all possible aggregate states $k^l\in\tilde{S}^l$ when the *l*-direction is active, $l=0,1$. If we define sets $\tilde{U}_s(\boldsymbol{k})=\tilde{U}_s\left(k^0,k^1,k\right)$, $\boldsymbol{k}=\left(k^0,k^1,k\right)\in\tilde{S}$, by the equality $\tilde{U}_s\left(k^0,k^1,k\right)=\left\{s(b)\middle|b\in\tilde{\tilde{U}}\left(k^0,k^1,k\right)\right\}$, then $\tilde{U}_s^l\left(k^l\right)=\tilde{U}_s\left(k^0,k^1,l\right)$, $k^l\in\tilde{S}^l,k^{1-l}\in\tilde{S}^{1-l},l=0,1$. The sets $\tilde{U}_s^l$ of all admissible special non-aggregate actions selected by the selection function *s*, when the *l*-direction is active, are then given by $\tilde{U}_s^l=\bigcup_{k^l\in\tilde{S}^l}\tilde{U}_s^l\left(k^l\right)=\left\{s(0),s(1),\ldots,s\left(r\left[f^l\left(M^l\right)\right]\right)\right\}$, that is $\tilde{U}_s^l=\left\{v_0,v_1,\ldots,v_{r\left(L^l\right)}\right\}$, $l=0,1$. Note that $s\left(r\left[f^l\left(M^l\right)\right]\right)=s\left(r\left[n_{M^l}^l\right]\right)=s\left(r\left[L^l\right]\right)=v_{r\left[L^l\right]}$. The set $\tilde{U}_s$ of all special, non-aggregate actions selected by selection function *s* for both directions is given by $\tilde{U}_s=\tilde{U}_s^0\cup\tilde{U}_s^1=\{v_0,v_1,\ldots,v_{B-1},v_B\}=\{0,v_1,\ldots,v_{B-1},L\}$ as $v_0=0$, $v_B=L$, and $L=\max\{L^0,L^1\}$. Apparently, $\tilde{U}_s\subset U$.

3.2 Transition, cost and time structure in aggregated decision models

After having partitioned the sets of possible values for the state variable and perhaps even for the action variable, we must state the transition, cost and time structure of the aggregated semi-Markov decision model of the traffic control process. That requires to determine the transition probabilities, the expected cost function and the expected stage length function for the aggregate state or state and action variables. We can do so using the concept of fixed weights, when the transition, cost and time structure of the original, non-aggregated decision model serves as the basis for the determination of the corresponding structure in the aggregated decision model. As it follows from probabilistic background [2] of the fixed-weight aggregation procedure, fixed weights λ_i^k correspond to conditional probabilities of occupying particular non-aggregate states *i* given that being in the aggregate state *k*.

3.2.1 Feasibility of actions in aggregated decision models

Before starting to specify prescriptions for transition probabilities and the expected cost and stage length functions in the aggregated model, we must take into account a fact that some action *u* declared as feasible in aggregate state component $k^l \in \tilde{S}^l$ can be infeasible in some non-aggregate state component $i^l \in S^l_{k^l}$, $l \in \{0,1\}$. Namely, an action $u = a \in \tilde{U}^l\left(k^l\right) \subset \tilde{U}^l \subset \tilde{U} = U$ need not be feasible in non-aggregate state component $i^l \in S^l_{k^l}$ in the state-aggregated decision model. Similarly it can happen that an action $u = s(b) \in \tilde{U}^l_s\left(k^l\right) \subset \tilde{U}^l_s \subset \tilde{U}_s \subset U$ with $b \in \tilde{\tilde{U}}^l\left(k^l\right) \subset \tilde{\tilde{U}}^l \subset \tilde{\tilde{U}}$ is infeasible in non-aggregate state component $i^l \in S^l_{k^l}$ in the state-and-action-aggregated decision model. When this occurs, an action *u* infeasible for state component i^l must be replaced by the nearest feasible action. Therefore a concept of the acceptance function [2] is employed to provide a proper feasible (i.e. acceptable) substitute. We use a simplifying notation for the sets of feasible actions in state components i^l given as follows: $U\left(i^l\right) = \{0\}$ if $i^l = 0$ and $U\left(i^l\right) = \{1,2,\ldots,i^l\}$ if $i^l \in \{1,2,\ldots,L\}$, where $l \in \{0,1\}$ indicates a current active direction. Then $U^l\left(i^l\right) = U\left(i^0, i^1, l\right) = U\left(i^l\right)$, $i^l \in S^l$, $l = 0,1$. Recall that the action space in the original, non-aggregated model is $U = \{0,1,\ldots,L\}$. The acceptance function $a_{i^l} : U \to U\left(i^l\right)$ providing a feasible counterpart $a_{i^l}(u) \in U\left(i^l\right)$ for action *u* is defined by the prescription $a_{i^l}(u) = \max \arg\min_{c \in U\left(i^l\right)} \{\mathrm{abs}(u - c)\}$, $u \in U, i^l \in S^l, l \in \{0,1\}$, where arg min denotes a subset of elements at which the minimum of a function is obtained. As the acceptance function returns the nearest feasible action in the set $U\left(i^l\right)$ to an action *u*, it is clear that $a_{i^l}(u) = u$ whenever $u \in U\left(i^l\right)$.

3.2.2 Model structure with state aggregation and with state-and-action aggregation

Now we are ready to state elements of the transition, cost and time structure of the decision model for feedback traffic control with lane restriction in the case of state aggregation as well as in the case of state-and-action aggregation. First we make a terminological convention. We will refer to aggregate state by the term meta-state and to aggregate action by the term meta-action. We use the abbreviation NADP for a non-aggregated decision process, SADP for a state-aggregated decision process and SAADP for a state-and-action-aggregated decision process. To attain a well-arranged look of equations, we use a modified notation for some symbols. The acceptance function for action $u \in U$ with respect to state component $i^l \in S^l$, where $l \in \{0,1\}$ is the indication of a current active direction, will be written in accordance with $a\left(u \middle| i^l\right) \equiv a_{i^l}(u)$. The subset of state component values for meta-state component $k^l \in \tilde{S}^l$, where superscript *l* denotes a direction, $l = 0,1$, will be denoted by the convention $S^l\left(k^l\right) \equiv S^l_{k^l}$. The fixed weight of state component $i^l \in S^l$ in relation to subset $S^l\left(k^l\right)$ of state components associated with meta-state component $k^l \in \tilde{S}^l$ for the *l*-direction, $l = 0,1$, will be represented by the symbol $\lambda^l\left(i^l \middle| k^l\right) \equiv \lambda^{l,k^l}_{i^l}$. This last notational convention invokes aforementioned association of the fixed weight with the conditional probability of occupying the *l*th state component i^l given that being in the *l*th meta-state component k^l, i.e., $\lambda^l\left(i^l \middle| k^l\right) \approx q^{l,k^l}_{\psi,n}\left(i^l\right) \equiv P\left\{X^l_n = i^l \middle| X^l_n \in S^l\left(k^l\right), \pi = \psi\right\}$, $i^l \in S^l\left(k^l\right)$, $k^l \in \tilde{S}^l$, $l = 0,1$, $n \in E$, where control policy π is represented by a policy $\psi \in \Pi^{\mathrm{SD}}$ with Π^{SD} denoting the set of all feasible stationary pure Markov policies. Policy ψ as a stationary deterministic (i.e. pure) Markov policy employed in the original NADP is a policy of the form $(d,d,\ldots) \equiv d^\infty$, where $d : S \to U$ is a decision function.

Probabilistic laws of motion for the modelled dynamic system evolving under risk are given by transition probabilities. Mutual independence of Poisson arrival processes of vehicles at each end of

the closure sector implies a separable product form in transition probabilities represented in the original NADP by a product of transition probabilities related to each state component. That still remains valid for transition probabilities in aggregated decision models as far as most actions are concerned. But in general, full separability is not possible. To describe stochastic dynamics in aggregated decision models, we are to specify transition probabilities $\tilde{p}_{\boldsymbol{k},\boldsymbol{m}}(a) \equiv \tilde{p}(\boldsymbol{k},\boldsymbol{m},a) = P\left\{\tilde{\boldsymbol{X}}_{n+1} = \boldsymbol{m} \middle| \tilde{\boldsymbol{X}}_n = \boldsymbol{k}, A_n = a\right\}$, $\boldsymbol{k},\boldsymbol{m} \in \tilde{S}, a \in \tilde{U}(\boldsymbol{k}) \subset U$, for a state-aggregated decision process, as well as transition probabilities $\tilde{\tilde{p}}_{\boldsymbol{k},\boldsymbol{m}}(b) \equiv \tilde{\tilde{p}}(\boldsymbol{k},\boldsymbol{m},b) = P\left\{\tilde{\boldsymbol{X}}_{n+1} = \boldsymbol{m} \middle| \tilde{\boldsymbol{X}}_n = \boldsymbol{k}, \tilde{A}_n = b\right\}$, $\boldsymbol{k},\boldsymbol{m} \in \tilde{S}, b \in \tilde{\tilde{U}}(\boldsymbol{k}) \subset \tilde{\tilde{U}}$, for a state-and-action-aggregated decision process. As $P\left\{\tilde{\boldsymbol{X}}_{n+1} = \boldsymbol{m} \middle| \tilde{\boldsymbol{X}}_n = \boldsymbol{k}, \tilde{A}_n = b\right\} = P\left\{\tilde{\boldsymbol{X}}_{n+1} = \boldsymbol{m} \middle| \tilde{\boldsymbol{X}}_n = \boldsymbol{k}, A_n = s(b)\right\}$, i.e. $\tilde{\tilde{p}}_{\boldsymbol{k},\boldsymbol{m}}(b) \equiv \tilde{\tilde{p}}(\boldsymbol{k},\boldsymbol{m},b) = \tilde{p}(\boldsymbol{k},\boldsymbol{m},s(b)) \equiv \tilde{p}_{\boldsymbol{k},\boldsymbol{m}}(s(b))$, for all $\boldsymbol{k},\boldsymbol{m} \in \tilde{S}, b \in \tilde{\tilde{U}}(\boldsymbol{k}) \subset \tilde{\tilde{U}}$, we can use symbol *u* to denote an action in transition probabilities $\tilde{p}(\boldsymbol{k},\boldsymbol{m},a)$ and $\tilde{p}(\boldsymbol{k},\boldsymbol{m},s(b))$. In this manner we obtain a unified statement of transition probabilities in the form of $\tilde{p}(\boldsymbol{k},\boldsymbol{m},u)$ both for a SADP, where $u = a \in \tilde{U} = U$, and for a SAADP, where $u = s(b) \in \tilde{U}_s \subset U$ with $b \in \tilde{\tilde{U}}$. Using the same reasoning (based on applications of the law of total probability) as in the one-dimensional case [2], we derive a formula for any of the transition probabilities $\tilde{p}(\boldsymbol{k},\boldsymbol{m},u)$, $\boldsymbol{k},\boldsymbol{m} \in \tilde{S}, u \in U$. We obtain, for $n \in E$, that

$$
\begin{aligned}
&\tilde{p}_{\boldsymbol{k},\boldsymbol{m}}(u) \equiv \tilde{p}(\boldsymbol{k},\boldsymbol{m},u) = \tilde{p}\left(k^0,k^1,k,m^0,m^1,m,u\right) \\
&= P\left\{\tilde{X}^0_{n+1} = m^0, \tilde{X}^1_{n+1} = m^1, \tilde{X}^2_{n+1} = m \middle| \tilde{X}^0_n = k^0, \tilde{X}^1_n = k^1, \tilde{X}^2_n = k, A_n = u\right\} \\
&\approx p_{k,m}(u) \sum_{i^k \in S^k\left(k^k\right)} \lambda^k\left(i^k \middle| k^k\right) \left\{ \left(\sum_{j^k \in S^k\left(m^k\right)} p^k_{i^k,j^k}\left(a\left(u \middle| i^k\right)\right) \right) \times \right. \\
&\qquad \left. \left[\sum_{i^{1-k} \in S^{1-k}\left(k^{1-k}\right)} \lambda^{1-k}\left(i^{1-k} \middle| k^{1-k}\right) \left(\sum_{j^{1-k} \in S^{1-k}\left(m^{1-k}\right)} p^{1-k}_{i^{1-k},j^{1-k}}\left(a\left(u \middle| i^k\right)\right) \right) \right] \right\}, \\
&\boldsymbol{k} = \left(k^0,k^1,k\right) \in \tilde{S}, \boldsymbol{m} = \left(m^0,m^1,m\right) \in \tilde{S}, u \in U,
\end{aligned}
\tag{4}
$$

where fixed weights $\lambda^l\left(i^l \middle| k^l\right)$ have to be chosen in accordance with the following constraints

$$
\begin{aligned}
&\lambda^l\left(i^l \middle| k^l\right) \geq 0, \quad i^l \in S^l\left(k^l\right), k^l \in \tilde{S}^l, l \in S^2 = \{0,1\}, \\
&\sum_{i^l \in S^l\left(k^l\right)} \lambda^l\left(i^l \middle| k^l\right) = 1, \quad k^l \in \tilde{S}^l, l \in S^2 = \{0,1\}.
\end{aligned}
\tag{5}
$$

To save space, the following notation has been used in (4): $p^l_{i^l,j^l}\left(a\left(u \middle| i^k\right)\right) \equiv p^l\left(i^l \xrightarrow{i^{1-l},k,a\left(u \middle| i^k\right)} j^l\right)$, $l \in \{k, 1-k\}$ and $p_{k,m}(u) \equiv p\left(k \xrightarrow{i^0,i^1,u} m\right)$, $k \in \{0,1\}$. General formula (4) together with constraints (5) serves for the determination of the transition structure both in the SADP and in the SAADP with any action $u \in U$. In case action *u* is some action chosen from the intersection of the sets $U^k\left(i^k\right)$ of feasible actions in all states components i^k forming the subset $S^k\left(k^k\right)$ of the partition of the set S^k of possible values for the *k*th state component, $k \in \{0,1\}$, then the acceptance function returns a value of its argument as a function value. That is, $a\left(u \middle| i^k\right) = u$ whenever *u* is feasible for all $i^k \in S^k\left(k^k\right)$. Then we get a completely separable product form for the transition probability from a meta-state $\boldsymbol{k}$ at a current decision epoch to a meta-state $\boldsymbol{m}$ at the next decision epoch

$$
\begin{aligned}
&\tilde{p}_{\boldsymbol{k},\boldsymbol{m}}(u) \equiv \tilde{p}(\boldsymbol{k},\boldsymbol{m},u) = \tilde{p}\left(k^0,k^1,k,m^0,m^1,m,u\right) \\
&= \tilde{p}^0_{k^0,m^0}(u)\,\tilde{p}^1_{k^1,m^1}(u)\,\tilde{p}_{k,m}(u), \quad \boldsymbol{k} = \left(k^0,k^1,k\right) \in \tilde{S},\ \boldsymbol{m} = \left(m^0,m^1,m\right) \in \tilde{S}, \\
&u \in \bar{U}^k\left(k^k\right) \equiv \bigcap_{i^k \in S^k\left(k^k\right)} U^k\left(i^k\right) \subset U^k \subset U, k^k \in \tilde{S}^k, k = 0,1,
\end{aligned}
\tag{6}
$$

where, together with constraints (5),

$$
\begin{aligned}
&\tilde{p}^l_{k^l,m^l}(u) = \sum_{i^l \in S^l\left(k^l\right)} q^{l,k^l}_{\psi,n}\left(i^l\right) p^l_{i^l,S^l\left(m^l\right)}(u) \approx \sum_{i^l \in S^l\left(k^l\right)} \lambda^l\left(i^l \middle| k^l\right) \sum_{j^l \in S^l\left(m^l\right)} p^l_{i^l,j^l}(u), \quad \tilde{p}_{k,m}(u) = p_{k,m}(u), \\
&u \in \bar{U}^k\left(k^k\right), k^l, m^l \in \tilde{S}^l, l \in \{0,1\}, k,m \in S^2 = \{0,1\}, n \in E.
\end{aligned}
\tag{7}
$$

The cost structure of the aggregated decision models is determined as soon as the expected cost function is specified. Then the expectation of the cost incurred in connection with the system operation in a review interval between two consecutive decision epochs is determined. Such statement of economic consequences of the system evolution over time is necessary to calculate the value of the criterion of interest with respect to a control policy considered. As the closure system acts in an environment of risk, the long-run expected average cost per unit time stands for the performance index of the system operation over an infinite planning horizon. To obtain criterion values for various feasible policies as well as to find out an optimal policy for taking actions in decision models with state or state-and-action aggregation, the expected cost function for aggregate state or state and action variables must be specified utilizing the cost structure of the original, non-aggregated decision model. As $\tilde{\bar{C}}(\boldsymbol{k},b) = E\left[\tilde{\bar{C}}\left(\tilde{\boldsymbol{X}}_n,\tilde{A}_n\right)\middle|\tilde{\boldsymbol{X}}_n = \boldsymbol{k},\tilde{A}_n = b\right] = E\left[\tilde{C}\left(\tilde{\boldsymbol{X}}_n,A_n\right)\middle|\tilde{\boldsymbol{X}}_n = \boldsymbol{k},A_n = s(b)\right] = \bar{\tilde{C}}\left(\boldsymbol{k},s(b)\right)$ holds for any $\boldsymbol{k} = \left(k^0,k^1,k\right) \in \tilde{S}$, $b \in \tilde{\tilde{U}}$, $s(b) \in \tilde{U}_s \subset U$, $n \in E$, it suffices to derive a statement specifying the expected cost function $\bar{\tilde{C}}(\boldsymbol{k},u)$ for any action $u \in U$ in order to obtain either the value of the expectation $\bar{\tilde{C}}(\boldsymbol{k},a)$ in a SADP, when $u = a \in \tilde{U} = U$, or the value of the expectation $\bar{\tilde{C}}\left(\boldsymbol{k},s(b)\right)$ in a SAADP, when $u = s(b) \in \tilde{U}_s \subset U$ with $b \in \tilde{\tilde{U}}$. The same reasoning as in the one-dimensional case [2], based here on successive applications of the law of total expectation by conditioning first on the values $i^0 \in S^0$ of the X^0_n-component and then on the values $i^1 \in S^1$ of the X^1_n-component of the state variable $\boldsymbol{X}_n = \left(X^0_n, X^1_n, X^2_n\right)$, provides the final formula

$$
\begin{aligned}
&\bar{\tilde{C}}(\boldsymbol{k},u) = \bar{\tilde{C}}\left(k^0,k^1,k,u\right) = E\left[\tilde{C}\left(\tilde{\boldsymbol{X}}_n,A_n\right)\middle|\tilde{\boldsymbol{X}}_n = \boldsymbol{k},A_n = u\right] \\
&= \sum_{i^0 \in S^0\left(k^0\right)} q^{0,k^0}_{\psi,n}\left(i^0\right)\left(\sum_{i^1 \in S^1\left(k^1\right)} q^{1,k^1}_{\psi,n}\left(i^1\right)\bar{C}\left(i^0,i^1,k,a\left(u\middle|i^k\right)\right)\right) \\
&\approx \sum_{i^0 \in S^0\left(k^0\right)} \lambda^0\left(i^0\middle|k^0\right)\left(\sum_{i^1 \in S^1\left(k^1\right)} \lambda^1\left(i^1\middle|k^1\right)\bar{C}\left(i^0,i^1,k,a\left(u\middle|i^k\right)\right)\right), \\
&\boldsymbol{k} = \left(k^0,k^1,k\right) \in \tilde{S}, u \in U, n \in E,
\end{aligned}
\tag{8}
$$

together with constraints (5).

An analogical procedure is used to obtain a statement for the expected stage length function in aggregated decision processes, which provides the expectation of the duration of a review interval, i.e. the expected length of the stage between two consecutive decision epochs. Like in the case of the expected cost function, the expected stage length function $\bar{\tilde{\tau}}(\boldsymbol{k},u)$ is a conditional expectation of the related quantity given that the system meta-state at a current decision epoch is $\boldsymbol{k} = \left(k^0,k^1,k\right) \in \tilde{S}$ and

the subsequent action is *u*. Action $u \in U$ can again represent either an action $u = a \in \tilde{U} = U$ in a SADP or an action $u = s(b) \in \tilde{U}_s \subset U$ with $b \in \tilde{\tilde{U}}$ in a SAADP. Since the following equality $\bar{\tau}\left(i^0, i^1, k, a\left(u \middle| i^k\right)\right) = T\left(k, a\left(u \middle| i^k\right)\right)$ holds for $i^0 \in S^0, i^1 \in S^1, k \in S^2 = \{0,1\}, u \in U$, the resulting formula is stated by

$$\bar{\bar{\tau}}(\boldsymbol{k}, u) = \bar{\bar{\tau}}\left(k^0, k^1, k, u\right) \approx \sum_{i^k \in S^k\left(k^k\right)} \lambda^k\left(i^k \middle| k^k\right) T\left(k, a\left(u \middle| i^k\right)\right), \ \boldsymbol{k} = \left(k^0, k^1, k\right) \in \tilde{S}, u \in U, \tag{9}$$

together with constraints (5).

As soon as the transition, cost and time structure of an aggregated decision model is determined, algorithmic methods [3, 4, 6] available in the theory of semi-Markov decision processes can be used to find out an optimal stationary pure Markov policy for taking actions in the corresponding aggregated decision process, either in a SADP or in a SAADP.

3.3 Disaggregation formulae for control policies

The fixed-weight aggregation procedure employed either in a SADP or in a SAADP only serves for overcoming obstacles associated with large-scale Markov/semi-Markov decision processes. An optimal policy worked out for an aggregated decision process must somehow be reinterpreted for the corresponding original, non-aggregated decision process. The final result of such a disaggregation procedure is a policy that is suboptimal for controlling behaviour in the original decision process. This counterpart operation to the aggregation procedure is required, because we need to have revealed cost-effective rules for feedback control of the original process.

The core idea of disaggregation is to employ an action, prescribed as optimal for a meta-state $\boldsymbol{k} = \left(k^0, k^1, k\right) \in \tilde{S}$ by a policy $\left(\tilde{d}^*\right)^\infty$ in a SADP or by a policy $\left(\tilde{\tilde{d}}^*\right)^\infty$ in a SAADP, for all the states $\boldsymbol{i} = \left(i^0, i^1, i\right) \in S$ such that their first two components i^0, i^1 are elements of the respective subsets $S^0\left(k^0\right), S^1\left(k^1\right)$ in the partitions of the sets S^0, S^1 of possible values of related components of the state variable, whilst the last component *i* of the state ***i*** is equal to the last component *k* of the meta-state ***k***. In order to ensure feasibility of control actions in the original, non-aggregated decision process, an application of the acceptance function to a suggested action must be carried out. A policy constructed in accordance with the aforementioned disaggregation procedure can be considered to form a near-optimal contingency plan for action selection in all possible state values $\boldsymbol{i} = \left(i^0, i^1, i\right) \in S$ that can be recognised at decision epochs as system evolves in time over an infinite planning horizon.

Once an optimal policy $\left(\tilde{d}^*\right)^\infty$ with decision function $\tilde{d}^* : \tilde{S} \to \tilde{U}$ is determined in a SADP, the accompanying suboptimal policy $\left(d'\right)^\infty$ with decision function $d' : S \to U$ for an original NADP can be obtained according to the disaggregation formula

$$\begin{aligned} &d'(\boldsymbol{i}) = d'\left(i^0, i^1, i\right) = a\left(\tilde{d}^*\left(k^0, k^1, k\right) \middle| i^k\right) = a\left(\tilde{d}^*(\boldsymbol{k}) \middle| i^k\right), \\ &\boldsymbol{k} = \left(k^0, k^1, k\right) \in \left\{\boldsymbol{k} = \left(k^0, k^1, k\right) \in \tilde{S} : i^0 \in S^0\left(k^0\right), i^1 \in S^1\left(k^1\right), k = i\right\}, \forall \boldsymbol{i} = \left(i^0, i^1, i\right) \in S. \end{aligned} \tag{10}$$

An optimal policy $\left(\tilde{\tilde{d}}^*\right)^\infty$ with decision function $\tilde{\tilde{d}}^* : \tilde{S} \to \tilde{\tilde{U}}$ determined in a SAADP suggests the accompanying suboptimal policy $\left(d''\right)^\infty$ with decision function $d'' : S \to U$ for an original NADP by

$$\begin{aligned} &d''(\boldsymbol{i}) = d''\left(i^0, i^1, i\right) = a\left(s\left(\tilde{\tilde{d}}^*\left(k^0, k^1, k\right)\right) \middle| i^k\right) = a\left(s\left(\tilde{\tilde{d}}^*(\boldsymbol{k})\right) \middle| i^k\right), \\ &\boldsymbol{k} = \left(k^0, k^1, k\right) \in \left\{\boldsymbol{k} = \left(k^0, k^1, k\right) \in \tilde{S} : i^0 \in S^0\left(k^0\right), i^1 \in S^1\left(k^1\right), k = i\right\}, \forall \boldsymbol{i} = \left(i^0, i^1, i\right) \in S. \end{aligned} \tag{11}$$

4 Conclusion

A policy $(d')^{\infty}$ or $(d'')^{\infty}$ resulting from the aggregation–disaggregation procedure represents an approximation to an optimal policy for the original large-scale semi-Markov decision process that models the feedback traffic control problem with lane restriction. A suboptimal policy $(d')^{\infty}$ or $(d'')^{\infty}$ prescribes that it is convenient, with respect to the costs incurred for closure system operation, to dispatch $d'(i^0, i^1, i)$ or $d''(i^0, i^1, i)$ vehicles to move through the closure sector along the active direction whenever i^0 vehicles are waiting in the queue of the 0-direction and i^1 vehicles are staying in the queue of the 1-direction at a current decision epoch if an active direction is just the *i*-direction. The employment of a suboptimal policy for feedback traffic control with lane restriction approaches the objective to minimise the consumption of delay time that vehicles spent waiting in the queues in front of the closure sector at both sector ends.

To avoid problems associated with the "curse of dimensionality", state aggregation is crucial. The accompanying action aggregation, if any, only reduces the number of possible alternatives for action choice in computational algorithms for optimal policy determination. On the other hand, action aggregation increases the risk of such an eventuality that a better solution remains unrevealed. But action aggregation is quite well justified to be applied, because it is natural to redefine the action space after action aggregation with the same level of coarseness as the sets in the state space have been done after state aggregation.

Because of computational tractability, the model has to operate with finite queue limits. But there are no restrictions limiting the lengths of the queues at both ends of the closure sector in the real situation on the road. The consequence is apparent: the higher the queue limits in a model the better the fidelity of the model in comparison with reality. Larger limits for numbers of vehicles registered in the model as waiting in the queues decrease the distortion in the description of the stochastic dynamics that governs the system evolution over time. Since the fixed-weight aggregation procedure enables one to built a decision model with a relatively large extent of possible non-aggregate values for particular components of the state variable, a suboptimal policy obtained after the aggregation–disaggregation procedure can be regarded to represent a better solution to the optimal traffic control problem than an optimal policy in the decision model without aggregation, where the extent of possible values for the state components is smaller.

References

[1] Bartl, O., A traffic control policy to treat the one-directional closure on the road. In: *Science, Education and Society. 11th International Scientific Conference. Section No. 2: Information and Communication Systems and Technologies.* (Žilina, SK, 17–19 Sept. 2003), University of Žilina, Žilina, 2003, pp. 21–29

[2] Bartl, O.: An aggregation procedure for large-scale Markov decision processes. In: *Proceedings of the 22nd International Conference Mathematical Methods in Economics.* (Brno, CZ, 15–17 Sept. 2004), Masaryk University Brno, Brno, 2004, pp. 9–15

[3] Puterman, M. L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, New York, 1994

[4] Tijms, H. C., *Stochastic Modelling and Analysis: A Computational Approach.* Wiley, Chichester, 1986

[5] White, C. C. and White, D. J., Markov decision processes. Invited review. *European Journal of Operational Research* **39**, 1989, 1–16

[6] White, D. J., *Markov Decision Processes.* Wiley, Chichester, 1993

This work was supported by Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences under grant No. 1/0498/03.

Analyzing the Path of Slovakia towards Monetary Union by Means of a Multi-Equation Gap Model

Michal Benčík
Institute of Monetary and Financial Studies, National Bank of Slovakia, I. Karvaša 1, 81325 Bratislava, Slovakia, e-mail michal.bencik@nbs.sk[1]

Abstract

After joining the European Union, Slovakia is preparing itself for the adoption of the euro. This process comprises bringing monetary policy and fiscal policy in line with the Maastricht criteria, while maintaining a low inflation rate and a stable nominal exchange rate during the two year pre-accession membership in ERM II. In this paper, we study the possibility of a consistent scenario of transition from the current state to full compliance with these criteria, avoiding nominal depreciation and, if it exists, to determine conditions for such development. We build a multi-equation structural gap model of transmission mechanism, incorporating the fiscal policy measure and long-run real appreciation due to Balassa-Samuelson effect. Using this model we compute an illustrative baseline scenario and four alternative scenarios, exploring the comparative dynamics of the model. The main value added of this contribution is development of a tool for studying necessary conditions for consistent transition towards of Slovakia monetary union membership..

Keywords

Euro adoption, transmission mechanism, econometric model, Taylor rule, exchange rate

1 Introduction and motivation

Chapter Slovak Republic entered the European Union and declared its desire to adopt the common currency as soon as possible. The adoption of euro shall be preceded by membership in the system ERM II. During the membership in ERM II, Slovakia shall besides keeping the nominal exchange rate close to parity also simultaneously implement changes in the economy and policy in order to fulfill the Maastricht criteria. These tasks require a steady development of the economy, followed by permanent changes in key economic indicators.

The purpose of this work is to adapt the gap models from the literature to the stylized facts for Slovakia and compute an illustrative scenario with consistent path both for real economy (output gap, net exports) as well as for inflation, exchange rate and interest rate from current state to the values complying with Maastricht criteria.

We develop a tool for policy analysis for Slovakia during the process of preparation to entry into monetary union. Our model follows roughly the structure of standard gap model for countries using inflation targeting, presented in Laxton and Scott (2000) and applied in many countries using inflation targeting, including the accession countries (for example Beneš et all, 2003), sometimes incorporating also the state space model for long run (equilibrium) values.

2 The description of the model

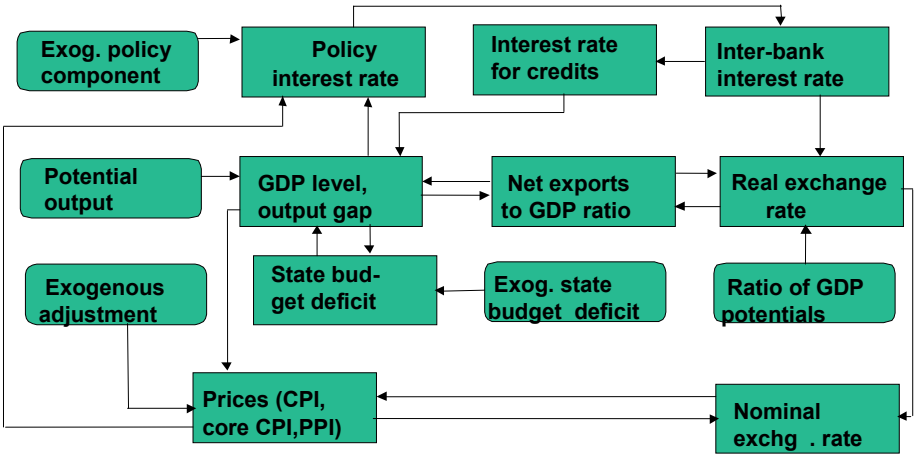
We use the quarterly time series for Slovak Republic, constructed by the Slovak statistical Office and National Bank of Slovakia. For foreign indicators, we use quarterly series from the database Olisnet,

[1] This material represents solely opinions of its author and these may not correspond to the official opinions of National Bank of Slovakia.

maintained by OECD. Most data series start in 1993, but due to high volatility and irregularity of development at the beginning of transition, the estimation periods were somewhat shorter than this. The last observation is for the first quarter 2004. The potential output and output gap ex post were computed from seasonally adjusted data prior to estimation, the equations were estimated from seasonally unadjusted data and seasonal dummies were included where necessary. This approach was chosen because methods for seasonal adjustment are often based on moving averages that could distort the estimation and lead to spurious significance of lagged variables. All databases were compiled and computations carried out using Eviews 5.0.We build a multi-equation gap model, reflecting the following relationships:

- Generalized Taylor rule
- Forward-looking Phillips curve
- Generalized real UIP condition reflecting the Balassa-Samuelson effect
- Generalized IS curve
- Net exports determination
- Adjustment of interest rate for credits to monetary policy
- Reaction of fiscal policy to business cycle.

A list of equations and variables is in the Appendix. The structure of the model is depicted in the picture:



THE STRUCTURE OF THE MODEL

3 Baseline scenario and alternative scenarios

After estimation, we computed an illustrative scenario of future development. We want this scenario to be interpretable form an economic point of view, but do not intend it as a true forecast. Due to instabilities in Philips curve, the paths of nominal variables were corrected in the following way:

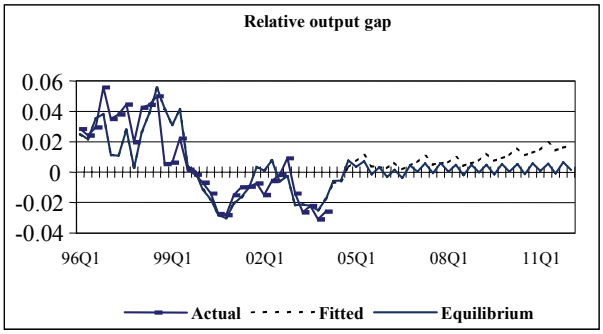
1. We defined a target series for core inflation, assuming actual values ex post and gradually declining to 2% per year ex ante,
2. We modified the model structure by
    a. Introducing an auxiliary equation making the inflation equal to its target,
    b. Rewriting the Philips curve so that there was the add-factor RESCPIC on the left-hand side and the difference between the core inflation and the fitted value of the original Philips curve on the right-hand side,
3. We simulated the modified model and stored the endogenous add-factor RESCPIC, we simulated the original model with the values of RESCPIC computed from the modified model

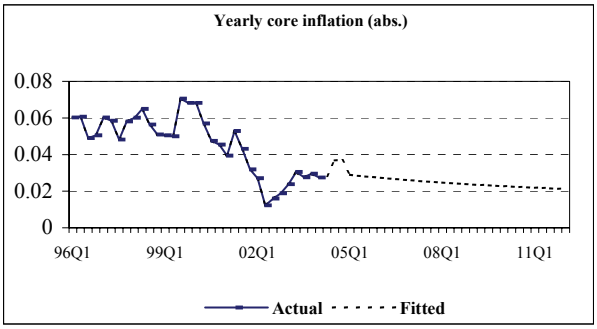
In this simulation, the core inflation assumed the target value. We assumed, in line with known requirements, that the fiscal policy will get mere restrictive and monetary policy will get more

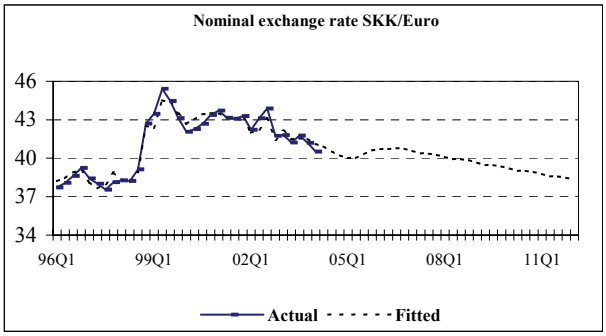
expansionary and reflected this in their exogenous parts. The below table lists the main exogenous variables and their way of extrapolation in the illustrative baseline scenario

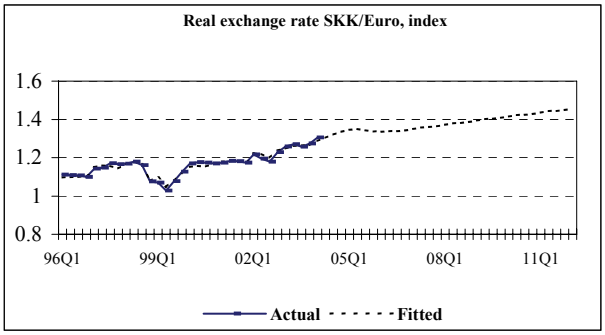
Variable	**Way of extrapolation**
Potential output for euro zone	Yearly growth by 2%
PPI index for euro zone	Yearly growth by 1.6% increasing later to 1.9%
LIBOR rate	Increasing up to 2% in 2004, 2% afterwards
Imports in Germany	Yearly growth 4.5 % in 2004, 5% afterwards
PPI index in USA	Yearly growth 1.5 % in 2004, 2% afterwards
Potential output in Slovakia	Yearly growth 4 %
Inflation due to administrative actions	Between 0.002 and 0.009 till 2007, 0 afterwards
Foreign direct investment inflow	10 billion SKK quarterly up to 2006, then halved

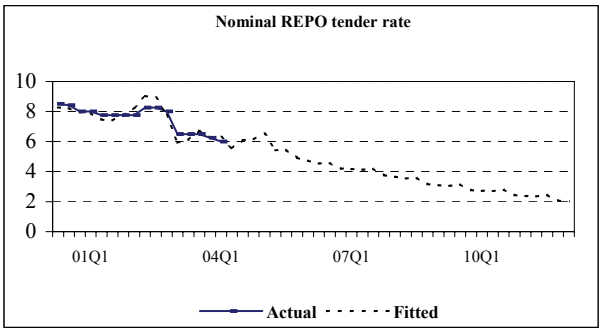
After applying these assumptions, the main endogenous variables made smooth transition form actual state in 2004 near to complying with Maastricht criteria in 2012 (the end of forecasting horizon.). Their paths are shown in following charts:

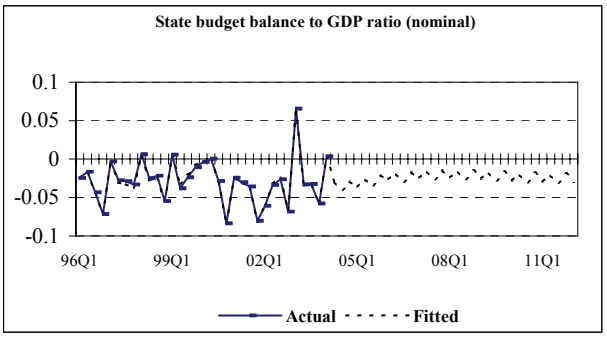












Eight alternative scenarios were computed for four variables, both with transitory (4 quarters) and permanent shocks:

- Increase by one percentage point in policy rate
- Decrease of state budget deficit by one percent of GDP
- Increase in imports in Germany by one percent
- Increase by 0.01 in real exchange rate• increase by one percent in potential output

All shocks begun in 1-st quarter of 1998, transitory shocks ended in 4-th quarter 1998.

The main results of alternative forecasts can be summarized as follows:

- Transitory shocks have only transitory effects.
- An exogenous permanent shift of the policy interest rate will be partially offset and causes a permanent shift in the output gap (the interest rate is supposed to deviate from the natural level) accompanied by a slight increasing of the state budget deficit, nominal and real appreciation and a permanent shift in the price *level*.
- An exogenous permanent tightening of fiscal policy will be offset only to a small extent and will cause a permanent drop in the output gap, interest rate, a permanent shift of the price level, a moderate permanent real and nominal appreciation.
- An exogenous permanent real appreciation will be amplified to its multiple. It will cause a permanent shift in the price level and relatively strong nominal appreciation.
- Permanent shift in the potential output will cause a permanent real and nominal appreciation and a permanent shift in the price level of a moderate size.

4 Conclusion

Before joining monetary union, during the membership in ERM II, Slovakia shall besides keeping the nominal exchange rate close to parity also simultaneously implement changes in the economy and policy in order to fulfill the Maastricht criteria. We constructed a model for simulating the effects of economic policy for this challenging period, based on gap models for countries in the regime of inflation targeting. We have assumed relaxing monetary policy, tightening fiscal policy quite favorable outside conditions. Our results show that it is possible to fulfill both these tasks, if the policies are coordinated. Simulations with varying assumptions about paths of exogenous variables showed fairly standard properties of the model.

References

[1] Beneš, J. – Hledik T. – Vávra D. - Vlček, J. (2003): 'The Quarterly Projection Model and its Properties' ,in The Czech National Bank's Forecasting and Policy Analysis System, edited by W. Coats, D. Laxton, and D. Rose, Prague, Czech National Bank

[2] Laxton, D. and Scott, A. (2000): 'On Developing a Structured Forecasting and Policy Analysis System Designed To Support Inflation Forecast Targeting (IFT)', in: InflationTargeting Experiences, England, Finland, Poland, Mexico, Brazil, Chile, The Central Bank of Republic of Turkey.

Appendix 1

COMPLETE LIST OF EQUATIONS OF THE MODEL

TAYLOR RULE
REPO2W=RESREPO+150*DLOG(CPIC(-1),0,4)+50*LYGAP(-1)

PHILIPS CURVE
DLOG(CPIC) = 0.74505*DLOG(CPIC(-1)) + 0.08223*DLCPIALL(1)
+ 0.09261*D(LYGAP(-1)) + 0.0154*D98Q3(-1) +(1-0.74505
-0.08223-0.10616)*DLOG(PPIF(-1))+0.01705*D97Q4+0.106161
INFLEX(TIME>33)-0.01414*D98Q3 +0.00636*(T1-0.25)-0.00668
*(T2-0.25)-0.01362*D1Q2(-1)-0.008*ISPERIOD("2000:2")+RESCPIC

CONSUMER PRICES
DLOG(CPIALL) = DLOG(CPIC)+INFLEX

CONTRIBUTION OF IMPORT PRICES TO CONSUMER PRICES

PPIF = SKKEURA*EUZPPI

CONSUMER PRICES - INFLATION RATE FOR LEADING -
DLCPIALL = DLOG(CPIALL)

PRODUCER PRICES -MANUFACTURING
PPIM=CPIC * QPPIM

PRODUCER PRICES
PPIALL=CPIALL * QPPI

REAL EFFECTIVE EXCHANGE RATE
DLOG(SKKEUROREALM) = 0.79898 + 0.24352*
DLOG(SKKEUROREALM(-1)) - 0.33230*(LOG(SKKEUROREALM(-1))
-LIGDP(-1)) - 0.06903*D98Q3(-3) - 0.06804*D98Q3(-1) +
0.00213*RINTDIF(-2) + 0.09258*QNXN(-3) + 0.07318
*PZISALDO(-1)/NAGDPN(-1) + 0.03464*ISPERIOD("1997:1") +
0.03462*ISPERIOD("2002:1")+RESSKKEUROREAL

SHARE OF SLOVAK GDP TO THAT OF EU12
LIGDP= (LYPOT)-LOG(GDPE121K6)

NOMINAL EXCHANGE RATE SKK/EURO
SKKEURA= PPIM/(SKKEUROREALM*EUZPPI*0.018801978896)

IS CURVE
DLOG(NAGDPR) = 0.00236 - 0.23144*LYGAP(-3) + 0.10336*T2 +
0.04934*T3S + 0.18397*D(QNXN(-2)) + 0.03593*D97Q4(-1) -
0.03213*ISPERIOD("2000:4") - 0.04029*T1 - 0.03703*T4
+ 0.03550*D98Q3(-2) + 0.01761*D98Q3(-3) - 0.17560*
SBFSAL/(NAGDPN/1000) - 0.12111*(0.01*DBCRAV-DLOG(PPIALL,0,4))
+RESNAGDPR

OUTPUT GAP
LYGAP=LOG(NAGDPR *(1.0567*T1 + 0.96197*T2 +
0.96565*T3 + 1.00600*T4))-LYPOT

SHARE OF NET EXPORTS ON GDP
D(QNXN) = -0.00941 - 0.46050*QNXN(-1) - 0.04991*T4 +
0.45634*DLOG(DEUMGSNA(-1)) - 1.00031*LYGAP - 0.07007
*ISPERIOD("2000:4") - 0.11382*DLOG(SKKEUROREALM(-2),0,4)
+RESQNX

BRIBOR RATE 3 MONTHS
DMMR3M=REPO2W+RESDMMR3M

INTEREST RATE FOR CREDITS

D(DBCRAV) = 0.06048*(DMMR3M(-1)-DBCRAV(-1)) + 0.31501
*D(DBCRAV(-1)) - 1.09106*ISPERIOD("2000:1")+RESDBCRAV

GDP IN CURRENT PRICES
NAGDPN =NAGDPR*CPIALL*QPY

REAL INTEREST RATE DIFFERENTIAL
RINTDIF=DMMR3M-100*(PPIALL/PPIALL(-4)-1)-LONDONUSD
+100*(FPUSAPPI/FPUSAPPI(-4)-1)

RATIO OF STATE BUDGET BALANCE TO GDP
SBFSAL/(NAGDPN/1000) = -0.02013 + 0.16204*LYGAP(-2) -
0.01161*T3 - 0.03520*T4 + 0.08584*(TIME=41)+RESQSBFSAL

Appendix 2 Variable list

CPIALL	CPI total 1995=100
CPIC	Core inflation, cumulated chain index
D1Q2	Dummy 2-nd quarter 2001=1
D97Q4	Dummy, 4-th quarter 1997 = 1
D98Q3	Dummy, 3-rd quarter 1998 = 1
DBCRAV	Nominal average interest rate for credits, %
DEUMGSNA	German imports of goods and services, basic index
DLCPIC	Core inflation, log difference (quarterly rate of change)
DMMR3M	BRIBOR rate 3 month, percentage points
EUZPPI	Euro zone producer price index
FPUSAPPI	Producer price index, USA
GDPE121k6	Real GDP, Euro zone, HP-filtered
INFLEX	Price shocks form deregulations and abolishment of subsidies
ISPERIOD (“date”)	Binary dummy with unit value in “date”, otherwise zero
INFLEX	Price shocks form deregulations and abolishment of subsidies
LIGDP	Ratio of Slovak real GDP to that of Euro zone
LONDONUSD	LIBOR 3 month USD rate, percentage points
LYGAP	Output gap, log
LYPOT	Potential output, log
NAGDPN	Nominal GDP
NAGDPR	Real GDP, 1995 prices
PPIALL	Industrial producer prices , level
PPIF	Contribution of foreign producer prices to inflation, level
PPIM	Industrial producer prices - manufacturing, level
QNXN	Ratio of nominal net exports of goods and services to nominal GDP
QPPI	Ratio of PPI to CPI
QPPIM	Ratio of PPI manufacturing to core inflation
QPY	Ratio of GDP deflator to CPI
REPO2W	Interest rate for two weeks REPO tender (taken as policy rate)
RESCPIC	Residual in behavioral equation (Philips curve)
RESDBCRAV	Exogenous residual for interest rate for credits
RESDMMR3M	Exogenous difference between REPO and 3 month BRIBOR
RESNAGDPR	Residual of behavioral equation (GDP)
RESQNX	Residual in behavioral equation (net exports share)
RESQSBFSAL	Residual in behavioral equation (budget deficit to GDP ratio)
RESREPO	Residual in Taylor rule
RESSKKEUROREAL	Residual in behavioral equation (exchange rate)
RINTDIF	Real interest rate differential (between 3 month BRIBOR and LIBOR
SBFSAL	State budget deficit
SKKEURA	Nominal exchange rate SKK /EURO
SKKEUROREALM	Real exchange rate SKK /EURO
T1, T2, T3, T3S, T4	Seasonal dummies (T3S has zero values till 3-rd quarter of 1998)
TIME	Linear time trend

Press Distribution Process with a Down-Up-Down Strategy

[a)]José M. Caridad y Ocerin, [b)] Francisco J. Rodríguez Aragón
[a)]*Dpt. of Statistics and Econometrics CórdobaETEA, Centro Andaluz de Prospectiva* ccjm@uco.es
[b)]*Dpt. of Distribution GELESA*[1] Franciscor@Gelesa.es
Spain

Abstract

Press distribution is done trough a network of selling points and it has to be adjusted to match supply and a random demand in each of them. Conflicting goals arouse, as each point demand has to be forecasted, so in a particular outlet there could be an excess of supply or a bigger than the expected demand. The companies in charge of the distribution generally use heuristic methods, based on their market knowledge and of past experiences, in some cases, with neural networks or time series models. Global production is estimated and then it is assigned to the network nodes point. A strategy for the distribution process is established based on a Down-Up-Down procedure, and a measure to evaluate the assignment process related to the company goals is proposed. This index is used to compare the annual evolution of sales and to determine the optimal ratio of the number of papers to be produced, related to the forecasted demand. A case study is presented related to a Spanish press distribution company.

Keywords: press distribution, efficiency, assignation of services, demand forecasting, evaluation of distribution, production policies.

Introduction

In the optimization process associated with press distribution there are three phases to be considered: goals associated with the distribution, the decisions about the number of copies to be supplied to each outlet, and finally, the control and evaluation process of the results achieved. This validation of the quality attained in the assignation is necessary, not only to assess the whole process, but mainly to formulate the new objectives in later distribution decisions. A function aimed at measurements of the efficiency is proposed; it informs about the degree of fulfilment of the objectives, according to a predefined economic criteria (as could be cost minimization). To define this function, it is necessary to consider to conflicting goals, associated to the distribution process in a network of outlets: the demand should be attended as much as possible, that is, enough number of copies have to be supplied to each of them, but without having an exaggerated excess of supply, as these have to be collected, with an additional cost. Thus, two variables have to be considered: the proportion of unsold copies, $d = D/S$, being D the number of returned papers and S the total supply, and $a = k_a/k$, where k is the number of outlets in the network and k_a those who outsold all their supply. In broad terms, if total supply, S, is increased, the proportion of selling points that run out of copies, a, decreases, and thus $a = a(d)$ is a decreasing function. Distributors do not get information about the possible excess of demand, so if and outlet does not return any copy, they assume that there is a deficit in the supply awarded to this particular selling point, although it could happens that the demand matched exactly the supply (Caridad, Rodríguez, 2005), unless this demand is estimated as proposed by Caridad and Rodríguez (2004).

The efficiency of the distribution process could be measured by a function $E = E(a,d)$ or simply taking into account the shape of the curve $a = a(d)$, as will be proposed later on. Of course some additional variables are linked to the efficiency function, but a and d are easily available, without additional costs, to any distributor, and are a common way to evaluate the process, to set goals for future time periods, and to compare and assess the distribution process over time.

Daily several decisions have to be adopted regarding the sales through a network of outlets. The *Up-Down* strategies are based on forecasting total sales, V, and then the print run, S. Then, this total

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

amount is distributed to the outlets, using different techniques, almost all based on past sales, with occasional corrections due to a particular demand, or a known event that affect a group of selling points. The distribution process is defined by the proportions of S, D = $\{p_j, j = 1,2,\ldots,k$, awarded to a particular point. There are different ways of forecasting V, deciding S and the distribution set, D. Total demand can be modelled using econometric tools, or time series techniques, and then, the distribution proportions are estimated with some predetermined method, like heuristically defined distribution tables, or some artificial intelligence tools, as proposed by Heskes (2002) and Heskes *et al* (2003), using Bayesian methods and neural networks in an integrated environment. Another way of proceeding, as proposed here, is a *Down-Up-Down* strategy. In Spain, each selling point is an independent shop that pays a fixed amount to the distributor, independently of the results of daily sales. As stated in Caridad and Rodríguez (2005), there are no penalties for returned copies, so there is no incentive from each outlet to adjust to the real demand, and in some cases, some of them ask for a systematic excess number of copies. It is thus the responsibility of the distribution firm to establish on a daily basis the distribution set D for the entire network. The benefits of the collaboration or the selling points to evaluate the demand are clear, but there are cost involved that can overcome these. Also, as Cebrián (2001) some other costs are to be taken into account: in 2001 paper costs increased over 25%. The optimization of costs and benefits are of primary concern to the distributor, as there could be relative advantages towards competitors.

In this paper there are three parts: first, a measure of efficiency in the distribution process is proposed, and then applied to a well known Spanish sports newspaper; then some ideas are given about the supply and distribution process for a journal distributed in Madrid area, and how goals about the variables a and d can be determined. Finally some proposal about de proposed *Down-Up-Down* alternatives could be implemented.

Evaluating the Distribution Process

In press distribution, the total number of copies, S_t, supplied a particular day t, is usually estimated using econometric modelling. Then, to each of the k outlets of a network of selling points a proportion $p_{jt} = s_{jt}/S_t$, $j = 1,2,\ldots,k$, is supplied. The distributor would aim to keep as low as possible the over or undersupply to each point, and in an ideal situation, to eliminate these deviations. This can not be achieved, as the demand in each outlet is a random variable, although some information is obtained every day about it: the sales on day t, v_{jt}, $j = 1,2,\ldots,k$, that are capped by the corresponding supply, s_{it}, $j = 1,2,\ldots,k$, and, the number of returned copies, d_{it}, $j = 1,2,\ldots,k$. From these data, an aggregated measure for the daily distribution is the proportions of the print run returned, d, and of the selling points, a, that have run out of stock:

$$d_t = \sum_{j=1}^{k} d_{jt} / S_t \qquad\qquad a_t = \sum_{j=1}^{k} k_{jt} / k_t$$

being $k_{jt} = 1$, if $d_{jt} > 0$, and 0 otherwise. Considering an interval of time, $t = 1,2,\ldots,T$, (usually a fiscal year), these two proportions are clearly related, and lie, approximately over an hyperbolic type function, that could be represented with a potential model as $a = \alpha d^{\beta} + \varepsilon$. This model arouses considering a Cobb-Douglas type evaluation function, as $y = \alpha d^{\beta} a + \varepsilon$ related to the efficiency of the distribution process.

Using daily data for a sample of one hundred outlets of a Spanish sport journal, this specification is plausible, as can be seen on figure 1, that produces the following estimated models:

$$\hat{a}_{2002} = 0.0025 d^{-2.5768} \qquad R^2 = 0.79$$
$$\hat{a}_{2004} = 0.0064 d^{-1.9335} \qquad R^2 = 0.78$$

For the same journal, it could be considered that the same specification is valid, as has been checked with different periods and publications.

The ideal situation is that all points on these scatter grams tend towards the origin. Thus, if one of these curves lies nearer to the axis, the distribution process associated to this period, could be considered more efficient that the other.

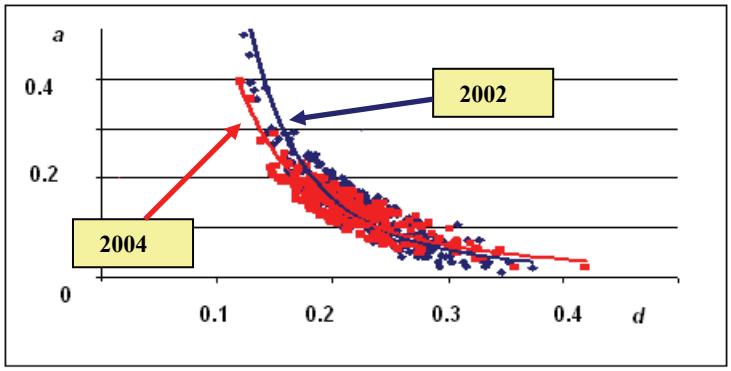


Figure 1: Assignation curves for a sport journal (2002-4)

As can be deduced from these data, the distribution process is unevenly efficient for each day, as could be expected. In some days, the supply overshot the demand, and the number of selling points that run out of stock is low, but, on the other hand, provoking a higher proportion of returned papers, and in some other days, a underestimation of global demand resulted in a higher proportion of outlets sold out, and probably, without been able to satisfy all their potential consumers.

The efficiency of the distribution process can be measured by the area under de curve $\hat{a} = \alpha d^{\beta}$, as shown in figure 2.

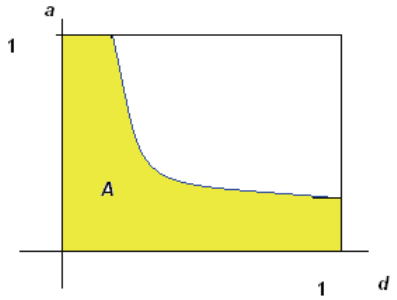


Figure 2: A measure of the efficiency of the distribution process.

Thus, this measure is

$$A = \alpha^{-1/\beta} + \int_{1/\alpha^{\beta}}^{1} \alpha d^{\beta} \mathrm{d}d = \alpha^{-1/\beta} + \alpha(1-\alpha^{-(1+1/\beta)})/(1+\beta)$$

or, alternatively, if the objective is to define a variable increasing with the efficiency, the efficiency measure is $E = A^{-1}$.

Applying this to data in figure 1, it can be obtained that

$$E_{2002} = 6.3216 \qquad E_{2004} = 6.8939$$

and, thus, an apparently better distribution process was in the year 2002.

One factor that has not been introduced is the average of sales in each outlet, $\mu = V/k$. This variable is related to the efficiency of any distribution, as it can be easily checked that it is easier to design a distribution process, D, to a set of selling points that have higher average sales, that to a network with lower a lower mean value, μ. Again, during a particular time period, the number of selling points surely changes, as well as the print run, affecting the difficulty of distributing a newspaper. Defining the ratio of average sales per outlet in two time periods, $r = \mu_1/\mu_2$, a new measure of efficiency to compare the efficiency E_1, an the period 1, with efficiency, E_2, corresponding to period 2, can be defined as

$$E_{2\bullet 1} = E_2^{r} = E_2^{\mu_1/\mu_2}$$

If the ratio $r > 1$, it should be easier to establish an 'efficient' distribution process (in period 2), that if $r < 1$, and, as $E > 1$, this *comparative measure of efficiency*, is penalized by this r value..

In the sport journal considered in our case study, the efficiency in 2004 compared with two years before, is $E_{2004\cdot 2002} = 6.8939^{25.5/27.4} = 6.0301$ corresponding to a slightly less efficient the distribution in 2004 than in 2002, against what was observed without this r-correction.

Any measure of relative efficiency should have several properties: $E_{t\cdot t'} = g(E_t, r_{t\cdot t'})$ has to be decreasing with $r = \mu_{t'}/\mu_t$, and increasing with E_t. Also, $g(E_t, 1) = E_t$. It is possible to define some alternative measures of efficiency to the proposed, as

$$E = \frac{1}{k}\sum_{j=1}^{k}(d_j^2 + a_j^2)$$

or some monotonous function of E, and later on introduce an r-correction, for comparative studies.

Objective Setting for a Distribution Process

The E measure can be used to set goals for the proportion d of unsold papers and for the proportion of outlets, a, that are going to sell all their supply. If C_D and C_A are the cost involved in increasing 1% these two proportions, and $c = C_D/C_A$ is their relative values, it is possible to set some goals to be attained in the distribution process, for a predefined level of efficiency E_0, minimising the cost function $C(d, a) = C_D d + C_A a$

with the restrictions $E_0 = E^r$ $\beta = \beta_0$

being β_0 the corresponding coefficient of the assignation curve from the previous period and $r = \mu_0/\mu$ the ratio of the expected mean sales per outlet, to the current sales. It is also assumed that the shape of this function would not change significantly from one period to the next.

This optimization problem can be solved by numerical methods, (Maron M. J. 1987), and it would provide the proportions d_0 and a_0 of the returned copies and of the outlets that will run out of papers; if these objectives are attained, they would minimize the distribution cost, with a predefined level of efficiency, E_0, in the distribution.

For obtaining the objectives for 2005, the 2004 assignation function is used $\hat{a} = 0.0064d^{-1.9335}$

For 2004an increase in mean sales was forecasted, from 25.5 copies two years before, to 27.4 in 2004, and again, for 2005, the average sale per outlet should increase by one unit. For establishing the 2005 goals, it was assumed that the assignation function would be of the form $\hat{a} = \alpha d^{-1.9335}$, as shown in the following figure, for different values of α.

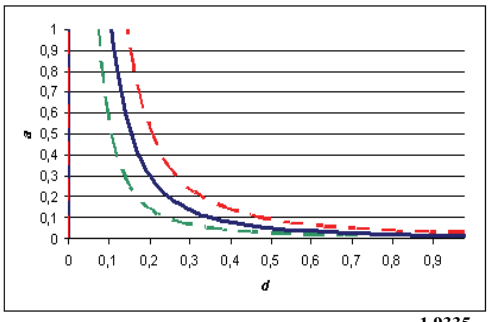


Figure 3. Forecasted assignation functions $\hat{a} = \alpha d^{-1.9335}$

Then, to compare with the 2004 function, the corresponding r correction is applied, with $r = 27.4/28.4$, and the efficiency measure for 2005 should be

$$E_{2005\bullet 2004} = \frac{1}{\alpha^{\frac{1}{1.9335}} - \frac{\alpha}{1-1.9335}\left(1-\alpha^{\frac{1-1.9335}{1.9335}}\right)} = 6.8939^{\frac{27.4}{28.4}} = 6.4408$$

The minimization of the cost function, $C = C_D d + C_A a$, with the restriction, $\hat{a} = \alpha d^{-1.9335}$, leads to the proportion of unsold units and of unattended outlets $d_{2005} = 23.40\%$, $a_{2005} = 12.10\%$.

The costs are considered equal: $C_D = C_A$. From these values, and from the sales forecast V_{2005}, the print run is deduced $S_{2005} = V_{2005}/(1 - 0.2340)$

The obtained proportions depend on the relative value associated to a returned copy, P_D, or to and unsold one due to an outlet that ran out of stock, P_A. The C_D and C_A values depend on these unitary costs. Different values of c originate the corresponding values of the objectives to be achieved for the proportion of returned papers and of the outlets that sold all their supply.

In case of a newspaper sold over a large geographical area, different goals for the proportions *a* and *d* can be formulated nor subsets of outlets, depending of variable costs of returning copies, or on the value attributed to unsold papers due to undersupply.

***Down-Up-Down* Strategies**

Instead of defining the distribution process by the proportions D = $\{p_j, j = 1,2,\ldots,k\}$ of the print run *S*, estimated from the total sales forecast, *V*, and the proportion, *d*, of returned copies, it is possible to follow a different strategy, that could lower both the rates of unsold copies, and of the number of outlets that sold all their supply. In the *Down-Up-Down* process, besides obtaining a forecast of future sales, the network is divided in *g* subsets of outlets, with geographical criteria. The rationale is that not far from selling points could be treated as intermediate units for forecasting demand, and, thus, calculating supply in a more precise way, as in case of a sold out point, a customer could choose to go to a nearby outlet. Also other criteria for forecasting demand could be to form groups of points with similar behaviour in the demand.

In the *Up-Down* strategy global sales obtained from aggregated sales and supply of each outlet

$$V_t = \sum_{j=1}^{k} v_{jt} \qquad\qquad S_t = \sum_{j=1}^{k} s_{jt}$$

for $t = 1,2,\ldots,n$, are estimated the forecast $\widehat{V}_{n+1}$ and the corresponding global supply

$$\hat{S}_{n+1} = f(\widehat{V}_{n+1}, d) = \widehat{V}_{n+1}/(1-d)$$

taking into account the expected or desired level of the proportion *d* of returned copies, assuming that this proportion is almost constant during the $n + 1$ period. In a more precise modelling, it is possible to allow this proportion to change with time.

In the *Down-Up-Down* strategy the *k* selling points are clustered in *g* groups, and if properly chosen, more precise data of future demand can be obtained. The number of outlets associated in group *i*, is designed by k_i, been $k = \Sigma k_i$. Using past sales and supply time series for each group

$$V_t^i = \sum_{j \in G_i} v^i_{\ jt} \qquad\qquad S_t^i = \sum_{j \in G_i} s^i_{jt}$$

and the objective proportion of returned papers, it is possible to forecast future sales and supplies, for each group, and aggregating these, the total forecast for sales and supply

$$\widehat{V}_t^G = \sum_{i=1}^{g} \widehat{V}_t^i \qquad\qquad \hat{S}_t^G = \sum_{i=1}^{g} \hat{S}_t^i \qquad\qquad t = n+1,\ n+2,\ldots$$

To obtain these values some classical time series models, like Arima, can be used, or neural networks, or, econometric models that include intervention variables. The obtained aggregated forecast can differ from the data obtained directly for the global sale and supply time series. In some case, it is possible to combine both forecasts, if this seems more precise. From the global forecast and from the predicted data for each group, the distribution proportions sets are obtained, for period $t = n + 1$, $D_G = \left\{ p_i = \hat{s}_i / \hat{S}, \quad i = 1,2,\ldots,g \right\}$

With these proportions are obtained the number of copies to be supplied to each one of the *g* groups, using the total supply finally decide, *S*.

Once the number of copies to be supplied to each group of outlets is decided, it is necessary to estimate the proportion of papers to be assigned to each selling point. This can be done using a methodology previously developed by Caridad, Rodriguez and Ceular (2004), optimising the expected cost of returned copies and of unsold papers in outlets that run out of them. This cost function, for any outlet, *j*, is

$$C(s_j) = P_D \sum_{i=1}^{s_j - 1} (s_j - i) f_j(i) + P_A \sum_{i=s_{jt}+1}^{+\infty} (i - s_j) f_j(i)$$

being *s* the supply, P_D and P_A, the loss associated to a returned or an unsold copy for lack of sufficient supply. The demand is assumed a random variable with probability function $f_j(x)$. As stated by Caridad *et al*, (2004), for outlets with average daily sales of less than 30 papers, the

Poisson and geometric distributions provide a good specification, while for outlets with larger sales; a Gaussian approximation facilitates the calculus. The minimisation of this function (for each outlet) leads to the optimum number of copies to be supplied to each point of sale generating the values $\{\hat{s}_{ij}, j = 1,2,\ldots,k_i, i = 1,2,\ldots,g\}$

and the corresponding proportions related to each group are obtained:

$$D_i = \left\{ p_{ij} = \hat{s}_{ij} / \hat{S}_i, \quad j = 1,2,...,k_i \right\} \qquad i = 1,2,\ldots,g.$$

The treatment of the demand distribution for each outlet can be improved taking into account that demand is not directly observable: sales are a truncated variable obtained from demand, as these can not be larger than supply. In Caridad and Rodriguez (2004[b]) the estimation of the demand parameters is done trough a maximum-likelihood method over truncated data.

To evaluate the feasibility of this approach, a real case is presented, using a group of seven points in a Madrid urban area. Daily data from a year of sales on week days from March 2004 are used. Aggregated data in this group are forecasted using an Arima model, taking into account the weakly cycle, and forecast of supply to be allocated during the following month, minimising costs, are obtained. Sales forecasts for this period are presented in the following table:

Day	Sales forecasts	Day	Sales forecasts	Day	Sales forecasts
1	77	8	81	15	84
2	68	9	65	16	67
3	77	10	80	17	66
4	66	11	68	18	79
7	76	14	73	21	82

Table 1: Daily sales forecast in a group of selling points

Using the proportion of returned copies for this newspaper, estimated at 21.61%, the supplies for each day is obtained dividing previous data by $1 - d = 1 - 0.2161$. In table 2, are presented the supply data, for weekdays of March 2005, minimising cost functions and the real number of copies send by the distributor using his own in house distribution process, bases on mean sales and moving averages over daily sales.

Day	Estimated supply	Real supply	Day	Estimated supply	Real supply	Day	Estimated supply	Real supply
1	98	104	8	103	100	15	107	101
2	87	95	9	83	93	16	85	91
3	98	95	10	102	111	17	84	91
4	93*	102	11	87	99	18	101	91
7	97	92	14	93	97	21	105	101

Table 2: Supply minimising distribution costs

As the group of outlets considered had not no low average daily sales, the distribution of the demand in each of them has been specified as of Poisson type. The relative cost of a returned versus an unsold copy is assumed $P_D/P_A = 1/5$, as it is done in the example newspaper. For each of the seven outlets, the proportion for each day during the period forecasted are shown on the following table

	PROPORTIONS				
Outlet	*Monday*	*Tuesday*	*Wednesday*	*Thursday*	*Friday*
Week 1					
1	0.214	0.215	0.215	0.218	0.240
2	0.214	0.204	0.215	0.218	0.200
3	0.111	0.108	0.118	0.109	0.120
4	0.051	0.075	0.075	0.069	0.070
5	0.068	0.086	0.075	0.079	0.080
6	0.205	0.183	0.183	0.188	0.180
7	0.137	0.129	0.118	0.119	0.110

Week 2					
1	0.210	0.220	0.217	0.216	0.240
2	0.210	0.209	0.217	0.216	0.200
3	0.109	0.110	0.120	0.108	0.120
4	0.050	0.066	0.076	0.069	0.070
5	0.076	0.088	0.076	0.078	0.080
6	0.202	0.187	0.185	0.186	0.180
7	0.143	0.121	0.109	0.127	0.110
Week 3					
1	0.248	0.217	0.213	0.218	0.238
2	0.248	0.207	0.213	0.218	0.198
3	0.119	0.109	0.117	0.109	0.119
4	0.059	0.076	0.074	0.069	0.069
5	0.089	0.087	0.085	0.079	0.089
6	0.238	0.185	0.181	0.178	0.178
7		0.120	0.117	0.129	0.109

Table 3: Forecasted proportions for the distribution process for a set of 7 outlets.

The real data of sales and returned papers for this period are available and can be used to asses the improvements resulting of applying the cost minimization process to a group of similar or nearby selling points. If sales are less than supply, it can be evaluated if the proposed distribution improves, with less returned copies, and if there are problems of undersupply. For sales equal to the quantity supplied, an improvement could derive from an increase in the supply, compared with the quantity decided by the distributor using his own methodology of the distribution. In the following table, it can be seen that the proposed method decreases the returned papers, increased sales and the number of outlets that run out of supply.

	Actual method			Proposed method		
	% of unsold units	No of outlets sold out, with smaller supply	No of outlets with smaller but sufficient supply	% of unsold units	No of outlets sold out, with smaller supply	No of outlets with a smaller but sufficient supply
Week 1	24.78	5	8	21.78	5	8
Week 2	24.80	3	5	21.68	5	12
Week 3	26.16	2	12	29.38	3	13

Table 4. Improvements in the distribution process

As it can be deduced from the previous table, the distribution process proposed produces better results than the originally used by the company in charge, as it estimated more precisely the supply to each outlet, reducing the amount of papers returned, without causing an shortage in any of the points considered. The lost sales for insufficient supply can not be obtained, as there is no information of this data, but it can be estimated using the methodology proposed by Caridad and Rodríguez (2004). In this case, the distribution process can be optimised taking into account a more precise variable that the proportion, a, of outlets that do not return copies (this could happens when the demand is equal to the supply, or if it exceeds it).

Final Comments

To summarize the information about the efficiency of press distribution in a network of selling points, two easy to obtain type of data are used: the number of unsold copies (that are returned to the distributor) and the proportion of outlets that run out of papers, in most cases, because the demand exceeds the supply. These two variables could be used to define a Cobb-Douglas type efficiency function, $y = f(a, d)$. A predefined constant value for this non observable measure,

originates the relations between *a* and *d*. In the short term these relations remain fairly stable, and can be related with average costs associated to the two conflicting situations: excess of demand or supply to a particular outlet. The efficiency function, *E*, is obtained from this function, $a = a(d)$. With an easy to apply cost minimisation process, objective values for these two variables, can be deduced, so the total print run is obtained from forecasted sales and from the unsold copies objective. A correction based on average sales in two time periods is considered to compare efficiency measures at two time intervals. Forecasting sales can be done by classical time series techniques (with possible interventions, to take into account known future events. These can be done for the whole network, or for groups of selling points, selected with geographical criteria, of being with homogenous demand. Again, the publisher policy has to be introduced, including the costs associated to an unsold copy or an unattended demand for one additional paper. The group of selling points supply is aggregated, and corrected with the global sales forecast, and then, the proportion of the print run to be send to each outlet is estimated.

Applying this process to part of a network of a sport newspaper in Madrid area, some improvements can be detected, as the proposed method obtain a lightly smaller proportion of unsold copies (24.32%) than the methodology used in this company (24.95%), when the objective aimed by the distributor was 23.4%. On the other side, the proportion of outlets that run out of papers decreases to less than 12%, against the 17.31% that was the real figure obtained. Also, in the selling points with enough supply, with the proposed method, this supply that would be deduced would be smaller, decreasing cost of unsold copies, and only in 11.54% of these outlets, the supply would be insufficient. This approach is been implemented in a network of more than two thousand selling points, that have to be clustered in homogenous demand groups.

References

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

Caridad y Ocerin, J.M. (2002). *Hybrid and classical forecasting methods: the demand of water in urban areas*. MME'02. pp. 33-45. Praga.

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

Caridad y Ocerin J. M.; Rodríguez Aragón F. J.; Ceular Villamandos N. (2004). *Demand estimation in press distribution*. MME'04. .42-55 Brno.

Cebrián J. L. (2002) *Fuertes en medio de la crisis*. Memoria Grupo Prisa 2002

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

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

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

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

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

Rodríguez Aragón, F. J. y Caridad y Ocerin, J. M. (2004[b]). *Demanda Insatisfecha en una Red de puntos de venta de prensa*. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba.

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

Rodríguez Aragón, F. J. y Caridad y Ocerin, J. M. (2004[d]). *Equilibrio de mercado en un punto de venta con demanda Normal*. Documento de Trabajo: D. de Estadística, Econometría, I. O. y Empresa. Universidad de Córdoba.

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

Moment problem and worst-case Value-at-Risk

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

Abstract

This paper deals with the problem of moments and its application to the calculation of the worst-case Value-at-Risk. To this purpose we exploit upper bounds for loss probability of univariate random variable with special properties, given expected value and variance. Subsequently, we suppose that except the first two moments of the distribution, we know further characteristics of the class of distributions. We assume symmetry and/or unimodality. The bounds are illustrated on the case of interbank exchange rate.

Keywords

exchange rate, problem of moments, worst-case Value-at-Risk

Introduction

This paper deals with a problem of stochastic programming under incomplete information on a probability distribution of random parameters. The basic choice of the set of feasible distributions $\mathcal{P}$ corresponds to specification of the distribution by its first and second order moments. Problem of moments is already discussed in [1] including results for unimodal ditributions of random variables.

Our main goal is to compare values of worst-case Value-at-Risk under different type of information on distribution of random parameters. We start with distribution determined by its first two moments and step by step we also suppose symmetry and/or unimodality. The framework for this article is paper [3], where the upper bounds for probability of loss were evolved in case of symmetry and unimodality. We also hold onto [5], where the essential results in duality are derived.

In the first section we define the problem of moments and introduce some known results. Then we interpret the definiton of worst-case VaR. In the following parts we present some known bounds for arbitrary distributions and in successive steps we show improvements of these bounds by adding more assumptions. We also correct the proof of lemma 2.3, aforesaid in [3]. Our main results are in the worst-case VaR formulas. An important advantage of these outcomes is the easy way of computation.

We illustrate the results obtained by applications of the discussed bounds on a currency market. We choose last year data and study increments in the interbank exchange rate between NOK and SEK.

1 Problem formulation

Consider a nonempty set $\Omega \subseteq \mathbb{R}$ and suppose that every finite subset of Ω is $\mathcal{F}$-measurable, where $\mathcal{F}$ is the Borel σ-algebra of Ω. For a given vector of functions $f = (f_0, \dots, f_n) : \Omega \to \mathbb{R}$ and a sequence $q = (q_0, \dots, q_n)$ we define the moment problem as

$$(P) \quad \sup_{\mu \in \mathcal{P}} E_\mu[\Phi(\omega)] \qquad \text{s.t.} \quad E_\mu[f(\omega)] = q. \tag{1}$$

[1] This work was supported by the Grant Agency of the Czech Republic (grant 201/05/H007).

Real valued functions $\Phi, f_i, i = 0, \ldots, n$, are assumed to be $\mathcal{F}$-measurable and μ-integrable, where $\mu : \mathcal{F} \to [0, 1]$ is a nonnegative measure on $(\Omega, \mathcal{F})$. We explicitly include the probability mass constraint by setting $q_0 \equiv 1, f_0 \equiv \mathbb{I}_\Omega$, where $\mathbb{I}$ is the indicator function. The set of all feasible probability distributions on $(\Omega, \mathcal{F})$ is denoted by $\mathcal{P}$ which is supposed to be a convex set.

The dual of the problem (1) can be written as

$$\inf_{y \in \mathbb{R}^{n+1}} y'q \qquad \text{s.t.} \quad E_\mu[y'f(\omega) - \Phi(\omega)] \geq 0 \quad \forall \mu \in \mathcal{P}. \tag{2}$$

The weak duality $val(D) \geq val(P)$ always holds. To fulfil the strong duality we need some additional assumtions, e.g. Slater condition.

Actually, it suffices to solve the problem (1) with respect to discrete probability distributions with a finite support on at most $n + 1$ points. First results of this type date to the mid 1950. For the proof see e.g. [5], lemma 3.1.

If we designate $\mathcal{T}_\delta$ as the set of all Dirac distributions $\delta_\omega, \omega \in \Omega$, then we can also write

$$\mathcal{P} = \text{cl}(\text{conv}(\mathcal{T}_\delta)). \tag{3}$$

Due to this outcome and representation (3) we can rewrite the dual problem (2) in following way

$$\inf_{y \in \mathbb{R}^{n+1}} y'q \qquad \text{s.t.} \quad y'f(\omega) - \Phi(\omega) \geq 0 \quad \forall \omega \in \Omega, \tag{4}$$

as the expected values from (2) are related now to Dirac distributions $\delta_\omega, \omega \in \Omega$.

1.1 Worst-case VaR

We define by $VaR_\alpha(X)$ the minimal level γ such that the probability the random parameter X achieves or exceeds γ is not greater than given $\alpha \in (0, 1]$. Then we solve the problem

$$VaR_\alpha(X) := \min_{\gamma \in \mathbb{R}} \gamma \qquad \text{s.t.} \quad P(X \geq \gamma) \leq \alpha. \tag{5}$$

When the distribution of X is perfectly known using its distribution function F_X we obtain the optimal value $\gamma^* = F_X^{-1}(1 - \alpha)$, where F_X^{-1} is the inverse of F_X, i.e. $F_X^{-1}(\alpha) := \inf\{x : F_X(x) \geq \alpha\}$. Specially if X is normally distributed with expected value μ_X and variance σ_X^2 we get $\gamma_N^* = \mu_X + \Phi^{-1}(1 - \alpha) \cdot \sigma_X$, where Φ is distribution function of $N(0, 1)$.

In most cases we do not exactly know the distribution of random variable X. We are only able to specify the set $\mathcal{P}_0 \subseteq \mathcal{P}$ of feasible probability distribution, e.g. $\mathcal{P}_0$ is the set of all distribution that fulfil moment conditions, or the set of symmetric or unimodal distributions. Then we apply the min-max strategy and define the worst-case VaR with respect to the set of probability distributions $\mathcal{P}_0$ as

$$VaR_\alpha^{wc}(X) := \min_{\gamma \in \mathbb{R}} \gamma \qquad \text{s.t.} \quad \sup_{\mu \in \mathcal{P}_0} P_\mu(X \geq \gamma) \leq \alpha. \tag{6}$$

2 Moment bounds

In this section we present upper bounds for loss probability of random variable X defined on $(\Omega, \mathcal{F})$. We assume that the only available information on the distribution is the knowledge of the first two moments – expected value and variance. In successive steps we add more assumptions such as symmetry and unimodality.

Consider the moment problem (1) and set $\Phi(x) := \mathbb{I}_{[X \geq \gamma]}$. We obtain the modification of inner problem (6)

$$\sup_{\mu \in \mathcal{P}} P_\mu(X \geq \gamma) \qquad \text{s.t.} \quad E_\mu[f(\omega)] = q, \tag{7}$$

where $\mathcal{P}, f$ and q are the same as above.

Hence, we are concerned with the following problem

$$\begin{aligned} \sup_{\mu\in\mathcal{P}} \quad & P_\mu(X \geq \gamma) \qquad \text{s.t.} \quad E_\mu \mathbb{I}_\Omega = 1, \\ & \qquad\qquad\quad E_\mu[X] = \mu_X, \\ & \qquad\qquad\quad E_\mu[X-\mu_X]^2 = \sigma_X^2, \end{aligned} \tag{8}$$

based on the knowledge of the first two moments.

Now we introduce some known bounds considering the class of probability distributions defined by a given expected value μ_X and variance σ_X^2.

Lemma 2.1 (one-sided Chebyshev bound)**.** *Consider the moment problem (8) with the set of feasible probability distribution $\mathcal{P}$ identified by μ_X and σ_X^2. Then the upper bound on the upper tail $P(X \geq \gamma)$ is given by*

$$\sup_{\mu\in\mathcal{P}} P_\mu(X \geq \gamma) = \begin{cases} \frac{\sigma_X^2}{\sigma_X^2+(\mu_X-\gamma)^2} & \text{for } \gamma > \mu_X, \\ 1 & \text{for } \gamma \leq \mu_X. \end{cases} \tag{9}$$

This result follows from [2].

2.1 Bounds for symmetric distributions

In this section we focus on the possibilities of improving the upper bound of the problem (8) by adding supplemental information on symmetry of the distribution of X.

Definition 2.1. Let $\Omega = I \subseteq \mathbb{R}$ be either a compact interval, or $I = \mathbb{R}$. We say that a distribution μ defined on $(\Omega, \mathcal{F})$ is M-symmetric if $\mu[M-x, M] = \mu[M, M+x]\ \forall x \in I_M$, where $I_M := \{x \geq 0 : M - x \in I \text{ and } M + x \in I\}$.

Let $\mathcal{P}_M^s$ denote the set of all M-symmetric probability distributions. Then it is easy to show that $\mathcal{P}_M^s$ is convex and closed. Due to (3), we can write $\mathcal{P}_M^s = \mathrm{cl}(\mathrm{conv}(\mathcal{T}_M^s))$, where $\mathcal{T}_M^s = \{\mu_x = \frac{1}{2}\delta_{M+x} + \frac{1}{2}\delta_{M-x}, x \in I_M\}$ is the set of M-symmetric Dirac distributions.

We can rewrite the dual problem (2) for M-symmetric distributions in the following way

$$\inf_{y\in\mathbb{R}^{n+1}} y'q \qquad \text{s.t.} \quad y'[f(M-x)+f(M+x)] - [\Phi(M-x)+\Phi(M+x)] \geq 0\ \forall x \in I_M. \tag{10}$$

We calculate here the expected values with respect to the distributions from the set $\mathcal{T}_M^s$.

If the bound obtained by solving (D^s) is achievable, then there exists an optimal distribution which is a convex combination of $n+1$ M-symmetric Dirac distributions. It holds under Slater condition, see [3], [5].

In our case, i.e. (8), M coincides with the expected value μ_X and $I = \mathbb{R}$.

Lemma 2.2. *Consider the problem (8) with $\mathcal{P} \equiv \mathcal{P}_{\mu_X}^s$. Then the upper bound on the upper tail $P(X \geq \gamma)$ identified by μ_X and σ_X^2 is given by*

$$\sup_{\mu\in\mathcal{P}} P_\mu(X \geq \gamma) = \begin{cases} \frac{1}{2}\min\{1, \frac{\sigma_X^2}{(\mu_X-\gamma)^2}\} & \text{for } \gamma > \mu_X, \\ 1 & \text{for } \gamma \leq \mu_X. \end{cases} \tag{11}$$

The proof can be found in [3].

2.2 Bounds for symmetric unimodal distributions

In the following section we add the assumption of unimodality of probability distribution of random variable X.

Definition 2.2. Let $\Omega = I \subseteq \mathbb{R}$ be either a compact interval, or $I = \mathbb{R}$. Distribution μ defined on $(\Omega, \mathcal{F})$ is said to be m-unimodal on $I \ni m$ if the corresponding distribution function is convex on the left of m and concave on the right of m.

Let $\mathcal{P}_m^u$ denote the set of continuous m-unimodal probability distributions. Then $\mathcal{P}_m^u$ is convex and $\mathrm{cl}(\mathcal{P}_m^u)$ is the set of all m-unimodal distributions.

For univariate real random variables we can find continuous transformation from the set of all probability distributions $\mathcal{P}$ to the set of all unimodal distributions $\mathrm{cl}(\mathcal{P}_m^u)$. This result has been presented e.g. in [1]. In consequence, we can write $\mathrm{cl}(\mathcal{P}_m^u) = \mathrm{cl}(\mathrm{conv}(\mathcal{T}_m^u))$, where $\mathcal{T}_m^u = \{\delta_{[x,m]} : x \in I, x \neq m\}$. By $\delta_{[a,b]}$ we denote probability distribution with uniform density on $[a, b]$. For other properties of unimodal distributions, see [6].

We obtain the convex set of all M-symmetric unimodal distributions as $\mathrm{cl}(\mathcal{P}_M^{su}) = \mathcal{P}_M^s \cap \mathrm{cl}(\mathcal{P}_M^u) = \mathrm{cl}(\mathrm{conv}(\mathcal{T}_M^{su}))$, where $\mathcal{T}_M^{su} = \{\delta_{[M-x,M+x]} : x \in I_M, x \neq 0\}$. For details see [3].

The corresponding dual problem (2) is

$$\begin{aligned} \inf_{y\in\mathbb{R}^{n+1}} \quad y'q \qquad \text{s.t.} \quad & y' \int_{M-x}^{M+x} f(z)\,\mathrm{d}z - \int_{M-x}^{M+x} \Phi(z)\,\mathrm{d}z \geq 0 \quad \forall x \in I_M, x \neq 0, \\ & y'f(M) - \Phi(M) \geq 0. \end{aligned} \tag{12}$$

The last condition arises from adding δ_M to the generating set $\mathcal{T}_M^{su}$ in order to obtain closure.

By analogy with symmetric case, the optimal distribution exists if the bound is achievable. Then it is a convex combination of $n+1$ M-symmetric unimodal distributions, possibly including a Dirac distribution at M.

In our case, i.e. (8), we have $M = \mu_X$ and $I = \mathbb{R}$.

Lemma 2.3. *Consider the problem (8) with $\mathcal{P} \equiv \mathcal{P}_{\mu_X}^{su}$. Then the upper bound on the upper tail $P(X \geq \gamma)$ identified by μ_X and σ_X^2 is given by*

$$\sup_{\mu\in\mathcal{P}} P_\mu(X \geq \gamma) = \begin{cases} \frac{1}{2}\min\{1, \frac{4}{9}\frac{\sigma_X^2}{(\mu_X-\gamma)^2}\} & \text{for } \gamma > \mu_X, \\ 1 & \text{for } \gamma \leq \mu_X. \end{cases} \tag{13}$$

Proof. This is a corrected proof from [3]. We substitute $q = (1, \mu_X, \sigma_X^2)$, $\Phi(x) = \mathbb{I}_{[x\geq\gamma]}$, $f_0(x) = \mathbb{I}_{[x\in I]}$, $f_1(x) = x$ and $f_2(x) = [x - \mu_X]^2$ to (12) and obtain the problem

$$\begin{aligned} \min_{y\in\mathbb{R}^3} \quad y_0 + \mu_X y_1 + \sigma_X^2 y_2 \qquad \text{s.t.} \quad & 2xy_0 + 2x\mu_X y_1 + \tfrac{2}{3}x^3 y_2 \geq \int_{\mu_X-x}^{\mu_X+x} \mathbb{I}_{[z\geq\gamma]}\,\mathrm{d}z \quad \forall x > 0, \\ & y_0 + \mu_X y_1 \geq \mathbb{I}_{[\mu_X\geq\gamma]}. \end{aligned} \tag{14}$$

We may assume, without loss of generality, that $\mu_x = 0$. We distinguish two cases:
$\gamma \leq 0$: We solve following problem

$$\begin{aligned} \min_{y\in\mathbb{R}^2} \quad y_0 + \sigma_X^2 y_2 \qquad \text{s.t.} \quad & y_0 + \tfrac{1}{3}x^2 y_2 \geq 1 \qquad\qquad 0 < x \leq -\gamma, \\ & 2xy_0 + \tfrac{2}{3}x^3 y_2 \geq x - \gamma \quad x > -\gamma, \\ & y_0 \geq 1. \end{aligned} \tag{15}$$

The first condition implies $y_2 \geq 0$. The optimum is achieved for $y_0 = 1$, $y_2 = 0$. The bound is equal to 1.
$\gamma > 0$: We reformulate the problem (14) as

$$\begin{aligned} \min_{y\in\mathbb{R}^2} \quad y_0 + \sigma_X^2 y_2 \qquad \text{s.t.} \quad & xy_0 + \tfrac{1}{3}x^3 y_2 \geq 0 \qquad\qquad 0 < x < \gamma, \\ & x(2y_0 - 1) + \tfrac{2}{3}x^3 y_2 \geq -\gamma \quad x \geq \gamma, \\ & y_0 \geq 0. \end{aligned} \tag{16}$$

The first constraint implies $y_2 \geq 0$. For $y_0 \geq \frac{1}{2}$, or $y_2 = 0$, the optimum is achieved for $y_0 = \frac{1}{2}$, $y_2 = 0$, and the bound is $\frac{1}{2}$.

In the case when $0 \leq y_0 < \frac{1}{2}$ and $y_2 > 0$ the second condition holds for $x \geq \gamma$ iff $g(\bar{x}) \geq 0$, where $\bar{x} = \sqrt{\frac{1-2y_0}{2y_2}}$ is the non-negative local minimum of the function $g(x) = \gamma + x(2y_0 - 1) +$

$\frac{2}{3}x^3y_2$. The requirement $g(\bar{x}) \geq 0$ is equivalent to the constraint $y_2 \geq \frac{2}{9}\frac{(1-2y_0)^3}{\gamma^2}$. The minimum is achieved at $y_0 = 0$ and $y_2 = \frac{2}{9\gamma^2}$, the optimal value is $\frac{2}{9}\frac{\sigma_X^2}{\gamma^2}$. □

Now we can handle the univariate worst-case VaR problem (6) for general, symmetric and/or unimodal distributions with known expected value and variance. The optimal values are solutions of the problems

$$\begin{aligned}
\gamma^* &:= \min_{\gamma > \mu_X} \gamma \qquad \text{s.t.} \quad \frac{\sigma_X^2}{\sigma_X^2 + (\mu_X - \gamma)^2} \leq \alpha, \\
\gamma_s^* &:= \min_{\gamma > \mu_X} \gamma \qquad \text{s.t.} \quad \frac{1}{2}\min\{1, \frac{\sigma_X^2}{(\mu_X - \gamma)^2}\} \leq \alpha, \\
\gamma_{su}^* &:= \min_{\gamma > \mu_X} \gamma \qquad \text{s.t.} \quad \frac{1}{2}\min\{1, \frac{4}{9}\frac{\sigma_X^2}{(\mu_X - \gamma)^2}\} \leq \alpha.
\end{aligned}$$

For $\alpha < \frac{1}{2}$ the minima are $\gamma^* = \mu_X + \sqrt{\frac{1-\alpha}{\alpha}}\sigma_X$, $\gamma_s^* = \mu_X + \frac{1}{\sqrt{2\alpha}}\sigma_X$ and $\gamma_{su}^* = \mu_X + \sqrt{\frac{2}{9\alpha}}\sigma_X$. If $\alpha \geq \frac{1}{2}$, the minima γ_s^* and γ_{su}^* are not achieved. There only exist infima and their values are μ_X. If $\gamma \leq \mu_X$ the solutions exist only for $\alpha = 1$.

In a similar way we may formulate the worst-case VaR for unimodal *non-symmetric* distributions.

3 Numerical results

This numerical study compares the worst-case VaR for arbitrary, normal, symmetric and symmetric unimodal distributions with given first two moments.

The assumption of normality is frequently used in practice. Therefore, we measure the differences between worst-case level if we relax a supposition of the normal distribution. The worst-case VaR value provides the maximal error, which can be caused by holding on to concrete probability distribution. The error is evaluated as the difference between the worst-case VaR and the VaR value provided by a specific distribution. Results for a random variable X with expected value $\mu_X = 0$ and variance $\sigma_X^2 = 1$ are summarized in Table 1.

α	γ^*	γ_s^*	γ_{su}^*	γ_N^*	$\gamma_{t_3}^*$	$\gamma_{t_4}^*$	γ_{log}^*
0.01	9.9499	7.0711	4.7140	2.3263	2.6315	2.6495	2.5334
0.02	7.0000	5.0000	3.3333	2.0537	2.0103	2.1203	2.1458
0.03	5.6862	4.0825	2.7217	1.8808	1.7035	1.8390	1.9165
0.04	4.8990	3.5355	2.3570	1.7507	1.5042	1.6496	1.7522
0.05	4.3589	3.1623	2.1082	1.6449	1.3587	1.5074	1.6234
0.1	3.0000	2.2361	1.4907	1.2816	0.9455	1.0841	1.2114

Table 1: Comparison of worst-case VaR for a random variable X with $\mu_X = 0$ and $\sigma_X^2 = 1$, where $\gamma_{t_3}^*$ and $\gamma_{t_4}^*$ are VaR for random variable with modified Student distribution, γ_{log}^* is computed for logistic distribution.

In the next example we study interbank exchange rate increments between Norwegian Kroner NOK and Swedish Krona SEK during the time period 1.1.2004 – 31.12.2004. Our data represent increments of daily closing values to insure data independence. We have chosen this two currencies because of their long-standing stability. Hence, there is a perspective to fulfil the assumptions of symmetry, respectively unimodality.

We apply the last 30 observations to estimate expected value and variance for a calculation of the next worst-case level. The results are presented in Figure 1. Real values of increments achieve or exceed the 5% worst-case value γ_{su}^* only in 3%, whereas the level γ_N^* in 6%. A possible improvement of these bounds is in using robust estimations of expected value and variance.

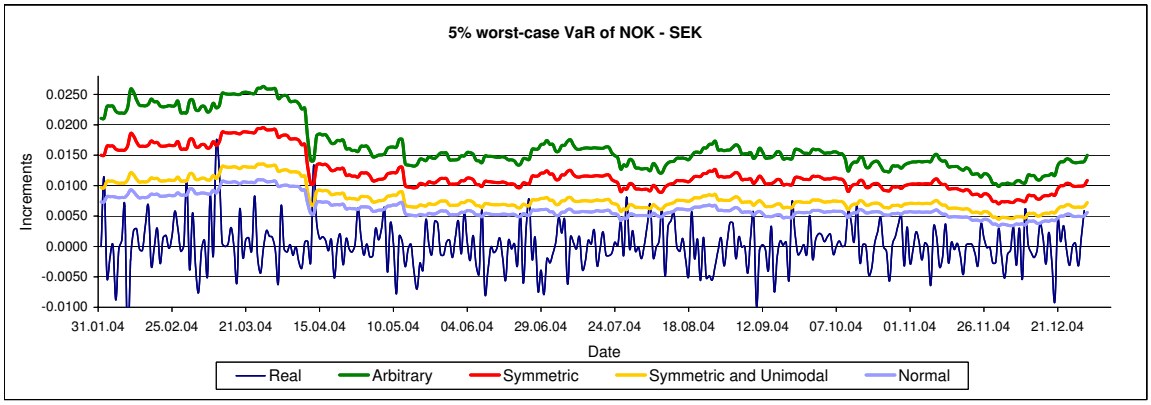


Figure 1: Worst-case VaR for increments of NOK - SEK interbank exchange rate.

The numerical experiments and tests of symmetry, unimodality and normality have been computed in R 1.7.0 using packages base and diptest, see [4], and in Microsoft Excel 2000 under Windows 2000.

Conclusion

In this paper we showed how the worst-case Value-at-Risk can be easily computed under several assumptions about the convex classes of distributions of random parameter. We always supposed knowledge of the first two moments - expected value and variance. Upper bounds for probability of loss were derived by using the dual formulation of problem (1) together with special properties such as symmetry and unimodality. We also presented the corrected version of the proof of lemma 2.3. Consequently, we exploited these bounds for development of the worst-case VaR levels. Finally, we applied these results to an interbank exchange rate. We compared these outcomes with Value-at-Risk computed for normally distributed random variable. The confrontation illustrates the fact that a lower information level concerning the class of distributions causes an increase of gap between worst-case VaR levels. It provides also the maximal error obtain by calculation under a specific distribution assumption.

References

[1] J. Dupačová: *Minimaxová úloha stochastického lineárního programování a momentový problém*, Ekonomicko matematický obzor, 1977, vol. 13, no. 3, pp. 279-307.

[2] K. Isii: *On the sharpness of Chebyshev-type inequalities*, Ann. Inst. Stat. Math., 1963, vol. 14, pp. 185-197.

[3] I. Popescu: *A semidefinite programming approach to optimal moment bounds for convex classes of distribution*, forthcoming in Mathematics of Operations Research, 2005.

[4] R Development Core Team: *R: A language and environment for statistical computing*, R Found. for Stat. Comp., Vienna, Austria, 2003, URL http://www.R-project.org.

[5] A. Shapiro: *On duality theory of conic linear problems*, Semi-infinite programming: recent advances, M.A. Goberna and M.A. Lopez, ed., Kluwer Acad. Publ., 2001, pp. 165-165.

[6] A. Shapiro: *Worst-case distribution analysis of stochastic programs*, publ. el. in Optim. Online, 2004, URL http://www.optimization-online.org/DB_HTML/2004/08/930.html.

Stochastic Dominance in the Choice of Optimal Ratios of the Non-Life Insurance Premiums

Bogdan Ciupek
Department of Operations Research, University of Economics, Katowice, Poland

Abstract

The paper suggests a method of a decision support process while setting optimal combinations of insurance premium rates by non-life insurance companies. This method allows determining optimal combination of gross insurance premium rates for any amount of insurance forms. Such an optimal combination will constitute the base for setting more detailed insurance premium tariffs for particular kinds of insurance.

Keywords

non-life insurance, insurance tariffs, decision making, stochastic dominance, mathematical modelling

Introduction

Basic function of insurance is to provide an insurance buyer with a real and reliable cover. Simultaneously, each insurance company realises its own individual goals that depend on the form and scope of its activities. In Poland most of insurance companies run their operations as joint stock companies. That influences greatly goals mentioned above. It is the maximisation of the profit originated from insurance activities that becomes the most important goal of an insurance company operating as a joint stock company. Moreover, each company aims at increasing its market share, reaching proper portfolio structure, reducing costs of insurance activities and diminishing costs of passive reinsurance. These goals are realised more or less efficiently. Additionally, there are significant differences between life insurance companies and non-life ones.

1. Formulation of the Problem and a Mathematical Model

Possibility of applying stochastic dominance to support the choice of optimal and conforming with the criteria accepted by an insurance company combination of insurance premiums is illustrated by the example. It has been assumed that the premiums have been calculated by means of actuarial methods and the approach suggested aims at showing an insurance company a combination of premium rates that realise predetermined goals to the maximum. It has been assumed that an insurance company wants to obtain the best values of the following ratios characterising correctness and security of its activities: productivity ratio, gross loss ratio, and retention ratio.
Productivity ratio is calculated according to the following formula:

$$\mathrm{PR} = \frac{\mathrm{TR}}{\mathrm{PED}} \cdot 100\% \tag{9}$$

TR – technical result: is a result obtained from the basic insurance activity. It constitutes a difference between due revenues originating from premiums and paid out benefits and changes of technical insurance reserves that are to cover potential indemnities and benefits including reinsurers' participation in the revenues from premiums, paid out indemnities and benefits, and costs of insurance activity.

PED – premium earned on the deductible: is a gross premium income from insurance contracts due in the reporting period decreased by reinsurers' participation in a premium and adjusted by changes of reserves for gross unexpired risk and reinsurers' participation in the change of reserves.
Gross loss ratio is calculated according to the following formula:

$$GLR = \frac{GIB + CR}{GPE} \cdot 100\% \qquad (10)$$

GIB – gross indemnities and benefits paid out: indemnities and benefits paid out by an insurance company in the reporting period on the basis of concluded insurance contracts.
CR – change in reserves for unpaid indemnities and benefits: increasing or decreasing of the amount of technical insurance reserves on the basis of unpaid indemnities and benefits.
GPE – gross premium earned: premium income in the reporting period adjusted by changes in premium reserves before taking reinsurers' participation into account.
Retention ratio is presented by the following formula:

$$RR = \frac{PID}{GPI} \cdot 100\% \qquad (11)$$

PID – premium income on the deductible: costs due in the reporting period because of concluded insurance contracts, regardless of the fact whether the amounts in question may refer to the whole or part of the following reporting period taking reinsurers' participation into account.
GPI – gross premium income: covers amounts due in the reporting period because of concluded insurance contracts, regardless of the fact whether the amounts in question may refer to the whole or part of the following reporting period taking reinsurers' participation into account.
The four-stage procedure is suggested in order to make the best choice of premium tariff sets:

- Stage I: determining a set of alternative premium tariff sets,
- Stage II: making distributions of alternatives evaluations towards attributes,
- Stage III: determining types of stochastic dominance applying among available alternatives towards each attribute,
- Stage IV: determining preferences on the set of alternatives.

In order to formulate a problem of an optimal choice of premium tariffs the following author's values of net premium rates for four kinds of insurance in seven variants (%) are assumed (in the problem being discussed third party insurance has been omitted because of its insignificant share in the portfolio of a discussed insurance company).
Sources (tables and graphs): author's data and calculations.

Table 1. Gross premium rates

No	**Life**	**MAT**	**Transport**	**Non-life**
1	6	2	12	2
2	4	3	8	2
3	3	4	10	2
4	2	2	10	1
5	3	3	12	1
6	2	1	6	3
7	5	4	8	1

For such premium rates the following author's values of all three insurance ratios a decision maker is interested in: productivity, gross loss and retention have been assumed.

Table 2. Values of technical insurance ratios for the Variant I of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	35	10	15	7
Loss	40	60	70	70
Retention	95	75	90	80

Table 3. Values of technical insurance ratios for the Variant II of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	25	15	-5	7
Loss	60	50	100	70
Retention	80	70	85	70

Table 4. Values of technical insurance ratios for the Variant III of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	15	30	10	10
Loss	70	30	80	70
Retention	70	90	85	70

Table 5. Values of technical insurance ratios for the Variant IV of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	10	10	10	0
Loss	80	60	80	100
Retention	90	75	85	80

Table 6. Values of technical insurance ratios for the Variant V of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	15	15	15	5
Loss	70	50	70	100
Retention	70	75	90	95

Table 7. Values of technical insurance ratios for the Variant VI of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	10	5	-10	15
Loss	80	80	120	60
Retention	90	100	100	95

Table 8. Values of technical insurance ratios for the Variant VII of gross premium rates

Criterion	Life	MAT	Transport	Non-life
Productivity	30	30	-5	0
Loss	50	30	100	100
Retention	90	90	85	95

2. Outcome and Interpretation of Solutions

On the basis of definitions determined by formulas 2-8 stochastic dominance among particular variants of net premium rates for the selected classes of insurance has been calculated. In order to evaluate stochastic dominance a random variable constructed on the basis of the values of ratios in question for particular kinds of insurance in the total premium income has been used. In the discussed example, for the matter of simplicity, the same shares in the premium income for all kinds of insurance have been assumed. As a result the distributions of stochastic dominance presented in tables 9-11 have been obtained:

Table 9. Stochastic dominance distribution for productivity ratio

Dominance	a1	a2	a3	a4	a5	a6	a7
a1	-	FSD	SISD	FSD	TSD	FSD	SSD
a2		-		SISD	TISD2	FSD	TSD
a3	TSD	FSD	-	FSD	SSD	FSD	SSD
a4		TSD		-		SSD	TSD
a5		SSD		FSD	-	FSD	TSD
a6						-	
a7		SISD	TISD2	SISD	SISD	SSD	-

Where:

- FSD denotes first grade stochastic dominance,
- SSD denotes second grade stochastic dominance,
- TSD denotes third grade stochastic dominance
- SISID denotes inverse second grade stochastic dominance
- TISD2 denotes inverse third grade stochastic dominance
-

Table 10. Distribution of stochastic dominance for gross loss ratio

Dominance	b1	b2	b3	b4	b5	b6	b7
b1	-	FSD	SSD	FSD	FSD	FSD	SSD
b2		-		FSD	FSD	FSD	SSD
b3	TISD2	SISD	-	FSD	FSD	FSD	SSD
b4				-		FSD	
b5				FSD	-	FSD	TSD
b6						-	
b7	TISD2	SISD	TISD2	SISD	SISD	SISD	-

Table 11. Distribution of stochastic dominance for retention ratio

Dominance	c1	c2	c3	c4	c5	c6	c7
c1	-	FSD	FSD	FSD	FSD		
c2		-					
c3		FSD	-				
c4		FSD	FSD	-	SSD		
c5		FSD	FSD	SISD	-		
c6	FSD	FSD	FSD	FSD	FSD	-	FSD
c7	FSD	FSD	FSD	FSD	FSD		-

Since an insurer is interested in the optimisation of premium rates taking into account all three criteria simultaneously, the following procedure to chose optimal combination has been applied:

1. For each criterion weight *gk* is determined; the sum of all weights for all criteria equals 1,
2. For each pair of alternatives a_i, a_j index of concordance is determined in the following way:

$$i(a_i, a_j) = \sum_{k=1}^{n} g_k d_k(a_i, a_j) \tag{12}$$

possible values of levels of concordance are determined according to where:

$d(a_i, a_j) = 1$ if alternative a_i dominates alternative a_j over criterion k in one of stochastic dominance

$d(a_i, a_j) = 0$ if there is no any stochastic dominance between each pair of alternative a_i and a_j over criterion k.

It is possible to introduce weights for interval [0; 1] for each kind of dominance. Diversification of weights allows to include additional information on the kind of stochastic dominance between alternatives in the analysis. In case of loss ratio criterion while calculating index of concordance only inverse stochastic dominance have been taken into consideration because an insurance company aims at minimising of this index. It is assumed that its utility function is a function of an INARA type.

Next, the level of concordance is fixed from which dominance of alternative a_i over alternative a_j from the point of view of criteria determined by a decision maker will be accepted. All the following formula where: $z_k = 1$ or $z_k = 0$.

$$p = \sum_{k=1}^{n} z_k g_k \qquad (13)$$

Level of concordance p determined by a decision maker may be included in the interval [0.5; 1] and it determines the number of attributes with which levels of concordance are set.

3. Using values obtained a matrix of preferences is created. It is the basis for the choice of optimal for an insurance company alternative as far as predetermined criteria are concerned.

In order to solve the problem of finding optimal combination of premium rates the following weights for particular criteria have been assumed: profitability: 0.5; loss: 0.3; retention: 0.2. These weights reflect insurance company preferences towards particular criteria.

Taking above assumptions into consideration the following matrix of concordance levels has been obtained:

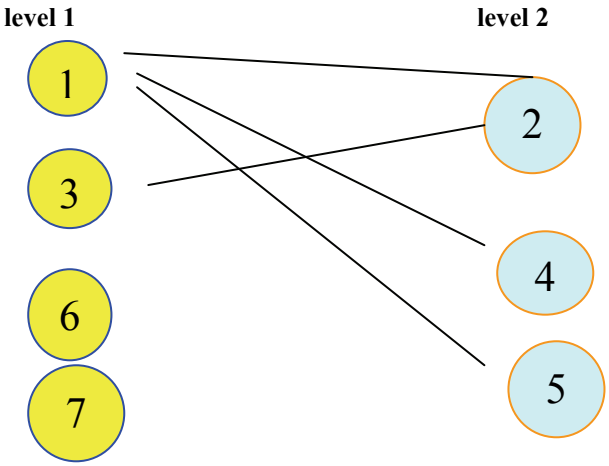
Table 12. Matrix of concordance levels

alternative	a1	a2	a3	a4	a5	a6	a7
a1	-	1.0	0.2	1.0	1.0	0.8	0.5
a2	0.0	-	0.0	0.3	0.3	0.8	0.5
a3	0.8	1.0	-	0.8	0.8	0.8	0.5
a4	0.0	0.7	0.2	-	0.2	0.8	0.5
a5	0.0	0.7	0.2	0.8	-	0.8	0.5
a6	0.2	0.2	0.2	0.2	0.2	-	0.2
a7	0.5	0.5	0.5	0.5	0.5	0.8	-

For p = 1 a graph has been created. It illustrates multiatribute stochastic dominance (see Diagram 7.1). The graph has been created applying the following principles:

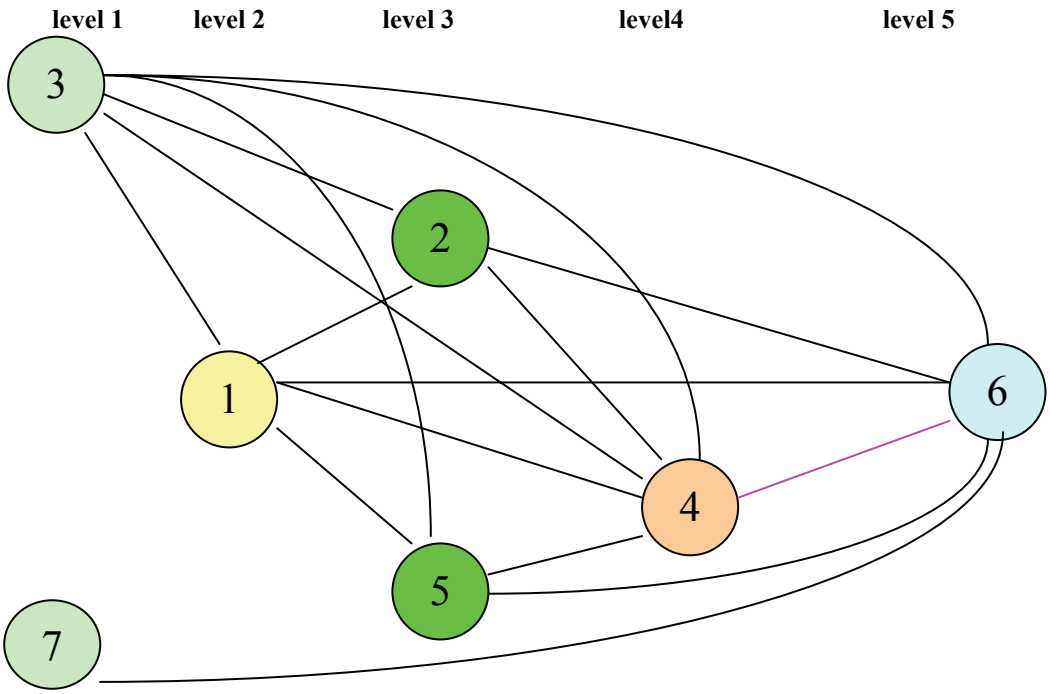
- alternatives with no outranking relation are at the same level,
- on the first level there are alternatives which are not outranked by any other alternative,
- on the middle levels there are alternatives which are outranked by one or more alternatives,
- on the last level there are alternatives which do not outrank any other alternative

Diagram 1. Outranking relation for the level p = 1. Source: On the basis of the Table 12.



According to the graph for p = 1 the first level is made up of alternatives 1, 3, 6, and 7 because none of them is outranked by another alternative, and there are no outranking relations among them. Alternative 2, 4 and 5 constitute second and last level since they do not outrank any other alternative and are outranked by all alternatives left. In this situation an insurance company may find it difficult to chose the best alternative because four of them are situated on the first level of the graph. Therefore, it is suggested to construct a similar graph for a lower level of concordance.
Next level of concordance to be taken into consideration is p = 0.8. In this case there is an outranking relation among comparable alternatives as far as two criteria: profitability and loss are concerned. Retention is not taken into account. Diagram 2 presents the graph obtained in this case.

Diagram 2. Outranking relations for the level p = 0.8. Source: On the basis of Table 12.



Only two alternatives: 3 and 7 are on this level of concordance value on the first level. They are not outranked by any other alternative and they do not outrank each other.

Second level is made up of alternative 1 which is outranked by alternatives 3 and 7 and outranks alternatives 2, 5, 4 and 6.

Alternatives 2 and 5 are on the third level. They are outranked by alternatives from levels 1 and 2. They outrank alternatives 4 and 6.

Fourth level is made up of alternative 4 which outrank only alternative 6 and is outranked by all alternatives left.

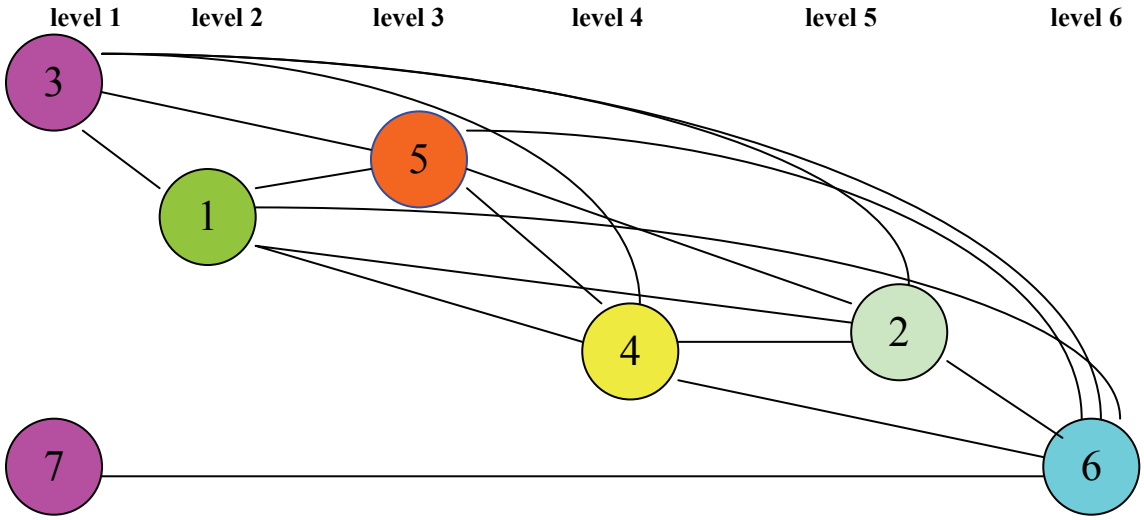
Alternative 6 is on the last fifth level. It is outranked by all alternatives. In this case insurance company's preferences as far as majority of alternatives may be determined. Only preference in relations to alternative 7 and alternatives of the third level remains unclear. However, alternative 7 is particularly interesting. It outranks only the worst alternative 6 and simultaneously is not outranked by any of alternatives left. Decreasing the level of concordance may explain the doubts.

Level p = 0.7 is the last level for the weights assumed for particular criteria. In this case outranking relations among alternatives will be determined taking two pairs of criteria into consideration:

- productivity and loss,
- productivity and retention.

For such level the graph presented on the Diagram 3 is obtained.

Diagram 3. Outranking relations for the level p = 0.7. Source: On the basis of Table 12.



Lowering the level of concordance to 0.7 enables a clear determination of preferences towards all alternatives but 7. The assessment of this alternative is questionable because it remains in unclear relation with left alternatives. It is still on the highest level not being outranked by any other alternative but at the same time it outranks only alternative 6 from the lowest level and outranked by all the other alternatives. Hence, a clear assessment of this alternative is not possible. Therefore alternative 3 should be considered the best alternative on this level of concordance. It also might be perceived as the best one on the level 0.8 where situation identical case of alternative 7 is observed.

Conclusions

Suggested method constitutes an alternative method of selecting insurance tariffs in non-life insurance. It may be a key tool supporting a decision maker representing an insurance company during a decision making process. Applying stochastic dominance makes it possible to chose such a combination of premium tariffs which will enable realisation of all goals predetermined by an insurance company to full extent. Simultaneously, possibility of direct comparing the results of choosing different variants of tariffs lets take into consideration both interests of an insurance company and its clients since an insurance company wishing to develop must take care of its present and potential clients.

References

[1] Ciupek B.: *Optymalizacja wyniku technicznego, struktury portfela i taryf składek zakładu ubezpieczę majątkowych*, Fundacja Warta, Warszawa, 2001

[2] Trzaskalik T., Trzpiot G., Zaraś K.: *Modelowanie preferencji z wykorzystaniem dominacji stochastycznych*. AE Katowice, 1998

[3] Trzpiot G.: *Wielowartościowe zmienne losowe w badaniach ekonomicznych*, AE Katowice, 1999

[4] Trzpiot G.: *Wielowartościowe dominacje stochastyczne* [In] Trzaskalik T. red. Metody i zastosowania badań operacyjnych. Część II, ss. 343-355, AE Katowice, 1999

Money's Function in the Czech Monetary Business Cycle[1]

Stanislav David, Osvald Vašíček
Masaryk University in Brno,
Faculty of Economics and Administration,
Department of Applied Mathematic and Informatics,
Lipová 41a, 602 00 Brno,
e-mail: davids@econ.muni.cz, osvald@econ.muni.cz

Abstract

Analysts have devoted considerable time and effort towards developing new and improved models for monetary policy evaluation. These newly developed models differ considerably in details. Do these existing models provide an accurate and complete description of money´s role in the monetary business cycle? That is the question addressed here. The paper begins by constructing small New Keynesian DSGE model. It goes on to estimate the model with quarterly time-series data of the Czech economy. Maximum likelihood estimates of the model's parameters suggest that money plays a nonessential role in the Czech monetary business cycle. The Kalman filter was used to evaluate negative log likelihood function of the model.

Keywords

Business cycle, Monetary policy, Linearized DSGE model, solution of DSGE model, Kalman Filter with log likelihood omptimalization

1. Introduction

Peter N. Ireland (see [7])has developed in recent years new approach of the maximum likelihood estimation of the business cycle models incorporating rational expectations based on the method of the Blanchard and Kahn. These models examine an accurate and complete description of money's role in the monetary business cycle.

Peter N. Ireland (see [7, 6]) has estimated the model with quarterly time-series data from the post-1980 United States economy. He concluded that the money plays a minimal role in shaping the dynamic behavior of output and inflation. It was a stimulus to verify result with the Czech Republic data.

The aim of the Section 2 is to construct a small, structural model with microfoundations that allows, but does not require, changes in the real money stock to directly affect the dynamics of output and inflation. The model is the same as the Ireland´s one. It is possible to describe the model in the same way as Ireland does: The theoretical model contains three equations summarizing the optimizing behavior of households and firms that populate the economy. The first of these resembles the IS curve in traditional Keynesian models, the second takes the form of a money demand relationship, and the third is a forward-looking version of the Phillips curve. Of course, real balances always enter the money demand function. But the cross-equation restrictions imposed by the model imply that real balances enter into the IS curve if and only if they also enter into the Phillips curve specification. Thus, according to the theory, if changes in the real stock of money have a direct impact on the dynamics of output and inflation, then that impact must come simultaneously through both the IS and the Phillips curve relationships.

[1]This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172.

The model reveals that assessing the importance of real balances in the forward-looking IS and Phillips curve specifications involves more than simply adding some measure of money to the equations and testing for the statistical significance of the associated coefficients. To isolate and quantify the effects of changes in real balances on output and inflation, the measure of money must be adjusted for shifts in money demand.

Section 3 represents the estimate of the model with quarterly time-series data from the post-1993 Czech republic economy

2. A Small, Structural Model of the Monetary Business Cycle

Description of the model has been taken over from the [7]. Here, the small, structural model developed in [3, 4] is modified to focus on the direct effects that changes in real money balances may have on the dynamics of output and inflation. This model elaborates on Rotemberg's framework (see [8]), in which monopolistically competitive firms face a quadratic cost of nominal price adjustment. Relative to the model in Ireland [3, 4], the one used here is generalized by allowing, but not requiring, real balances to appear in the IS and Phillips curve specifications. At the same time, however, the model is simplified, following [8] and [11], by abstracting from the process of capital accumulation.

The economy consists of a representative household, a representative finished goods-producing firm, a continuum of intermediate goods-producing firms indexed by $i \in [0, 1]$, and a monetary authority. During each period $t = 0, 1, 2, \dots$, each intermediate goods-producing firm produces a distinct, perishable intermediate good. Hence, intermediate goods may also be indexed by $i \in [0, 1]$, where firm i produces good i. The model features enough symmetry, however, to allow the analysis to focus on the behavior of a representative intermediate goods-producing firm that produces the generic intermediate good i.

The representative household seeks to maximize the expected utility function

$$E \sum_{t=0}^{\infty} \beta^t a_t \{u[c_t, (M_t/P_t)/e_t] - \eta h_t\},$$

with $1 > \beta > 0$ and $\eta > 0$ subject to the budget constraint

$$M_{t-1} + T_t + B_{t-1} + W_t h_t + D_t \geq P_t c_t + B_t/r_t + M_t,$$

which must be satisfied for all $t = 0, 1, 2, \dots$. In the utility function, c_t, M_t/P_t, and h_t denote the household's consumption, real balances, and labor supply during period t. The preference shocks a_t and e_t follow the autoregressive processes

$$\ln(a_t) = \rho_a \ln(a_{t-1}) + \epsilon_{at} \tag{1}$$

and

$$\ln(e_t) = (1 - \rho_e) \ln(e) + \rho_e \ln(e_{t-1}) + \epsilon_{et}, \tag{2}$$

where $1 > \rho_a > 0, 1 > \rho_e > 0, e > 0$, and the zero-mean, serially uncorrelated innovations ϵ_{at} and ϵ_{et} are normally distributed with standard deviations σ_a and σ_e. As shown below, the shocks a_t and c_t translate, in equilibrium, into disturbances to the model's IS and money demand curves.

In the budget constraint, the household's sources of funds include M_{t-1}, nominal money carried into period t, T_t, a lump-sum nominal transfer received from the monetary authority at the beginning of period t, and B_{t-1}, the value of nominal bonds maturing during period t. The household's sources of funds also include labor income $W_t h_t$, where W_t denotes the nominal wage, and nominal dividend payments D_t received from the intermediate goods-producing firms. The household's uses of funds consist of consumption c_t of the finished good, purchased at the nominal price Pt, newly-issued bonds of value B_t/r_t, where r_t denotes the gross nominal interest rate, and the money M_t to be carried into period $t + 1$. It is convenient in what follows to let $m_t = M_t/P_t$ denote the household's real balances and $\pi_t = P_t/P_{t-1}$ denote the gross inflation rate during period t. During each period $t = 0, 1, 2, \dots$, the representative finished goods-producing firm uses $y_t(i)$ units of each intermediate good $i \in [0, 1]$, purchased at the nominal

price $P_t(i)$, to manufacture y_t units of the finished good according to the constant-returns-to-scale technology described by

$$\left[\int_0^1 y_t(i)^{\theta-1/\theta} di\right]^{\theta/(\theta-1)} \geq y_t,$$

with $\theta > 1$. The finished goods-producing firm maximizes its profits by choosing

$$y_t(i) = [P_t(i)/P_t]^{-\theta} y_t,$$

which reveals that θ measures the constant price elasticity of demand for each inter-mediate good. Competition drives the finished goods-producing firm's profits to zero in equilibrium, determining P_t as

$$P_t = \left[\int_0^1 P_t(i)^{1-\theta} di\right]^{1/(1-\theta)}.$$

During each period $t = 0, 1, 2, \ldots$, the representative intermediate goods-producing firm hires $h_t(i)$ units of labor from the representative household to manufacture $y_t(i)$ units of intermediate good i according to the linear technology

$$z_t h_t(i) \geq y_t(i).$$

The aggregate productivity shock z_t follows the autoregressive process

$$\ln(z_t) = (1 - \rho_z)\ln(z) + \rho_z \ln(z_{t-1}) + \epsilon_{z_t}, \tag{3}$$

where $1 > \rho_z > 0, z > 0$, and the zero-mean, serially uncorrelated innovation e_{zt} is normally distributed with standard deviation σ_z. In equilibrium, this supply-side disturbance acts as a shock to the Phillips curve. Since the intermediate goods substitute imperfectly for one another in producing the finished good, the representative intermediate goods-producing firm sells its output in a monopolistically competitive market: the firm acts as a price-setter, but must satisfy the representative finished goods-producing firm's demand at its chosen price. And here, as in [8], the intermediate goods-producing firm faces a quadratic cost of adjusting its nominal price, measured in terms of the finished good and given by

$$\frac{\phi}{2}\left[\frac{P_t(i)}{\pi P_{t-1}(i)} - 1\right]^2 y_t,$$

with $\phi > 0$, where π measures the gross steady-state inflation rate. This cost of price adjustment makes the intermediate goods-producing firm's problem dynamic: it chooses $P_t(i)$ for all $t = 0, 1, 2, \ldots$ to maximize its total market value. At the end of each period, the firm distributes its profits in the form of a nominal dividend payment $D_t(i)$ to the representative household. In a symmetric equilibrium, all intermediate goods-producing firms make identical decisions, so that $y_t(i) = y_t, h_t(i) = h_t, P_t(i) = P_t$, and $D_t(i) = D_t$ for all $i \in [0, 1]$ and $t = 0, 1, 2, \ldots$. In an equilibrium, the first-order conditions describing the optimizing behavior of the representative household and intermediate goods-producing firm can be approximated by

$$\hat{y}_t = E_t\hat{y}_{t+1} - \omega_1(\hat{r}_t - E_t\hat{\pi}_{t+1}) + \omega_2[(\hat{m}_t - \hat{e}_t) - (E_t\hat{m}_{t+1} - E_t\hat{e}_{t+1})] + \omega_1(\hat{a}_t - E_t\hat{a}_{t+1}), \tag{4}$$

$$\hat{m}_t = \gamma_1\hat{y}_t - \gamma_2\hat{r}_t + \gamma_3\hat{e}_t \tag{5}$$

and

$$\hat{\pi}_t = (\pi/r)E_t\hat{\pi}_{t+1} + \psi[(1/\omega_1)\hat{y}_t - (\omega_2/\omega_1)(\hat{m}_t - \hat{e}_t) - \hat{z}_t]. \tag{6}$$

In (4)-(6), $\hat{y}_t$, $\hat{m}_t$, $\hat{\pi}_t$, $\hat{r}_t$, $\hat{a}_t$, $\hat{e}_t$, and $\hat{z}_t$ denote the percentage (logarithmic) deviations of y_t, m_t, pi_t, r_t, a_t, e_t, and z_t from their steady-state values y, m, π, r, 1, e, and z. All of the parameters in (4)-(6) ought to be nonnegative; each ultimately depends on the underlying parameters describing private agents' tastes and technologies. And in addition to the cross-equation restrictions that appear explicitly in (4)-(6), the constraints

$$\gamma_1 = (r - 1 + yr\omega_2/m)(\gamma_2/\omega_1) \tag{7}$$

and

$$\gamma_3 = 1 - (r-1)\omega_2 \tag{8}$$

must be satisfied.

Equation (4) generalizes forward-looking IS curve by allowing real balances m_t to enter the specification. Intuitively, (4) represents a log-linearized version of the Euler equation that links the household's marginal rate of intertemporal substitution to the real interest rate. When utility is nonseparable, real balances affect the marginal rate of intertemporal substitution; hence, they also appear in the IS curve.

Equation (5), meanwhile, takes the form of a money demand relationship, with income elasticity γ_1 and interest semi-elasticity γ_2.

Equation (6) is a forward-looking Phillips curve that also allows real balances m_t to enter the specification when ω_2 is nonzero. According to the model, therefore, real balances belong in a correctly-specified IS curve if and only if they also belong in a correctly-specified Phillips curve.

Seven variables enter into (1)-(6). Hence, the model is closed by adding a seventh equation an interest rate rule for monetary policy of the form

$$\hat{r}_t = \rho_r \hat{r}_{t-1} + \rho_y \hat{y}_{t-1} + \rho_\pi \hat{\pi}_{t-1} + \epsilon_{rt} \tag{9}$$

where ρ_r, ρ_y, and ρ_π, are nonnegative parameters, and where the zero-mean, serially uncorrelated policy shock ϵ_{rt} is normally distributed with standard deviation σ_r. Like Taylor's rule (see [9]), (9) calls for the monetary authority to adjust the short-term nominal interest rate in response to deviations of output and inflation from their steady-state levels. Equation (9) generalizes Taylor's rule, however, by adding a term involving the lagged interest rate: when ρ_r is nonzero, the interest rate adjustment to output and inflation occurs gradually over time.

3. **Estimation results**

For the solving the model we have used Czech economy quarterly data from year 1995. The solution to this system is using the method of Blanchard and Kahn (see [1]), takes the form of a state-space econometric model. The model's parameters are estimated by maximum likelihood, as described by Hamilton (see [2]). It is interesting that the results are quite similar to the Ireland ones because the length of the data is smaller and they are different.

Due to the reasons shown in [5] are the parameters ω_1, and ψ fixed prior to estimation. Hence, table 1 displays maximum likelihood estimates of the model's remaining parameters, holding $\omega_1 = 1$ fixed at the value that implies the same level of risk aversion as a utility function that is logarithmic in consumption and holding $\psi = 0.1$ fixed at the value used previously in [4]. Although 18 parameters appear in table 1, only 16 of these are estimated independently, since the restrictions (7) and (8) are imposed. The standard errors correspond to the square roots of the diagonal elements of minus one times the inverted matrix of second derivatives of the maximized log likelihood function.

The parameter ρ_y, measuring the interest rate response to changes in output, tends to be zero, suggesting that Czech national bank policy has concentrated mainly on controlling inflation during the post-1995 sample period. The estimates $\rho_r = 0.6851$ and $\rho_\pi, = 0.4518$ imply a considerable amount of interest rate smoothing and a vigorous policy response to inflation, with a long-run elasticity exceeding unity. This last feature allows the interest rate rule to be consistent with the existence of a unique rational expectations equilibrium.

The estimates of $\ln(y)$, $\ln(m)$, $\ln(\pi)$, and $\ln(r)$ help the model match the average level of each variable in the data. These parameters are estimated in logs, rather than levels, to avoid scaling problems in the numerical routine that maximizes the likelihood function. The estimates of ρ_α, ρ_e and ρ_z indicate that the exogenous IS, money demand, and Phillips curve shocks are quite persistent, while the sizable estimates of σ_α, σ_e and σ_z imply that all three of these shocks are important in explaining fluctuations in the data.

Once again, these results point to the statistical adequacy of popular specifications, in which real balances do not enter the IS and Phillips curves.

Figure 1 presents the impulse response of each variable detrended output, detrended real balances, inflation, and the interest rate to three of the model's shocks: the preference shock ϵ_{at}, the technology shock ϵ_{zt} and the policy shock ϵ_{rt}. The figure omits the impulse responses to the money demand shock ϵ_{et} since, as explained above.

A one standard deviation IS shock ϵ_{at} increases output by about 1 percent and increases the annualized inflation rate by about 0.3 percent; under the estimated policy rule, the monetary authority responds with an 0.25 percent increase in the annualized nominal interest rate. Response to the technology is the main difference of our economy. A one standard deviation technology shock ϵ_{zt} increases output by nearly 0.4 percent, decreases the annualized inflation rate by 0.25 percent, and induces the monetary authority to lower the annualized interest rate by 0.13 percent. Finally, a one standard deviation policy shock ϵ_{rt} is generating a 0.8 percent fall in output and a 0.15 percent fall in annualized inflation.

Parameter	Unconstrained Estimate Using M2	Standard Error
ω_2	6.4077e-007	0.20727
γ_1	0.019722	0.054606
γ_2	0.97161	0.39471
γ_3	0.98028	0.0075457
ρ_r	0.68518	0.072444
ρ_y	0.059645	0.034253
ρ_p	0.45187	0.13179
$\ln(y)$	5.8681	0.0061532
$\ln(m)$	6.9484	0.017134
$\ln(p)$	0.010426	0.0027163
$\ln(r)$	0.020095	0.0043083
ρ_a	0.89098	0.035143
ρ_e	0.90357	0.056138
ρ_z	0.81581	0.091734
σ_a	0.0259	0.0059937
σ_e	0.012285	0.0012629
σ_z	0.010057	0.0024255
σ_r	0.003941	0.00050789

TABLE 1. Estimated parameters

4. Conclusions

The maximum likelihood method with Blanchard - Kahn extension has been used for the estimation of a small, structural model of the business cycle - the model of the Peter N. Ireland. The model has been reestimated with Czech republic post 1995 data.

The results obtained here indicate that the Czech Republic data gives similar results as post-1980 United States data. The money plays also minimal role in the monetary business cycle. The impulse response is also quite similar except a technology shock. The results suggests that real balances fail to enter the IS and Phillips curve equations that govern the dynamics of output and inflation.

When we see that the money's role in the monetary business cycle appears limited it is possible to formulate another hypothesis, that we will release the money from the model framework. This approach is estimated for the Czech economy for example in [10].

References

[1] O. J. Blanchard and Ch. M. Kahn, The Solution of Linear Difference Models Under Rational Expectations, Econometrics, vol. 48, 1305-1311, July 1980

[2] J. D. Hamilton, Time Series Analysis, Princeton University Press,1994

[3] P. N. Ireland, A Small, Structural, Quarterly Model for Monetary Policy Evaluation, Carnegie-Rochester Conference Series on Public Policy, vol. 47, pp. 83-108, December 1997
[4] P. N. Ireland, Interest Rates, Inflation, and Federal Reserve Policy Since 1980, Journal of Money, Credit, and Banking, vol. 32, Part 1, pp. 417-434, August 2000
[5] P. N. Ireland, Money's Role in the Monetary Business Cycle, Working Paper 8115. Cambridge, National Bureau of Economic Research, February 2001
[6] P. N. Ireland , MATLAB programs that to reproduce the econometric work in paper "Money's Role in the Monetary Business Cycle,", Source: http://www2.bc.edu/~irelandp/#programs, 2002
[7] P. N. Ireland, Money's Role in the Monetary Business Cycle, The Journal of Money, Credit and Banking, Vol. 36, No. 6, pp. 969-983, December 2004, related with programs found on: http://www2.bc.edu/ irelandp/
[8] J. J. Rotemberg, Sticky Prices in the United States, Journal of Political Economy, vol. 90, pp. 1187-1211, December 1982
[9] J. B. Taylor, Discretion Versus Policy Rules in Practice, Carnegie-Rochester Conference Series on Public Policy, Vol. 39, pp. 195-214., December 1993
[10] O. Vašíček and K. Musil, A Model Interpretation of the Czech Inflation Targeting and the Monetary Policy, So far not publicated
[11] M. Woodford, Price-Level Determination Under Interest-Rate Rules, Princeton University, April 1999

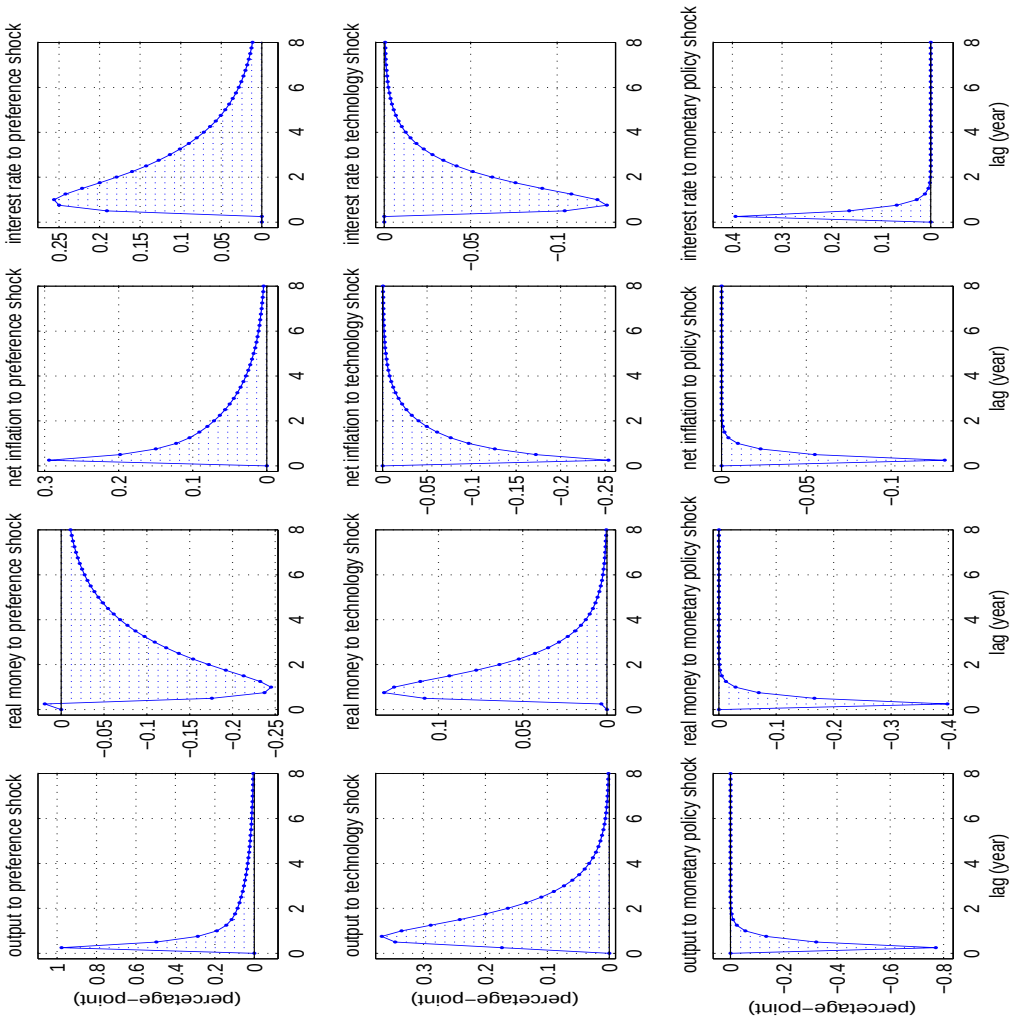


FIGURE 1. Impulse responses

Estimating priorities in AHP with interval pair-wise judgements

Dimitris K. Despotis and Dimitris Derpanis
University of Piraeus, Department of Informatics
80 Karaoli &Dimitriou Street, 18534 Piraeus, Greece
{despotis, dderp}@unipi.gr

Abstract

In this paper we deal with the problem of priority elicitation in the analytic hierarchy process (AHP) on the basis of approximate pair-wise comparison judgements. We propose a min-max goal programming formulation to derive priorities in the case that the preference judgements are provided as interval numbers. By applying variable transformations we develop a linear programming model that is capable of deriving priorities from both consistent and inconsistent interval judgements. The proposed method is illustrated by numerical examples.

Key words
Analytic hierarchy process, Priority setting, Interval judgements.

1 Introduction

A key issue addressed in multicriteria decision analysis (MCDA) is the assignment of priorities (weights) to decision elements. The analytic hierarchy process (AHP) introduced by Saaty [7] is one of the most widely used approaches for deriving such priorities through pair-wise comparisons of decision elements. The AHP proceeds in four steps: (a) break down the decision problem into a hierarchy of decision elements (general goal, criteria, sub-criteria, alternatives); (b) construct the pair-wise comparison matrices for the decision elements in each level of the hierarchy with respect to one decision element at a time in a level immediately above it; (c) derive local priorities for the decision elements from the pair-wise matrices and (d) synthesize the local priorities to derive global priorities of the alternatives with respect to the general goal of the problem.

In the AHP context, a comparison matrix is an *nxn* positive reciprocal matrix $R=(r_{ij})$ of paired comparisons of n decision elements in a certain level of the hierarchy with respect to a decision element in a level immediately above it. In the original AHP, each entry r_{ij} represents a judgement concerning the perceived dominance (relative importance or preference) of decision element i over j and is provided as a crisp number in the bounded discrete scale $(1/9 \le r_{ij} \le 9)$. The basic method proposed by Saaty for deriving the priorities $w = (w_1, w_2, ..., w_n)$ of the n decision elements from the matrix R is the eigenvector method but there are several other scaling methods to assess these priorities.

An important issue addressed in the literature [1], [8] is the approximate articulation of preferences in the AHP context. In such a situation, the decision maker provides a range of values (interval) $[l_{ij}, u_{ij}]$, instead of a single point r_{ij} on the scale, to express her/his preference of a decision element i over an element j.

In this paper we focus on the steps (b) and (c) of the AHP, when preferences are stated by means of interval pair-wise judgments. In the second section we provide a brief review of the relative literature. In the third section we develop and illustrate our approach to deriving priorities from interval pair-wise judgments. The paper ends with some concluding remarks.

2 Dealing with approximate preferences in AHP

Approximate preferences in AHP have been dealt first by considering the entries r_{ij} in R as fuzzy numbers with triangular [12] or trapezoidal [2] membership functions. Saaty and Vargas [8] introduced the interval numbers to handle approximate preferences and used a simulation approach to derive priority intervals from paired comparisons matrices with interval numbers. Recently Mikhailov [6] introduced linear or non-linear membership functions to derive crisp from interval pair-wise comparisons.

With interval judgments, the decision maker can make statements such as "the element *i* is at least three but no more that seven times as preferable as the element *j*". Such pair-wise comparisons are collected on a matrix *R* having the general form

$$R = \begin{bmatrix} 1 & [l_{12} \;\; u_{12}] & [l_{13} \;\; u_{13}] & \cdots & [l_{1n} \;\; u_{1n}] \\ \left[\frac{1}{u_{12}}, \frac{1}{l_{12}}\right] & 1 & [l_{23} \;\; u_{23}] & \cdots & [l_{2n} \;\; u_{2n}] \\ \left[\frac{1}{u_{13}}, \frac{1}{l_{13}}\right] & \left[\frac{1}{u_{23}}, \frac{1}{l_{23}}\right] & 1 & \cdots & [l_{3n} \;\; u_{3n}] \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \left[\frac{1}{u_{1n}}, \frac{1}{l_{1n}}\right] & \left[\frac{1}{u_{2n}}, \frac{1}{l_{2n}}\right] & \left[\frac{1}{u_{3n}}, \frac{1}{l_{3n}}\right] & \cdots & 1 \end{bmatrix}$$

where l_{ij} and u_{ij} are the lower and upper bounds defined on the scale (the Saaty's scale for example) that the decision maker uses to express the relative importance of the element *i* over the element *j*. The matrix R is reciprocal in the sense that $l_{ji} = 1/u_{ij}$ and $u_{ji} = 1/l_{ij}$.

The preference programming method of Arbel [1] is a linear programming approach to derive priorities from such a matrix R of interval numbers. An extension of Arbel's method is given by Salo and Hamalainen [10], [11]. According to the preference programming method the priority vector $w = (w_1, w_2, ..., w_n)$ is obtained as a solution to the following set of linear inequalities

$$S = \left\{ w = (w_1, w_2, ..., w_n) / \; l_{ij} \le w_i / w_j \le u_{ij}, i, j = 1, ..., n, \; w_1 + w_2 + ... + w_n = 1 \right\}$$

Arbel [1] suggested that the feasible region *S* itself can be viewed as a representation of the decision maker's preferences on the decision elements that he compares. Salo and Hamalainen [10], [11] use a linear programming technique to compute the minimum and the maximum values that each priority (w_i, i=1,...,n) can attain. The resultant intervals are used to express the decision maker's preferences. The feasible region S is non-empty, if the above system of inequalities is solvable, i.e. if there exists at least one priority vector $w = (w_1, w_2, ..., w_n)$ such that the ratios w_i / w_j, for all *i* and *j* lie in the corresponding intervals [l_{ij}, u_{ij}]. This is the case of fully consistent intervals or, in other words, the case of a fully consistent comparison matrix *R*. However, in case of inconsistent comparisons in *R*, the feasible region *S* is empty and Arbel's method is not applicable. The extended region approach [9] and the lexicographic programming approach [4] are two alternative techniques to derive priorities from inconsistent matrices of paired comparisons. Lee et al. [5] introduce uncertainty in the comparisons and propose a stochastic model and an iterative process to determine the priorities. In the following we develop an alternative min-max goal programming approach for priority setting in the presence of interval judgments, regardless of their consistency.

3 A min-max approach to priority setting

As mentioned in the previous section, the interval judgments $[l_{ij}, u_{ij}]$ in the matrix *R* give rise to the following system of inequalities:

$$\begin{aligned} & l_{ij} \le w_i / w_j \le u_{ij} \ , \ i = 1,...,n-1, \ j = i+1,...,n \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{1}$$

Initially, let as assume that the system (1) is solvable, i.e. the matrix *R* of interval judgments is fully consistent. The above system (1) can then be expressed as follows:

$$\begin{aligned} & w_i = w_j l_{ij} + w_j s_{ij} (u_{ij} - l_{ij}) \ , \ s_{ij} \in [0,1] \ , \ i = 1,...,n-1, \ j = i+1,...,n \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{2}$$

The system (2) is non-linear due to the variable s_{ij}. This new variable is used to express the ratio w_i / w_j in terms of the left and the right extreme of the interval $[l_{ij}, u_{ij}]$. With the constraint that the values of the variable s_{ij} are in [0,1], we adopt the assumption the ratio w_i / w_j lies in the interval $[l_{ij}, u_{ij}]$. If we replace in (2) $w_j s_{ij}$ with the variables p_{ij} ($p_{ij} = w_j s_{ij}$), the new variables p_{ij} meet the conditions $0 \le p_{ij} \le w_j$, as it is $s_{ij} = p_{ij} / w_j$ with $w_j > 0$ and, in case of consistent interval judgements, it is also $0 \le s_{ij} \le 1$ for every *i* and *j*. With these variable transformations the non-linear system (2) takes the following linear form (see [3] for more details, although in a different context):

$$\begin{aligned} & w_i = w_j l_{ij} + p_{ij} (u_{ij} - l_{ij}) \\ & p_{ij} - w_j \le 0 \\ & p_{ij} \ge 0 \qquad\qquad i = 1,...,n-1, \ j = i+1,...,n \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{3}$$

Notice that for p_{ij}=0 it is $w_i / w_j = l_{ij}$ and for p_{ij}=w_j it is $w_i / w_j = u_{ij}$. As mentioned above the solution space of the system (2) is non-empty only in the consistent case. However, with the formulation introduced above we can handle also the case of inconsistent interval judgements that is the case where the system (3) has no feasible solution. We relax the assumption that all the ratios w_i / w_j lie in the intervals $[l_{ij}, u_{ij}]$ by allowing these weight ratios to lie outside the intervals. Particularly, to model the situation that a weighs ratio exceeds the upper bound of the interval, the corresponding auxiliary variable p_{ij} should be allowed to take values beyond w_j. Similarly, the variables p_{ij} should be allowed to take negative values, in order to model the situation that the weights ration exceeds the lower bound of the interval. Let *z* be a non-negative variable expressing the required extension of either side of the range of values $[0, w_j]$ that p_{ij} can

take. With the introduction of the variable z the following linear program is solvable even in the case that the interval judgments are inconsistent.

$$
\begin{aligned}
&\min z \\
&w_i - w_j l_{ij} - p_{ij}(u_{ij} - l_{ij}) = 0 \\
&p_{ij} - w_j - z \le 0 \\
&-p_{ij} - z \le 0 \qquad\qquad i = 1,...n-1,\ j = i+1,...,n \qquad\qquad (4) \\
&\sum_{j=1}^{n} w_j = 1 \\
&w_j \ge 0, z \ge 0 \\
&p_{ij}\ free
\end{aligned}
$$

After calculating the priorities $w = (w_1, w_2, ..., w_n)$ by the model (4), the priority ratios w_i / w_j may or may not lie in the intervals $[l_{ij}, u_{ij}]$. In case of inconsistencies, some of the priority ratios may exceed the upper bound of the interval, others my lie below the lower bound. In model (4) however, the priorities are estimated in a manner that the maximal of these deviations is minimized. That is, in case of inconsistent interval judgments, the ratio w_i / w_j that violates the concerned interval, it comes as close as possible to the upper bound u_{ij} from the right or to the lower bound l_{ij} from the left. This is a min-max goal programming approach to deriving the priorities. The value of z is an indication of inconsistency. It gets a zero value in case of consistent interval judgments and a strictly positive value in case of inconsistencies in the matrix of paired interval comparisons. Moreover, the higher is the value of the variable z in the optimal solution the higher is the inconsistency in the matrix.

To illustrate the proposed approach, consider the following interval judgment matrix (only the upper triangular part of the matrix is presented) where four decision elements A, B, C, D are compared in pairs [4]:

	A	B	C	D
A	1	[1,2]	[1,2]	[2,5]
B		1	[2,5]	[4,5]
C			1	[2,3]
D				1

Solving model (4) for the preference data in the above matrix we get with the optimal solution the priorities w_1=0.3636, w_2=0.3636, w_3=0.1818, w_4=0.0909. The optimal value of the variable z is zero. This is an indication that the interval judgments in the above matrix are fully consistent, that is all the ratios of the estimated priorities are in the corresponding intervals. It is obvious that in case of consistent judgments, the priorities that we obtain by the model (4) are not unique. Indeed, in case of consistent intervals, the system (1) is solvable and any possible solution of this system is an optimal solution of model (4). One can seek however for characteristic solutions such as the solution that maximizes the priority w_1, for example (w_1=0.3704, w_2=0.3704, w_3=0.1852, w_4=0.0741).

Modifying the intervals in the above matrix, for example setting for the paired comparison (A, D) the interval [2,3] and for the pair (B, C) the interval [3,5] we get an

inconsistent comparison matrix. Solving the model (4) for the new matrix we get the unique priorities w_1=0.3478, w_2=0.3913, w_3=0.1594, w_4=0.1014. The optimal value of the variable z =0.0435. This verifies that the comparison matrix is inconsistent. If we perturb further the matrix by assuming for the comparison of (C, D) the interval [6,8], we improve the inconsistency of the matrix and we get the unique priorities w_1=0.3103, w_2=0.4138, w_3=0.2069, w_4=0.0690 with z=0.1034.

4 Conclusions

Approximate articulation of preferences is the means to overcome the decision maker's inability, due to fuzziness or uncertainty, to provide point estimates on the scale when he/she compares criteria or decision alternatives in the AHP framework. In such a setting crisp priorities must be estimated from interval pair-wise comparisons. The approach proposed in this paper for solving this problem is based on a min-max goal programming formulation that enables the estimation of local priorities for the decision elements regardless the consistency of the matrix of pair-wise comparisons. Moreover, the underlying priorities derive through an optimality criterion. A measure of inconsistency is also obtained that enables the analyst to locate the inconsistencies in the process of preference elicitation and probably provide advice to the decision maker to raise these inconsistencies.

References

[1] Arbel, A., Approximate articulation of preference and priority derivation, *European Journal of Operational Research* 43 (1989) 317-326.

[2] Buckley, J.J., Fuzzy hierarchical analysis, *Fuzzy Sets and Systems* 17 (1987) 233-247.

[3] Despotis, D.K, Smirlis, Y.G., Data envelopment analysis with imprecise data, *European Journal of Operational Research* 140 (2002) 24-36.

[4] Islam, R., Biswal, M.P., Alam, S.S., Preference programming and inconsistent interval judgments, *European Journal of Operational Research* 97 (1997) 53-62.

[5] Lee, M., Pham, H., Zhang, X., A methodology for priority setting with application to software development process, *European Journal of Operational Research* 118 (1999) 375-389.

[6] Mikhailov, L., A fuzzy approach to deriving priorities from interval pairwise comparison judgements, *European Journal of Operational Research* 159 (2004) 687-704.

[7] Saaty, T.L., The analytic hierarch process, McGraw-Hill, New York, 1980.

[8] Saaty, T.L., Vargas, L.G., Uncertainty and rank order in the analytic hierarchy process, *European Journal of Operational Research* 32 (1987) 107-117.

[9] Salo, A.A., Inconsistency analysis by approximately specified priorities, Mathematical and Computer Modelling 17 (1993) 123-133.

[10] Salo, A.A., Hämäläinen, R.P., Processing interval judgments in the analytic hierarchy process, in: Goicoechea, Duckstein, Zionts (Eds.), Multicriteria Decision Making, Proceedings of the IX International Conference on MCDM, Springer, Berlin, 1991, p.p. 359-372.

[11] Salo, A.A., Hämäläinen, R.P., Preference programming through approximate ratio comparisons, *European Journal of Operational Research 82* (1995) 458-475.

[12] Van Laarhoven, P.G., Pedrycz, W., A fuzzy extension of Saaty's priority theory, Fuzzy Sets and Systems 11 (1983) 229-241.

Application of Data Envelopment Analysis for Efficiency Evaluation in the Health Services

Martin Dlouhý, Lenka Flusserová
Faculty of Informatics and Statistics, University of Economics Prague, W. Churchill Sq. 4, 130 67 Prague, the Czech Republic, e-mail: dlouhy@vse.cz.

Abstract

Data Envelopment Analysis (DEA) is a method of efficiency evaluation of production units. The strengths of DEA are: (1) the method is able to deal with multiple inputs and outputs; (2) for technically inefficient unit DEA identifies the peers that are real production units, (3) software for DEA is now available, which makes it is easy to carry out all calculations and present results. The main weaknesses of DEA are: (1) a deterministic nature of method. Because DEA does not include an error term in the model, there is possibility of a measurement error; (2) hypothesis testing is much less developed in DEA in a comparison to econometric analysis. DEA proved to be a very attractive method of efficiency evaluation for health economists throughout the world. DEA is now mentioned in the textbooks of health economics. There is no doubt that the number of applications in the health services will be growing.

Keywords

Data envelopment analysis, efficiency evaluation, technical efficiency, health services.

1 The origins of DEA

M. J. Farrell (1957), in his seminal paper on the measurement of efficiency of production, promoted the ideas to specify the production frontier as the most pessimistic piecewise linear envelopment of the data and to construct efficiency measures based on radial uniform contractions or expansions from inefficient observations to the frontier. At that time, Farrell was not able to solve his model for multiple inputs and outputs. This problem was solved in 1978 by Abraham Charnes, William W. Cooper, and Edwardo Rhodes, who formulated model called Data Envelopment Analysis (DEA). However, Førsund and Sarafoglou (2002) reminded that independent discoveries in research were rather the rule than exception and showed that the DEA model had been already known to agriculture economists at the University of California in Berkeley in 1971, but at that time the model had stayed overlooked.

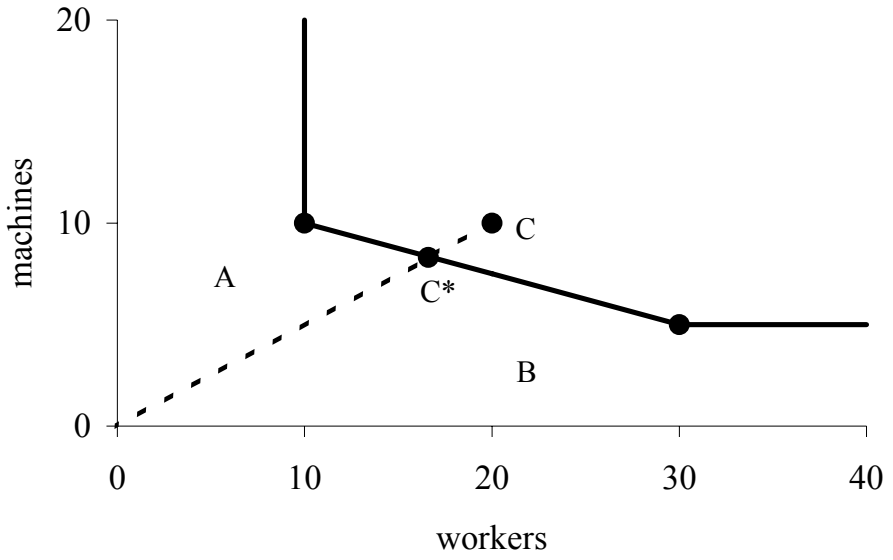
The success story of DEA began since the publication of highly influential paper by Charnes, Cooper, and Rhodes in 1978. Since then a great variety of DEA models with various extensions and modifications has been developed. These extensions, modifications and applications can be found, for example, in textbooks Charnes, Cooper, Lewin, and Seiford (1994), Cooper, Seiford, and Tone (2002), Jablonský, Dlouhý (2004). A good source of information about DEA is Ali Emrouznejad's Data Envelopment Analysis Page (http://www.deazone.com/). The journals that published most articles about DEA are *European Journal of Operational Research*, *Journal of Productivity Analysis*, *Journal of the Operational Research Society*, and *Management Science* (Tavares, 2002). During two decades DEA has become a very popular method. Tavares (2002) registered in his DEA database 3,230 publications, which had been written by 2,152 authors from 49 countries. Many applications were from health care because the method is attractive to health economists due to its ability to deal with multiple inputs and outputs. Hollingsworth (2003) reviewed 188 papers on frontier efficiency measurement in health care and found that the DEA-based methods dominated the literature.

2 Mathematical Formulation of DEA Model

DEA uses quantities of inputs consumed and outputs produced to calculate the relative technical efficiency of a decision-making unit. The relative technical efficiency of the unit is defined as the ratio of its total weighted output to its total weighted input or, vice versa, as the ratio of its total weighted input to its total weighted output. DEA allows each unit to choose its own weights of inputs and outputs in order to maximize its efficiency ratio. A technically efficient unit is able to find such weights that it lies on the production frontier. The production frontier represents the maximum amounts of output that can be produced by given amounts of input (in the output maximization model) or, alternatively, the minimum amounts of inputs required to produce the given amount of output (in the input minimization model). No production unit can achieve efficiency higher than 100%.

For each production unit DEA (a) calculates the efficiency score; (b) determines the relative weights of inputs and outputs; and (c) identifies peers for each unit that is not technically efficient. The peers of an inefficient unit are technically efficient units with similar combinations of inputs and outputs. The peers serve as benchmarks, which show potential improvements that the inefficient unit can attain. Because the peers are real production units, the efficiency improvements are attainable. A calculation of the efficiency score by DEA is showed graphically in Figure 1. Let us suppose that there are three production units *A*, *B*, and *C*, that produce the same level of output with two inputs—workers and machines. The production units *A* and *B* are technically efficient and they both lie on the production frontier. The production unit *C* uses more inputs than it is technically necessary, therefore, this unit lies above the production frontier and is technically inefficient. The units *A* and *B* are the peers which show unit *C* how to reduce both inputs to be technically efficient. A hypothetical unit *C** is a linear combination of units *A* and *B*. Notice however that unit *C* can achieve technical efficiency by moving to any position on the production frontier. The hypothetical unit *C** shows the alternative in which all inputs are reduced proportionately.

Figure 1: Technical Efficiency in DEA



Now let us continue with a mathematical formulation of the DEA model. The mathematical formulation of the so-called CCR model (**C**harnes, **C**ooper, and **R**hodes, 1978), which assumes the constant returns to scale, is:

$$\text{maximize} \quad \frac{\sum_{i=1}^{r} u_i y_{iq}}{\sum_{j=1}^{m} v_j x_{jq}},$$

$$\text{subject to} \quad \frac{\sum_{i=1}^{r} u_i y_{ik}}{\sum_{j=1}^{m} v_j x_{jk}} \le 1, \quad k = 1, 2, \ldots, n, \qquad \textbf{(1)}$$

$$u_i \ge \varepsilon, i = 1, 2, \ldots, r, \qquad v_j \ge \varepsilon, j = 1, 2, \ldots, m,$$

where x_{ik} is the amount of input i used by unit k, y_{ik} is the amount of jth output produced by unit k, weights u_i and v_j are variables in the model, ε is infinitesimal constant. The value of objective function is the efficiency score of unit q. This program (1) is an input-oriented version of the CCR model. To solve the program (1) by the linear programming, the program has to be transformed to the linear form:

$$\text{maximize} \quad \sum_{i=1}^{r} u_i y_{iq},$$

$$\text{subject to} \quad \sum_{i=1}^{r} u_i y_{ik} \le \sum_{j=1}^{m} v_j x_{jk}, \quad k = 1, 2, \ldots, n, \qquad \textbf{(2)}$$

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

$$u_i \ge \varepsilon, i = 1, 2, \ldots, r, \qquad v_j \ge \varepsilon, j = 1, 2, \ldots, m.$$

The model (2) has to be formulated and solved for each unit $q=1, 2, ..., n$. For the computation, it is useful to consider a dual model. The formulation of the dual model in the matrix form is:

$$\begin{aligned} &\text{minimize} && \theta_q - \varepsilon(\mathbf{e}^T\mathbf{s}^+ + \mathbf{e}^T\mathbf{s}^-), \\ &\text{subject to} && \mathbf{X}\boldsymbol{\lambda} + \mathbf{s}^- = \theta_q \mathbf{x_q}, \qquad \textbf{(3)} \\ & && \mathbf{Y}\boldsymbol{\lambda} - \mathbf{s}^+ = \mathbf{y_q}, \\ & && \boldsymbol{\lambda} \ge \mathbf{0}, \mathbf{s}^+ \ge \mathbf{0}, \mathbf{s}^- \ge \mathbf{0}, \end{aligned}$$

where θ_q is the efficiency score, $\boldsymbol{\lambda}$ is the vector of variables, $\mathbf{s}^+$, $\mathbf{s}^-$ are vectors of slack variables, $\boldsymbol{x_q}$ and $\boldsymbol{y_q}$ are vectors of inputs and outputs of the evaluated unit, $\mathbf{X}$ is a matrix of m inputs and $\mathbf{Y}$ is a matrix of r outputs, $\mathbf{e}^T = (1, 1, \ldots, 1)$, and ε is infinitesimal constant. The production unit q is technically efficient if the optimal value of variable θ^*_q is one and all slack variables s_i^{+*} and s_i^{-*} equal zero (the symbol "*" denotes the optimal value). After the formulation of the CCR model, many other models were developed. According to the returns to scale, the theory distinguishes four alternative DEA models (see Table 1).

Table 1: Returns to Scale in DEA Models

Returns to scale	**Formulation**
Constant returns to scale (CRS)/CCR model	$\boldsymbol{\lambda} \ge \mathbf{0}$
Variable returns to scale (VRS)/BCC model	$\boldsymbol{\lambda} \ge \mathbf{0}$, $\mathbf{e}^T\boldsymbol{\lambda} = 1$
Non-increasing returns to scale (NIRS)	$\boldsymbol{\lambda} \ge \mathbf{0}$, $\mathbf{e}^T\boldsymbol{\lambda} \le 1$
Non-decreasing returns to scale (NDRS)	$\boldsymbol{\lambda} \ge \mathbf{0}$, $\mathbf{e}^T\boldsymbol{\lambda} \ge 1$

3 Applications

In this section, we present some examples of application of DEA in the health services. For example, Ozgen and Ozcan (2002) examined market competition and facility characteristics that can be related to technical efficiency in the production of multiple dialysis outputs. They used the 1997 U.S. data on 791 Medicare-certified freestanding dialysis facilities. In their model, they included three outputs—outpatient dialysis, dialysis training, and home dialysis treatment—and nine inputs. Technical efficiency scores were estimated by DEA, using an input-oriented variable-returns-to-scale model (known as the BCC model). A binary variable of efficiency status was then regressed (multivariate logistic regression analysis) against market and facility characteristics and control factors. A significant association was found between efficiency and the type of ownership, with non-profits and government-owned facilities to be more likely inefficient producers. On the other hand, the membership in large chains was associated with decreasing efficiency, which suggests that growing consolidation in the industry may not be a strategy for gaining better technical efficiency. A limitation of the study is the absence of the case-mix measure.

Hollingsworth (2003) reviewed 188 papers on frontier efficiency measurement in health care. The review mainly included DEA-based methods that dominated the literature, but there was increasing number of parametric techniques, such as stochastic frontier analysis. Results confirmed that the public provision demonstrated less variability in efficiency than the private one and contradicted the hypothesis that the private market provision of services is more efficient than the public provision. It may be also important that the European hospitals showed less variability in efficiency than the U.S. hospitals. However, the incomparability of studies, the inability to measure the real output of the health care industry are big questions to be addressed in future research.

Dlouhý and Pánková (2003) investigated overall hospital performance over a period of four years. They sought to identify potential inefficiencies and overall trends in efficiency of hospitals in the region of West Bohemia, the Czech Republic, 1996–1999. They used two inputs: labor and capital. Medical staff represented the labor. This input was defined as the sum of the number of physicians and that of paramedical personnel, which were both assessed in terms of the annual average of full time equivalents (the inpatient-care departments only). The average bed capacity was used as a proxy measure of the capital. A single output measure of inpatient care provided by the hospital was the annual number of admissions. They assumed there were no technological changes during the observed period. An intertemporal production frontier was estimated. Since the demand for inpatient care (output) was considered as given, a natural objective of the local government or hospital management is the input minimization. That led to application of the input-minimization DEA model. The findings suggested that there were potential savings relating to the operation of hospitals. However without any quality measure of care or of case-mix, the high productivity could mean, in reality, a low quality of care or overloaded medical staff. The overall hospital performance decreased during the observed period with a possible reversal in the last year. A change in the reimbursement mechanism was suggested as one of the possible explanations of that trend. The introduction of strict budgeting in July 1997, which replaced less regulated fee-for-service system, resulted in a significant impact on the behavior of hospitals. Trends for individual hospitals, however, showed large variations that could not be easily interpreted.

Dlouhý (2001) discussed DEA as an alternative tool for a health-system performance evaluation. The work was inspired by the *World Health Report 2000* (World Health Organization, 2000), which was presented as an expert analysis of the influence of health systems in the daily lives of people. Drawing from a range of experiences and analytical tools, the *Report* traced the evolution of health systems, explored their diverse characteristics, and uncovered a unifying framework of shared goals and functions. Using this as a basis for analysis, the *Report* broke a new ground in presenting for the first time an index of national health system's attainment and an index of performance relative to potential. The measures developed in the *Report* were based on three health-system goals: the level and distribution of health, the responsiveness of the system to the legitimate expectations of the population, and fairness of contribution to financing the health system. An important objective was to

see what one can learn from the best. The *Report* aims to stimulate a debate about better ways of measuring health systems performance and thus finding a successful new direction for health systems to follow.

The *Report* identified five objectives (achievements) of the health system whose relative weights were obtained through a survey of 1,006 respondents (Table 2).

Table 2: Measures of Health System Attainment and Their Relative Weights

Objective (measure)	Weights
Health - level (disability-adjusted life expectancy)	25.0%
Health - distribution (index of equality of child survival)	25.0%
Responsiveness - level (composite score)	12.5%
Responsiveness - distribution (score)	12.5%
Fairness of financial contribution (index)	25.0%

The *attainment* considers the level of outcomes, the *performance* considers the resources used. In the *Report*, resources are measured by per capita health expenditures in international dollars. Ranking of countries in terms of their health system performance suggests that, at very low levels of expenditure, performance is worse and more varied than at high spending levels, even when performance is judged relative to countries' human resources and health expenditures. The findings of the *Report* also showed that while much achievement depends greatly on how much a health system spends, it was possible to achieve health quality, respect for persons, and financial fairness even at low resource levels.

The *Report* suggests that assessing of *relative performance* requires a scale, one end of which establishes an upper limit or *frontier*, corresponding to the most that could be expected of a health system. This *frontier* represents the level of attainment which a health system might achieve, but which no country surpasses. As we know the previous chapters, similar terms—*relative performance* and *frontier*—are used in DEA. Both methodologies can deal with multiple inputs and outputs; both can also offer estimation of performance score and of the (production) frontier. On the other hand the approaches differ in the possibility of country's individual weights, which is allowed in DEA, but not allowed in the *Report*. Paradoxically, the *Report*, when describing the stewardship function, emphasized the role of democratic government to select national preferences (understand preferences as weights) to the objectives; but the *Report* evaluated health systems on the basis of universal weights for all countries. Therefore, information on individual weights, which may be obtained by DEA, should be of great interest. Another advantage of DEA is a possibility to determine for each inefficient health system a group of peers. The group of peers consists of countries with the same preferences, but with higher levels of health system performance.

The data on 191 countries came from the statistical annex of the *World Health Report 2000*. The CCR model with five outputs (Table 2) and one input (per capita health expenditures in international dollars) was used. The Czech Republic was found to be one of 50 countries with efficient health systems. This should be understood in the way that there are weights (preferences) according to which the Czech Republic has the best health system in the world.

4 Conclusion

The strengths of DEA are: (1) the method is able to deal with multiple inputs and outputs; (2) For technically inefficient unit DEA identifies the peers that are real production units, so the efficiency improvements should be attainable; (3) software for DEA is now available, which makes it is easy to carry out all calculations and present results. The main weaknesses of DEA are: (1) a deterministic nature of method. Because DEA does not include an error term in the model, there is possibility of a measurement error; (2) hypothesis testing is much less developed in DEA in comparison to econometric analysis. That increases the danger that researcher will include variables that are not

important and exclude important ones, choose a wrong type of model, or select a wrong model orientation (input/output).

Efficiency evaluation in health services is a theoretical as well as practical problem. DEA proved to be a very attractive method of efficiency evaluation for health economists throughout the world. DEA is now mentioned in the textbooks of health economics (e. g., Folland, Goodman, Stano, 2004) and health services research (e. g., Wan, 1995). There is no doubt that the number of applications in the health services will be growing.

The work was supported by project no. 402/03/1360 "Models of multiple criteria decision making and data envelopment analysis as support tools for economic decisions", funded by the Grant Agency of the Czech Republic

References

[1] Charnes, A., Cooper W. W., Lewin, A. Y, Seiford, L. M. (Eds.), 1994. *Data envelopment analysis: theory, methodology and applications*, Boston: Kluwer Academic Publishers.

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

[3] Cooper, W. W., Seiford, L.W., Tone, K., 2000. *Data envelopment analysis*. Boston: Kluwer Academic Publishers.

[4] Dlouhý, M., 2001. An alternative of health system performance measurement in the context of World Health Report 2000. *The Seventh International Conference on Theoretical and Practical Aspects of Public Finance 2001*, Prague: University of Economics Prague.

[5] Dlouhý, M., Pánková, V., 2003. Hospital performance and trends, In M.S. Rauner, K. Heidenberger (Eds.), *Quantitative Approaches in Health Care Management*. Frankfurt am Main: Peter Lang.

[6] Farrell, M. J., 1957. The measurement of productive efficiency of production. *Journal of the Royal Statistical Society*, Series A, 120(III):253–281.

[7] Folland, S., Goodman, A. C., Stano, M., 2004. *The economics of health and health care*. Fourth Edition, Upper Saddle River: Prentice Hall.

[8] Førsund, F. R., Sarafoglou, N., 2002. On the origins of data envelopment analysis. *Journal of Productivity Analysis*, 17:23–40.

[9] Hollingsworth, B., 2003. Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, 6:203–218.

[10] Jablonský, J., Dlouhý, M., 2004. *Metody hodnocení efektivnosti produkčních jednotek* [Methods of efficiency evaluation of production units]. Prague: Professional Publishing.

[11] Ozgen, H., Ozcan, Y. A., 2002. A national study of efficiency for dialysis centers: An examination of market competition and facility characteristics for production of multiple dialysis outputs. *Health Services Research*, 37(3):711–732.

[12] Tavares, G., 2002. *A bibliography of data envelopment analysis (1978–2001)*. RUTCOR Research Report 01-02, Piscataway, NJ: Rutgers University. Retrieved from http://rutcor.rutgers.edu/pub/rrr/reports2002/1_2002.pdf

[13] Wan, T. T. H., 1995. *Analysis and evaluation of health care systems: an integrated approach to managerial decision making*. Baltimore: Health Professions Press.

[14] World Health Organization, 2000. *The World Health Report 2000—Health systems: improving performance*. Geneva.

An Application of Quasi-Hierarchy Approach to Decision Making Under Uncertainty

Cezary Dominiak
Department of Operations Research
University of Economics, Katowice, Poland

Abstract

Scenario planning is very popular and useful technique for decision aiding under uncertainty. The decision analysis problem became complex when it is necessary to take into consideration more than one criterion. In this paper the multicriteria decision aiding procedure under uncertainty is proposed. Our proposal may be used in such cases when decision maker can give information about his preferences in the way of hierarchy of goals with tolerance limits. The proposed procedure is illustrated by simple numerical example of strategic decision-making problem in the food processing enterprise.

Keywords

Decision analysis, scenario planning, multicriteria decision under uncerntainty, strategic decision making

INTRODUCTION

Decision making under uncertainty is a very important area of decision theory. Uncertainty implies that in the certain situation a person does not possess the information which quantitatively and qualitatively is appropriate to describe, prescribe or predict deterministically and numerically a system, its behavior or other characteristics [Zimmerman 2000]. Thus uncertainty relates to a state of the human mind i.e. lack of complete knowledge about something [Stewart 2004].

In earlier works term "Risk" was applied to the situations in which probabilities of outcomes are known objectively, recently term "Risk" means a chance of something bad happening (Fishburn 1984). The term "Uncertainty" is applied to the problems in which exist alternatives with several possible outcomes.

The sources of uncertainty may be divided into two main groups: internal sources of uncertainty and external ones. Internal sources of uncertainty are created by imprecision of human judgments concerned with specification of preferences or values or to assessment of consequences of actions [Stewart 2004]. In the MCDA approach we can find a wide range of methods and techniques to deal with uncertainty created by internal factors: **s**ensitivity analysis (e.g. [Rios Insua 1990]), fuzzy set approach (e.g. : [Klir and Fogler 1988], [Chang and Wang 1997], rough set approach (e.g. [Greco, Matarazzo, Slowinski 1999].

External uncertainty refers to lack of knowledge about the consequences of our choices [Stewart 2004]. For those types of problems the following methods are applied: probabilistic models and expected utility (e.g. [Kahnemann and Tversky 1976], [Bazerman 2002] [Rosquist 2001]), pairwise comparisons based on stochastic dominance (e.g. [D'Avignon and Vincke 1988], [Martel and Zaras 1995]). The risk measures as surrogate

criteria are also applied (e.g. [Millet and Wedley 2002], [Sarin and Weber 1993] [Jia and Dyer 1996]. In such problems where we have to take into account external uncertainty the scenario planning approach may be applied (e.g.: [Klein G., Moskowitz H., Ravindran A 1990], [Goodwin and Wright 1997], [Pomerol 2001], [Urli and Nadeau 2004]).

In this paper we propose the hierarchy and quasi-hierarchy approach to multi criteria decision making problems under uncertainty which is based on the scenario planning.

2. SCENARIO PLANNING

Scenario planning was developed as a technique for facilitating the process of identifying uncertain and uncontrollable factors, which may impact on the consequences of decisions in the strategic management context. Scenario analysis is widely accepted as important component of strategic planning. Scenario planning may be treated as a process of organizational learning, distinguished by an emphasis on the explicit and ongoing consideration of multiple futures. Five principles, which should guide scenario construction:

- at least two scenarios are required to reflect uncertainty,
- each scenario must be plausible, meaning it can be seen to evolve in a logical manner from the past and present
- each scenario must be internally consistent
- scenarios must be relevant to the DM's concerns and they must provide a useful comprehensive and challenging framework against which the DM can develop and test strategies and action plans
- The scenarios must produce a novel perspective on the issues of concern to the DM (Van der Heijden 1996**)**

3. Proposed procedures for decision aiding

3.1 Problem formulation

We consider the "traditionally understood" problem of decision making under uncertainty, so we assume we don't know the probabilities of the states of nature. Thus we have discrete set of alternatives and discrete set of scenarios have been selected for the purposes of evaluating alternatives. For single criterion problems we can apply decision rules, but here we consider the existence of more than one criterion.

Let

n – denotes the number of alternatives,
m – denotes the number of scenarios,
K – is the number of criteria

For the simplicity let us assume that values of all criteria are maximized. Let vector

$$A_i^k = [a_{i1}^k, \ldots, a_{im}^k]$$

denotes the vector of values of k-th criterion of alternative i-th. Matrix

$$A^k = [a_{ij}^k]_{n \times m}$$

is matrix which consist of n vectors A_i^k, and this matrix shows values of k-th criterion for all alternatives in each considered scenario. Moreover let

$$\overline{a}_i^k = \min_{j=1,\ldots,m} a_{ij}^k$$

is the worst value of k-th criterion for i-th alternative.

If DM is able to formulate his preferences in the form of order of criteria the hierarchy and quasi-hierarchy approach may be applied. The proposed procedures are based on traditional max-min decision rule.

3.2 The Hierarchy Approach

Let us assume that DM is able to order criteria from the most important to the least important. The lower number of criterion means the higher importance.

$$k_1 \rangle k_2 \rangle \ldots\ldots \rangle k_K$$

Step 1

$$\widetilde{A}^{k_1} = \{A_i : \max_{i=1,\ldots m} \overline{a}_i^{k_1}\}$$

Step t (t=2,...K)

$$\widetilde{A}^{k_t} = \{\widetilde{A}^{k_{t-1}}{}_i : \max_{i=1,\ldots m} \overline{a}_i^{k_t}\}$$

Example 1

We consider the problem with four alternatives. Each alternative is described by three criteria and four scenarios are taken into consideration. The tables A^1, A^2, A^3 show the values of criterion for each alternative and scenario.

A^1:

	S_1	S_2	S_3	S_4	Min
A_1	11	**10**	13	15	**10**
A_2	10	**8**	14	13	8
A_3	15	13	12	**10**	**10**
A_4	12	10	**8**	**7**	7

A^2:

	S_1	S_2	S_3	S_4	Min
A_1	10	15	**9**	13	9
A_2	13	**11**	12	15	11
A_3	16	**13**	14	20	**13**
A_4	18	19	17	**16**	16

A^3:

	S_1	S_2	S_3	S_4	Min
A_1	**10**	11	14	15	10
A_2	18	16	**12**	13	12
A_3	**14**	15	19	20	14
A_4	16	**15**	17	19	15

In the first step (criterion k_1) we delete alternatives A_2, A_4 from the set of alternatives. Then (on the base of values of criterion k_2) we conclude that alternative A_3 should be suggested as a decision. After the second step of procedure the set of alternatives consisted of one object thus the third criterion did not influence on the final decision at all. This is typical disadvantage of the hierarchy approach which may be relaxed by using the tolerance limits and the quasi-hierarchy.

3.3 Quasi-hierarchy Approach

DM describes the tolerance limit for each criterion. Let q_t denotes the tolerance limit described for t^{th} criterion. Thus in the first step we should find the subset which is defined as follows:

Step 1

$$\widetilde{A}^{k_1} = \{A_i : \overline{a}_i^{k_1} \geq \max_{i=1,\ldots m} \overline{a}_i^{k_1} - q_1\}$$

Step t (t=2,...K)

In steps $k_2,\ldots,k_K$ we find the following subsets of the set of alternatives

$$\widetilde{A}^{k_t} = \{\widetilde{A}^{k_{t-1}}{}_i : \overline{a}_i^{k_t} \geq \max_{i=1,\ldots m} \overline{a}_i^{k_t} - q_t\}$$

Example 2

We consider the data from example 1. Let us assume that Decision Maker described the tolerance limits as follows: $q_1=2$, $q_2=3$, $q_3=1$.

Thus in the first step of procedure we delete alternative A_4 from the set of considered alternatives. Then we take into account the second criterion and alternative A_1 is deleted. In the last step we compare the alternatives A_2 and A_3 and on the basis of values of the third criterion we conclude that alternative A_3 should be suggested as a final decision.

4. Conclusions

In this paper we discussed the problem of decision making under uncertainty and the main approaches to MCDA under uncertainty were shortly described. The scenario planning is recognized as a useful technique to deal with uncertainty, especially for the strategic management decisions.

The hierarchy and quasi – hierarchy procedures were presented which may be used for multicriteria decision aiding under uncertainty. Proposed procedures were illustrated by simple numerical examples.

In our opinion proposed procedures may be easy applied and may be useful to support the decision aiding process in real-life strategic management.

References:

[Bazerman 2002] Bazerman M.H.: *Judgment in Managerial Decision Making*, John Wiley & Sons, New York 2002.

[Chang and Wang 1997] Chang N.B., Wang S.: *A fuzzy goal programming approach for the optimal planning of metropolitan solid waste management systems*. European Journal of Operations Research, 99:303-321,1997.

[D'Avignon and Vincke 1988] D'Avignon G.R.D and Vincke P.: *An outranking method under uncertainty.* European Journal of Operations Research, 36: 311-321, 1988.

[Fishburn 1984] Fishburn P.C.: *Foundations of risk measurement. I. Risk or probable loss*. Management Science, 30:396-406,1984.

[Goodwin and Wright 1997] Goodwin P. and Wright G.: *Decision Analysis for Management Judgement*, John Wiley & Sons,1997.

[Greco, Matarazzo, Slowinski 1999] Greco S., Matarazzo B., Slowinski R.: *Rough sets theory for multicriteria decision analysis.* European Journal of Operations Research, 129:1-47, 2001.

[Jia and Dyer 1996] Jia J. and Dyer J.S.: *A standard measure of risk and risk-value models*, Management Science, 42:1691-1705, 1996.

[Kahnemann and Tversky 1976] Kahnemann D. and Tversky A.: *Prospect theory An decision analysis under risk*, Econometrica 47:263-291, 1976.

[Klir and Fogler 1988] Klir G.J. and Fogler T.A.: *Fuzzy sets. Uncertainty and Information*. Prentice Hall, New Jersey 1988.

[Klein 1990] Klein G., Moskowitz H., Ravindran A.: *Interactive multiobjective optimization under uncertainty*. Management Science, 36: 58-75, 1990.

[MARTEL AND ZARAS 1995] MARTEL J.M AND ZARAS K. *Stochastic dominance in multicriterion analysis under risk. theory and decision*, 39:31-49, 1995.

[Millet and Wedley 2002] Millet I. and Wedley W.C.: M*odeling risk and uncertainty with the ahpAHP.* Journal of Multi-Criteria Decision Analysis, 11:97-107, 2002

[Pomerol 2001] Pomerol J.C.: *Scenario Development and Practical Decision Making Under Uncertainty. Decision Support Systems*, 31: 197-204, 2001.

[Rios Insua 1990] Rios Insua D. : *Sensitivity Analysis in Multi Objective Decision Making*, Lecture Notes in Economic and Mathematical System, vol. 347, Springer, Berlin 1990.

[Rosquist 2001] Rosquist T.: *Simulation and multi-attribute utility modeling of life cycle profit.* Journal of Multi-Criteria Decision Analysis, 10:205-218, 2001.

[Sarin and Weber 1993] Sarin R.K. and Weber M.: *Risk-value models*. European Journal of Operations Research, 70:135-149, 1993.

[Stewart 2004] Stewart D.J.: *Uncertainties in MCDA*, in: Multi Criteria Decision Analysis eds Greco P., Springer 2004.

[Urli and Nadeau 2004] Urli B. and Nadeau R.: *PROMISE/scenarios: An interactive method for multiobjective stochastic linear programming under partial uncertainty*. European Journal of Operations Research, 155:361-372, 2004

[Van der Heiijden 1996] Van der Heiijden K.: Sceanrios: *The Art of Strategic Conversation*, John Wiley & Sons,1996

[Zimmermann 2000] Zimmermann H.: An *application-oriented view of modeling uncertainty*. European Journal of Operations Research, 122:190-198, 2000.

Properties of the S-shape Value Function and its Applications

Renata Dudzińska and Donata Kopańska-Bródka
Department of Operations Research
University of Economics in Katowice, Poland

Abstract

In decision making an attitude toward risk of the decision maker is very important. In decision theory some attitudes are distinguished: risk aversion, risk seeking and risk neutral, but the research of the decision makers' behaviors indicates that the attitude toward risk is changeable. Many people can be both risk averters and risk seekers, depending on the range of monetary values being considered and an individual's attitudes can vary over time. The value functions for such decision makers' are the s-shape curves. Properties of the s-shaped value functions differ from the concave-shaped function. According to Prospect Theory, the value function by which investors make investment decision is an s-shaped function depending on his preference and his wealth. In the paper we will show some properties of the s-shape value function and its application in investment decisions.

Keywords

Prospect theory, s-shape value function, portfolio

1 Introduction

Decision makers during the decision process take into account various criteria. In XVIII century Daniel Bernoulli [2] has proposed to choose that decision which maximizes the utility. He assumed that the utility function is logarithmic [4], because this function satisfies the assumption of diminishing marginal utility observed in real decision problems.

The Bernoulli's concept applies to action with sure outcome, but it hasn't take into account the preferences to risky actions or probability of outcomes. The next step was done by von Neumann and Morgenstern, who have given axioms of utility theory [10, 7, 8] and they have showed that the concept of the expected utility is based on the preferences of rational decision makers.

In the empirical research on the criteria of rational decision making, the researchers found that decision makers don't make their decisions according to the theory of the expected utility (see Allais paradoxes [1]). They change their preferences to risky actions depending on the circumstances. The first utility functions which meet changeable attitude to risky actions were proposed by Friedman and Savage [3] and Markowitz [6]. These functions are concave in some interval of wealth and convex in the other.

In all above concepts the domain of utility functions is level of wealth. The next step in the development of utility theory was the work of Kahneman and Tversky [5]. They named their concept as Prospect Theory and they have stated that decision makers evaluate the decision depending on their actual level of wealth and on that if the decision will bring gain or loss. The decision makers are risk avers if the decision brings gain and they are risk prone if the decision makes loss. In 1992 Tversky and Kahneman published the extension of Prospect Theory which they named Cumulative Prospect Theory [9]. It applies to any finite prospect and can be extended to continuous distributions.

2 Prospect Theory

Kahneman and Tversky in [5] have divided the decision process into two phases: editing and evaluation. In the editing phase decision options are being organized and reformulated. In this phase outcomes are defined as gain and lose relative to some neutral reference point, next they are simplified and decomposed into a riskless and risky component. The prospect is denoted as $(x_1,p_1;\ldots; x_n,p_n)$, which means that prospect yields outcome x_i with probability p_i (i=1,…,n) where $p_1+\ldots+p_n=1$.

In the following phase decision maker evaluate each of the prospects, and choose the prospect with highest value. The value V of the prospect is expressed in terms of two scales, π and v. The scale π assigns to each probability p a decision weight $\pi(p)$. This decision weight reflects the impact of p on the over-all value of the prospect. The scale v assigns to each outcome x a number $v(x)$ which reflects the subjective value of that outcome. It is assumed that $v(0)=0$, $\pi(0)=0$ and $\pi(1)=1$.

The value V of the prospect $(x_1,p_1;\ldots; x_n,p_n)$ is defined in the following manner:

- if $p_1+\ldots+p_n=1$ and all outcomes x_i are strictly positive or strictly negative than
  $$V(x_1,p_1;\ldots; x_n,p_n)=v(x_{min})+\sum_{\substack{i \\ x_i>x_{min}}}\pi(p_i)\left[v(x_i)-v(x_{min})\right] \quad (1)$$
  where $x_{min}=\min\{x_i\}$
- if $p_1+\ldots+p_n<1$ or some outcomes x_i are positive and some are negative or some equal to zero than
  $$V(x_1,p_1;\ldots; x_n,p_n)=\pi(p_1)v(x_1)+\ldots+\pi(p_n)v(x_n) \quad (2)$$
  which is the mean of values $v(x_i)$ weighted by $\pi(p_i)$.

Example 1.

Let's consider the investment in a share. For our share we estimated the outcomes and their probability: we can get the outcome 2% with probability 0.4; the outcome 4% with probability 0.5 and the outcome 5% with probability 0.1. Because the probabilities sum up to 1 and all outcomes are strictly positive, the value of this prospect is calculated as in formula (1):

$V(2\%,0.4; 4\%,0.5; 5\%,0.1) = v(2\%) + \pi(0.5)\,[v(4\%)-v(2\%)] + \pi(0.1)\,)\,[v(5\%)-v(2\%)]$

Example 2.

Let's consider another share, which have the outcome -2% with probability 0.4; the outcome 4% with probability 0.5[1]. Because the probabilities don't sum up to 1 (or one of the outcome equals to 0), the value of this prospect is calculated as in formula (2):

$V(-2\%,0.4; 4\%,0.5) = \pi(0.4)\,v(-2\%) + \pi(0.5)\,v(4\%)$

2.1 The Value Function and its properties

In Prospect Theory outcomes are understood as gains and losses, rather than the final states. Gains and losses are the changes in the wealth from a reference point. This point can be interpreted as a current wealth of a decision maker.

On the basis of their research, Kahneman and Tversky [5] stated that decision makers do not perceive gains and losses in the same manner. They hypothesized that the value function for changes of wealth is concave above the reference point ($v''(x)<0$ for $x>0$) and often convex below this point ($v''(x)>0$ for $x<0$). It is consistent with diminishing marginal value of gains and losses (with the increase of their absolute magnitude). The concavity of the value function for gains means that decision makers are risk avers and they choose certain decision and do not take risky action connected with even higher gain. Whereas considering losses decision makers are risk-prone (convexity of the value function) and they choose to take risky loss than certain loss. Sometimes decision makers reveal opposite attitude, but it is connected with some additional circumstances.

The other very important feature of the value function is that decision makers feel the loss more than the gain of the same absolute value. Thus, the value function is steeper for the losses than for the gains.

[1] It is assumed that we yield the outcome 0 with probability 0.1.

The figure 1 presents the example of the s-shape value function with their properties.

To sum up, the value function in the Prospect Theory is:

- defined for gains and losses,
- concave for gains and convex for losses,
- steeper for losses and than for gains.

2.2 The Weighting Function and its properties

The other scale, in terms of which prospects are evaluated, is represented by the weighting function. Each outcome has its decision weight, which depends on the probability, but is not interpreted as the probability of that outcome. Decision weight measures the impact of event on the desirability of prospect, and not merely the perceived likelihood of this event.

The weighting function π is an increasing function of p and $\pi(0)=0$ and $\pi(1)=1$. Furthermore, for small probabilities $\pi(p)>p$, it means that the decision makers overweight low probabilities. This property Kahneman and Tversky [5] called subadditivity.

The next property of the weighting function is called subcertainty. It states that for all $p\in(0,1)$, $\pi(p)+\pi(1-p)<1$. As a consequence of this property we have that $\pi(p)<p$ for high probabilities.

The figure 2 presents hypothetical weighting function with its properties.

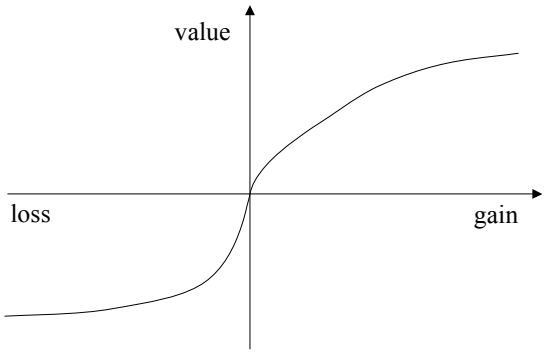


Figure 1. The s-shape value function

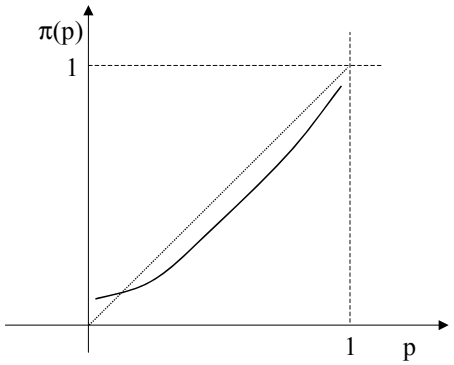


Figure 2. The hypothetical weighting function

3 Empirical research

In our research we have applied the concept of the Prospect Theory to the data from the Warsaw Stock Exchange. All calculations were done on the basis of the daily rates of return for 79 stocks quoted on the market all over the year 2004.

We have assumed that each stock is a prospect with three possible outcomes: mean of positive rates of return, mean of negative rates of return and zero rate of return. For each outcome we have calculated probabilities as a number of positive (or negative or zero) rates of return to the number of all rates of return.

Tversky and Kahneman has proposed [9] to use the following function as a value function

$$v(x)=\begin{cases} x^{\alpha} & \text{for } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{for } x < 0 \end{cases} \tag{3}$$

where

x – the outcome,

α – the parameter equaling to 0.88,

β – the parameter equaling to 0.88,

λ – the parameter equaling to 2.25.

The values of the parameters were assumed on those levels as in the paper [9].

We use the following form of the weighting function proposed by Tversky and Kahneman

$$\pi(p) = \frac{p^{\gamma}}{\left(p^{\gamma} + (1-p)^{\gamma}\right)^{\frac{1}{\gamma}}} \tag{4}$$

where

p – the probability,

γ – the parameter equaling to 0.65.

In the paper [9] the value of the parameter γ were estimated on the two levels depending on that if probability concerns positive outcome or negative one. For positive outcome γ equaled to 0.69 and for negative one equaled to 0.61. In our research we have taken the mean of those two values.

For each prospect (stock) we have calculated expected value of the rate of return, its variability coefficient and the value of the prospect. On the basis of calculations we have build three rankings (table 1).

Table 1. Positions of the stocks in rankings by decreasing expected value $E(R_i)$, increasing variability coefficient γ_i and decreasing value of the prospect $V_i(\mathbf{x},\mathbf{p})$ (no. 1 for the "best" stock)

$E(R_i)$	Stock	γ_i	$V_i(\mathbf{x},\mathbf{p})$
1	VISTULA	**1**	**1**
2	TIM	**2**	22
3	WOLCZANKA	**3**	30
4	PROCHNIK	23	66
5	EKODROB	22	71
6	POLIMEXMS	**6**	11
7	HYDROTOR	**5**	**8**
8	BORYSZEW	**4**	**3**
9	KABLE	12	27
10	ALMAMARKET	11	36
11	MCI	20	56
12	DUDAZM	**7**	**7**
13	ELEKTRIM	24	69
14	GETIN	25	48
15	ELEKTROEX	33	65
16	WANDALEX	**8**	33
17	PAGED	13	42
18	RAFAKO	16	35
19	STALPROFI	10	13
20	SWARZEDZ	30	64
21	ATLANTIS	42	70
22	ELBUDOWA	18	23
23	MOSTALZAB	47	75
24	KRUK	28	47
25	STALEXP	36	68
26	LUBAWA	27	32
27	REMAK	34	54
28	MNI	29	34
29	CERSANIT	**9**	**4**
30	PKNORLEN	14	17
31	KOGENERA	32	46
32	MOSTALEXP	45	62
33	BACA	21	10
34	BPHPBK	19	21
35	IMPEXMET	35	38
36	PGF	15	**9**
37	COMARCH	17	**5**
38	MILLENNIUM	38	37
39	ROLIMPEX	39	39
40	INTERIA	54	67
41	HUTMEN	52	59
42	TPSA	26	16
43	PEKAO	37	26
44	KRUSZWICA	55	51
45	BZWBK	31	19
46	BRE	43	28
47	KGHM	51	58
48	OPTIMUS	58	55
49	ROPCZYCE	50	43
50	COMPLAND	40	18
51	FARMACOL	41	14
52	BUDIMEX	49	31
53	KREDYTB	46	12
54	JELFA	48	15
55	LENTEX	57	41
56	AGORA	59	25
57	NETIA	56	20
58	INGBSK	44	**2**
59	SOFTBANK	60	50
60	HANDLOWY	53	**6**
61	GROCLIN	61	29
62	KETY	62	44
63	WILBO	79	60
64	SNIEZKA	78	24
65	POLIFARBC	77	49
66	ORBIS	74	40
67	KOPEX	76	63
68	PROKOM	73	57
69	IBSYSTEM	75	76
70	PROSPER	71	52
71	KOMPAP	72	73
72	DEBICA	67	45
73	SWIECIE	68	53
74	STRZELEC	70	72
75	REDAN	66	61
76	MIESZKO	69	77
77	HOOP	64	74
78	STERPRO	65	78
79	TRASTYCHY	63	79

For the first ten stocks from each ranking we have build equally weighted portfolios. The first portfolio $\mathbf{P_E}$ consists of shares of VISTULA, TIM, WOLCZANKA, PROCHNIK, EKODROB, POLIMEXMS, HYDROTOR, BORYSZEW, KABLE and ALMAMARKET. The second portfolio $\mathbf{P_\gamma}$ consists of shares of VISTULA, TIM, WOLCZANKA, BORYSZEW, HYDROTOR, POLIMEXMS, DUDAZM, WANDALEX, CERSANIT and STALPROFI. The third portfolio $\mathbf{P_{V(x,p)}}$ consists of shares

of VISTULA, INGBSK, BORYSZEW, CERSANIT, COMARCH, HANDLOWY, DUDAZM, HYDROTOR, PGF and BACA.

We have assumed that each portfolio is bought on 31st of December 2004 and is to be sold in the consecutive 39 working days starting from the 3rd of January 2005 and for those days we have calculated the rates of return. The obtained data are showed in the table 2.

Table 2. Rates of return for portfolios bought on the 31st of December 2004 and sold on the consecutive days.

Date	$\mathbf{P_E}$	$\mathbf{P_\gamma}$	$\mathbf{P_{V(x,p)}}$
01/03/2005	-0,0024	-0,0101	-0,0001
01/04/2005	-0,0026	-0,0097	-0,0023
01/05/2005	-0,0164	-0,0277	-0,0093
01/06/2005	-0,0244	-0,0328	-0,0149
01/07/2005	-0,0258	-0,0384	-0,0187
01/10/2005	-0,0261	-0,0454	-0,0209
01/11/2005	-0,0405	-0,0680	-0,0274
01/12/2005	-0,0500	-0,0702	-0,0289
01/13/2005	-0,0471	-0,0618	-0,0265
01/14/2005	-0,0553	-0,0702	-0,0340
01/17/2005	-0,0585	-0,0690	-0,0350
01/18/2005	-0,0839	-0,0800	-0,0345
01/19/2005	-0,0747	-0,0748	-0,0342
01/20/2005	-0,0698	-0,0842	-0,0371
01/21/2005	-0,0623	-0,0794	-0,0369
01/24/2005	-0,0651	-0,0832	-0,0421
01/25/2005	-0,0635	-0,0831	-0,0404
01/26/2005	-0,0654	-0,0781	-0,0335
01/27/2005	-0,0617	-0,0765	-0,0261
01/28/2005	-0,0580	-0,0724	-0,0250
01/31/2005	-0,0471	-0,0593	-0,0202
02/01/2005	-0,0241	-0,0522	-0,0162
02/02/2005	-0,0346	-0,0507	-0,0134
02/03/2005	-0,0534	-0,0582	-0,0198
02/04/2005	-0,0388	-0,0528	-0,0121
02/07/2005	-0,0324	-0,0582	-0,0188
02/08/2005	-0,0512	-0,0638	-0,0278
02/09/2005	-0,0605	-0,0658	-0,0267
02/10/2005	-0,0685	-0,0705	-0,0319
02/11/2005	-0,0679	-0,0697	-0,0309
02/14/2005	-0,0833	-0,0662	-0,0291
02/15/2005	-0,0920	-0,0701	-0,0314
02/16/2005	-0,0899	-0,0707	-0,0207
02/17/2005	-0,0835	-0,0591	0,0030
02/18/2005	-0,0897	-0,0673	-0,0027
02/21/2005	-0,0844	-0,0702	-0,0025
02/22/2005	-0,0843	-0,0703	0,0025
02/23/2005	-0,0617	-0,0658	0,0006
02/24/2005	-0,0650	-0,0593	0,0078

Almost all (except for four) of the rates of return are negative. It's due to the general tendency on the Stock Exchange in the beginning of the 2005 year. Comparing rates of return of three portfolios for each day it's worth noting that portfolio $\mathbf{P_{V(x,p)}}$, consisting of the "best" shares according to the value of the prospect, has the highest rate of return.

4 Conclusion

The aim of our research was an attempt of assessment of shares quoted on the Warsaw Stock Exchange, no the basis of assumptions of Prospect Theory and the comparison with its assessment according to the mean-variance approach.

In order to apply the concept of Prospect Theory we have taken the construction of the probability distribution of the rate of return dividing them into desired (generating gains), undesired and neutral.

In our research equally weighted portfolio consisting of the shares with the highest value of prospect (according to Prospect Theory) had the highest rate of return comparing to other portfolios.

References

[1] Allais M., Hagen O. (Eds). (1979). Expected Utility Hypotheses and the Allais' Paradox. Reidel Publishing Company. Dordrecht.

[2] Bernoulli D. (1954). Exposition of a new theory on the measurement of risk. Econometrica. Vol. 22. pp. 23-36. (English translation).

[3] Friedman M., Savage L.J. (1948). The Utility Analysis of Choices Involving Risk. Journal of Political Economy, Vol. 56, pp. 279-304.

[4] Jensen N.E. (1967). An introduction to Bernoullian utility theory. Swedish Journal of Economics. pp. 163-183.

[5] Kahneman D., Tversky A. (1979). Prospect Theory: An Analysis of Decision Under Risk. Econometrica. Vol. 47, pp. 263-291.

[6] Markowitz H. (1952). The Utility of Wealth. Journal of Political Economy. Vol. 60, pp. 151-158.

[7] Schoemaker P.J.H. (1982). The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. Journal of Economic Literature. Vol. 20, pp. 529-563.

[8] Starmer C. (2000). Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk. Journal of Economic Literature. Vol. 38, pp. 332-382.

[9] Tversky A., Kahneman K. (1992). Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty. Vol. 5, pp. 297-323.

[10] von Neumann J., Morgenstern O. (1944). Theory of Games and Economic Behavior.

Microeconomic Analysis of Equilibrium in Network Industries in Context of Influence of Regulated Prices

Eleonora Fendeková and Michal Fendek
University of Economics, Bratislava, Slovakia

Abstract

The aim of the paper is to examine the equilibrium conditions in the market of network industries. With regard to analysis of equilibrium in network industries models it is important to point out that except for competition policy protection the state fulfils another specific task – regulation of network industries. The state influences proportional relations between price and supply of network industry production.

The conditions for equilibrium of network industries and methods of their regulations will be examined in the paper. The stress will be laid on the regulation on the base of returns – Rate of Return Regulation (ROR). Attention will be paid to the ways of calculation reasonable profit in regulated industries.

Keywords

Network industries, regulated prices, reasonable profit in regulated industries, rate of return regulation, Averch-Johnson model.

1 Introduction

In connection with the transition of the Slovak economy to a market system much attention has been drawn in the Slovak economic theory to the study and analysis of general principles of functioning of market economy structures. In more analytically oriented studies the attention was paid to the observation of general economic equilibrium models of market environment. With regard to the above-mentioned trends a significant place has been (and still is) assigned to the microeconomic equilibrium models of market structures.

Microeconomic market equilibrium models are not a general cure for all, often objectively very sensitive issues of this transition. A branch with its production potential of limited number of producers on the supply side accounts for an ideal object for application of branch equilibrium models. By analysis of branch structures of national economics network industries should bear a special care, because they represent, in many cases, key production and service branches.

The existence of pure monopoly in network industries increases the role of regulation mechanisms in connection with objectification and increase in their social effectiveness. The objective of regulation mechanisms is to find an reasonable proportion between price and product supply of network industry under assumption of the existence of a competitive market.

Each condition of disequilibrium could have a negative impact on the existing conditions as well as on the potential development of the economic system. A perfect competition as a theoretical concept of market structure is only an ideal hypothesis of a “fair” market environment in current developed economics, when all producers and customers have approximately similar hope for a successful performance of their economic activities. One of the most important attributes of a harmonically developing economic system is a state-guaranteed protection of competition policy principles. It is obvious that the existence of market structures of imperfect competition evokes negative effects for competition policy.

It is inevitable to say that theories about positive social and economic effects of concentration, like:

reduction of production cost;
support of innovative processes and technical development;
increasing of employment;
growth of living standard condition and social benefits of employees;
boost of state budget revenues

have its reasoning only in short-term period. In long-term course their concrete effect is usually indemonstrable.

With regard to analysis of equilibrium in network industries models it is important to point out that except for competition policy protection the state fulfils another specific task – regulation of network industries. The state influences proportional relations between price and supply of network industry production.

From a social viewpoint the problem lies in a fact, that monopolies realise a maximum profit at higher prices and lower level of output than companies operating in competition environment. The main aim of regulation of monopolies is to bring a proportion between price and supply closer to a situation that would occur in a competition market.

Under regulation we more and more understand a regulation in a wider term, it means, determining of market structure, including ownership, regulation of natural monopolies behaviour.

Regulation of pure monopolies with regard to the May 2004 entry of Slovak Republic into the European Union has become a discussed issue.

2 Rate of Return Regulation

A traditional methodological tool for price regulation, which is applied by price regulators at determining the price of products of network industries is the regulation on the basis of the rate of return, through which the production prices in the most developed economies are regulated, i.e. of electricity, gas and other companies.

The aim is to ensure the regulated subject to determine the price of production or services for his customer so that he can cover from its revenue all its *reasonable and providently arisen costs* as well as the regulated recoverability of its *provident* investment.

Let us now derive the allowable rate of cost return for investments *RoR* of the regulated firm analytically. Let us suppose that the firm produces a homogenous product in production volume q, which it realizes at a relevant market for the price p. Let us further suppose that the firm uses two production factors, namely labour force with consumption level L by the labour price w and the capital with consumption level K by the capital price r.

The profit of the firm is generally defined as the difference between yields and costs in form

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

where
$t(q) = p \times q$ – function of proceeds of the firm, $t: R \rightarrow R$
$n(q) = n_v(q) + n_F$ – function of the total costs of the firm, $n: R \rightarrow R$
$n_v(q)$ – function of variable costs of the firm, $n_v : R \rightarrow R$
n_F - fixed costs of the firm, $n_F \in R$

If we substitute the general cost function by a cost function on the basis of consumption of production factors, we get a profit function in the following form

$$\pi(q) = p \times q - w \times L - r \times K$$

If we further express the production volume q on the basis of the production function in the form

$$q = f(K, L)$$

and the production price p on the basis of the price-demand function in the form

$$p = p(q)$$

so we can express the profit function in the form

$$\pi(q) = p(q) \times q - w \times L - r \times K$$

and after further modification in the form

$$\pi(q) = p(f(K, L)) \times f(K, L) - w \times L - r \times K$$

A non-regulated firm can set its endogenous decision parameters in any way. So it chooses an optimum output volume q^*, an acceptable optimum price p^* and corresponding consumptions of the production factors labour L and capital K in order to reach the maximum profit. It calculates the optimum output and optimum price by solving the following task of mathematical programming

$$\pi(q) = p(f(K, L)) \times f(K, L) - w \times L - r \times K \rightarrow \max$$
$$K, L \in R_{\geq 0}$$

So in this case the non-regulated firm has no formal obstacles for setting the parameters guaranteeing its maximum profit. On the other hand the regulated form must respect the boundaries defined by the regulator. The mode of price regulation on the basis of the rate of return consists in the fact that through the exogenously defined control variable *RoR* the allowable level of the quotient of the proceeds of the firm $p \times q$ reduced by its non-capital expenditures $L \times w$ and of the volume of consumed capital K is regulated.

The firm can optimise, respectively freely determine the consumption levels of labour L, capital K by the market prices of production factors w, r on one side and on the other side the level of its production q and the production price p. But the firm must respect the rate of return defined by the regulator, i.e. the validity of the relation

$$RoR \geq \frac{p \times q - w \times L}{K} \tag{1}$$

Let us now explore more in detail the relation between the rate of return of the capital expenditure and the profit of the regulated subject. The profit can be analytically expressed as the difference between the proceeds and the costs of the firm in the form

$$\pi(q) = p \times q - w \times L - r \times K \tag{2}$$

Let us deduct from both sides of the relation (1) the price of the capital r. We get the relation

$$RoR - r \geq \frac{p \times q - w \times L}{K} - r$$

After another modification we get

$$RoR - r \geq \frac{p \times q - w \times L}{K} - \frac{r \times K}{K}$$

$$RoR - r \geq \frac{p \times q - w \times L - r \times K}{K} \tag{3}$$

From the comparison of the relations (2) and (3) we get the relation

$$RoR - r \geq \frac{\pi(q)}{K}$$

$$(RoR - r) \times K \geq \pi(q) \tag{4}$$

We can see from the relation (4), that the regulated subject can set its system parameters only so, that its reached profit does not exceed the value of the capital evaluated by the difference between the rate of return defined by the regulator *RoR* and the price of capital *r*.

The regulated firm can set its controlled, respectively endogenous decision parameters only in a way in which it respects the condition determined by the regulator. It determines the regulated volume of output q_R, the acceptable regulated price p_R and the corresponding consumptions of production factors labour L and capital K so that it reaches the maximum profit and at the same time it respects the condition of the regulator (4) about not-exceeding the reasonable profit level. The regulated output and the regulated price are calculated by the solution of the following task of mathematical programming

$$\pi(q) = p(f(K,L)) \times f(K,L) - w \times L - r \times K \rightarrow \max \tag{5}$$

by the boundaries

$$p(f(K,L)) \times f(K,L) - w \times L - r \times K - (RoR - r) \times K \leq 0 \tag{6}$$

$$K, L \in R_{\geq 0} \tag{7}$$

The solution of this optimisation task is the optimum level of consumption of the production factors labour L^* and capital K^*, on the basis of which subsequently the regulated optimum level of output q_R^* is quantified using the production function based on the relation

$$q_R^* = f(K^*, L^*)$$

and the regulated optimum price p_R^* with the use of the price-demand and production function based on the relation

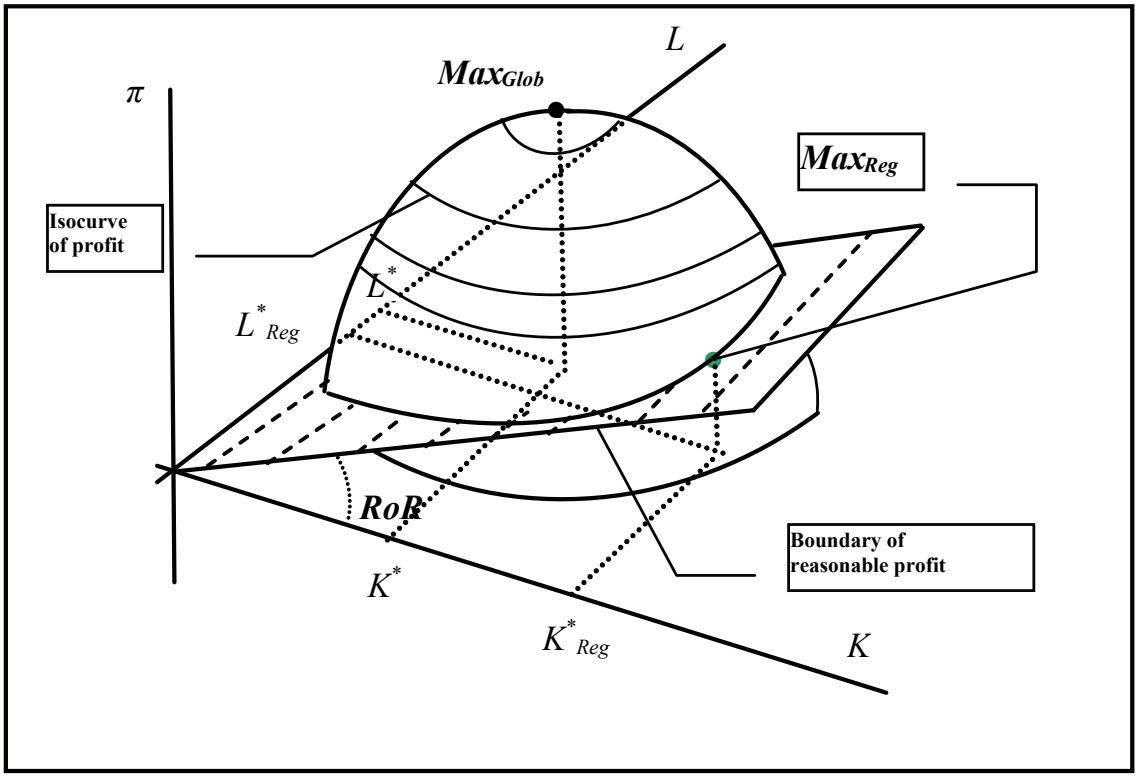
$$p_R^* = p(q_R^*) = p(f(K^*, L^*)$$

whereby the rate of return of the capital of the firm defined by the parameter *RoR,* i.e. the exogenous control parameter determined by the regulator is respected.

We show the geometrical interpretation of determining the optimum regulated price and the supply of the firm in the conditions of respecting the determined rate of return of the capital in Figure 1.

In the situation when the firm would not be regulated and it would have an exclusive position at a relevant market, it would choose such an optimum volume of consumption of the variable input *labour* L^* and *capital* K^*, that would ensure a maximum profit $\pi(q) = p(f(K^*, L^*)) \times f(K^*, L^*)$. On the basis of the optimum consumption of variable inputs it would determine its optimum supply $q^* = f(K^*, L^*)$ and the optimum price of production $p^* = p(q^*) = p(f(K^*, L^*)$. The function of profit would reach in this case its global maximum, to which in Figure 4 corresponds the point $\boldsymbol{Max_{Glob}}$.

Figure 1



In case the firm is regulated, it can choose only such a combination of production factors, so that the corresponding volume of supply and price of production generates the so-called *reasonable profit*, i.e. this relation is valid

$$(RoR - r) \times K \geq p(f(K, L)) \times f(K, L) - w \times L - r \times K$$
$$(RoR - r) \times K \geq \pi(q)$$

So in the end result the regulated form can produce in such a way that its *reasonable profit* does not exceed *RoR – r* multiple of the level of the variable input capital. This condition is called the *boundary of reasonable profit of a regulated firm* in professional literature.

On Figure 1 the *boundary of reasonable profit* is represented as a plane passing the coordinate axis of the production factor labour *L (K=0)* and *RoR - r> 0* represents the *tangent* of the angle, which the plane of boundary forms with the positive half-axis *K*. The bounding plane intersects the elipsoid of

the values of the profit of the firm, whereby the points below this plane and on this plane represent the profit allowable under the regulation and the points above this plane the non-allowable profit of the firm.

The regulated form maximizes its profit by the solution of the optimisation task (5), (6), (7). On Figure 1 the point of maximum allowable profit of the regulated firm is the point $\boldsymbol{Max_{Reg}}$, whereby this maximum allowable profit is reached by the regulated firm for the optimum regulated value of the production factor, from which it produces the optimum regulated volume of output $q^*_R = f(L^*_{Reg}, K^*_{Reg})$, which it realizes on the market for the price $p^*_R = p(q^*_R)$.

However, this form of price regulation contains one serious risk, that it often motivates the firm to use a greater volume of the variable input capital than in a non-regulated firm.

3 Conclusion

Based on the analysis of the behaviour of the firm in the conditions of regulation on the basis of the return of the used capital we have shown that in this regulation scheme the firm has the tendency to react to the tightening of the regulation conditions by the increase of the volume of used capital. However, the increase of the volume of the used capital is far from the aim of the system of regulation. The objective of regulation is rather to influence the values of other indicators important for the firm and for the economy, such as the volume of production, the level of product sale, respectively the cost level.

There are of course also other forms of price regulation, which influence the reasonable profit of the firm directly on the basis of the volume of its production, the level of product sale of the regulated firm, respectively on the basis of the amount of its total costs. The aim is to support the effective development of the regulated subject by help of regulation mechanisms.

References

[1] Fendeková, E.: Oligopoly v odvetví. Ekonóm, Bratislava, 2005

[2] Scherer, F.M.-Ross, D.: Industrial Market Structure and Economic Performance. Boston: Houghton Miffin Company, 1990.

[3] Tirole, J.: The Theory of Industrial Organisation. Massachusetts, Cambridge: The MIT Press, 1997.Act No. 291/2002 Coll. dated May 21, 2002 about the State Treasury and modification and completion of some other acts

Doc. Ing. Eleonora Fendeková, CSc., Department of Business Economics, University of Economics Bratislava, Dolnozemská 1, 852 35 Bratislava, Slovakia, e-mail:nfendek@dec.euba.sk
Prof. Ing. Michal Fendek, CSc., Department of Operations Research and Econometrics, University of Economics Bratislava, Dolnozemská 1, 852 35 Bratislava, Slovakia, e-mail:fendek@dec.euba.sk

Estimating Technical Efficiency of Human Capital Production in the Italian University with Correction for Student Characteristics: the Case of Florence University

Guido Ferrari [1] and Tiziana Laureti [2]
[1] Università di Firenze, Viale Morgagni 59, 50134 FI; ferrari@ds.unifi.it
[2] Università della Tuscia (VT), Via del Paradiso 47, 01100 VT; laureti@unitus.it

Abstract

The aim of this paper is to estimate the technical efficiency of human capital formation in the Italian University by using a stochastic frontier model in which both random terms are heteroscedastic. Using individual level data, we allow the variance of the single parameter half normal distribution to be a function of a set of explanatory variables related to student specific characteristics in order to incorporate exogenous influence on efficiency. Moreover, since the mix of resources and the internal organization varies from faculty to faculty, we incorporate heteroscedasticity in the two sided error term. The application of the methodology to the Florence University graduates revealed that technical efficiency measures are extremely sensitive to the proposed correction for heteroscedasticity.

Keywords

Stochastic frontier, Technical efficiency, Exogenous variables, Heteroscedasticity,

1 Introduction

The issue of the measurement of efficiency of human capital formation in higher education has been the focus of considerable research activities especially in the United States, and more recently, in the United Kingdom and Italy. The increase in studies of this type can be attributed, among other factors, to the development of parametric and non- parametric techniques for estimating frontier efficiency, which have provided researchers with both the ability and flexibility necessary for modelling the complex production processes and cost structures within higher education (HE) institutions.

Non parametric frontier analysis, namely Data Envelopment Analysis (DEA), is the most recent methodology used to examine the problems of performance measurement of HE institutions. The DEA has been primarily applied to HE at the departmental level and the university level.

In the UK, Tomkins and Green (1988) looked at accounting departments' cost efficiency, followed by Beasley (1990), on chemistry and physics departments. More recently, paying particular attention to the subject of research performance assessment, Johnes (1995) examined the scale and technical efficiency of economics departments, while Beasley (1995) used DEA-based methodologies to examine the trade-offs between the resources used by departments in the production of research and teaching output. Finally, Athanassopoulos and Shale (1997) applied DEA to evaluate cost efficiency and outcome efficiency of universities.

Surveying the US literature on higher education institution efficiency behaviour, Ahn, Charnes and Cooper (1988) estimated technical and scale efficiency for doctoral-granting institutions, using DEA. In order to obtain an efficient "student satisfaction" ranking Breu and Raab (1994), suggested applying DEA to the "best" 25 U.S. News and World Report-ranked universities.

Most of the work on performance measurement of HE institutions, using parametric techniques, in particular Stochastic Frontier (SF) models, consists in estimating cost functions with the purpose of analysing economies of size and scope. Key studies include Johnes (1997), Johnes, Oskrochi, and Crouchley (2002). Using a simultaneous estimation technique, Stevens (2001) also analysed the determinants of the vector of cost inefficiencies estimated by the SF approach. This approach is applied by Robst (2001) to cost data for HI in US too.

In Italy, this field of research is relatively new, and attention is especially focused on the efficiency evaluation of HE at departmental level. In particular, the efficiency of Venice University academic departments was examined by Rizzi (2002) and the empirical analysis was performed both with a DEA and a deterministic frontier model. Also, Pesenti e Ukovich (1999) used DEA to assess the efficiency of the Departments of the University of Trieste, considering both research and teaching matters. Ferrari and Laureti (2002) proposed a suitable approach to measure the frontier efficiency of the university education system at the individual student level of data.
Since the students' characteristics (e.g. background, previous education, family social and economical status) strongly affect the results of the process, Ferrari and Laureti (2002) suggested a model of human capital production process, in which the basic production unit is the individual student who produces himself/herself as a graduate by using factors provided by the faculty and the student's own effort and capabilities. With the aim of identifying a flexible tool in the non-parametric frontier literature it was applied an output DEA oriented model which enables us both to adequately measure the graduates' technical efficiency and the peculiar characteristics of the university education process to emerge. Using the program evaluation procedure, Ferrari and Laureti (2002) decomposed the total efficiency measure into a within-faculty component and a between-faculty component, in order to identify the sources of the graduates' efficiency variation.
To overcome the limits of the non-parametric methodologies, which assumes all deviations from the frontier are the results of inefficiency, Ferrari and Laureti (2004) adopted a stochastic frontier production function model, which has been adjusted to account for student characteristics.
In particular, Ferrari and Laureti (2004) suggested the incorporation of exogenous influences directly into the inefficiency error component by parametrizing the mean of the pre truncated distribution which describe the one sided error term. Models of this form have been proposed, among others, by Battese and Coelli (1995) and Kumbhakar, Ghosh and McGuskin (1991).
We must consider, however, that the efficiency measures frequently used are based on residuals obtained from the estimation of a stochastic frontier. With this regard, Caudill and Ford (1993) observed that residuals are sensitive to specification errors, particularly in stochastic frontier models, and that this sensitivity will be passed on to the efficiency measures. In order to correct for heteroscedasticity, we should test for it and, if present, correct for heteroscedasticity in the one sided error term.
In this paper we suggest the estimation of a stochastic frontier model (Aigner, Lovell and Schimdt, 1977), under the assumption of heteroscedasticity in both error terms. This allows consider both the influence of students' characteristics and the impact of the faculty resources on efficiency.
The remainder of the paper is organized as follows. In Section 2, we provide the model specification. Section 3 presents the empirical results for the University of Florence, after describing the data. Section 4 contains some concluding comments.

2. The Educational Production Frontier Model

In order to investigate the impact of faculty and student characteristics on efficiency, conducting an analysis at individual level data, we suggest the stochastic production frontier model $\ln y_i = \ln f(\mathbf{x}_i;\boldsymbol{\beta}) + v_i - u_i$, in which both *v,* which captures the effects of statistical noise, and *u,* which reflects technical inefficiency of production are heteroscedastic. The two error component are specified by the following distributional assumptions $v_i \sim N(0,\sigma_{vi}^2)$, with $\sigma_{vi}^2 = h(\mathbf{w}_i;\boldsymbol{\gamma})$ and $u_i \sim N^+(0,\sigma_{ui}^2)$ with $\sigma_{ui}^2 = g(\mathbf{z}_i;\boldsymbol{\delta})$.
The log likelihood function is

$$\ln L = \text{constant} - \frac{1}{2}\sum_i \ln\left[h(\mathbf{w}_i;\boldsymbol{\gamma}) + g(\mathbf{z}_i;\boldsymbol{\delta})\right] + \sum_i \ln\Phi\left(-\frac{\varepsilon_i\lambda_i}{\sigma_i}\right) - \frac{1}{2}\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}$$

where $\sigma_i^2 = \sigma_{vi}^2 + \sigma_{ui}^2 = h(\mathbf{w}_i;\boldsymbol{\gamma}) + g(\mathbf{z}_i;\boldsymbol{\delta})$ $\lambda_i = \frac{\sigma_{ui}}{\sigma_{vi}} = \frac{\sqrt{g(\mathbf{z}_i;\boldsymbol{\delta})}}{\sqrt{h(\mathbf{w}_i;\boldsymbol{\gamma})}}$.

The log likelihood function can be maximized to obtain estimates of $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$. This provides estimates of two functions $\hat{\sigma}_u^2 = g\left(\mathbf{z}_i;\hat{\boldsymbol{\delta}}\right)$ and $\hat{\sigma}_{vi}^2 = h\left(\mathbf{w}_i;\hat{\boldsymbol{\gamma}}\right)$ which we have substituted into the conditional mode $M\left(u_i|\varepsilon_i\right) = \begin{cases} -\varepsilon_i\left[\dfrac{1}{1+\sigma_{vi}^2/\sigma_{ui}^2}\right] & if\ \varepsilon_i \le 0 \\ 0 \qquad otherwise \end{cases}$, which is in turn it has been substituted into the expression $TE_i = \exp\left\{-\hat{u}_i\right\}$ to obtain unbiased estimates of the technical efficiency of each graduate.

We would like to emphasize that relaxing the constant-variance property of the single parameter half normal distribution, as proposed by Caudill, Ford and Gropper (1995), we allow the variance to be a function of a set of explanatory variables related to student specific characteristics (exogenous variables), that is $\sigma_{ui}^2 = g\left(\mathbf{z}_i;\boldsymbol{\delta}\right)$. This allows efficiency, which also depends on the variance of the half normal distribution to depend on exogenous variables. In this way we can solve two problems at once: incorporating exogenous influences on efficiency and correcting for one source of heteroscedasticity.

As argued earlier, in the process of human capital formation there are student specific factors that influence technical efficiency, then their effect will show up as graduate – specific parameters of the distribution of *u*. In this case, σ_{ui}^2 must be expressed as a function of producer specific variables $\mathbf{z}_i$ and estimated by generalizing the maximum likelihood techniques.

In order to implement the maximum likelihood procedure, a parametric specification for $g\left(\mathbf{z}_i;\boldsymbol{\delta}\right)$ must be supplied. Among the few specifications proposed in the literature, we will adopt the specification $\sigma_{ui} = \sigma g\left(\mathbf{z}_i;\boldsymbol{\delta}\right)$, which, if $\mathbf{z}_i$ includes an intercept, can be simplified to $\sigma_{ui} = g\left(\mathbf{z}_i;\boldsymbol{\delta}\right)$. This functional form, suggested by Caudill, Ford and Gropper (1995), has some computational advantages because it automatically constrains $\sigma_{ui} > 0$ and it is easily constrained to yield the homoscedastic case, thus making a likelihood ratio test possible.

Moreover, incorporating heteroscedasticity[1] in the two sided error term, that is $\sigma_{vi}^2 = h\left(\mathbf{w}_i;\boldsymbol{\gamma}\right)$ we can consider that the educational products are "delivered" in highly different ways, since the mix of resources (e.g., human resources, physical plant, intellectual and technological materials) and the internal organization varies from faculty to faculty.

3. An Empirical Application: the case of Florence University

The efficiency analysis requires a full data set of the performance and personal characteristics of individuals leaving university in a given year. Such a data set, compiled by the Statistics Department "G.Parenti", is available for the 1998 graduates from the University of Florence[2].

An output measure and six input measures[3] are constructed using this data for each of the 2,236 graduates, which results after a preliminary analysis checking for outliers.

Among potential explanatory variables in $\sigma_{ui}^2 = g\left(\mathbf{z}_i;\boldsymbol{\delta}_i\right)$ we have considered gender, macro-region of residence, working status while attending university, type of diploma and if the student has repeated a year at school. As already mentioned, the vector $\mathbf{w}$, in $\sigma_{vi}^2 = h\left(\mathbf{w}_i;\boldsymbol{\gamma}\right)$, includes an intercept term and a set of dummy variables assigning graduates to one of the faculties.

---

[1] The misspecification resulting from not incorporating heteroscedasticity in the ML estimation of our frontier can cause parameters estimators to be inconsistent as well as invalidating standard techniques of inference.

[2] In addition, data on the didactic inputs of faculties during the period 1990-1998 at the university were collected from the University of Florence administrative archives and the Central Library.

[3] The selected inputs are: (i) average number of full and associate professors and researchers per graduate; (ii) average number of university pieces of equipment (ANE); (iii) average number of seats in lecture halls per student (ANS); (iv) average number of journals and reviews in the library per student (ANJ); (v) average number of books in the library per student (ANB); (vi) final high-school mark (HSM). The output measure is constructed using the ratio of average grade of the exams related to the effective length of study compared to the obligatory length

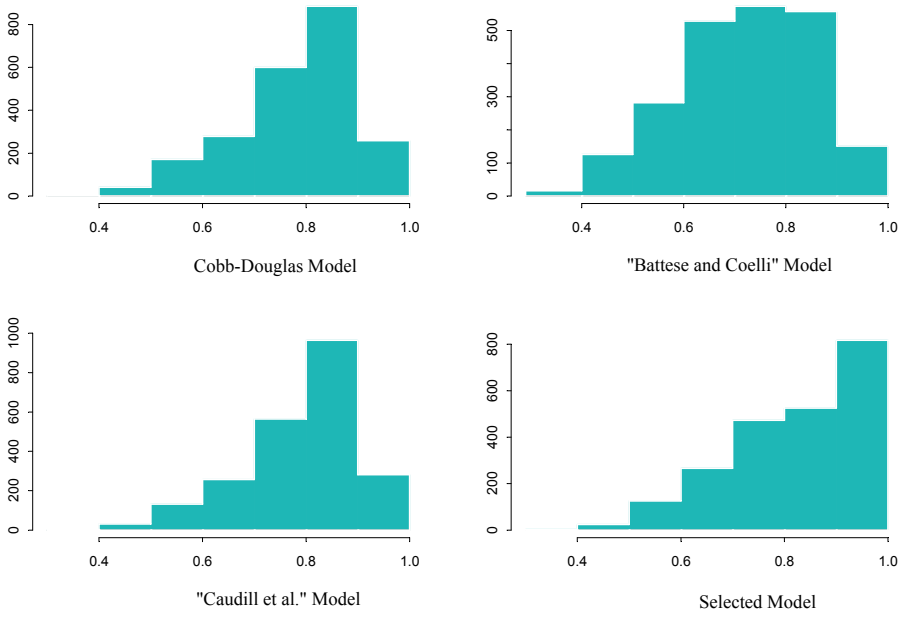
It should be stressed that there are considerable differences in resource usage among the faculties, which reflect very clear differences in the way that the graduate producing process is implemented within academic disciplines[4]. For example, the minimum value for the average number of equipment per student (0.037) is associated with a graduate within the Law faculty, while the maximum (1.863) is associated with a graduate in the Medicine faculty.

The maximum likelihood estimates of the parameters of the stochastic production frontier models were obtained using the STATA 8.1 computers software. Due to the large number of parameters estimated in the production frontier model the results are not presented here. Likelihood ratio test were used to test the different specifications. The simple Cobb-Douglas model, which does not incorporate student characteristics (the hypothesis that there are no technical inefficiency effects), is rejected.

The test for absence of heteroscedasticity in both error terms gives a value of the chi-squared statistic of 126, which exceeds the critical value of χ^2_{19} at any of the usual levels of significance. The null hypothesis that there is homoscedasticity in the one sided error term is rejected as well. The test that there is no heteroscedasticity in the two-sided error term gives a value of the chi-squared statistic equal to 84, which exceeds the critical value of χ^2_{10} at any of the usual levels of significance. These results forcefully support the presence of heteroscedasticity in both error term.

Technical efficiency measure was estimated for the selected model, and for three other models for comparison purpose. Figure 1 shows the histograms of technical efficiencies obtained from the four models.

Figure 1 Empirical Distribution of Technical Efficiencies



The various models clearly produce different empirical distributions. In Table 1 we present the average efficiency of graduates belonging to different faculties, obtained from different models.

[4] We can present here a Table showing the average figures of the variables by faculty due to lack of space.

Table 1 Technical efficiency measures for the various models

Faculty	Cobb Douglas Model (1)	Battese and Coelli Model (2a)	Battese and Coelli Model (2b)	Caudill et.al Model (3)	Selected model (4)
Agricultural Sciences	0.764	0.705	0.855	0.770	0.827
Architecture	0.757	0.678	0.863	0.755	0.790
Economics	0.751	0.682	0.935	0.764	0.785
Pharmacology	0.824	0.784	0.945	0.841	0.913
Law	0.795	0.739	0.931	0.812	0.838
Engineering	0.814	0.770	0.918	0.828	0.867
Humanities	0.803	0.751	0.919	0.817	0.867
Medicine and Surgery	0.752	0.717	0.920	0.783	0.807
Biology and Mathematics	0.756	0.716	0.942	0.783	0.849
Political Sciences	0.799	0.731	0.939	0.804	0.842
Sciences of Education	0.792	0.735	0.913	0.798	0.866
Totals	**0.779**	**0.719**	**0.915**	**0.790**	**0.830**

The aim was to examine the effects on technical efficiency measures of using different specifications. The models considered were: 1) the Cobb Douglas functional form ignoring technical inefficiency effect ; 2) the frontier model with technical inefficiency effects through parametrizing the mean, the Battese and Coelli model, which was estimated both for the entire university (a) and for each faculty separately (b); 3) the model which introduce student characteristics by parametrizing the variance of the one sided error term, therefore acconting for heteroscedasticity, and lastly but not the last 4) the selected model which acconts heteroscedasticity in both error terms. Although the mean efficiency measure of the graduates varies by subject studied, it can be easily seen, referring to university as a whole, that not accounting for heteroscedasticity leads to an underestimation of technical efficiencies.

4. Some Concluding Remarks

In this paper, we suggested the use of a doubly heteroscedastciy stochastic frontier model to measure technical efficiencies of graduates in the Italian University system. We also applied this model to the case of Florence University graduates. In this first rough analysis we found that the model specification is strongly supported by the data. This suggests that we might obtain inconsistent estimators if we do not adopt the appropriate specification. Moreover the graduate efficiecy measures are found to be extremely sensitive to the proposed correction. The use of the suggested model allows to obtain comparable technical efficiency scores across faculties, since they are corrected for student characteristics and for the subject area. Therefore, given the relevant role that higher education plays in human capital formation, analyses of the comparable technical efficiency of the university system may provide a valid background for public policy discussion and government policy-making, particularly in the face of radical changes such as those introduced in Italy.

References

[1] Aigner, D.J, Lovell, C.A.K. e Schimdt P.(1977) "Formulation and estimation of stochastic frontier production function models", Journal of Econometrics, Vol.6, pp.21-37

[2] Ahn, T., Charnes, A., & Cooper, W.W. (1988). Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. Socio-economicPlanning Sciences 22(6), 259-269.

[3] Athanassopoulos, A.D, and Shale, E. (1997) Assessing the Comparative Efficiency of Higher Education Institutions in the UK by Means of Data Envelopment Analysis, Education Economics Vo. 5, N.2,117-134.

[4] Beasley, J.E. (1990), Comparing University Departments, OMEGA, Int.J.Mgmt.Sci.2, 171-183.

[5] Beasley, J.E. (1995) Determining Teaching and Research Efficiencies. Journal of the Operational Research Society, 46, 441-452.

[6] Battese, G.E., Coelli, T.J. (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data, *Empirical Economics* 20, 325-332.

[7] Breu, T.M. and Raab, R.L. (1994), Efficiency and Perceived Quality of the Nation's "Top 25" National Universities and National Liberal Arts Colleges: An Application of Data Envelopment Analysis to Higher Education. Socio-Economic Planning Sciences, 28, 33-45.

[8] Caudill, S.B. and Ford, J.M. (1993) "Biases in Frontier Estimation due to Heteroscedasticity", *Economics Letters*, Vo.41, pp.17-20.

[9] Caudill, S., Ford J.M. and Gropper (1995) Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity, *Journal of Business & Economic Statistics*, January 1995, Vol.13, n.1.

[10] Ferrari G. and Laureti (2002) Evaluating the Efficiency of Human Capital formation in the Italian University: Evidence from Florence, WP 2002/16 Florence University Press.

[11] Ferrari, G. and Laureti T.(2004) "Multi-Factor Efficiency Analysis of the Florence University Graduates" Proceedings of XLII Riunione Scientifica SIS.

[12] Johnes, G. (1995) Scale and teaching efficiency in the production of economic research, Applied Economics Letters, Vol.2, pp.7-11.

[13] Johnes, G. (1997). Costs and industrial structure in contemporary British higher education. *EconomicJourna*l, 107, p. 727-737.

[14] Izadi, H., Johnes, G., Oskrochi, R., & Crouchley, R. (2002). Stochastic frontier estimation of a CES cost function: The case of higher education in Britain. Economics of Education Review, (21),p 63-71.

[15] Kumbhakar,S.C., Ghosh, S., McGuskin, J.T.(1991) A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business and Economics Statistics*, 9, 279-286.

[16] Kumbhakar S.C. and Lovell C.A.K. (2000) *Stochastic Frontier Analysis*, Cambridge University Press.

[17] Pesenti, R. Ukovich W. (1999) Evaluating academic activities using DEA Internal Report, DEEI - University of Trieste, Italy

[18] Reifschneider D. and Stevenson R. (1991) Systematic Departures from the Frontier: A Framework for the Analysis of Firm inefficiency, International Economic Review, Vol.32, No.3, August, pp. 715-723.

[19] Rizzi D. (2002) *L'efficienza* dei dipartimenti dell'Università Ca' Foscari di Venezia via DEA e DFA , *Economia pubblica* 3 Pg. 103-136 Vol. 31

[20] Robst, J. (2001). Cost efficiency in public higher education institutions. Journal of Higher Education, 72(6), p. 730-750.

[21] Stevens, P.A. (2001). The determinants of economic efficiency in English and Welsh universities.National Institute of Economic and Social Research, Discussion paper number 185.

[22] Tomkins, C. and Green, R. (1988) An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting, Financial Accountability and Management, Vo.44, pp.147-164.

Modeling of combinatorial auctions in network economy

Petr Fiala, Lenka Flusserová

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Prague 3
pfiala@vse.cz, lflusser@vse.cz

Abstract

The network economy is a term for today's global relationship among economic subjects characterized by massive connectivity. Today network systems provide the infrastructure and foundation for the functioning of societies and economies. They come in many forms and include physical networks such as transportation and logistical networks, communication networks, energy networks, as well as more abstract networks as economic, financial, social, and knowledge networks. The paper presents modeling of auctions in network economy. Auctions are important market mechanisms for the allocation of goods and services. Combinatorial auctions are those auctions in which bidders can place bids on combinations of items. Combinatorial auction is an appropriate instrument for selling network capacity. The winner determination problem in a combinatorial auction for selling network capacity is formulated. Efficient algorithms for multicommodity network problems can be used.

Keywords

Network economy, combinatorial auction, winner determination problem, multicommodity network flows

1 Introduction

The paper presents modeling of combinatorial auctions in network economy. It is an integration of two characteristic elements of today's economic reality, the network economy and auction mechanisms. The network economy is a term for today's global relationship among economic subjects characterized by massive connectivity. Network industries play a crucial role in modern life. Today network systems provide the infrastructure and foundation for the functioning of societies and economies. Auctions are important market mechanisms for the allocation of goods and services. Auctions are preferred often to other common processes because they are open, quite fair, easy to understand by participants, and lead to economically efficient outcomes. Their popularity is also caused by expanding of e-commerce. Design of auctions (see [5]) is a multidisciplinary effort made of contributions from economics, operations research, computer science, and other disciplines.

There is a possible classification of auctions by different aspects:
1. Traded items (indivisible, divisible, pure commodities, structured commodities).
2. Participants' roles in auctions (one-sided, multilateral auctions).
3. Objectives of auctions (optimization, allocation rules, pricing rules).
4. Complexity of bids (simply, related bids).
5. Organization of auctions (single-round, multi-round, sequential, parallel, price schemes).

For auctions for selling network capacity it is useful to use so called combinatorial auctions. Combinatorial auctions refer to auctions in which participants are allowed to bid on combinations of items. A classical problem of combinatorial auctions is the winner determination problem. The problem can be formulated as an integer programming problem and is well-known to be NP-hard. In the paper the winner determination problem for selling network capacities is formulated. The model is based on combinatorial auctions with a network structure of items.

2 Network economy

The network economy (see [2], [3], [7]) is a term for today's global relationship among economic subjects characterized by massive connectivity. The central act of the new era is to connect everything to everything in deep web networks at many levels of mutually interdependent relations, where resources and activities are shared, markets are enlarged and costs of risk are reduced. The connection is enabled by an explosive development of information and communication technologies. Network connections enable tighter relations between firms and stakeholders. New technologies provide a permanent feedback that enables activity modifications and quick responses and therefore fundamentally change business models. Network industries play a crucial role in modern life. Today network systems provide the infrastructure and foundation for the functioning of societies and economies. They come in many forms and include physical networks such as: transportation and logistical networks, communication networks, energy networks, as well as more abstract networks comprising: economic and financial networks, environmental networks, social, and knowledge networks. Many important non-network industries share many essential economic features with network industries. These non-network industries are characterized by strong complementary relations.

The reality of today's networks includes features:

- large-scale nature and complexity,
- increasing congestion,
- complementarity,
- externalities,
- switching costs,
- alternative behaviors of users of the networks,
- interactions between the networks themselves.

Many of today's networks are characterized by both a large-scale nature and complexity of the network topology. Congestion is playing an increasing role in not only transportation networks but also in telecommunication networks. The crucial relationship in networks is the complementarity between the pieces of the network. Complementarity turns to be a crucial factor in the markets for information goods. Networks exhibit positive externalities. The value of a unit increases with the expected number of units to be sold. Costs of switching to a different service or adopting a new technology are significant. The decisions made by the users of the networks, in turn, affect not only the users themselves but others, as well, in terms of profits and costs, timeliness of deliveries, the quality of the environment, etc. The behavior of the users of the networks themselves may be non-cooperative. An example is behavior of users of transport or telecommunication networks, where optimization from single users' perspective may not be optimal from a system one. This situation is illustrated by the famous Braess's paradox, where an addition of a new link with identical demand faces the increasing costs for all users.

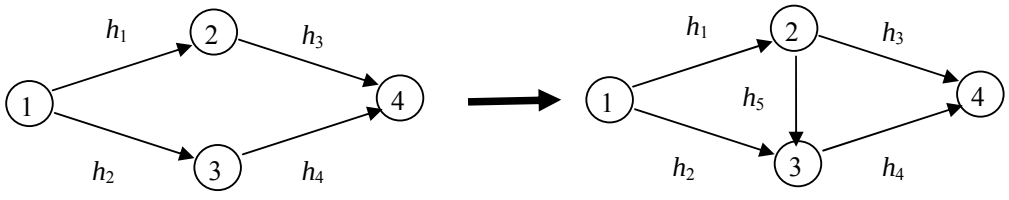


Fig. 1

The original network (Fig. 1) consists of four nodes 1, 2, 3, 4 and four edges h_1, h_2, h_3, h_4, the origin and the destination of the network are presented by nodes 1 a 4. There are two paths from the origin to the destination of the network $C_1 = \{h_1, h_3\}$ a $C_2 = \{h_2, h_4\}$. Let we suppose the costs on the edges depending on the flow quantities are

$$n_1(x_1) = 10\, x_1,\ n_2(x_2) = x_2 + 50,\ n_3(x_3) = x_3 + 50,\ n_4(x_4) = 10\, x_4$$

and the total required network flow $X = 6$.

In the case of user-optimization, the equilibrium solution is given by the situation, where all paths connecting the origin-destination pair have equal and minimal costs and therefore no user has any incentive to switch this path. The equilibrium solution is given by flows on edges

$$x^*_1 = 3,\ x^*_2 = 3,\ x^*_3 = 3,\ x^*_4 = 3$$

and by associated path costs

$$n(C_1) = 83,\ n(C_2) = 83.$$

A new edge h_5 joining node 2 to node 3 with user cost $n_5(x_5) = x_5 + 10$ is added. The change creates new path $C_3 = \{h_1, h_5, h_4\}$. The original solution is no longer equilibrium. The new equilibrium solution has the flow of amount 2 for all three paths. Edge flows are

$$x^*_1 = 4,\ x^*_2 = 2,\ x^*_3 = 2,\ x^*_4 = 4,\ x^*_5 = 2$$

and the associated path costs are

$$n(C_1) = 92,\ n(C_2) = 92,\ n(C_3) = 92.$$

Costs grow up for every user of the network from the value 83 to the value 92. This cost increase is caused by the fact that edges h_1 and h_4 are shared by two paths and the flows and cost are increasing on these edges. The addition of a path connecting an origin-destination pair that shares no links with the original connection will never result in Braess's paradox.

3 Combinatorial auctions

An auction mechanism is denoted as a combinatorial auction, if combinations of items are traded and not single items only (see [6]). Combinatorial auctions are increasingly considered as an alternative to simultaneous single-item auctions. The advantage of combinatorial auctions is the more precise expression of bidder's preferences. This advantage is primarily important in the case of complementary items. The items are complementary, if the utility of set of items is greater than a sum of utilities of single items. Two items A and B are complementary, if it holds

$$u(\{A, B\}) > u(\{A\}) + u(\{B\}).$$

From different types of combinatorial auctions we present an auction of indivisible items with one seller and several buyers. Let us suppose that one seller offers a set G of m items, $j = 1, 2, \ldots, m$, to n potential buyers. Items are available in single units. A bid made by buyer i, $i = 1, 2, \ldots, n$, is defined as

$$N_i = \{S, p_{i,S}\},$$

where

$S \subseteq G$, is a combination of items,

$p_{i,S}$ is the offered price by buyer i for the combination of items S.

The objective is to maximize the revenue of the seller given the bids made by buyers. Constraints establish that no single item is allocated to more than one buyer and that no buyer obtains more than one combination. The winner determination problem belongs to NP-hard problems.

Notations and variables are introduced for model formulation:

$\delta_{j,S} = 1$, if item $j \in S$, $\delta_{j,S} = 0$, if item $j \notin S$,

$x_{i,S}$ is a bivalent variable specifying if the combination S is assigned to buyer i ($x_{i,S} = 1$).

The winner determination problem can be formulated as follows

$$\sum_{i=1}^{n} \sum_{S \subseteq G} p_{i,S}\, x_{i,S} \rightarrow \max$$

subject to

$$\sum_{i=1}^{n} \sum_{S \subseteq G} \delta_{j,S}\, x_{i,S} \leqq 1, \quad \forall j \in G,$$

$$\sum_{S \subseteq G} x_{i,S} \leqq 1, \quad \forall i,\ i = 1, 2, \ldots, n,$$

$$x_{i,S} \in \{0, 1\}, \ \forall\ S \subseteq G, \ \forall i,\ i = 1, 2, \ldots, n.$$

For general solving of the winner determination problem is proposed dynamic programming (see [4]). Authors also consider several restrictions on allowable bids that make the problem computationally manageable.

4 Auctions on networks

The traded commodities can be network capacities, which enable various types of flows in network industries. A classical example is capacity of telecommunication networks, where the capacities of certain links are supplied and demanded. The objective is the optimal combination of segments in required paths. Principles of combinatorial auctions are useful for selling network capacities. The utility of the path capacity is greater than a sum of utilities of edge capacities.

For illustration we present a basic formulation of the winner determination problem in a combinatorial auction for selling network capacities. Let $G = (U, H)$ be a network, where U is a set of nodes and H a set of edges. To each edge $h_j \in H$, $j = 1, 2, \ldots, m$, is associated a capacity $k(h_j)$. Capacities of the network are owned by a single seller and there are n buyers, potentially interested in path capacities. The combinatorial aspect of the problem ensues from the fact that buyers desire to obtain path capacities (combinations of edges), rather than on individual capacities of edges. We suppose that buyer i, $i = 1, 2, \ldots, n$, submits a single bid specified by following specifications

$$N_i = \{Z_i, K_i, G_i, k_i, p_i\},$$

where

Z_i, K_i is an origin-destination pair of nodes specifying the path for buyer i ,

G_i is a subgraph of the graph G specifying edges for possible paths,

k_i is the required capacity for the path,

p_i is the offered price of buyer i for the combination.

Other notations and variables are introduced for model formulation:

C_i is the set of all paths between Z_i and K_i in the subgraph G_i ,

c is a path from the set C_i ,

$\delta(h_j, c) = 1$, if $h_j \in c$, $\delta(h_j, c) = 0$, if $h_j \notin c$,

y_c is a variable specifying the capacity on the path $c \in C_i$,

x_i is a bivalent variable specifying if bid N_i is winning ($x_i = 1$).

The winner determination problem in a combinatorial auction for selling network capacities can be formulated as follows

$$\sum_{i=1}^{n} p_i x_i \quad \rightarrow \quad \max$$

subject to

$$\sum_{c \in C_i} y_c = k_i x_i, \; i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} \sum_{c \in C_i} \delta(h_j, c)\, y_c \leqq k(h_j), \; h_j \in H,$$

$$x_i \in \{0, 1\}, \; i = 1, 2, \ldots, n,$$

$$y_c \geqq 0, \; c \in C_i, \; i = 1, 2, \ldots, n.$$

When paths are specified by buyers, and single units of capacity are available on edges a requested by buyer, the model is equivalent to the model of the winner determination problem, in which items are edges and combinations are paths. The particularity of the model lies in the fact that buyers do not need indicate a specific path along which the capacity should be allocated. It is auctioneer's task of routing the requested capacities in order to determine the winning bids. The model could be solved by commercial mixed integer programming software. However, the model can be formulated as a multicommodity flow problem. Effective algorithms can be used for solving multicommodity network problems (see [1]).

Illustrative example

Fig.2 presents a network with following capacities

$k(h_1) = 3$, $k(h_2) = 3$, $k(h_3) = 3$, $k(h_4) = 3$, $k(h_5) = 3$, $k(h_6) = 3$, $k(h_7) = 3$, $k(h_8) = 3$, $k(h_9) = 3$.

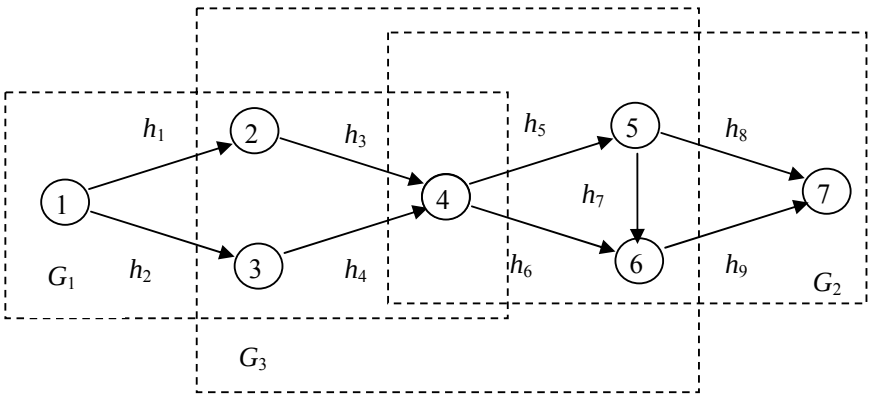


Fig.2

Let we suppose three bidders with their bids

$N_1 = \{Z_1 = 1, K_1 = 4, G_1, k_1 = 3, p_1 = 10\}$,
$N_2 = \{Z_2 = 4, K_2 = 7, G_2, k_2 = 3, p_1 = 10\}$,
$N_3 = \{Z_3 = 3, K_3 = 6, G_3, k_3 = 3, p_3 = 10\}$.

This simple example is an illustration of terms and notations. There is the optimal solution of the example

$$x_1 = 1,\ x_2 = 1,\ x_3 = 1,\ \sum_{i=1}^{n} p_i x_i = 30.$$

The paths are identified for particular bids

$C_1 = \{h_1, h_3\}$,
$C_2 = \{h_6, h_9\}$,
$C_3 = \{h_4, h_5, h_7\}$.

5 Conclusions

Analysis and optimization of network economy functioning are challenges for application of modeling approaches. Auctions are important market mechanisms for the allocation of goods and services. Recently research and applications of combinatorial auctions are significantly increasing. In the paper a basic model for selling network capacity is formulated. The model is based on combinatorial auctions. The advantage for solving this model is a possibility of utilization algorithms for multicommodity network problems. This basic model can be extended for other types of auctions on network structures. Integration of findings from economics, operations research and computer science is promising for interesting results.

The research project was supported by Grant No. 402/05/0148 from the Grant Agency of Czech Republic „Network economy – modeling and analysis.“

References

[1] Ahuja, R. K., Magnanti, T. L., Orlin, J. B. (1993): Network flows: Theory, algorithms, and applications. Prentice Hall, Englewood Cliffs.

[2] Economides, N. (1996): The Economics of Networks. International Journal of Industrial Organization 14, no. 2.

[3] Fiala P. (2005): Síťová ekonomika – výzva pro dnešek. MOMAN 05, Praha.

[4] Pekeč, A., Rothkopf, M. H. (2003): Combinatorial auction design. Management Science 49, no. 11, 1485-1503.

[5] Rothkopf, M. H., Park, S. (2001): An elementary introduction to auctions. Interfaces 31, no. 6, 83-97.

[6] Rothkopf, M. H., Pekeč, A., Harstad, R. M. (1998): Computationally manageable combinatorial auctions. Management Science 44, 1131-1147.

[7] Shapiro, C., Varian, H. (1999). Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press, Boston.

Modeling of interacting agents in network economy

Petr Fiala, Václav Kořenář

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Prague 3

pfiala@vse.cz, korenar@vse.cz

Abstract

The paper presents models for analyzing network economy. There is possibility to analyze behavior of the network by aggregating production units. Modeling of micro units in a stochastic and dynamic framework uses Markov processes to model interacting micro units. Specifications of decision processes for micro units determine stochastic mechanisms for the units to change their decision over time. The approach can be used for network models as aggregation of micro units and for determination how uncertainty in the model affects macro behavior. Using simulation we are able to find out how each unit of the network can influence the others, their decisions and behavior.

Keywords

Network economy, interacting agents, Markov process

1 Introduction

The paper presents models for analyzing network economy. The network economy (see [4], [6]) is a term for today's global relationship among economic agents characterized by massive connectivity. The central act of the new era is to connect everything to everything in deep web networks at many levels of mutually interdependent relations. Network systems contain both positive and negative feedbacks. These processes are different and vary in strength. A variety of feedback processes create complex system behavior. Dynamic models try to reflect changes in real or simulated time and take into account that the network model components are constantly evolving.

There is possibility to analyze "macro" behavior of the network by aggregating "micro" units. Modeling of "micro" units in a stochastic and dynamic framework aims:

1. use of Markov processes to model interacting microeconomic units,
2. focus on the multiplicity of micro states,
3. analysis of multiple equilibria,
4. introduction of hierarchical and network structures .

We can model "micro" units as interacting Markov processes and the dynamic version of other discrete-choice models are put in this framework. Specifications of decision processes for micro units determine stochastic mechanisms for the units to change their decision over time. The approach can be used for "macro" models as aggregation of micro units and the multiplicity determines how uncertainty in the model affects macro behavior. It is also possible to use simulation models and methods for better understanding the relationships among the units of the network. Using simulation we are able to find out how each unit of the network can influence the others, their decisions and behavior.

There are two approaches in the analysis of network externalities. The "macro" approach assumes that network externalities exist, and attempts to model their consequences. The

"micro" approach attempts to find the cause of network externalities. Some goods and services generate more value when more users consume the same goods and services. Many products are required to be used with other products at the same time. They have little or even no value if they are used in isolation. The consumers using these products constitute networks in which the utility derived from consumption of these goods or services increases as additional consumers purchase the same goods and services. A market characterized by such properties is called a network market in which there exist positive consumption externalities termed network effects. Network connections bring important effects. Networks established for the purpose of sharing or creating new information provide better, more complete information as more units join and use them. The attractiveness to users of networks increases as they increase in size. When the value of a product to one user depends on how many users there are, this product exhibits network effects. The result of these effects is that the self-regulating behavior of demand and price, which is the foundation of traditional microeconomics, no longer applies. Under traditional "negative feedback" economic theory, if demand approaches supply, prices are forced up. This causes demand to decrease and supply to increase, eventually bringing prices back down to their equilibrium level. But information intensive networks are governed by a very different set of "positive feedback" dynamics. Technologies subject to strong network effects tend to exhibit long lead times followed by explosive growth.

The paper presents models of interacting agents in network economy. The models explore a sector economy where microeconomic agents move from one sector to another sector. A traditional approach for two-sector economy is based on financial comparisons. An alternative approach focuses on the random partitions of the set of agents (firms) into clusters (sectors), uses Markov processes, and utilizes the conditional probability specifications for new entries and exits to derive equilibrium distributions for cluster sizes.

2 Model of two-sector economy

A traditional approach to model two-sector economy (see [3]) has analyzed several economic problems, such as that of how to optimally allocate capital stocks among two sectors, and of assessing the effects of exchange rate changes to induce entries or exits of firms.

It is assumed that the economy is closed; the total number of firms is fixed at N. All firms are assumed to be indistinguishable, and choose between two alternatives of either producing goods one or goods two. There are N firms in the economy, where $N - n$ is the number of firms producing goods one, n is the number of firms producing goods two. The prices of the two goods are normalized to be 1 and p, that is, p is the price of goods two in terms of goods one.

The economy as a whole receives the revenue

$$R(p, n) = f_1(N - n) + p f_2(n),$$

where f_1 and f_2 are the production function of sector 1 and 2 respectively, given as functions of the number of firms. If p is sufficiently high, then it is more profitable for firms in sector one to move to sector two.

There is an entry cost associated with changing sectors

$$h(p, n) = h_0[f_1(N - n + 1) - f_1(N - n)].$$

This assumes that the move takes one period to complete, and that the cost of moving is proportional to the lost production because one firm in sector one shuts down its production to start producing goods two. In a move in the opposite direction, the loss of move is given by

$$l(p, n) = l_0[f_2(n) - f_2(n - 1)].$$

Here h_0 and l_0 are some positive constants.

Let $V(p, n_2)$ be the value function of this economy. It is the discounted expected present value of the stream of profits when the economy optimally allocates the firms between the two sectors. Denoting the discount rate by, it is given by

$$\rho\, V(p, n) = R(p, n) + E(\frac{dV}{dt}).$$

This expression is a typical one in finance. Regarding the expected value of the economy as financial asset, the right-hand side, which is the sum of the revenue and the capital gain (loss) term, must equal to the return from holding the asset on the left-hand side.

Now, one firm moves from sector one to two when p satisfies

$$V(p, n - 1) = V(p,n) - h(p, n),$$

and

$$V_p(p, n - 1) = V_p(p, n) - h_p(p, n).$$

Denote this p by p^+_n. The first equation states that at this critical or switch-over price, the values with n-1 and n minus the cost of moving are the same. The second equation is a technical gradient-matching condition without which the assumption that V is the optimal value is violated. The other switch-over price p^-_n at which a firm moves back to sector one from sector two is similarly expressed. It is assumed that p moves as a Brownian process, and solved for V, and then derives these prices. When $p(t)$ cuts the schedule p^+_n from below, then a move of firm from sector one to two occurs, that is n changes to $n + 1$. Similarly for a move from n to $n - 1$ which occurs at price p^-_n .

3 Models of multi-sector economy

An alternative approach (see [1], [2]) focuses on the random partitions of the set of firms into clusters induced by subsets formed by firms of the same types, and utilizes the conditional probability specifications for new entries and exits to derive equilibrium distributions for cluster sizes. It is used the master equation (backward Chapman-Kolgomorov equation) as the dynamic equation for the probabilities of state vectors.

Given the total number of agents, N, and the number of possible types, K, both of which are assumed in this paper to be known and finite for ease of explanation, we examine how the N-set, that is, the set $\{1, 2, \ldots, N\}$ is partitioned into K clusters, or subsets, where K *is* finite, but large.

For the analysis continuous-time Markov chains are used. Define a state vector X_t which takes on the value $\mathbf{n} = (n_1, n_2, \ldots, n_K)$, called frequency vector, where n_i is the number of agents of type i, $i = 1, 2, \ldots, K$, $N = n_1 + n_2 + \cdots + n_K$.

Continuous-time Markov chains are specified by transition rates. In the model we need to specify entry rates, exit rates and rates of type changes. Over a small time interval Δ, rates are multiplied by the length of interval to approximate the conditional probabilities up to $O(\Delta)$.

Entry rates by an agent of type j is given by

$$w(\mathbf{n}, \mathbf{n} + \mathbf{e_j}) = \varphi_j(n_j, \mathbf{n}),$$

where $\mathbf{e}_j$ is a vector with the only nonzero element of one at component j.

Exit rates of an agent of type k specified by

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e_k}) = \psi_k(n_k, \mathbf{n})$$

and transition rates of type i agent changing into type j agent by

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e_i} + \mathbf{e_j}) = \lambda_{i,j}(n_i, n_j, \mathbf{n}).$$

With transition rates between states specified, the dynamics for the probability is given by the following master equation, where x, y , and z refer to some states

$$\frac{dp(x,t)}{dt} = \sum_y w(y,x)p(y,t) - \sum_z w(x,z)p(x,t).$$

A special example of interest has the transition rates:

$$w(\mathbf{n}, \mathbf{n} + \mathbf{e}_k) = c_k(n_k + h_k),$$

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e}_j) = d_j n_j ,$$

$$w(\mathbf{n}, \mathbf{n} - \mathbf{e}j + \mathbf{e}k) = \lambda_{,jk}\, d_j\, n_j c_k(n_k + h_k),$$

with $\lambda_{,jk} = \lambda_{,kj}$, and where $j, k = 1, 2, \ldots K$. We assume that $d_j \geq c_j > 0$, and $h_j > 0$.

The first transition rate specifies entry rate of type k agents, and the second that of the exit rate by type j agents and the last specifies the transition intensity of changing types by agents from type j to type k. In the entry transition rate specification $c_k n_k$ stands for network effects which make it easier for others to join the cluster, and $c_k h_k$ stands for the innovation effects which are independent of the cluster size. These transition rates for type changes are in [5]).

The jump Markov process thus specified has the stationary distribution

$$\pi(\mathbf{n}) = \prod_{j=1}^{K} \pi_j(n_j),$$

where

$$\pi_j(n_j) = (1 - g_j)^{-h_j} \binom{-h_j}{n_j} (-g_j)^{n_j},$$

where $g_j = c_j/d_j$.

These expressions are derived straightforwardly by applying the detailed balance conditions to the transition rates (see [5]).

The partition vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ is introduced, where v_i is the number of clusters with exactly i agents. Then hold the relations

$$\sum_{i=1}^{n} v_i = k_n \leq K \quad , \quad \sum_{i=1}^{n} i v_i = n \ ,$$

where k_n is the number of clusters formed by n agents.

For simplification it is supposed that $h_j = h$ and $g_j = g$ for all j. Then the stationary distribution can be expressed as follows

$$\pi(\mathbf{n}) = \binom{-Kh}{n} \prod_{j=1}^{K} \binom{-h}{j}^{v_j} .$$

The parameter $\mu = Kh$ is introduced, where K is very large and h is very small. The negative binomial expression

$$\binom{-h}{j}^{v_j}$$

approaches

$$\left(\frac{-h}{j}\right)^{v_j} (-1)^{jv_j}$$

as h is very small.

The stationary distribution according to the partition vector v can be expressed as

$$\pi(v) = \binom{-\mu}{n} (-1)^n \frac{K!}{v_1! v_2! \ldots v_n! (K-k_n)!} \prod_{j=1}^{K} \left(\frac{h}{j}\right)^{v_j} .$$

In the limit of K becoming infinite and h approaching 0 while keeping Kh at μ we get so called Ewens distribution

$$\pi(v) = \frac{n!}{\mu^{[n]}} \prod_{j=1}^{n} \left(\frac{\mu}{j}\right)^{v_j} \frac{1}{v_j!},$$

where $\mu^{[n]} = \mu(\mu+1)\ldots(\mu+\text{n}-1)$ is so called ascending factorial.

Ewens distribution has a single positive parameter μ. Its value influences the number of clusters formed by the agents. Smaller values of μ tend to produce a few large clusters, while large values produce a large number of clusters of smaller sizes.

Example
A value of parameter μ influences the number of clusters. We will analyze two extreme situations.

The first situation - n agents forming a single cluster:
The probability of is given by

$$\pi(v_j = 0; j = 1,2,\ldots,n-1; v_n = 1) = \frac{(n-1)!}{(\mu+1)(\mu+2)\ldots(\mu+n-1)}.$$

The second situation- n agents forming n clusters with one agent:
The probability of is given by

$$\pi(v_1 = n; v_j = 0; j = 2,3,\ldots,n) = \frac{\mu^{n-1}}{(\mu+1)(\mu+2)\ldots(\mu+n-1)}.$$

With the parameter μ tends to zero the probability of the first situation tends to 1 and the probability of the second situation tends to 0. If the parameter μ is much greater than n the probability of the first situation tends to 0 and the probability of the second situation tends to 1.

4 Conclusions

Analysis and optimization of network economy functioning are challenges for application of modeling approaches. The paper compares a traditional approach on two-sector economy and an alternative approach on multi-sector economy.

The alternative approach based on continuous-time Markov chains provides more detailed analysis of the situation. The approach is appropriate for the analysis of network economy including network effects. The paper presents some simple models for the network economy but the approach is promising for more detailed analysis.

The research project was supported by Grant No. 402/05/0148 from the Grant Agency of Czech Republic „Network economy – modeling and analysis“.

References

[1] Aoki, M. (1998): New Approaches to Macroeconomic Modeling: Evolutionary Stochastic Dynamics, Multiple Equilibria, and Externalities as Field Effects. Cambridge University Press, Cambridge.

[2] Aoki, M. (2002) Aggregate Behavior and Fluctuations in Economics. Cambridge University Press, Cambridge.

[3] Dixit, A., (1989): Entry and exit decisions of firms under fluctuating real exchange rates. Journal of Political Econ. 97, 620–637.

[4] Economides, N. (1996): The Economics of Networks. International Journal of Industrial Organization 14, no. 2.

[5] Kelly, F. (1979): Reversibility and Stochastic Networks. J.Wiley, New York.

[6] Shapiro, C., Varian, H. (1999). Information Rules: A Strategic Guide to the Network Economy. Harvard Business School Press, Boston.

Estimation of Binary Choice Model with Panel Data

Zuzana Fíglová
Department of Econometrics, University of Economics Prague
W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic

Abstract

The aim of this paper is to estimate the binary choice model with panel data. Binary choice model is based on random utility theory of consumer behavior where each consumer maximizes his utility through the consumption of goods and services. Panel data represents data where multiple cases were observed at two or more time periods. The model was applied to data from the Household Budget Surveys 2000-2003 in order to analyze choice behavior of households. We analyzed the role of income as a determinant of ownership of PC and its change through the observed period of time.

Keywords

binary choice model, panel data, household budget survey, durable goods

1 Binary Choice Model with Panel Data

Panel data analysis is a powerful analytical tool for studying both the space and time dimension of the data. In this study we combined a panel data approach with discrete choice models. Discrete choice models have become very useful and popular in a various areas as consumer behavior, marketing research, energy, psychology, telecommunications, biomedical research, medicine etc.

We can define discrete choice (or qualitative response) models as models, where the dependent variable takes a discrete and finite set of values. That means consumer or decision maker faces up to a choice among a set of alternatives. In this set, the number of alternatives is finite, the alternatives are mutually exclusive, and finally the set of alternatives is exhaustive. We will concentrate on models, where dependent variable takes a dichotomous value and this model is called binary choice model.

Let consider the data consist of observations of n subjects, and the i^{th} subject has T_i observations. Denote the t^{th} observation of the i^{th} subject by $(y_{it}, \boldsymbol{X}_{it})$, where y_{it} is a binary value and $\boldsymbol{X}_{it} = (x_{it0}, x_{it1})^T$ is the covariate matrix with column vectors x_{it0} and x_{it1} associated with values 0 or 1 of the dependent variable y_{it}. In the context of consumer behavior mentioned above, y_{it} denotes ownership of the durable goods chosen by the i^{th} subject (in our case household by decile) at the t^{th} time of asking. If the individual-specific effect α_i, is assumed to be fixed we can define the following fixed effect regression model based on a latent-variable formulation (Wooldridge, 2002)

$$y_{it}^{*} = \boldsymbol{x}_{it}^{T}\boldsymbol{\beta} + \alpha_i + \eta_{it} \qquad i = 1,2,\ldots,n, \quad t = 1,2,\ldots,T, \tag{1}$$

where y_{it}^{*} is a latent variable (not observed by the researcher). A consumer chooses $y_{it} = 1$ if the latent variable is positive and 0 otherwise, hence

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^{*} > 0, \\ 0 & \text{if } y_{it}^{*} \le 0. \end{cases} \tag{2}$$

The latent variable can be y_{it}^{*} interpreted as the utility difference between choosing $y_{it} = 1$ and 0. In the binary logit model y_{it} can be modeled as independent across all subjects and across all repeated observations and the probability that y_{it} takes on the value 1 is

$$P_i = E_i(y_{it} = 1 \left| \boldsymbol{x}_{it}, \alpha_{\boldsymbol{i}} \right.) = \frac{\exp(\boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha_{\boldsymbol{i}})}{1 + \exp(\boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha_{\boldsymbol{i}})}. \quad (3)$$

The data under the study are grouped or replicated data. That means we have to obtain the relative frequencies (details e.g. Gujarati, 2003) $\hat{P}_i = n_i / N_i$, where N_i is the number of families and n_i among whom are durable goods owners. The stochastic disturbance term u_i has to follow the normal distribution with zero mean and variance equal $1/[N_iP_i(1–P_i)]$ and it means the disturbance term in the logit model is heteroskedastic. Instead of applying ordinary least square (OLS) method we have to use the weighted least square (WLS), we have to transform each variable by $w_i = N_i\hat{P}_i(1-\hat{P}_i)$. In our case N_i is fairly large, so $\hat{P}_i$ will be a reasonably good estimator of P_i. When we use $\hat{P}_i$, we can obtain the estimated logit as

$$\hat{L}_i = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right) = \boldsymbol{x}_{it}^T \boldsymbol{\beta} + \alpha_{\boldsymbol{i}}. \quad (4)$$

The expression $P_i/(1 – P_i)$ is *odds ratio* in favor of the occurrence of an event. We have to be careful about all the conclusions that will be valid strictly speaking if the sample is reasonably large.

We used the fixed effects model, that has constant slope but intercepts differ according to the cross-sectional (group) unit–in this case, the household group by deciles. While the intercept is cross-section (group) specific and in this case differs from household to household, it may or may not differ over time. Because *i*-1 dummy variables are used to designate the particular group, this model can be estimated by the *Least Square Dummy Variable* (LSDV) estimator. This is computed by least squares regression of $y_{it}^* = (y_{it} - \bar{y}_i)$ on the same transformation of $\boldsymbol{x}_{it}$ where the averages are group specific means. The individual specific dummy variable coefficients can be estimated using group specific averages of residuals.

Formulation of the fixed effects model depends on the assumptions we make about the intercept, the slope coefficients and the error term. The first model represents the fixed effect model with constant slope coefficients and we allow for intercepts to vary between households (MODEL 1). The second fixed effects model has constant slope coefficients and intercepts differ according to time (MODEL 2). In the last model we suppose the slope coefficients are constant but the intercept vary over household as well as time (MODEL 3).

2 Analysis of ownership of PC

Ownership of durable goods analyzed, for example, Hušek and Moravová (2002) who applied binary logit and probit models to data from the *Household Budget Survey* for the year 2000 in the Czech Republic. They estimated and predicted the percentage of the households with an ownership of a mobile phone. They used an income per household member and main economic activity of the head of household (pensioner, self-employed, farmer, employee) as independent variables.

The general approach to information technologies in the Czech Republic was described in the State Information Policy, approved by the Government Resolution No. 525 of 31 May 1999. The Czech Republic has adopted several laws that constitute a legal framework for development of the information society (Act on Digital Signatures; Act on Electronic Commerce etc.) In the Czech Republic, 45% of the population has access to a personal computer (PC) and about 20% use their PC daily. There has been a higher and higher PC ownership rate due to the lower prices and new sales strategies for PCs have led to an increase especially in the number of home PCs. Our aim is to apply

the LSDV model and analyze the role of income as a determinant of ownership of PC and its change through the observed period of time.

The data were obtained from *Household Budget Surveys* that were carried out by the Czech Statistical Office. The main objective of these surveys is to provide a detailed view of spending patterns of various kinds of households. We analyzed the observed period 2000-2003. It contains the data from households distributed by net income per person in deciles, ownership of major household durable goods such as refrigerator, microwave oven, personal computer etc. Our pooled data set contains a total of 10 household groups x 4 years = 40 observations. In other words ten household groups by deciles are followed for four years and are sampled annually.

We estimated econometric models to investigate whether or not household group by decile or time raise the probability that a household owns PC. We specified three models, in the MODEL 1 as a continuous dependent variable we used ownership rate (*OR*) of PC in households by deciles. Households are distributed in ten deciles by their net money income per person. The households with the lowest net income are included in the first decile and we labeled them as *HD1* and the households with the highest income are named as *HD10*. Independent dummy (0-1) variables in MODEL 1 are households (*HD1*,…,*HD10*) and because we wanted to avoid falling into the dummy-variable trap (i.e. the situation of perfect collinearity) there is no dummy for *HD1*. We are treating HD1 as the base, whose intercept value is given by β_0. We formulate this MODEL 1 as follows

$$OR_{it} = \beta_0 + \beta_1 HD2_i + \beta_2 HD3_i + \beta_3 HD4_i + \beta_4 HD5_i + \beta_5 HD6_i + \beta_6 HD7_i + \beta_7 HD8_i + \beta_8 HD9_i + \beta_9 HD10_i + u_{it}$$

where $HD2_i$ takes a value of 1 if the observation belongs to *HD2* and zero otherwise, etc. The results from estimated model are displayed in Table 1.

Table 1 Estimated coefficients in logit model with panel data (MODEL 1)

Variable	Logit model	
Intercept	-7.7974	(-10.2967)
HD2	0.1174	(1.0530)
HD3	-0.0981	(-0.5982)
HD4	-0.2951	(-1.6714)
HD5	-0.0778	(-0.4812)
HD6	-0.0785	(-0.4382)
HD7	0.0175	(0.1099)
HD8	0.2003	(1.3358)
HD9	0.1054	(0.6852)
HD10	0.2748	(1.6930)
	$R^2 = 0.3322$ $\quad \overline{R}^2 = 0.1320$	
	$F_{(9,30)} = 1.6587$ (0.1435)	

[a] In parenthesis are *t*–values of estimated coefficients for significance level $\alpha = 0.05$.

As we can see from Table 1 all estimated regression coefficients except intercept in MODEL 1 are not statistically significant and there is a high value of R^2 but *F*-test leads to the rejection of the hypothesis that dummy variables *HD2*,…,*HD10* have effect on ownership of PC in households.
In the MODEL 2 we included as dependent dummy variables year 2000-2003 (*YEAR01, YEAR02, YEAR03*) without dummy variable for the year 2000 and the results consist Table 2.

$$OR_{it} = \alpha_0 + \alpha_1 YEAR01 + \alpha_2 YEAR02 + \alpha_3 YEAR03 + u_{it}$$

Table 2 Estimated coefficients in logit model with panel data (MODEL 2)

Variable	Logit model	
Intercept	-9.6422	(-22.7378)
YEAR01	0.1688	(2.1689)
YEAR02	0.2961	(3.9455)
YEAR03	0.5424	(7.3203)
	$R^2 = 0.6146$ $\overline{R}^2 = 0.5825$	
	$F_{(3,36)} = 19.1371$ (0.0000)	

[a] In parenthesis are *t*–values of estimated coefficients for significance level α = 0.05.

Table 2 contains the results from estimated logit regression MODEL 2. The estimated regression coefficients are highly statistically significant. As we can see from the value of the *F*-test we can reject the null hypothesis at the significance level α = 0.05 of no effect of dummy variables *YEAR01, YEAR02* and *YEAR03* on ownership of PC in households. The positive signs of the coefficients mean dummy variables increase the probability of ownership of PC. We can not reject the hypothesis of heteroskedasticity in MODEL 2 when applying the White test with ignoring the cross–product terms, $\chi^2_{(6)} = 14.1110$ with *p*-value equals 0.0284 and $F_{(6,29)} = 2.6343$ with *p*-value equals 0.0366. We can not also reject the hypothesis of no autocorrelation using the Lagrange multiplier test where $\chi^2_{(1)} = 6.5103$ with *p*-value equals 0.0107 and $F_{(1,35)} = 6.8038$ with *p*-value equals 0.0133.

The MODEL 3 is presented in Table 3 as a combination of MODEL 1 and MODEL 2 with all included dummy variables.

$$OR_{it} = \alpha_0 + \alpha_1 YEAR01 + \alpha_2 YEAR02 + \alpha_3 YEAR03 + \beta_1 HD2_i + \beta_2 HD3_i + \beta_3 HD4_i + \beta_4 HD5_i + \beta_5 HD6_i + \beta_6 HD7_i + \beta_7 HD8_i + \beta_8 HD9_i + \beta_9 HD10_i + u_{it}$$

Table 3 Estimated coefficients in logit model with panel data (MODEL 3)

Variable	Logit model	
Intercept	-8.9157	(-27.4564)*
YEAR01	0.1678	(3.6464)*
YEAR02	0.2928	(6.4186)*
YEAR03	0.5474	(12.4035)*
HD2	0.0201	(0.4410)
HD3	-0.2241	(-3.4099)*
HD4	-0.4047	(-5.7606)*
HD5	-0.1969	(-3.0407)*
HD6	-0.2713	(-3.7029)*
HD7	-0.1141	(-1.7856)
HD8	0.0786	(1.3029)
HD9	-0.0218	(-0.3528)
HD10	0.0720	(1.0662)
	$R^2 = 0.9076$ $\overline{R}^2 = 0.8665$	
	$F_{(12,27)} = 22.1021$ (0.0000)	

[a] In parenthesis are *t*–values of estimated coefficients for significance level α = 0.05.

The results of estimated logit MODEL 3 from the Table 3 shows that dummy variables *YEAR01, YEAR02* and *YEAR03* are statistically significant at the significance level α = 0.05. Dummy variables *HD3, HD4, HD5* and *HD6* have the impact on dependent variable that is ownership of PC. Their negative signs mean that the dummy variables decrease the probability of ownership by household. In MODEL 3 we can reject the hypothesis of heteroskedasticity when using the White test with ignoring the cross–product terms (there were not enough observations), $\chi^2_{(24)} = 22.9730$ with *p*-value equals 0.5214 and $F_{(24,2)} = 0.1124$ with *p*-value equals 0.9987. We can reject the hypothesis of no

autocorrelation using the Lagrange multiplier test where $\chi^2_{(1)} = 0.5634$ with p-value equals 0.4529 and $F_{(1,26)} = 0.3715$ with p-value equals 0.5475.

We can interpret the estimated slope coefficient of *HD4* shows if a household is classified by its net income in fourth decile, the weighted log odds in favor of ownership a PC goes down by 0.4047 units. The interpretation of the odds is more understandable interpretation that the previous one of logit. For a unit increase in a variable (in our case for a household in fourth decile), the weighted odds in favor of ownership a PC decreases by $e^{-0.4047} = 0.6672$ or about 66.72%.

We can also compute the probability of owning PC for a certain year and for a certain type of household. Suppose we want to compute the probability for the year 2001 (*YEAR01* = 1) and the household in fourth decile (*HD4* = 1) and all other dummy variables are equal to zero. We obtain the estimated logit $\hat{L}_i = -1.3218$ and we have

$$-1.3218 = \ln\left(\frac{\hat{P}_i}{1-\hat{P}_i}\right).$$

Therefore

$$\frac{\hat{P}_i}{1-\hat{P}_i} = e^{-1.3218} = 0.2667,$$

and the estimated probability $\hat{P}_i = 0.2105$. That is the probability of owning a PC is about 21 percent. Similar for the year 2000 and the household in first decile the estimated probability of owning a PC is equal $\hat{P}_i = 0.2374$.

3 Conclusion

In this study we were concerned about estimation and calculating the probability in binary logit model with panel data. Panel data consists of observations on the same cross-sectional units over several periods of time. They increase the sample size and are better suited to analyze the dynamics of change. We applied the *Least Square Dummy Variable* approach and analyzed the data about the household's ownership of a PC. There is a strong evidence of increasing ownership rate of PC every observed year. There could be also included more variables as education, number of children or age in the model. Another approaches could be used for analysis, for example the random effects model that differs from the fixed effects model only in the specification of the disturbance term.

References

[1] Czech Statistical Office (www.czso.cz)

[2] Gujarati, D.: *Basic Econometrics*, McGraw Hill Publishers, 2003.

[3] Hsiao, C.: *Analysis of Panel Data*, Cambridge University Press, 2003.

[4] Hušek, R., Moravová, J.: Analýza a prognóza na základě modelů binární diskrétní volby. *Politická ekonomie*, 2002, L/4, pp. 579–593.

[5] Hušek, R., Pelikán, J.: *Aplikovaná ekonometrie. Teorie a praxe.* Professional Publishing, 2003.

[6] Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*, MIT Press, 2002.

Multiple relations of composite commodities[1]

Jiří Frank, Martin Gavalec
University of Hradec Králové, Czech Republic
jiri.frank@uhk.cz, martin.gavalec@uhk.cz

Abstract

Applications of composite-commodity theorem to a multiple relation is studied. Two tradable inputs are related in the complementarity form and, at the same time, each of these inputs is related in the input-output form to a common final product. Possible effects of the described multiple relations, in comparison to dually related goods are examined on the basis of relevant simultaneous equations.

Keywords

Related goods, composite-commodity theorem, substitution, complementarity, input/output relation

1 Introduction

The fundamental principle of related goods in terms of demand for a group of such goods dates back to Hicks in the form of his composite-commodity theorem saying "when the relative prices of a group of commodities can be assumed to remain unchanged, they can be treated as a single commodity" ([1], p. 50).

Applications of composite-commodity theorem have been rare so far and have been typically focused (Chen and others [2], [3]) on examining individual possible different relations of pairs of commodities (specified as substitutive or complementary, input-output connected or joint products, and some other not generally applicable relations). All these examinations have been performed separately, according to these mentioned types of related commodities. And moreover, what has not been taken into consideration in these distinctions of relevant relations is a possible interpretation of related

[1]This work was partly supported by Czech Science Foundation, # 402/04/0642

commodities as tradable factors of production, among which the dominance of complementarity exists ([1], p. 98).

Therefore, a broader variety of possible composite commodities can surely be constructed and assessed [4], and we can also speak in this connection about multiple relations of composite commodities.

2 Variety of relations between tradable commodities

In the above mentioned generalized but specific vision of traded commodities in terms of shared demand, we have to realize that there is a broader variety of relations among those traded commodities than envisaged by above mentioned specifications in the published literature. Individual traded commodities are generally mostly related to each other, and they can be observed and classified in this respect also with regard to their position or role in some relevant production function.

Two or more traded commodities can be mutually related as standard inputs (e.g. raw materials, components, intermediate products) for a specific final product in the form of substitution or in the form of complementarity. Under the term 'standard', continuum of goods is supposed satisfying the relevant demand.

Some traded commodities, on the other hand, can be characterized in their supply as discrete goods due to their, let us say, specific role as specific inputs in the relevant production function (supplies of capital goods, transfer of process technology).

Finally, there is the relevant final product presented itself as a traded commodity having a relationship to other commodities which are possible inputs (standard or specific) in the relevant production function.

In this contribution, we consider a broader specific application, in which a multiple relation is studied. The description and analysis of the corresponding model are presented in details in the following two sections.

3 Multiple relation model

A specific case of a multiple relation is studied in this section. A tradable commodity C_1 is related as a common output to two input commodities C_2, C_3, which, at the same time, are related to each other in a complementarity relation.

The combined multiple relation consisting of two parallel input-output relations and one complementarity relation can be described by the following system of linear equations

$$\begin{aligned} Q_1 &= a_{10} - a_{11}P_1 + u_1 \\ Q_2 &= a_{20} + a_{21}P_1 - a_{22}P_2 - a_{23}P_3 + u_2 \\ Q_3 &= a_{30} + a_{31}P_1 - a_{32}P_2 - a_{33}P_3 + u_3 \end{aligned}$$

$$\begin{aligned} S_1 &= b_{10} + b_{11}P_1 - b_{12}P_2 - b_{13}P_3 + v_1 \\ S_2 &= b_{20} + b_{21}P_1 + b_{22}P_2 + v_2 \\ S_3 &= b_{30} + b_{31}P_1 + b_{33}P_3 + v_3 \end{aligned}$$

where Q_1, Q_2, Q_3 (S_1, S_2, S_3) denote the demand (supply) volumes of commodities C_1, C_2, C_3. Variables P_1, P_2, P_3 are the prices of commodities and the variables u_1, u_2, u_3 (v_1, v_2, v_3) are random disturbances in the demand (supply) function for each commodity. Variables u_1, u_2, u_3 and v_1, v_2, v_3 are called nonsystematic risks. All the risk variables are stochastically independent of each other, and their expectation values are equal to zero, i.e. $E(u_i) = E(v_i) = 0$ for $i = 1, 2, 3$.

The values of all coefficients in the model are assumed to be positive. The negative influence of coefficients a_{11}, a_{22}, a_{33} in the model is shown by the minus sign. Similar convention is used for coefficients a_{23}, a_{32}, expressing the complementarity relation between C_2 and C_3 and coefficients b_{12}, b_{13} expressing the parallel input-output relations.

Using matrix notation, the above model can be written in a compact way, as a special case of a more general model

$$\mathbf{Q} = \mathbf{a}_0 + \mathbf{AP} + \mathbf{u} \tag{1}$$

$$\mathbf{S} = \mathbf{b}_0 + \mathbf{BP} + \mathbf{v} \tag{2}$$

with vectors

$$\mathbf{Q} = \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \quad \mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \\ S_3 \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$$

$$\mathbf{a}_0 = \begin{bmatrix} a_{10} \\ a_{20} \\ a_{30} \end{bmatrix} \quad \mathbf{b}_0 = \begin{bmatrix} b_{10} \\ b_{20} \\ b_{30} \end{bmatrix} \quad \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

Matrices $\mathbf{A}, \mathbf{B}$ in the considered model have the form

$$\mathbf{A} = \begin{bmatrix} -a_{11} & 0 & 0 \\ a_{21} & -a_{22} & -a_{23} \\ a_{31} & -a_{32} & -a_{33} \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_{11} & -b_{12} & -b_{13} \\ b_{21} & b_{22} & 0 \\ b_{31} & 0 & b_{33} \end{bmatrix}$$

The equilibrium state of the model is characterized by the equality of demand and supply, i.e. by the equation $\mathbf{Q} = \mathbf{S}$, which implies

$$\mathbf{a}_0 + \mathbf{AP} + \mathbf{u} = \mathbf{b}_0 + \mathbf{BP} + \mathbf{v}$$

Hence, the equilibrium prices $\mathbf{P}^\star$ fulfill the matrix equation

$$(\mathbf{A} - \mathbf{B})\mathbf{P}^\star = \mathbf{b}_0 - \mathbf{a}_0 + \mathbf{v} - \mathbf{u}$$

Denoting $\mathbf{M} := \mathbf{A} - \mathbf{B},\ \mathbf{c}_0 := \mathbf{b}_0 - \mathbf{a}_0,$ and $\mathbf{w} := \mathbf{v} - \mathbf{u}$, we get

$$\mathbf{MP}^\star = \mathbf{c}_0 + \mathbf{w}$$

We may assume that the equilibriuim prices $\mathbf{P}^\star$ are uniquely determined. Then matrix $\mathbf{M}$ is regular and prices $\mathbf{P}^\star$ can be described by formula

$$\mathbf{P}^\star = \mathbf{M}^{-1}(\mathbf{c}_0 + \mathbf{w}) \tag{3}$$

To express the formula (3) in more detail, we compute first the inverse matrix $\mathbf{M}^{-1}$. By the well-known formula using algebraic minors of matrix $\mathbf{M}$, we have

$$\mathbf{M}^{-1} = \frac{1}{|\mathbf{M}|}\begin{bmatrix} |M_{11}| & -|M_{21}| & |M_{31}| \\ -|M_{12}| & |M_{22}| & -|M_{32}| \\ |M_{13}| & -|M_{23}| & |M_{33}| \end{bmatrix} =$$

$$= \frac{1}{|\mathbf{M}|}\begin{bmatrix} m_{22}m_{33} - m_{23}m_{32} & m_{13}m_{32} - m_{12}m_{33} & m_{12}m_{23} - m_{13}m_{22} \\ m_{23}m_{31} - m_{21}m_{33} & m_{11}m_{33} - m_{13}m_{31} & m_{13}m_{21} - m_{11}m_{23} \\ m_{21}m_{32} - m_{22}m_{31} & m_{12}m_{31} - m_{11}m_{32} & m_{11}m_{22} - m_{12}m_{21} \end{bmatrix}$$

The detailed formula for $\mathbf{P}^\star$ can be then obtained by inserting the original coefficients from expressions

$$\mathbf{M} = \begin{bmatrix} -a_{11} - b_{11} & b_{12} & b_{13} \\ a_{21} - b_{21} & -a_{22} - b_{22} & -a_{23} \\ a_{31} - b_{31} & -a_{32} & -a_{33} - b_{33} \end{bmatrix} \qquad \mathbf{c}_0 = \begin{bmatrix} b_{10} - a_{10} \\ b_{20} - a_{20} \\ b_{30} - a_{30} \end{bmatrix}$$

$$P_1^\star = |\mathbf{M}|^{-1}\Big((m_{22}m_{33} - m_{23}m_{32})(c_{10} + w_1) + (m_{13}m_{32} - m_{12}m_{33})(c_{20} + w_2) + \\ + (m_{12}m_{23} - m_{13}m_{22})(c_{30} + w_3)\Big)$$

$$P_2^\star = |\mathbf{M}|^{-1}\Big((m_{23}m_{31} - m_{21}m_{33})(c_{10} + w_1) + (m_{11}m_{33} - m_{13}m_{31})(c_{20} + w_2) + \\ + (m_{13}m_{21} - m_{11}m_{23})(c_{30} + w_3)\Big)$$

$$P_3^\star = |\mathbf{M}|^{-1}\Big((m_{21}m_{32}-m_{22}m_{31})(c_{10}+w_1)+(m_{12}m_{31}-m_{11}m_{32})(c_{20}+w_2)+ \\ +(m_{11}m_{22}-m_{12}m_{21})(c_{30}+w_3)\Big)$$

The equilibrium trade volumes $\mathbf{Q}^\star$ are then simply computed by putting the value of vector $\mathbf{P}^\star$ into the formula $\mathbf{Q}^\star = \mathbf{a}_0 + \mathbf{A}\mathbf{P}^\star + \mathbf{u}$. We get

$$\mathbf{Q}^\star = \mathbf{a}_0 + \mathbf{A}\mathbf{M}^{-1}(\mathbf{c}_0 + \mathbf{w}) + \mathbf{u} \tag{4}$$

The detailed formula for $\mathbf{Q}^\star$ is computed analogously as above.

4 Comparison of variances

In this section we compare the variances of equilibrium prices $\mathbf{P}^\star$ and equlibrium trade volumes $\mathbf{Q}^\star$ in the multiple relation model of the section 3 with those of equilibrium prices $\mathbf{P}^{\star\star}$ and equlibrium trade volumes $\mathbf{Q}^{\star\star}$ in the binary relation model with independent commodity C_1 and with the complementarity relation between commodities C_2, C_3. The binary relation model is described by the system of equations

$$\begin{aligned} Q_1 &= a_{10} - a_{11}P_1 + u_1 & \qquad S_1 &= b_{10} + b_{11}P_1 + v_1 \\ Q_2 &= a_{20} - a_{22}P_2 - a_{23}P_3 + u_2 & \qquad S_2 &= b_{20} + b_{22}P_2 + v_2 \\ Q_3 &= a_{30} - a_{32}P_2 - a_{33}P_3 + u_3 & \qquad S_3 &= b_{30} + b_{33}P_3 + v_3 \end{aligned}$$

The equilibrium values $\mathbf{P}^{\star\star}$ and $\mathbf{Q}^{\star\star}$ are computed analogously as in section 3. Denoting

$$\widetilde{\mathbf{M}} = \begin{bmatrix} -a_{11}-b_{11} & 0 & 0 \\ 0 & -a_{22}-b_{22} & -a_{23} \\ 0 & -a_{32} & -a_{33}-b_{33} \end{bmatrix}$$

we have

$$\mathbf{P}^{\star\star} = \widetilde{\mathbf{M}}^{-1}(\mathbf{c}_0 + \mathbf{w}) \tag{5}$$

$$\mathbf{Q}^{\star\star} = \mathbf{a}_0 + \mathbf{A}\widetilde{\mathbf{M}}^{-1}(\mathbf{c}_0 + \mathbf{w}) + \mathbf{u} \tag{6}$$

Using the expressions (3), (5), we get for $P_1^\star$ and $P_1^{\star\star}$

$$\begin{aligned} \mathrm{Var}(P_1^\star) &= |\mathbf{M}|^{-2}\big(|M_{11}|^2\,\mathrm{Var}(w_1) + |M_{21}|^2\,\mathrm{Var}(w_2) + |M_{31}|^2\,\mathrm{Var}(w_3)\big) \\ \mathrm{Var}(P_1^{\star\star}) &= |\widetilde{\mathbf{M}}|^{-2}\big(|M_{11}|^2\,\mathrm{Var}(w_1)\big) \end{aligned}$$

Similarly, for $Q_1^\star$ and $Q_1^{\star\star}$

$$\mathrm{Var}(Q_1^\star) = |\mathbf{M}|^{-2}\Big(a_{11}^2\big(|M_{11}|^2\,\mathrm{Var}(w_1) + |M_{21}|^2\,\mathrm{Var}(w_2) + \\ + |M_{31}|^2\,\mathrm{Var}(w_3)\big) + \mathrm{Var}(u_1)\Big)$$

$$\mathrm{Var}(Q_1^{\star\star}) = |\widetilde{\mathbf{M}}|^{-2}\Big(a_{11}^2\big(|M_{11}|^2\,\mathrm{Var}(w_1)\big) + \mathrm{Var}(u_1)\Big)$$

As the matrix M is regular, we may assume, without any loss of generality, that $|\mathbf{M}| > 0$. Then it is easy to compute

$$|\widetilde{\mathbf{M}}| - |\mathbf{M}| = b_{12}\,|M_{12}| - b_{13}\,|M_{13}|$$

Let us assume that the following condition is fulfilled

$$b_{12}\,|M_{12}| - b_{13}\,|M_{13}| > 0 \tag{7}$$

Then we have $|\widetilde{\mathbf{M}}| > |\mathbf{M}| > 0$, $|\widetilde{\mathbf{M}}|^2 > |\mathbf{M}|^2$ and $|\widetilde{\mathbf{M}}|^{-2} < |\mathbf{M}|^{-2}$ and, as a consequence, the inequalities

$$\mathrm{Var}(P_1^\star) > \mathrm{Var}(P_1^{\star\star}), \quad \mathrm{Var}(Q_1^\star) > \mathrm{Var}(Q_1^{\star\star})$$

hold true. This means that under the condition (7) the volatility of the price $P_1^\star$ in the multiple-related model is larger than the volatility of the price $P_1^{\star\star}$ in the model without the input-output relation. However, if the condition (7) is not satisfied, the relations between prices $\mathrm{Var}(P^\star)$ and $\mathrm{Var}(P^{\star\star})$, as well as the relations between the traded volumes $Q_1^\star$ and $Q_1^{\star\star}$, are ambiguous. Similar conclusion can be stated about the prices and the traded volumes of the commodities C_2, C_3.

5 Conclusion

This contribution is a natural extension of possible applications of composite-commodity theorem perceived earlier in the published literature, even after taking into consideration other extensions.

It seems that only in our understanding of multiple relations of composite commodities it could be possible to create such a set of relations which wouldl have implications both on the demand side and the supply side, where prices of commodities (on the supply side) in the position of exogenous variables would be influenced in one certain way while quantities of commodities (on the demand side) in the position of exogenous variables

would be influenced in some other way. But it is still a long way to go, to be able to make proof of this general hypothesis by using our instruments.

So far, we have tested two above defined relations of traded commodities making together a first multiple relation and have created its relevant model in the framework of composite commodities. We have found that the volatility of prices and traded volumes in the considered multirelational model is larger than in the model with one single relation. However, this observation is true only when the coefficients of the model satisfy one simple condition, and without this condition the relation between models is ambiguous. Our considerations indicate that similar results could be true also for further types of multirelational models.

References

[1] J.R. Hicks, *Value and Capital*, Oxford University Press, Oxford, 1st ed. 1939, reprinted 1941.

[2] J.R. Chen, *Modelling of Related Commodities and Its Implications for Stabilization Policy of the World Market*, Journal of Agricultural Economics **55** (1994), 139–166.

[3] J.R. Chen, H. Stocker, *Composite Commodities and the Prebisch-Singer Thesis*, World Development **25** (1997), 1863-1871.

[4] D. Moro, *Aggregation without Separability : Composite Commodity Theorems in Quantity Space*, Economics Letters **71** (2001), 67-73.

Applications of Gini's Mean Difference to Portfolio Analysis

Agata Gluzicka and Donata Kopańska - Bródka
Department of Operations Research, University of Economics in Katowice

Abstract

Gini's mean difference is a measure of dispersion defined as expected absolute pairwise difference of all variates within a distribution. The Gini mean difference is a popular measure of the income inequality in the economics literature.
According to Markowitz, investors make investment decisions employed variance is a measure of risk (mean-variance model), but Yitzhaki presents models in which risk is measured by Gini's mean difference (mean – Gini model).
This paper is a review across different definitions of Gini's mean differences and extended Gini's index. The autors will show more significant properties of this coefficient. The Gini's mean will be applied to construct optimal portfolios for companies listed on Warsaw Stock Exchange.

Keywords

Gini mean differences, the dominance criteria, efficient portfolio

1 Introduction

In 1912 Gini proposed the coefficient as a measure of the variability of any statistical distribution of probability. He based his coefficient on the average of the absolute difference between pairs of observations. Yitzhaki (1983) introduced an extended version of Gini's mean, which was a modification based on levels of risk aversion. He also defined two efficiency criteria to sort uncertain prospects into efficient sets. Shalit and Yitzhaki (1984) developed the Gini's mean model of portfolio analysis and derived the Gini's mean equilibrium pricing of risky assets. Now in economics literature we can find many applications of Gini's mean as a measure of risk.

2 Properties of Gini's mean difference

Gini's mean difference is an index of the variability of a random variable. It is based on the expected value of the absolute difference between every pair of realizations of the random variable. Gini's mean difference is a measure of dispersion defined as the expected absolute pairwise difference of all variables within a distribution. The first time this coefficient was introduced by Gini [3], who expressed this coefficient as:

$$\Delta = E|X - Y|,$$

where X and Y are independent random variable with cumulative distribution F.
Kendall and Stuart [3] defined Gini's mean difference (Δ) in two different cases - to take into account the type of distributions.
If F is a discrete distribution then we can the Gini's mean difference defined as:

$$\Delta = \frac{1}{T^2}\sum_{k=1}^{T}\sum_{i=1}^{T}\left|X_i - X_k\right| f(X_i) f(X_k)$$

where: X_i, X_k are the consecutive observations of variable X, T is the number of observations and f(X) is a density function of distribution F.
In the case when distribution F is continuous, the Gini's mean difference can be expressed as:

$$\Delta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|X - Y\right| dF(X) dF(Y)$$

for independent random variable X and Y.
Yitzhaki [2, 5] proposed a few modifications of the Gini's mean difference. He also introduced the dominance criteria, which are expressed in terms of 1/2Δ, rather then Δ. Consequently, in portfolio applications 1/2Δ, noted as Γ, is used rather the full Gini's mean.
Let F(X) and f(X), respectively, represent the cumulative distribution and density function of variable X, and assume that there exist $a \geq -\infty$ and $b \leq +\infty$ such that F(a)=0 and F(b)=1. For the random variables X and Y, Yitzhaki defined the Gini's mean difference as follows:

$$\Gamma = \frac{1}{2}\int_a^b \int_a^b |X - Y| f(X) f(Y) dX dY$$

If the distribution is discrete with T possible values, then:

$$\Gamma = \frac{1}{T^2}\sum_{j=1}^{T}\sum_{k>j}^{T} |X_j - X_k|$$

In the case when we have only one random variable, we can define the Gini's mean difference as follows [5]:

$$\Gamma = \int_a^b [1 - F(X)] dX - \int_a^b [1 - F(X)]^2 dX$$

or

$$\Gamma = \overline{X} - a - \int_a^b [1 - F(X)]^2 dX \text{ for finite value a.}$$

where F(X) is the cumulative distribution of variable X and $\overline{X}$ is the expected value of the distributions.
The Gini's mean can also be defined as the covariance between the rate of return and its probability distribution [4]:

$$\Gamma = 2\text{cov}\,[r, F(r)]$$

where r is the rate of return and F(r) is the cumulative distribution function.
Some interesting properties [5] of the Gini's mean difference follow immediately from these definitions. The Gini's coefficient is nonnegative and bounded from above by ($\overline{X} - a$). If X is a given constant then Gini's coefficient Γ = 0. Furthermore, its maximum value is reached for the distribution:

$$f(X) = \begin{cases} \frac{1}{2} & X = a \\ \frac{1}{2} & X = b \\ 0 & X \neq a, X \neq b \end{cases}$$

If X is a normally distributed variable with $\overline{X}$ and σ^2, then Gini's mean difference is following:

$$\Gamma = \sigma / \sqrt{\Pi}$$

3 The relationship between Gini's mean and stochastic dominance

Yitzhaki has defined two efficiency criteria to sort uncertain prospects into efficient and inefficient sets. He has shown that the Gini's mean efficiency conditions are necessary conditions for first and second-degree stochastic dominance. There are that criteria:

Theorem 1. Let X_1 and X_2 be a two independent random variable**.** The conditions

$$\overline{X}_1 \geq \overline{X}_2$$

and

$$\overline{X}_1 - \Gamma_1 \geq \overline{X}_2 - \Gamma_2$$

are necessary conditions for X_1 to dominate X_2 according to first and second stochastic dominance rules [5].

Theorem 2. Let X_1 and X_2 be a two random variable with equal expected return. Assume also that the cumulative distributions $F_1(X)$ and $F_2(X)$ intersect at most once. Then the condition:

$$\overline{X}_1 - \Gamma_1 > \overline{X}_2 - \Gamma_2$$

is a sufficient condition for X_1 to dominate X_2 according to stochastic dominance rules [5].
Γ_1 and Γ_2 denote Gini's mean differences for variable X_1 and X_2 respectively,

4 The Gini's mean in investment decisions

In the Gini's mean models, the investor use the portfolio's Gini's index as the measure of risk to be minimized, subject to a given mean return. Gini's mean difference is expressed as twice the covariance between the variable and its cumulative distribution. Thus, Gini's mean is simply [4]:

$$\Gamma=2\mathrm{cov}[X; F(X)]$$

In practice, we analyze only a population sample and estimate the Gini's mean from these data points.
Let's consider a portfolio, whose rate of return Y is obtained by $Y = \sum_{i=1}^{N} w_i R_i$ where R_i is the rate of return of the i-th security, and w_i denotes the weight. Then the Gini's mean is written as:

$$\Gamma = 2\mathrm{cov}[Y; F]=2\mathrm{cov}[\sum_{i=1}^{N} w_i R_i ; F]=2\sum_{i=1}^{N} w_i \mathrm{cov}[R_i; F]=2w^T K[R; F]$$

where w is a vector of portfolio weights (share of security), K[R; F] is a vector of covariance of assets rates of return with the cumulative distribution function (F) of the rates of return of portfolio. In applications, the empirical cumulative distribution function is used as an estimator for the cumulative distribution. To obtain this estimator we need to rank the rates of return of the portfolio in increasing order and divide the rank of each observation by the number of observations.
We have the following optimization problem:

$$2w^T K[R; F] \rightarrow \text{minimize}$$

subject to:

$$w^T \mu=\mu_p$$
$$\mathbf{1}=w^T \mathbf{1}$$
$$w \geq 0$$

where **1** is a vector of ones, μ_p is a given level of the mean the rate of return of portfolio and μ is a vector of assets means of rates of return.
We can solve the problem numerically for different values of μ_p and thus obtain the efficient set.
In our research we constructed efficient sets using two models. The first one is based on Mean-Variance approach (MV model). This model is following:

$$V^2(Y) \rightarrow \text{minimize}$$

subject to:

$$w^T\mu = \mu_p$$
$$\mathbf{1} = w^T\mathbf{1}$$
$$w \geq 0$$

where $V^2(Y)$ is the variance of rate of return of portfolio.
In order to analyze the differences between M-V approach and Gini's mean model in portfolio selection problem the empirical test was performed.
Our study included a comparison of the M-V and Gini's mean efficient sets, the level of diversification and a test of the efficiency. Weekly data (i. e. weekly rates of return) from selected stock for the period 2000-2004 were used to generate an M-V efficient set of portfolios based on the Markowitz model and the Gini's mean model formulation. The output from these models gave, at each specified value of expected rate of return, the corresponding portfolio with the smallest variance or the Gini's mean difference. We constructed portfolios in two groups – SPOŻYWCZE and TECHWIG. In both groups we had eleven securities. We constructed the next portfolio for different level of portfolio mean rate of return.
Our results are presented the following figures and tables:

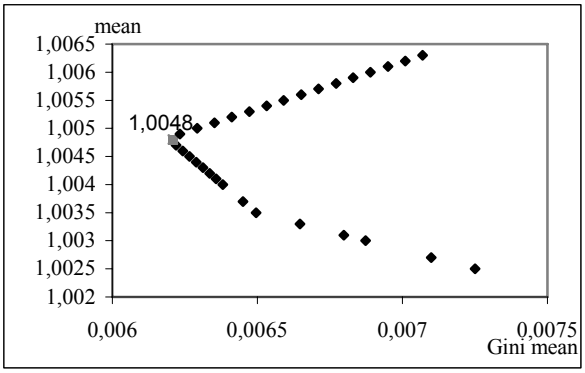


Figure 1. The relationship between the Gini's mean and the mean of portfolio in SPOŻYWCZE group

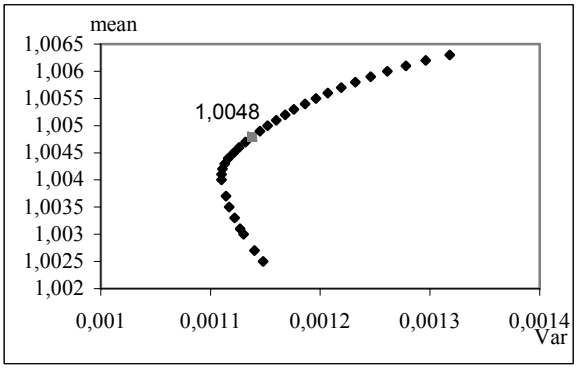


Figure 2. The relationship between the variance and the mean of portfolio in SPOŻYWCZE group

μ_p	1,0048	1,0049	1,005	1,0051	1,0052	1,0053	1,0054	1,0055	1,0056	1,0057	1,0058	1,0059	1,006	1,0061	1,0062	1,0063
Level of diversification – MV model	6	6	6	6	6	5	5	5	5	4	4	4	4	4	3	1
Level of diversification – MG model	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Number of common securities in portfolios	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 1. Comparison of the levels of diversification of selected securities from the SPOZYWCZE group

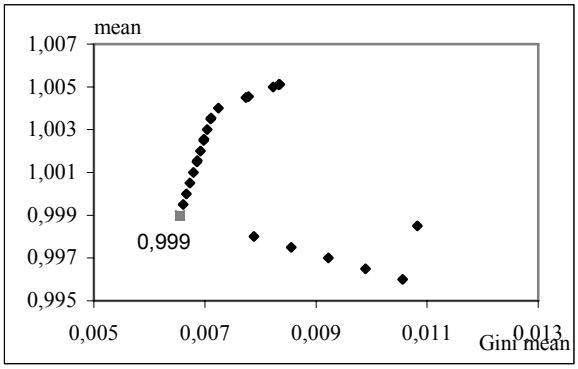


Figure 3. The relationship between the Gini's mean and the mean of portfolios in TECHWIG group

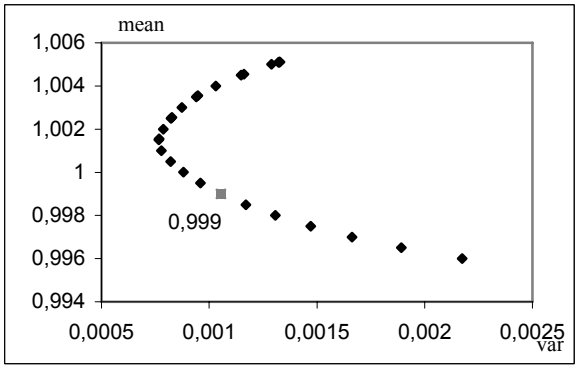


Figure 4. The relationship between the variance and the mean of portfolio in TECHWIG group

μ_p	1,0048	1,0049	1,005	1,0051	1,0052	1,0053	1,0054	1,0055	1,0056	1,0057	1,0058	1,0059	1,006	1,0061	1,0062	1,0063
Level of diversifications – MV model	6	6	6	6	6	5	5	5	5	4	4	4	4	4	3	1
Level of diversifications – MG model	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Number of common securities in portfolios	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 2. Comparison of the levels of diversification of selected securities from the TECHWIG group

5 Conclusion

The main results of this study are as follows:

We compared the efficiency frontier of MG and MV portfolios. The gray point at figure 1 is a MG portfolio of global minimum Gini's mean. At figure 2, we indicated portfolio with the same rate of return. This portfolio is a one of efficient MV portfolios. If we compared the portfolios with the same rate of return in TECHWIG group we obtained different result. At figure 3 by gray point we indicated the MG portfolio of global minimum Gini's mean. The MV portfolio with the same rate of return is not efficient portfolio (gray point at figure 4).

The MV portfolio with the highest required rate of return consists of only one component. The MG portfolio with the same highest rate of return consists of the shares of two securities. There two compared portfolios have the same one component (table1, table 2).

In consideration of portfolio composition, the optimal portfolios according to the Gini's mean model are less diversified than portfolios selected by the M-V criteria.

	Number of portfolio	The rates of return of i-th MV portfolios	The rates of return of i-th MG portfolios
Data from 1st week	1	0,9992	1,0033
	2	0,9997	1,0002
	3	1,0005	1,0012
	4	1,0012	1,0023
Data from 2nd week	1	0,9998	1,0002
	2	1,0008	1,0045
	3	1,0020	1,0057
	4	1,0034	1,0068

Table 3. The real rates of return for indicated portfolios

For different four MV and MG efficient portfolios, with the same required rate of return, we compared the real rates of return, which were calculated on the base of the data from the next few weeks after the last observation (starting from 3rd January 2005). In SPOŻYWCZE group the MG portfolios had better rates of return than MV portfolios. The results are presented in table 3. For portfolios in TECHWIG group the differences between rates of return of MG portfolio and MV portfolio are not significant.

Our results suggest that the Gini's mean model can be better choice than the well know Mean-Variance model.

6 References

[1] Dorfman R. "A formula for the Gini coefficient", The Review of Economics and Statistics (February 1979), pg. 146-149
[2] Okunev J. "The generation of mean Gini efficient sets", Journal of Business Finance and Accounting (January 1991), pg.209-218
[3] Shaffer D., "Estimating the Gini Hedge Ratio", Managerial Finance (2003), vol. 29, pg.73-84
[4] Shalit H., Yitzhaki S. "The mean-Gini efficient portfolio frontier", The Journal of Financial Research (2005), vol. XXVIII, pg. 59-75
[5] Yitzhaki S., Shalit H., "Mean-Gini, Portfolio theory and the Pricing of Risky Assets", The Journal of Finance (December 1984), vol. XXXIX, pg. 1449-1468
[6] Yitzhaki S. "Stochastic Dominance, Mean Variance and Gini's Mean Difference", The American Economic Review (March 1982), vol. 72, pg.178-185

The Knowledge Granules of Multicriteria European Countries Classification Problem

Iwona Gruszka
The Karol Adamiecki University of Economics, Department of Operational Research, Katowice
gruszka@ae.katowice.pl

Abstract

Presently, in the age of the process of European Union enlargement, one of the most important problem is to answer the questions about the classification of EU members by the economic characteristics into the groups according to The World Bank Organization. The aim of this work is to extract the knowledge granules of multicriteria European countries classification problem.

Keywords

Multicriteria Classification Problem, Rough Sets, Granulation of Knowledge

1. Introduction

Presently, in the age of the European Union creation process, one of the most important problem is to answer the questions about the classification of EU's members by the economic characteristics into the groups according to The World Bank Organization.

The classical methods that allow solving the problem come from the statistics and econometrics areas. However they are very useful but often they can't be used because of the real data does not satisfy the strong assumption like multinormal distribution condition. The methods from artificial intelligence pool could be used, e.g. the rough sets theory, instead of statistical and econometrical once.

The rough sets theory was introduced by Pawlak (1982). The mathematical base on rough sets approach is a binary relation on universe of objects. In the classic rough sets theory there is an indiscernibility relation. As an equivalence relation it permits to divide the universe of objects on equivalence classes called elementary sets and forms basic granules of knowledge of the universe. However, from the other point of view, it is quiet natural to extend the indiscernibility concept to situations where ones objects dominates to another ones by considering the criteria which domains are preferentially ordered. For this reason Greco, Materazzo and Słowiński (1996, 1999) have proposed an extension of the rough set theory. This innovation is based on substitution of the indiscernibility relation (reflexive, symmetric, transitive) by a dominance relation (reflexive, transitive).

The aim of the analysis is to extract the knowledge granules of multicriteria European countries classification problem. The extraction was conduced based on dominance rough sets approach in the „4eMka" system designed in Poznań University of Technology, Poland, Institute of Computing Science, by the Słowiński group.

2. The rough sets approach based on dominance relation to multicriteria classification problem[1]

Rough set based data analysis starts from a data table, called a decision table. The decision table contains data about objects of interests evaluated in terms of some criteria. We distinguish in the decision table condition (C) and decision (d) criteria. The condition criteria provide as the information about the universe of considered objects, while the decision criteria describes the classification decision made by decision maker (an a priori classification). The decision table describes decision in terms of conditions that must be satisfied in order to carry out the specified decisions. With every decision table we can associate a set of decision rules, which forms a decision model and which can be seen as a logical description of approximations of decisions.

Let $S=<U,Q,V>$ be a decision table, where:

U – a finite set of objects (universe), $U = \{x_1,...,x_m\}, m \in N$

C – a finite set of condition criteria, $C = \{c_1,...,c_n\}, n \in N$

d – a decision criterion,

V_C – a domain of criterion c,

[1] Described on the basis of Greco, Materazzo, Slowinski (1996), (1996a), (1999).

V – a criterion's domain: $V = \bigcup_{c \in C} V_c$.

Let $Y = \{Y_t : t \in T\}, T = 1,...,k, k \in N$, be a set of decisional classes of U , such that each $x \in U$ belongs to one and only one of class $Y_t \in Y$. We assume that for every $r, s \in T$, such that $r > s$, each object of Y_r is preferred to each object of Y_s [2].

Let us also consider the following upward and downward cumulated classes, respectively:

$$Y_t^{\leq} = \bigcup_{s \leq t} Y_s \,, \tag{1}$$

$$Y_t^{\geq} = \bigcup_{s \geq t} Y_s \,. \tag{2}$$

It is said that x dominates y with respect to C (notation $xD_c y$) if x is at least good as y according to each criteria $c \in C$. By the domination relation there is possible to describe the granules of the knowledge of universe U in form of n – dimensional cones . Given $x \in U$, let us describe the *C*- dominating set of x and *C*-dominated set of x respectively:

$$D_C^+(x) = \{y \in U : yD_C x\}, \tag{3}$$

$$D_C^-(x) = \{y \in U : xD_C y\}. \tag{4}$$

We can define the *C*-lower and the *C*-upper approximation of $Y_t^{\geq}$ (notation $\underline{C}Y_t^{\geq}$) and $Y_t^{\geq}$ (notation $\overline{C}Y_t^{\geq}$), $t \in T$, respectively as:

$$\underline{C}Y_t^{\geq} = \{x \in U : D_C^+(x) \subseteq Y_t^{\geq}\}, \tag{5}$$

$$\overline{C}Y_t^{\geq} = \bigcup_{x \in Y_t^{\geq}} D_C^+(x). \tag{6}$$

Analogously, we define the *C*-lower and the *C*-upper approximation of $Y_t^{\leq}$ (notation $\underline{C}Y_t^{\leq}$) and $Y_t^{\leq}$ (denotation $\overline{C}Y_t^{\leq}$), $t \in T$, respectively as:

$$\underline{C}Y_t^{\leq} = \{x \in U : D_C^-(x) \subseteq Y_t^{\leq}\}, \tag{7}$$

$$\overline{C}Y_t^{\leq} = \bigcup_{x \in Y_t^{\leq}} D_C^-(x). \tag{8}$$

The *C*-boundaries (*C*-doubtful regions) of $Y_t^{\geq}$ and $Y_t^{\leq}$, $t \in T$ are defined as

$$BN_C(Y_t^{\geq}) = \overline{C}Y_t^{\geq} \setminus \underline{C}Y_t^{\geq}, \tag{9}$$

$$BN_C(Y_t^{\leq}) = \overline{C}Y_t^{\leq} \setminus \underline{C}Y_t^{\leq}. \tag{10}$$

From the approximations of cumulated decisions rules the rules of decisional model are induced. In the „4eMka" system, designed in The Institute of Computer Sciences at The Poznań University of Technology in Poland. The rules can take following forms (Grzymała-Busse (1992), Stefanowski (2001)).

3. Application

The decisional table of considering the real multicriteria European countries classification problem contains 733 countries (15 countries evaluated in 35 periods of time, and 13 countries evaluated in 14 periods of time) constituting the objects of universe, and 6 conditional criteria: unemployment rate, gross saving, final consumption expenditure of general government at current prices, imports of goods and services, of goods and services. The decisional criterion here is the gross domestic product per capita, which is the common macroeconomic indicator used in measuring the level of economic development. The a priori partition of universe is like follow:

- Y_H - the decisional class of high income countries (525 countries),
- Y_{UM} - the decisional class of upper middle income countries (73),
- Y_{LM} - the decisional class of lower middle income countries (136).

[2] The preference relation is presented in Roy (1985).

The countries where evaluated in related periods of time:

1. Italy, France, Luxemburg, Belgium, Germany, Netherlands, Denmark, Ireland, United Kingdom, Greece, Spain, Portugal, Austria, Finland, Sweden: 1960, 1965, 1970, 1971, 1972, 1973, 1974, 1975, 1976, 1977, 1978, 1979, 1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004,
2. Poland, Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Slovakia, Slovenia, Bulgaria, Romania, Turkey: 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004.

All the data is given from the Eurostat sources
(http://europa.eu.int/comm/economy-finance/publications/statistical-en.htm,ECFIN/ 174/2004-EN).
There is not possible to present all the granules. Author is presenting the diagrams of capacity of the granules (Figures 1., 2., and 3.), then.

Fig.1. The capacity of $D_C^+(x)$ granules, where x is equal: Italy, France, Luxemburg, Belgium, Germany, Netherlands, Denmark, Ireland (Fig.1a), United Kingdom, Greece, Spain, Portugal, Austria, Finland, Sweden (Fig.1b).

Fig.2. The capacity of $D_C^-(x)$ granules, where x is equal: Italy, France, Luxemburg, Belgium, Germany, Netherlands, Denmark, Ireland (Fig.2a), United Kingdom, Greece, Spain, Portugal, Austria, Finland, Sweden (Fig.2b).

Fig.3. The capacity of $D_C^+(x)$ granules, where x is equal: Poland, Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Slovakia, Slovenia, Bulgaria, Romania, Turkey (Fig.3a), and the capacity of $D_C^=(x)$ granules, where x is equal: Poland, Czech Republic, Estonia, Cyprus, Latvia, Lithuania, Hungary, Slovakia, Slovenia, Bulgaria, Romania, Turkey (Fig.3b).

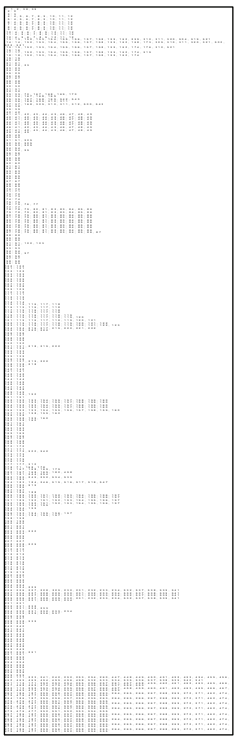


Fig. 1a

Fig. 1b

Fig. 2a

Fig. 2b

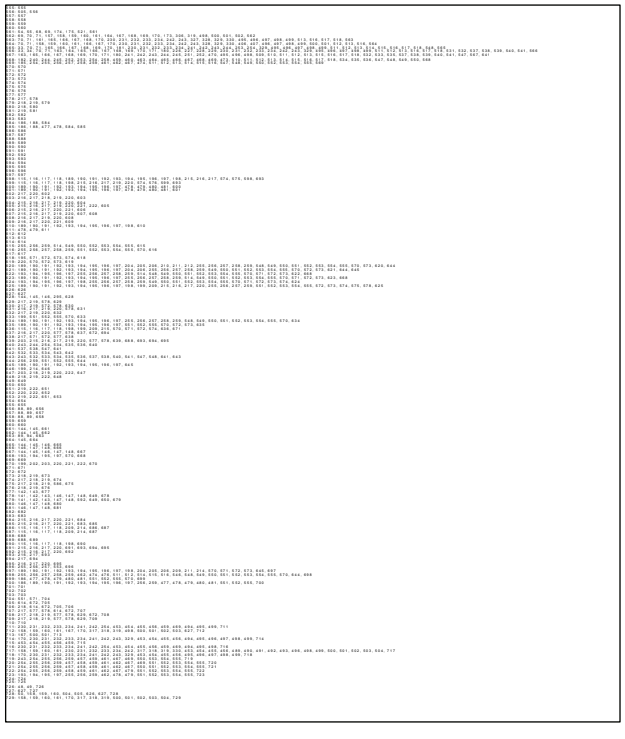


Fig. 3a

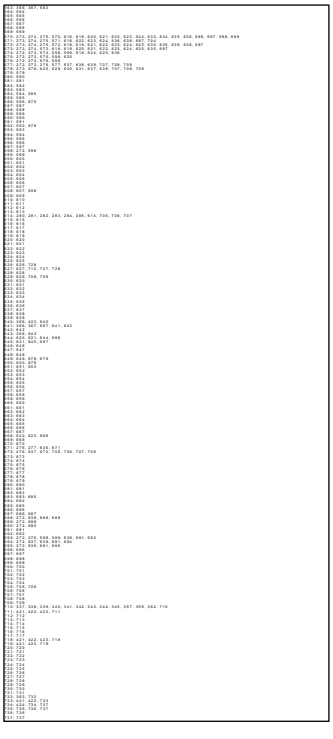


Fig. 3b

4. Conclusions

The rough sets theory is very useful tool to discovery the knowledge of granules in European countries classification problem. The next paper will present the decisional model induced from the approximations of decisional classes made by the $D_C^+(x)$ and $D_C^-(x)$ granules.

References

[1] Greco S., Materazzo B., Slowinski R. (1996) Rough Approximation of a Preference Relation by Dominance Relation, *ICS Research Report 16/96*, 13.

[2] Greco S., Materazzo B., Slowinski R. (1996) On Joint Use of Indiscernibility, Similarity and Dominance Rough Approximation of Decision Classes, *ICS Research Report 16/96*, 11.

[3] Greco S., Materazzo B., Slowinski R. (1999), The Use of Rough Sets and Fuzzy Sets in MCDM, *Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications*, Kluwer Academic Publihers Boston/Dordrecht/London

[4] Grzymała-Busse J.W. (1992) LERS- A System for Learning from Examples Based on Rough Sets, *Intelligent Decision Support. Handbook of Applications and Advances of The Rough Sets Theory*, Slowinski R. (Eds.), Kluwer Academic Publisher, 15.

[5] Pawlak Z. (1982), Rough sets, International Journal of Computer and Information Sciences 11, 15.

[6] Roy B. (1985), Methodologie Multicritere d'Aide a la Decision, Editions Economica, Paris.

[7] Stefanowski J. (2001) Algorytmy indukcji reguł decyzyjnych w odkrywaniu wiedzy, *Rozprawy 361*, Wydawnictwo Politechniki Poznańskiej, Poznań.

Power Indices in Voting by Count and Account

Midori Hirokawa[1] and Milan Vlach[2]

[1] Faculty of Economics, Hosei University
4342 Aihara Machida, Tokyo 194-0298, Japan
mhiro@mt.tama.hosei.ac.jp
[2] The Kyoto College of Graduate Studies for Informatics
7 Monzen-cho Tanaka Sakyo-ku, Kyoto 606-8225, Japan
m_vlach@kcg.ac.jp

Abstract. We consider a game that reflects both simple majority voting (voting by count) and weighted majority voting (voting by account). Such a compound game has a long history and recently found application in decision-making procedures used in some stockholders' meetings and in conducting rehabilitation procedures in financially troubled companies. We show that the typical power indices have the following property: the difference in the compound game between the power index of a voter with greater weight and the power index of a voter with lower weight is not exceeding the analogous difference in the corresponding weighted majority game.

1 Introduction

We consider a compound decision rule of a weighted majority and a simple majority. Each player has a weight and a coalition is winning only when the sum of the weights of the members of the coalition exceeds half of the total weight of all players and, simultaneously, the number of the members of the coalition exceeds half of the number of all players.

Such a compound game has a long history, see Peleg [4] for an example of its use by some Jewish communities in Europe in the seventeenth century. Moreover, applications have been found in decision-making procedures used in some stockholders' meetings and in conducting rehabilitation procedures in financially troubled companies, see Hirokawa and Xu [2].

In the decision making by voting it is worth examining effective power of players endowed with weights. Suppose, for example, that a party A (*resp.,* B, C) has 20% (*resp.,* 40%, 40%) of total votes. This is equivalent to (30%, 30%, 40%) vote distribution in a simple majority voting because, in both cases, the sets of coalitions that can pass the bill is the same. Research on power analysis has been intensive in the past decades and several kinds of power indices has been proposed; for example, see Brams, Lucas and Straffin [1].

We show that two typical power indices have the following property: the difference in the compound game between the power index of a voter with greater

weight and the power index of a voter with lower weight is not exceeding the analogous difference in the corresponding weighted majority game. Peleg [4] presents a similar result on power shift if measured by Shapley-Shubik power index. Here we prove a stronger result, and show that it holds also for the Banzhaf power index. This justifies the claim that, for a voter with a relatively lower weight, the voting by count and account is better than the corresponding weighted majority game.

2 Models of Voting Games and Power Indices

Consider two games, $G = (N, \mathcal{W})$ and $G^* = (N, \mathcal{W}^*)$ where $N = \{1, \cdots, n\}$ is the set of players, and $\mathcal{W}$ and $\mathcal{W}^*$ are the sets of winning coalitions defined as follows: each player i has a weight w_i, and

$$\mathcal{W} = \{S \in \mathcal{P}(N) \,|\, w(S) > \bar{w}\}, \quad \text{and}$$

$$\mathcal{W}^* = \{S \in \mathcal{P}(N) \,|\, w(S) > \bar{w},\ \sharp S > \lfloor \tfrac{n}{2} \rfloor\}.$$

where $\mathcal{P}(N)$ is the set of all subsets of N, $w(S) = \sum_{i \in S} w_i$, and $\bar{w} = \frac{w(N)}{2}$.[1]

In game $G^* = (N, \mathcal{W}^*)$, a coalition is winning only when the sum of the weights of the members of the coalition exceeds half of total weight of all players and, simultaneously, the number of the members of the coalition exceeds half of the number of all players. According to the terminology of Peleg [4], we call G^* *a voting by count and account*, because it reflects both simple majority voting (voting by count) and weighted majority voting (voting by account).

For power indices, the following concept of a *swing voter* has an important role. A player is said to be a swing voter for a given winning coalition if and only if he belongs to the coalition, but if he leaves it then the resulting coalition is losing. Note that for a given winnnig coalition, there may exist several swing voters, and a player may be a swing player for several coalitions. For game G (resp., G^*) and for a given $k \in N$, let $\mathcal{W}(k)$ (resp., $\mathcal{W}^*(k)$) be the set of winning coalitions where k is a swing voter. Formally:

$$\mathcal{W}(k) = \{S \in \mathcal{P}(N) \,|\, k \in S \ \&\ S \in \mathcal{W} \ \&\ S \backslash \{k\} \notin \mathcal{W}\}\ ;$$

$$\mathcal{W}^*(k) = \{S \in \mathcal{P}(N) \,|\, k \in S \ \&\ S \in \mathcal{W}^* \ \&\ S \backslash \{k\} \notin \mathcal{W}^*\}.$$

In order to see the relation between $\mathcal{W}(k)$ and $\mathcal{W}^*(k)$, define $\mathcal{W}^+(k)$ and $\mathcal{W}^-(k)$ as follows:

$$S \in \mathcal{W}^-(k) \quad \text{iff} \quad w(S) > \bar{w} \ \&\ k \in S \ \&\ \sharp S > \lfloor \tfrac{n}{2} \rfloor \ \&\ w(S \backslash \{k\}) \leq \bar{w}\ ;$$

$$S \in \mathcal{W}^+(k) \quad \text{iff} \quad k \in S \ \&\ \sharp(S \backslash \{k\}) = \lfloor \tfrac{n}{2} \rfloor \ \&\ w(S \backslash \{k\}) > \bar{w}.$$

Note that by the definitions of $\mathcal{W}(k)$ and $\mathcal{W}^*(k)$:

[1] For a real number x and a set X, $\lfloor x \rfloor$ denotes the greatest integer which does not exceeds x, and $\sharp X$ denotes the cardinal number of the set X.

$$
\begin{aligned}
S \in \mathcal{W}(k) \quad &\text{iff} \quad w(S) > \bar{w} \ \ \& \ \ k \in S \ \ \& \ \ w(S\backslash\{k\}) \le \bar{w} \ ; \\
S \in \mathcal{W}^*(k) \quad &\text{iff} \quad w(S) > \bar{w} \ \ \& \ \ k \in S \ \ \& \ \ \sharp S > \lfloor \tfrac{n}{2} \rfloor \\
&\qquad\qquad \& \ \ [\, w(S\backslash\{k\}) \le \bar{w} \ \text{ or } \ \sharp(S\backslash\{k\}) = \lfloor \tfrac{n}{2} \rfloor \,].
\end{aligned}
$$

It can easily be seen that :

$$
\mathcal{W}^*(k) = \mathcal{W}^+(k) \cup \mathcal{W}^-(k), \ \ \mathcal{W}^+(k) \cap \mathcal{W}^-(k) = \emptyset,
$$
$$
\mathcal{W}^-(k) \subset \mathcal{W}(k), \ \text{ and } \ \mathcal{W}^+(k) \cap \mathcal{W}(k) = \emptyset.
$$

In other words: $\{\mathcal{W}^+(k), \mathcal{W}^-(k)\}$ is a partition of $\mathcal{W}^*(k)$; $\mathcal{W}^-(k)$ is the set of winning coalitions where the player k marginally contributes to the winning by the number of his votes; and $\mathcal{W}^+(k)$ is the set of winning coalitions where the player k marginally contributes to the winning just by the number of the members. Note that $\mathcal{W}^-(k)$ is a subset of $\mathcal{W}(k)$, and $S \in \mathcal{W}^+(k)$ has no common element with $\mathcal{W}(k)$.

In what follows we are concerned with the following two power indices.

Shapley-Shubik power index Let g_φ be defined by

$$
g_\varphi = (\sharp S - 1)!(n - \sharp S)!/n!
$$

for each $S \in \mathcal{P}(N)$. For each $k \in N$, the Shapley-Shubik power index φ_k (*resp.,* φ_k^*) of player k in game G (*resp.,* G^*) is

$$
\varphi_k = \textstyle\sum_{S \in \mathcal{W}(k)} g_\varphi(S) \quad (\ \textit{resp.,}\ \varphi_k^* = \textstyle\sum_{S \in \mathcal{W}(k)} g_\varphi(S)\ ).
$$

Banzhaf power index Let g_β be defined by

$$
g_\beta(S) = 1
$$

for each $S \in \mathcal{P}(N)$. For each $k \in N$, the Banzhaf power index β_k (*resp.,* β_k^*) of player k in game G (*resp.,* G^*) is

$$
\beta_k = (\textstyle\sum_{S \in \mathcal{W}(k)} g_\beta(S))/2^{n-1} \quad (\ \textit{resp.,}\ \beta_k^* = (\textstyle\sum_{S \in \mathcal{W}^*(k)} g_\beta(S))/2^{n-1}\ ).
$$

3 Power Shift Theorem

In this section, we compare players' power in G with that in G^*. To simplify the notation, we assume that $0 \le w_1 \le w_2 \le \cdots \le w_n$.

Lemma 1 If $i > j$, then there are four one-to-one mappings f, f^*, f^-, f^+ with the following properties:

(a) $f : \mathcal{W}(j) \to \mathcal{W}(i)$ and $\sharp f(S) = \sharp S$ for each $S \in \mathcal{W}(j)$;

(b) $f^* : \mathcal{W}^*(j) \to \mathcal{W}^*(i)$ and $\sharp f^*(S) = \sharp S$ for each $S \in \mathcal{W}^*(j)$;

(c) $f^- : \mathcal{W}(j)\backslash\mathcal{W}^-(j) \to \mathcal{W}(i)\backslash\mathcal{W}^-(i)$ and
$\sharp f^-(S) = \sharp S$ for each $S \in \mathcal{W}(j)\backslash\mathcal{W}^-(j)$;

(d) $f^+ : \mathcal{W}^+(i) \to \mathcal{W}^+(j)$ and $\sharp f^+(S) = \sharp S$ for each $S \in \mathcal{W}^+(i)$.

Proof For $i, j \in N$, define $f_{ij} : \mathcal{P}(N) \to \mathcal{P}(N)$ as follows:

$$f_{ij}(S) = \begin{cases} S & \text{if} \quad i \in S, \\ (S\backslash\{j\}) \cup \{i\} & \text{if} \quad i \notin S. \end{cases}$$

Note that f_{ij} has the following two properties: for each $S \subset N$, if $j \in S$, then $\sharp f_{ij}(S) = \sharp S$; for each $S, S' \subset N$, if $j \in S \cap S'$ and $i \neq j$, then $S \neq S'$ implies $f_{ij}(S) \neq f_{ij}(S')$.

Let f (*resp.*, f^*, f^-) be the restriction of f_{ij} to $\mathcal{W}(j)$ (*resp.*, $\mathcal{W}^*(j)$, $\mathcal{W}(j)\backslash\mathcal{W}^-(j)$) and f^+ be the restriction of f_{ji} to $\mathcal{W}^+(i)$. We show that f is a one-to-one mapping satisfying (a). The satisfaction of (b), (c), and (d) can be verified analogously. Also all claims for f^*, f^-, and f^+ can be proved in a similar way.

(a) Since $j \in S$ for each $S \in \mathcal{W}(j)$, we have $\sharp f(S) = \sharp S$ for each $S \in \mathcal{W}(j)$. Let S and S' be different members of $\mathcal{W}(j)$. Since $j \in S \cap S'$ and $i \neq j$, we also have $f(S) \neq f(S')$. Thus f is one-to-one.

It remains to prove that $f(S) \in \mathcal{W}(i)$ for each $S \in \mathcal{W}(j)$. Let $S \in \mathcal{W}(j)$. Then we have

$$w(S) > \bar{w} \text{ and } w(S\backslash\{j\}) \leq \bar{w}.$$

(i) If $i \in S$, then $w(S\backslash\{i\}) \leq w(S\backslash\{j\})$ because $i > j$ implies $w_i \geq w_j$. Since $f(S) = S$ in this case, we obtain $i \in f(S)$ and

$$w(f(S)) = w(S) > \bar{w} \text{ and } w(f(S)\backslash\{i\}) = w(S\backslash\{i\}) \leq w(S\backslash\{j\}) \leq \bar{w}.$$

Therefore $f(S) \in \mathcal{W}(i)$, provided $i \in S$.

(ii) If $i \notin S$, then $f(S) = (S\backslash\{j\}) \cup \{i\}$, and we obtain $i \in f(S)$. Moreover, since $i > j$ implies $w_i \geq w_j$, we have $w((S\backslash\{j\}) \cup \{i\}) \geq w(S)$. It follows $w(f(S)) \geq w(S) > \bar{w}$. Note that $f(S)\backslash\{i\} = S\backslash\{j\}$ which implies $w(f(S)\backslash\{i\}) = w(S\backslash\{j\}) \leq \bar{w}$. □

Lemma 2 Let g be a nonnegative real-valued function g on $\mathcal{P}(N)$ such that $g(S) = g(S')$ whenever $\sharp S = \sharp S'$. Then, if $i > j$,

$$\textstyle\sum_{S\in\mathcal{W}(j)} g(S) \leq \sum_{S\in\mathcal{W}(i)} g(S) \text{ and } \sum_{S\in\mathcal{W}^*(j)} g(S) \leq \sum_{S\in\mathcal{W}^*(i)} g(S).$$

Proof According to Lemma 1(a), there exists a one-to-one mapping f of $\mathcal{W}(j)$ into $\mathcal{W}(i)$ such that $\sharp f(S) = \sharp S$ for each $S \in \mathcal{W}(j)$. Since $g(S) = g(S')$ whenever $S = S'$, we have $\sum_{S\in\mathcal{W}(j)} g(S) = \sum_{S\in\mathcal{W}(j)} g(f(S))$. However, $\{f(S)|S \in \mathcal{W}(j)\}$ is a subset of $\mathcal{W}(i)$ and f is a one-to-one mapping. Therefore, $\sum_{S\in\mathcal{W}(j)} g(f(S)) \leq \sum_{S\in\mathcal{W}(i)} g(S)$. Consequently, we have $\sum_{S\in\mathcal{W}(j)} g(S) \leq \sum_{S\in\mathcal{W}(i)} g(S)$.

Similarly from Lemma 1(b), we obtain $\sum_{S\in\mathcal{W}^*(j)} g(S) \leq \sum_{S\in\mathcal{W}^*(i)} g(S)$. □

Lemma 3 If $i > j$, then

$$\sum_{S\in\mathcal{W}^*(i)} g(S) - \sum_{S\in\mathcal{W}^*(j)} g(S) \le \sum_{S\in\mathcal{W}(i)} g(S) - \sum_{S\in\mathcal{W}(j)} g(S)$$

for each nonnegative real-valued function g on $\mathcal{P}(N)$ such that $g(S) = g(S')$ whenever $\sharp S = \sharp S'$.

Proof For each $k \in N$, we observe that $\mathcal{W}^+(k) \cap \mathcal{W}^-(k) = \emptyset$ and $\mathcal{W}^*(k) = \mathcal{W}^+(k) \cup \mathcal{W}^-(k)$. Therefore

$$\begin{aligned}&\sum_{S\in\mathcal{W}^*(i)} g(S) - \sum_{S\in\mathcal{W}(i)} g(S)\\ &= \sum_{S\in\mathcal{W}^+(i)} g(S) + \sum_{S\in\mathcal{W}^-(i)} g(S) - \sum_{S\in\mathcal{W}(i)} g(S).\end{aligned}$$

Since $\mathcal{W}^-(i) \subset \mathcal{W}(i)$, we obtain

$$\sum_{S\in\mathcal{W}^-(i)} g(S) - \sum_{S\in\mathcal{W}(i)} g(S) = -\sum_{S\in\mathcal{W}(i)\setminus\mathcal{W}^-(i)} g(S).$$

Consequently

$$\sum_{S\in\mathcal{W}^*(i)} g(S) - \sum_{S\in\mathcal{W}(i)} g(S) = \sum_{S\in\mathcal{W}^+(i)} g(S) + \sum_{S\in\mathcal{W}(i)\setminus\mathcal{W}^-(i)} g(S).$$

It follows from Lemma 1(c), (d) that

$$\begin{aligned}&\sum_{S\in\mathcal{W}^+(i)} g(S) \le \sum_{S\in\mathcal{W}^+(j)} g(S) \text{ and}\\ &\sum_{S\in\mathcal{W}(i)\setminus\mathcal{W}^-(i)} g(S) \ge \sum_{S\in\mathcal{W}(j)\setminus\mathcal{W}^-(j)} g(S).\end{aligned}$$

Therefore we have

$$\begin{aligned}\sum_{S\in\mathcal{W}^*(i)} g(S) - \sum_{S\in\mathcal{W}(i)} g(S) &\le \sum_{S\in\mathcal{W}^+(j)} g(S) - \sum_{S\in\mathcal{W}(j)\setminus\mathcal{W}^-(j)} g(S)\\ &= \sum_{S\in\mathcal{W}^*(j)} g(S) - \sum_{S\in\mathcal{W}(j)} g(S).\end{aligned}$$

which is equivalent to the required inequality. □

Theorem If $i > j$, then we have :

$$\varphi_i^* \ge \varphi_j^*, \quad \varphi_i^* - \varphi_j^* \le \varphi_i - \varphi_j, \quad \text{and} \quad \beta_i^* \ge \beta_j^*, \quad \beta_i^* - \beta_j^* \le \beta_i - \beta_j.$$

Proof Note that $g_\varphi(S)$ depends only on the cardinal number of S and $g_\beta(S)$ is a positive constant. Thus we obtain the desired results as a straightforward consequence of Lemma 2 and 3. □

According to this theorem, the indices φ^* and β^* are monotone. Note that from Lemma 2, we can also obtain $\varphi_i \ge \varphi_j$ and $\beta_i \ge \beta_j$ whenever $i > j$. The latter two monotone properties are well known; see, for example, Lucas [3]. We have shown that they are valid also for the compound game under consideration.

The theorem also shows a power shift from "bigger" voters to "smaller" voters, *i.e.,* bigger (resp., smaller) voters are given less (more) power in the compound game than in the simple weighted-sum game. Hence, in a voting by count and account, a smaller voter is in more advantageous position than in the corresponding simple weighted majority voting.

Finally we compare our result with Peleg's work. Peleg [4] shows a similar power shift theorem for the Shapley-Shubik power index. In particular, he shows that $\sum_{k=1}^{i} \varphi_k^* \geq \sum_{k=1}^{i} \varphi_k$ for $i = 1, \cdots, n$. In contrast, we show that every difference of two players' value is less in the compound game than in the corresponding game of simple majority voting. It turns out that our result implies Peleg's result but the converse implication is not true. In this sense, our result shows that, voting by count and account is even more egalitarian than according to Peleg [4]'s result when measured by the Shapley-Shubik power index. Moreover, we obtained analogous results for the Banzhaf power index. It is of interest to establish which classes of power indices have similar properties.

References

1. Brams, S.J., Lucas, W.F., Straffin P.D. (eds.): Political and Related Models. Springer-Verlag (1983).

2. Hirokawa, M., Xu, P.: Small Creditors Power in Civil Rehabilitation - A Compound Game of a Simple Majority Game and a Weighted Majority. Mimeo, Hosei University (2005).

3. Lucas, W.J.: Measuring Power in Weighted Voting. In Brams, S.J., Lucas, W.F., Straffin P.D. (eds.): Political and Related Models, 183-238. Springer-Verlag (1983).

4. Peleg, B.: Voting by Count and Account. In Selten R. (ed.) Rational Interaction: Essays in Honor of John C. Harsanyi, 41-51. Springer-Verlag (1992).

Quantitative analysis of economy model using method of moments*

Miroslav Hloušek **
Faculty of Economics and Administration of Masaryk University in Brno,
Department of Applied Mathematic and Computer Science, Lipová 41a,
602 00 Brno, email: hlousek@econ.muni.cz

Abstract
This paper deals with calibration of economy model using method of moments. Real data of United States are used. The time series are decomposed into the trend and the cycle component using Hodrick-Prescott filter. Estimation of the historical standard deviations and autocorrelations is made. The model equations are converted into reduced form of VAR model. The properties of the model in terms of moments are computed. The parameters are properly set to replicate the moments in data. The results is demonstrated on behavior of the model using impulse responses.

Keywords
method of moments, closed economy model, calibration, VAR model, Hodrick-Prescott filter, Schur decomposition

1 Introduction

The ideal way to match theory and facts is a controversial issue. The main idea of this paper is to calibrate a model of closed economy with rational expectations to fit some broad stylized facts about fluctuations in the U.S. economy. This analysis is inspired by Söderstrom (2005). The method of calibration is based on minimization of the deviation between unconditional moments of reduced form of VAR model and the moments in data.

2 Stylized facts

This section presents some stylized fact for the U.S. economy in terms of the standard deviations and autocorrelations. These facts serve as a benchmark with which we compare the implications of the theoretical model. Following time series were chosen as the real counterparts of variables in the model: real gross domestic product, inflation rate based on consumer prices index and nominal interest rate on 3-month Treasury bills. We use

* This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172.
** I thank Jaromír Beneš, Osvald Vašíček and Jan Vlček for useful comments and suggestions.

quarterly, seasonally adjusted data for the period from 1970Q1 to 2004Q4. The data series were obtained from FRED database of Federal Reserve Bank of St. Louis [1] and are shown in Figure 2 in Appendix.

The time series are decomposed into the trend and the cycle component using Hodrick-Prescott filter. The optional parameter λ in Hodrick-Prescott filter that expresses the ratio of the variance of the cyclical shock and the variance of the trend shock is set to $\lambda = 1600$ in accordance with the recommendation of the authors. The cyclical component estimated on data corresponds to the deviation of a variable from steady state level in the model. The properties of the data – the standard deviations and autocorrelations up to three lags are shown in Table 1.

Table 1: Standard deviations and autocorrelations in U.S. data

	Standard deviation	Autocorrelations		
		1 lag	2 lags	3 lags
Output	1.58	0.86	0.69	0.48
Inflation	1.45	0.87	0.70	0.50
Interest rate	1.35	0.79	0.53	0.40

3 Model

Suppose a model of closed economy with rational expectations. The model is characterized by four equations that are expressed in log-linearized form:

$$y_t = \alpha y_{t-1} - \beta r_t + \eta_t \tag{1}$$

$$\pi_t = \gamma \pi_{t-1} + (1-\gamma) E_t \pi_{t+1} + \delta y_t + \chi_t \tag{2}$$

$$r_t = i_t - E_t \pi_{t+1} \tag{3}$$

$$i_t = \lambda\, i_{t-1} + \kappa\, y_t + \phi E_t \pi_{t+1} + \xi_t, \tag{4}$$

It is a gap model, all variables are expressed as deviation (gap) from steady state level. For simplicity, further references to the variables are without the attribute gap. The variable y_t is the output, r_t is the real interest rate, i_t is the nominal interest rate, π_t is the rate of inflation. $E_t\, a_{t+1}$ denotes conditional expectations.[2] The parameters have these properites: $\alpha \in (0,1)$, $\beta > 0$, $\gamma \in (0,1)$, $\delta > 0$ $\kappa > 0$ and $\phi > 1$. Finally, η_t, χ_t and ξ_t represent exogenous shocks.

The equation (1) relates aggregate spending to lagged values of output and the real interest rate; it corresponds to the aggregate demand equation. The equation (2) is an inflation-adjustment equation in which current inflation depends on lagged and expected future inflation and output. This equation (often referred to as Phillips curve or aggregate supply) is based on multiperiod overlapping nominal contracts. The equation (3) is an

[1] Available on "http://research.stlouisfed.org/"

[2] Expectation of a_{t+1} formed upon information available in period t.

identity for the real interest rate. The equation (4) is a monetary rule; the nominal interest rate (instrument of monetary policy) depends on lagged interest rate, current output and expected inflation. There occur three types of shocks: aggregate demand shock, aggregate supply shock and monetary policy shock. For the sake of simplicity, the equation (3) can be inserted into equation (1) and we will work only with three equations and three variables.

4 Schur decomposition and reduced VAR model

The model's equations are put in the form:

$$\mathcal{A}E_t x_{t+1} = \mathcal{B}x_t + \mathcal{C}\epsilon_t, \tag{5}$$

where $\mathcal{A}, \mathcal{B}$ are square matrices of parameters, $\mathcal{C}$ is matrix of coefficients of exogenous shocks ϵ_t. $E_t x_{t+1} = E(x_{t+1}|\Upsilon_t)$, where $E(.)$ is mathematical expectation operator, Υ_t is the information set at t. Vector x_t contains two types of variables: predetermined and non-predetermined. The value of predetermined variable is known at time t, the non-predetermined is not.

If we suppose the shock occurs only in the first period we are able to compute the trajectory of predetermined variables in an easier way. After Schur decomposition the system of equations can be converted into reduced form of VAR model

$$\hat{x}_t = A\hat{x}_{t-1} + B\epsilon_t$$

where $\hat{x}_t$ is a vector of predetermined variables in time t, ϵ_t is a vector of shocks for that holds $\epsilon_t'\epsilon_t = \Sigma$, A and B are matrices of parameters of reduced form. The matrices are derived from Schur decomposition, for more details see Hloušek (2003), and their composition is as follows

$$A = Z_{11}{S_{11}}^{-1}T_{11}{Z_{11}}^{-1}$$

$$B = Z_{11}{S_{11}}^{-1}T_{11}(-Z_{11})^{-1}Z_{12}(-T_{11})^{-1}D_2 + Z_{11}{S_{11}}^{-1}T_{12}(-T_{22})^{-1}D_2 + Z_{11}D_1$$

Based on the reduced form the unconditional moments – standard deviations and autocorrelations are computed.[3] The standard deviations are diagonal elements of covariance matrix $\Omega(0)$. The autocorrelations are diagonal elements of matrix $R(k)$ according to the

[3]The unconditional covariance does not depend on time, therefore

$$\underbrace{\operatorname{cov}\hat{x}}_{\Omega} = A\underbrace{\operatorname{cov}\hat{x}}_{\Omega}A' + B\underbrace{\operatorname{cov}\epsilon}_{\Sigma}B'$$

and using formula for vectorization $\operatorname{vec}(XYZ) = (Z' \otimes X)\operatorname{vec} Y$, the result is

$$\operatorname{vec}\Omega = (A \otimes A)\operatorname{vec}\Omega + \operatorname{vec}(B\Sigma B')$$

$$\operatorname{vec}\Omega = [I - (A \otimes A)]^{-1} + \operatorname{vec}(B\Sigma B')$$

following formula

$$\Omega(k) = E\hat{x}_t'\hat{x}_{t-k} \quad \text{equivalently} \quad \Omega(k) = A^k\Omega(0)$$

$$R(k) = S\,\Omega(k)S \qquad\qquad s_{ij} = \begin{cases} \frac{1}{\sqrt{\omega(k)_{ij}}} & i = j \\ 0 & \text{otherwise} \end{cases}$$

5 Calibration

The parameters and the standard deviations of shocks (σ_η, σ_χ, σ_ξ) in the model are properly set to replicate the moments in the data. The values of parameters are naturally restricted by their domains of definiton. The criterion is to minimize sum of squared deviations between the moments of the model and the moments in the data. The moments implied by the model with the "best" setting of parameters are shown in Table 2.

Table 2: Standard deviations and autocorrelations of calibrated model

	Standard deviation	Autocorrelations		
		1 lag	2 lags	3 lags
Output	1.02	0.84	0.67	0.52
Inflation	1.26	0.83	0.67	0.52
Interest rate	1.88	0.78	0.57	0.39

The interest rate is much more volatile and output and inflation are more stable in the model than in the data. In terms of autocorrelations the parameterization is very close to the actual data. All variables are rather less persistent compared with the data. Exceptions are the third lag of output and inflation and the second lag of interest rate that show more persistence in the model. Despite these differences the result seems to be satisfactory.

The "best" setting of the parameters is shown in Table 3. The behavior of agents concerning of output is characterized by large inertia (parameter α in aggregate demand equation) which can be explained by habits in consumption. The forward-looking term in the Phillips curve $(1-\gamma)$ is quite large. It indicates that economic agents take future inflation into account when they are making predictions. An interesting thing is that

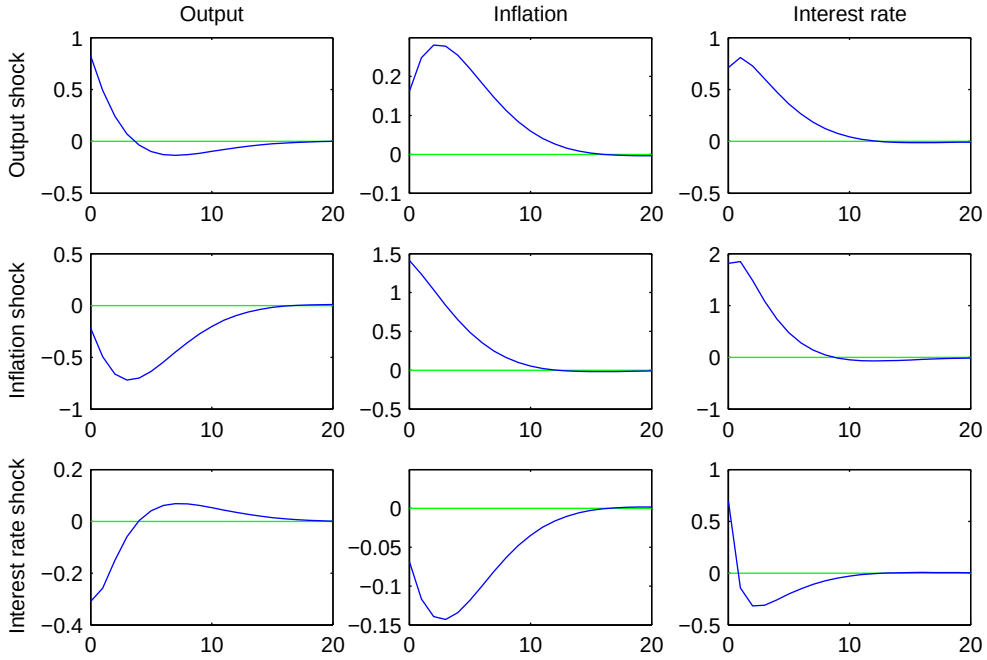
Table 3: Parametrs and the standard deviations of shocks

Agregate demand		Phillips curve		Monetary rule	
α	0.83	γ	0.65	λ	0.30
β	0.39	δ	0.09	κ	0.22
				ϕ	1.55
σ_η	0.52	σ_χ	0.49	σ_ξ	0.62

central bank has more preference for interest rate smoothing than output stabilization ($\lambda > \kappa$). Such behavior suggests that central bank concerns about stability on financial markets and the payment system. The result also indicates that central bank does not pursue "strict" inflation targeting (both λ and κ are non-zero).

To illustrate the behavior of the calibrated model, Figure 1 shows how the economy responds to unit shocks at time $t = 0$. Althought we have chosen not to match the cross-correlations in the data, these impulse responses look quite reasonable. After output shock η_t (shock in aggregate demand) the central bank increases interest rate to change the positive output gap into a negative gap in order to fight the inflationary impulse. This is done fairly quickly and inflation returns to the equlibrium after approximately fifteen periods. Shock in the Phillips curve χ_t causes jump of inflation that is higher than 1 % because of forward-looking term in the Phillips curve. Again, the monetary policy response opens up a negative output gap to fight with inflation. The reaction of interest rate is sizable in this case and ensures to reduce inflation after twelve quarters. Unexpected one-shot increase of the interest rate (shock in monetary policy rule ξ_t) causes recession in the economy. The output responds immediately, the maximum effect on inflation comes after four quarters. However, the interest rate is immediately moved into negative values to offset the gaps in both variables.

Figure 1: Impulse response of calibrated model



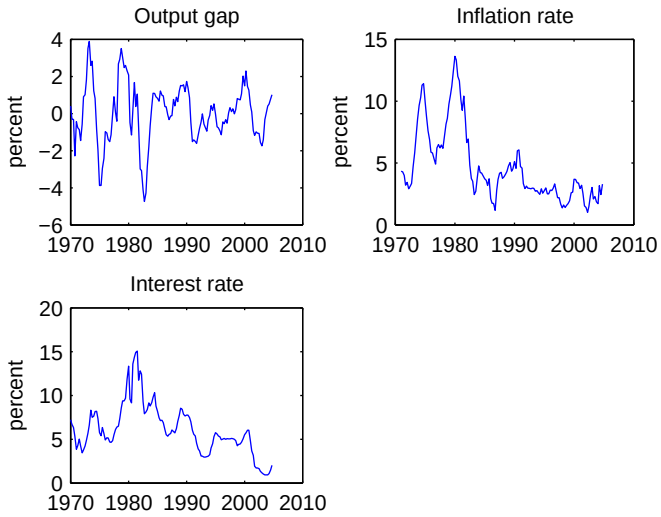
6 Conclusion

This analysis indicates that our calibrated model provides a reasonable description of the U.S. economy and the method of moments is a useful tool for matching model to data. However, the calibration depends on sample period, model specification and matching criteria. The next research will be therefore aimed to issue how these choices affect the results, i.e. the robustness of the model and the method.

7 Appendix

The data series are shown in Figure 2. For transparency, the output is already detrended and the gap is shown, the inflation rate and the interest rate are depicted in original values.

Figure 2: Data series, 1970Q1–2004Q4



References

[1] HAMILTON, J. D.: *Time Series Analysis.* Princeton: Princeton University Press, 1994.

[2] HLOUŠEK, M.: Model of closed economy with rational expectations. *Mathematical Methods in Economics 2003.* Prague: Czech University of Agriculture in Prague, 2003, Faculty of Economics and Management, pp. 112-116.

[3] SÖDERSTROM, U.,SÖDERLIND, P., VREDIN, A.: New-Keynesian Models and Monetary Policy: A Reexamination of the Stylized Facts, January 2005, Forthcoming in *Scandinavian Journal of Economics.*

Estimation in chance-constrained problem

Michal Houda

Institute of Information Theory and Automation,
Academy of Science of the Czech Republic,
Pod vodárenskou věží 4, CZ-182 08 Praha 8
houda@karlin.mff.cuni.cz

Abstract. Many engineering and economic applications make use of the stochastic programming theory. Major part of models require a complete knowledge of distribution of random parameters, but this assumption is rarely accomplished. We then need to study behaviour of optimal solutions when the distribution changes slightly. In our contribution we consider the chance constrained problem; we recapitulate some known theoretical results about stability and estimation of the problem. We concentrate especially on results from stochastic and robust programming; we try to outline a link between empirical and robust estimates of chance-constrained problem.

1 Introduction

Optimization problems take their important part in both economic and technical disciplines. Input parameters of these problems, modelled as deterministic in the first approach, are random by nature. The randomness of some of them could be freely ignored – to be exact, replaced by some deterministic versions, e.g. average values – without registering any notable modification of resulting solutions. On the other side, there are examples where this is not possible at all – for example, replacing random parameters with their expectations in the classical Kall's linear programming example ([11]) leads to a solution which is not feasible with 75% probability. This is an important task of the post-optimization sensitivity analysis: it has to find a way how to play the game with the randomness.

If the randomness cannot be ignored, we have to incorporate it in the model in a convenient way. Stochastic programming theory is searching for an optimal solution of the optimization model where random parameters are taken into account. If a realization of the random parameter is not known in advance (i. e., before some decision is taken), we cannot work with the random variable itself but its probability distribution. Moreover, a *complete knowledge* of the distribution is required.

[1] This research is supported by the project Nonlinear Dynamical Economic Systems: Theory and Applications registered under No. 402/03/H057, by the project Stochastic Decision Approaches in Nonlinear Economic Models registered under No. 402/04/1294, and by the project Validation of Economic Decision Models and Results registered under No. 402/05/0115 by the Grant Agency of the Czech Republic.

2 Chance-constrained problem

Consider a general constrained problem in the form

$$\min_{x\in X} c'x \quad \text{subject to} \quad f(x;\xi) = \max_{i=1,\dots,m} f_i(x;\xi) \le 0 \tag{1}$$

where $\xi \in \Xi \subset \mathbb{R}^s$ is a random (uncertainty, instance) parameter, $X \subset \mathbb{R}^n$ is a closed convex set, and vector functions $f_i : X \times \Xi \to \mathbb{R}$ are convex for each fixed $\xi \in \Xi$. Without loss of generality, we assume that the objective is linear and independent of ξ.

If a realization ξ of the random parameter is known and fixed, a solution of (1) could be easily obtained by means of deterministic optimization. Two main philosophies are developed in the literature in order to deal with uncertainty of random parameter. First, the *chance-constrained approach* considers ξ as a random variable and checks the constraints in (1) to be fulfilled with a given probability $1 - \beta \in [0; 1]$ – the problem then reads

$$\min_{x\in X} c'x \quad \text{subject to} \quad x \in X_\beta := \big\{x \in X;\ \mu\{f(x;\xi) > 0\} \le \beta\big\} \tag{2}$$

where $\mu \in \mathcal{P}(\Xi)$ is the probability distribution of the random vector ξ. The chance-constrained optimization has a long history, starting at least by the work of Charnes and Cooper [7]; an extensive presentation of many results is given in Prékopa's book [12]. Problem (2) is not necessarily a convex optimization problem even if function f is convex in x for all ξ (see [12] again). Another difficulties are arising when we evaluate the probability measure in X_β – it involves a multidimensional integral.

The probability distribution is rarely known completely. Instead, various approximation and estimating techniques are used in practice. Here, let us present an approach using empirical distribution function. Consider a set of iid (independent identically distributed) samples $\xi_1, \dots, \xi_N$ from the original distribution μ. The *empirical distribution function* is a random variable defined by $\mu_N := \frac{1}{N}\sum \delta_{\xi_i}$ where δ_ξ denotes the unit mass at ξ. The problem (2) is now approximated by replacing probability distribution μ by μ_N, and the problem reads

$$\min_{x\in X} c'x \quad \text{subject to} \quad x \in X_N := \big\{x \in X; \frac{1}{N}\,\mathrm{card}\{i; f(x;\xi_i) > 0\} \le \beta\big\}. \tag{3}$$

Stability theory of stochastic programming answers a question how far is the optimal solution of approximated problem (3) with respect to the optimal solution of the original chance-constrained problem (2). A measurability issues arise when studying convergences properties of the problem. For this reason we establish some notions to overcome these issues. For details, see [13].

A class $\mathcal{F}$ of measurable functions from Ξ to $\bar{\mathbb{R}}$ is called *μ-permissible* if there is a countable subset $\mathcal{F}_0$ of $\mathcal{F}$ such that for each $F \in \mathcal{F}$ there exists a sequence F_j in $\mathcal{F}_0$ converging pointwise to F and such that the integral $\mu F_j := \int_\Xi F_j(\xi)\mathrm{d}\mu(\xi)$

also converge to the integral μF. μ-permissibility of the class is somehow a technical assumption to ensure the measurability of studied objects. Next, define a distance between the probabilities measures $\mu_N(\cdot)$ and μ (Zolotarev pseudometric) by

$$d_{\mathcal{F}}(\mu_N(\cdot), \mu) = d_{\mathcal{F}_0}(\mu_N(\cdot), \mu) = \sup_{F \in \mathcal{F}} |(\mu_N(\cdot) - \mu)F|. \tag{4}$$

$\mathcal{F}$ is a *μ-Glivenko-Cantelli class* if it is μ-permissible and the sequence of distances $\{d_{\mathcal{F}}(\mu_N(\cdot), \mu)\}$ converges to zero almost surely. This is desirable property of the correctly selected probability metric.

Theorem 50 of Römisch work [13] – as a direct consequence of the general stability theorem formulated with respect to a special form of Zolotarev pseudometric – gives us, under some assumptions (μ-permissible class $\mathcal{F}$, metric regularity for X_β, existence and uniqueness of optimal solution) – an upper bound for the convergence rate of distance between the optimal solutions. For example, if $f(x;\xi)$ is linear, then the bound reads

$$\mathrm{P}\big\{\sqrt{N} d(\sup_{x \in X_N^*} d(x; X^*) > \varepsilon\big\} \leq \hat{K} \varepsilon^{d_f} e^{-2 \min\{\delta, \hat{L}^{-1} \Psi_\mu^{-1}(\varepsilon)\}^2} \tag{5}$$

where X_N^* is the (localised) solution set of estimated problem (3), X^* is the solution set of the chance-constrained problem (2), δ, $\hat{K}$ and $\hat{L}$ are positive constants (arising from the stability theory), $\Psi_\mu^{-1}(\cdot)$ is the inverse associated growth function of the objective (see [13] for a precise definition), and d_f is VC dimension of the constraints $f(\cdot;\cdot)$ (see Section 4). The notable thing is that the approximated solution converge exponentially to the original solution of the chance-constrained problem (2).

3 Robust optimization problem

An alternative approach leading to solve an uncertain program is the so-called "worst-case" analysis which also adopted the name of *robust optimization*. In this approach one looks for a solution which is feasible for all instances of ξ, i. e. solving the problem

$$\min_{x \in X} c'x \quad \text{subject to} \quad f(x;\xi) \leq 0 \text{ for all } \xi \in \Xi. \tag{6}$$

This framework was introduced by Ben-Tal and Nemirovski [1], [2], [3] and developed by other authors in various directions (see e. g. [8], [9], [4] and references therein). The robust convex programming problem is convex, but it involves an infinite number of constraints and it is numerically hard to solve again. Robust optimization theory proposes a method of approximation of the problem (6) by the following so-called *sampled problem*; instead of requiring constraints to be satisfied for all instances of ξ, we are satisfied by only finite (but sufficiently large) number of samples $\xi_1, \ldots, \xi_N$:

$$\min_{x \in X} c'x \quad \text{subject to} \quad f(x;\xi_i) \leq 0 \text{ for } i = 1, \ldots, N. \tag{7}$$

A question arises for this problem how the conditions of the nominal problem (1) are satisfied. This question of the robust post-optimization analysis can be formulated as follows: we are searching for minimal sample size N such that an optimal solution of (7) is feasible solution of chance constrained problem (2) (i. e. violates conditions of the nominal problem only with a small probability). In this context we speak about two sources of error in the sampling method:

- with a probability α, the feasible (or optimal) solution of the sampled problem (7) is not feasible solution of the chance-constrained problem;
- with a probability β, the feasible solution of (7) can still violate nominal constraints $f(x;\xi) \leq 0$.

A result was obtained by Calafiore and Campi [6], [5] for convex functions $f(x;\xi)$. If for given β and α

$$N \geq \frac{2n}{\beta} \ln \frac{1}{\beta} + \frac{2}{\beta} \ln \frac{1}{\alpha} + 2n \tag{8}$$

then the optimal solution of the sampled problem (7) is feasible solution to the chance-constrained problem with the probability $1 - \alpha$. Hence the sample size of $\mathcal{O}(n/\beta)$ is enough to achieve this goal.

The complete knowledge of the distribution μ is required in order to generate the samples ξ_i. Again, we only know some approximation to μ in practice. This question is treated by Erdoğan and Iyengar [10]: consider a set of "close" probability distributions $\mathcal{Q} := \{\mu;\ \rho_P(\mu;\mu_0) \leq \delta\}$, where $\delta > 0$ is a prescribed constant, $\mu_0 \in \mathcal{P}(\Xi)$ a suitable *central* probability measure and ρ_P is the Prohorov metric defined on $\mathcal{P}(\Xi)$ by

$$\rho_P(\mu;\nu) := \inf\{\varepsilon;\ \mu B \geq \nu(B^\varepsilon) + \varepsilon \ \text{ for all measurable } B \subset \Xi\} \tag{9}$$

where $B^\varepsilon := \{x \in X :\ \inf_{z\in B}\|x - z\| \leq \varepsilon\}$. Prohorov metric is a smallest distance "in probability" between two random variables with the distributions μ and ν. See again [10] and references therein for relations to other probability metrics.

Now we are ready to formulate a so-called *ambiguous chance-constrained problem*; the constraint now have to be satisfied with a high probability for all sufficiently close measures:

$$\min_{x\in X} c'x \quad \text{subject to} \quad x \in \bar{X}_\beta := \left\{x \in X :\ \mu\{f(x;\xi) > 0\} \leq \beta \quad \forall \mu \in \mathcal{Q}\right\} \tag{10}$$

We approximate this problem by a *robust sampled problem* defined as follows

$$\min_{x\in X} c'x \quad \text{subject to} \quad f(x;z) \leq 0 \quad \forall z : \|z - \xi_i\| \leq \delta,\ i = 1,\ldots,N \tag{11}$$

and again try to find a sufficient sample size N ensuring that an optimal solution of the robust sampled problem (11) is feasible for ambiguous chance-constrained problem (10). This has been done in [10]: the optimal solution of the robust sampled problem (11) is feasible solution to the ambiguous chance-constrained problem (10) with the probability $1-\gamma$ if the sample size is of order $\mathcal{O}(n/(\beta-\gamma))$.

4 VC-dimension and computational learning theory

Another question of robust optimization is the behaviour of all the feasibility set of the problems (7) and (11), not only of their optimal solutions. Here we need in addition a notion of Vapnik-Chervonenkis (VC) dimension and computational learning theory (see e. g. [14]).

Let $x \in X$ be a decision vector, denote $C_x := \{\xi \in \Xi : \ f(x;\xi) \leq 0\}$ (*concept* or *classification*) a set of instances which are "feasible" for this decision. Learning X_β means to find concepts $C_f := \{C_x;\ x \in X\}$ with $\mu(\xi \notin C_x) \leq \beta$, i. e. (in real world), to approximate $f(x;\cdot)$ with a finite number of samples. Denote this set of samples by $S := \{\xi_1, \ldots, \xi_N\} \subset \Xi$ and denote $\Pi_f(S) := \{(I_{C_x}(\xi_1), \ldots, I_{C_x}(\xi_N)\}$ a set of dichotomies induced by C_f ($I_A(\cdot)$ is an indicator function of the set A). If $\operatorname{card} \Pi_f(S) = 2^N$ we say that S is *shatered* by C_f – this means that $\Pi_f(S)$ contains all possible results and hence no information is given about C_f. *Vapnik-Chervonenkis dimension* d_f of $f(\cdot;\cdot)$ (or C_f) is the maximal cardinality of S that is shatered by C_f.

VC dimension measures the complexity of concepts C_f. We usually assume that $d_f < +\infty$ – this means that some information is finally given with a sufficiently large number of samples. But the VC dimension could be rather large, as seen for example in the third case below:

- in the linear case with a single constraint, $d_f \leq n$;
- in the linear case with multiple constraints, $d_f \leq \mathcal{O}(n^2)$;
- in the case of pointwise maximum from (1), $d_f \leq \mathcal{O}(10^m \max_i d_{f_i})$.

To end the robust programming part of the paper, let apply this construct to our questions of feasibility; one can deduce following sample size orders (see [10]):

- the worst case number of samples required to learn X_β is $\mathcal{O}(d_f/\beta)$.
- the worst case number of samples required to learn $\bar{X}_\beta$ is $\mathcal{O}(d_f/(\beta - \gamma))$.

5 Conclusion and open questions

We have made a short review of two different approaches to estiomation in uncertain optimization problems where some input parameters are considered as random variables. First, ve have mentioned the chance-constrained problems and its approximation through empirical distribution functions. Known stability results relate to the difference between original optimal solution and the optimal solution of the approximated problem. Second approach is based on the robust optimization theory: we replace an infinite number of constraints (required to be satisfied for all instances of random parameters) by a finite one (constraints are satisfied by finite number of samples). Robust optimization then looks for the sample size which is needed to satisfy the constraints with a high probability, instead to measure distance between original and approximated optimal solution.

Some questions arise and are planned to be explored in the future research:

– how the optimal values of (robust) sampled problem relate to the optimal value of the chance-constrained problem or its approximation;
– both result have in common the dependence on the complexity of constraints expressed via VC dimension; it is of interest to estimate an upper bound of VC dimensions for some special types of constraints and problems (e. g. quadratic ones);
– if the independence of samples is crucial and if there is a way to use somehow dependent samples;
– probability metrics can be adopted to the problem via stochastic programming stability theorems; this way we could modify the robust sampled problem for ambiguous chance-constrained problem and make some conclusions about achieved results.

References

1. Ben-Tal, A., and Nemirovski, A.: Robust convex optimization. Mathematics of Operations Research **23** (4) (1998), 769–805.
2. Ben-Tal, A., and Nemirovski, A.: Robust solutions of uncertain linear programs. Operations Research Letters **25** (1) (1999), 1–13.
3. Ben-Tal, A., and Nemirovski, A.: Lectures on Modern Convex Optimization. SIAM, Philadelphia, PA (2001).
4. Bertsimas, D., and Sim, M.: Robust conic optimization. To appear in Mathematical Programming.
5. Calafiore, G., and Campi, M. C.: Decision making in an uncertain environment: the scenario based optimization approach. In: Multiple Participant Decision Making (J. Andrýsek, M. Kárný, and J. Kracík, eds.), Advanced Knowledge International (2004), 99–111.
6. Calafiore, G., and Campi, M. C.: Uncertain convex programs: randomized solutions and confidence levels. Mathematical Programming **102** (1, Ser. A) (2005), 25–46.
7. Charnes, A., and Cooper, W. W.: Chance constrained programming. Management Science **6** (1959), 73–79.
8. El Ghaoui, L., and Lebret, H.: Robust solutions to least-squares problems with uncertain data. SIAM Journal of Matrix Analysis Applications**18** (4) (1997), 1035–1064.
9. El Ghaoui, L., Oustry, F., and Lebret, H.: Robust solutions to uncertain semidefinite programs. SIAM Journal of Optimization**9** (1) (1998), 33–52.
10. Erdoğan, E., and Iyengar, G.: Ambiguous chance constrained problems and robust optimization. CORC Technical Report TR-2004-10, Columbia University (2004). To appear in Mathematical Programming.
11. Kall, P.: Stochastic Linear Programming. Springer-Verlag, Berlin (1976).
12. Prékopa, A.: Stochastic Programming. Kluwer Academic Publishers, Dordrecht, and Akadémiai Kiadó, Budapest (1995).
13. Römisch, W.: Stability of stochastic programming problems. In: Stochastic Programming (A. Ruszczyński, and A. Shapiro eds.), Handbooks in Operations Research and Management Science Vol. 10, Elsevier, Amsterdam (2003), 483–554.
14. Vapnik, V. N.: The Nature of Statistical Learning Theory. Springer Verlag, New York (1995).

Sensitivity Analysis in Piecewise Linear Programming Models

Milan Houška, Martina Beránková
Czech University of Agriculture Prague

Abstract

A piecewise linear programming (PLP) model is usually solved by transformation to a linear programming (LP) model. In the case of convex optimization model the standard simplex method can be used for obtaining results.
Stability of solution and sensitivity on changes of initial parameters cannot be tested by the standard method. It does not respect some implicit relationships between variables, which arise from the transformation of PLP model to LP model.
The article suggests alternative sensitivity analysis, which respects these relationships. It describes a method for sub-optimal solutions finging and RHS constants components sensitivity analysis. Principles of the suggested method can be also used for cost coefficients sensitivity analysis.

Keywords

Convex Optimization Models, Piecewise Linear Programming, Linear Programming, Simplex Method, Sensitivity Analysis

1 Introduction

This article is focused on convex optimization models and especially on sensitivity analysis of their results. There is suggested alternative sensitivity analysis for RHS constants, which respects relationships among variables derived from the transformation of PLP into LP model. A problem of PLP model sensitivity analysis is that there is no relevant information about it in any kind of literature. The purpose of this article is to show one of the possible way how to solve this lack of information sources. Suggested approach keeps specifics of PLP, but is more simplied than general known methods.

2 Definitions

A general piecewise linear programming model is defined as follows

$$\max\left\{z(x)\middle|\ \sum_{j=1}^{n} a_{ij}x_j \le b_i,\ i=1,\ 2,\ ...,\ m;\ \ x_j \ge 0,\ j=1,\ 2,...,n\right\}, \tag{1}$$

where
$z(x) = z_1(x_1) + z_2(x_2) + \ ... \ + z_j(x_j) + \ ... \ + z_n(x_n)$,
and

$$\begin{aligned}
&z(x_j) = s_{j1}x_j + t_{j1} && x_1 \in \langle 0; k_{j1} \rangle,\\
&z(x_j) = s_{j2}x_j + t_{j2} && x_1 \in \langle k_{j1}; k_{j2} \rangle,\\
&\vdots && \\
&z(x_j) = s_{jp_j}x_j + t_{jp_j} && x_1 \in \langle k_{j(p_j-1)}; k_{1p_1} \rangle,\\
&z(x_j) = s_{j(p_j+1)}x_j + t_{j(p_j+1)} && x_1 \in \langle k_{jp_j}; \infty), \ j\ =\ 1,\ 2,\ ...\ ,\ n
\end{aligned} \tag{2}$$

The model (1) is a convex optimization model, if following conditions are fulfilled:

- the feasible solution set is convex and coherent,
- for each partial objective function is considered to be

$$s_{j1} > s_{j2} > \ldots > s_{j(p_j+1)},\ j = 1,\ 2,\ \ldots,\ n \tag{3}$$

Following [1], the model (1) can be transformed as follows

$$\max\left\{ \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n}\sum_{k=1}^{p_j} c_{jk}^{-} r_{jk}^{-} + \sum_{j=1}^{n}\sum_{k=1}^{p_j} c_{jk}^{+} r_{ik}^{+} \;\middle|\; \begin{array}{l} \sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = 1,\ 2,\ \ldots,\ m,\ j = 1,\ 2,\ldots,n, \\ x_j + r_{jk}^{-} - r_{jk}^{+} = k_{jk},\ k = 1,2,\ldots p_j \\ x_j \ge 0,\ r_{jk}^{-} \ge 0,\ r_{jk}^{+} \ge 0, \\ r_{jk}^{-} \cdot r_{jk}^{+} = 0 \end{array} \right\}, \tag{4}$$

where

$\sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = 1,\ 2,\ \ldots,\ m$ are the original constraints,

$x_j + r_{jk}^{-} - r_{jk}^{+} = k_{jk},\ k = 1,2,\ldots p_j$ are added boundary point constraints with r_{jk}^{-} (r_{jk}^{+}) left (right) distance variables,

c_j and $c_{jk}^{+/-}$ are cost coefficients and

$x_j \ge 0,\ r_{jk}^{-} \ge 0,\ r_{jk}^{+} \ge 0$ are non-negativity constraints of all variables.

The non-linear constraitns

$r_{jk}^{-} \cdot r_{jk}^{+} = 0$

can be ignored in convex optimization models. It is not difficult to show that violating these constraints makes worse value of the objective function. Thus, the model (4) can be solved by standard simplex method.

3 Sensitivity analysis of optimal solution for PLP model

Optimal solution of convex model PLP can be found by standard simplex method. LP transformed model (4) cannot implicitly reflect rules for the distance variables entering into basis. These rules don't have any impact to optimal solution finding, but it modifies interpreatation of optimal solution as well as the sensitivity analysis of the transformed LP model. That's why the standard methods for sensitivity analysis can't be used.

First it is necessary to define terms "basis replacement" and "change of the solution structure". The standard LP model uses these terms like equivalent. When the basis is replaced, it leads to the change of the solution structure. This thesis is not exactly valid for PLP models.

Let's define X as the set of row indexes of decision basic variables and D as the set of row indexes of slack basic variables in the final simplex table. Then we define the term "**basis replacement"** as one step of simplex algorithm and the term "**change of the solution structure"** as the basis replacement with one change in X or D set.

Following part of the article will be focused on two aspect of optimal solution analysis:

- range of feasibility of non-basic variables
- one of the RHS constant change

3.1 Feasibility range of non-basic variables

In standard LP model, the feasibility range for non-basic variable is being used

$$0 \le x_k \le \min\left(\frac{\beta_i}{\alpha_{ik}}\right) \quad \forall i : \alpha_{ik} > 0 \tag{5}$$

Nevertheless, the relationship (5) does not respect original PLP specifics. Let x_p is a non-basic variable. Its maximum value determined by original constraints is equal to x_p^{max}. When the boundary point k_{p1} is lower than x_p^{max}, by the relationship (5) we obtain the feasible values range for x_p

$x_p \in \langle 0; k_{p1} \rangle$

If $x_p > k_{p1}$, then the basic variable r_{11}^- value is negative. So the basis replacement would be required, but the solution structure stays the same. Relationship (5) determines too narrow range of feasible values for x_p variable.

Thus, it is necessary to modify the relationship (5). If the relationship

$$0 \le x_k \le \min\left(\frac{\beta_i}{\alpha_{ik}}\right) \quad \forall i \in \{X \cup D\} : \alpha_{ik} > 0 \tag{6}$$

is used, the changes in the set of basic distance variables will be ignored, because they do not cause the changes of the solution structure.

Using (6) the values of basic decision variables and slack variables will be non-negative. But when

$$\exists i: \alpha_{ik} > 0 \wedge \min_{i=1}^{m}\left(\frac{\beta_i}{\alpha_{ik}}\right) < \min_{p \in \{X \cup D\}}\left(\frac{\beta_p}{\alpha_{pk}}\right), \tag{7}$$

value of some distance variable will be negative. Let the basic variable $r_{p1}^- < 0$. Then the following assignment has to be done:

1) The other distance variable connected to the same boundary point (non-basic, r_{p1}^+) is equal to their absolute value, so $r_{p1}^+ = |r_{p1}^-|$, (8)
2) Negative value of distance variable is changed to zero, so $r_{p1}^- = 0$. (9)

Let's generalize the statements above. During increasing the x_p value, it may be necessary to exceed several boundary pouint values. Then several pairs of distance variables values have to be changed by (8) and (9).

Define the set R as a set of row indexes of the analysed simplex table. Then the pairs of distance variables values, for which rows is valid

$$\alpha_{pk} > 0 \wedge \min_{\substack{i \in \{X \cup D\} \\ a_{ik} > 0}}\left(\frac{\beta_i}{\alpha_{ik}}\right) > \frac{\beta_p}{\alpha_{pk}} \quad \forall p \in R \tag{10}$$

are changed by (8) and (9).

3.2 Sensitivity analysis for the RHS constant change

In standard LP model, the relationship for sensitivity analysis for the RHS vector change is as follows

$b_i \in \langle b_i - \underline{\lambda}, b_i + \overline{\lambda} \rangle$,

where

$$\underline{\lambda} = \max_{\alpha_{ik} > 0}\left(\frac{-\beta_i}{\alpha_{ik}}\right) \tag{11}$$

$$\overline{\lambda} = \min_{\alpha_{ik} < 0}\left(\frac{-\beta_i}{\alpha_{ik}}\right).$$

The base of the sensitivity analysis for PLP models is the same. For PLP models it is necessary to extend (11) by following procedure:

1) Determine i: i is a row index of $\max_{\alpha_{ik} > 0}\left(\frac{-\beta_i}{\alpha_{ik}}\right)$
2) If $i \notin \{X \cup D\}$, then go to 3, else the structure of solution is changed and $\underline{\lambda} = \frac{-\beta_i}{\alpha_{ik}}$.
3) Do the test of dual feasibility of solution for i-th row and then determine the entering variable.

4) If the entering variable is decision or slack variable, then the solution structure is changed and $\underline{\lambda} = \frac{-\beta_i}{\alpha_{ik}}$. Otherwise, solution structure is not changed and i-th row will be ignored, then go to 1).

Procedure for $\overline{\lambda}$ is analogical to previous one.

4 Example

Let's solve the model

$x_1 + x_2 \leq 6$

$z = z_1 + z_2 \rightarrow \max$

where

$z_1 = -x_1$	for $x_1 \in \langle 0;2 \rangle$	$z_2 = 4x_2$	for $x_2 \in \langle 0;1 \rangle$
$z_1 = -2x_1 + 2$	for $x_1 \in \langle 2;4 \rangle$	$z_2 = 2x_2 + 2$	for $x_2 \in \langle 1;8 \rangle$
$z_1 = -4x_1 + 10$	for $x_1 \in \langle 4;\infty)$	$z_2 = -x_2 + 14$	for $x_2 \in \langle 8;\infty)$

Transforming the model into (4), we obtain

$$\max \left\{ -x_1 - r_{11}^{+} - 2r_{12}^{+} + 4x_2 - 2r_{21}^{+} - 3r_{22}^{+} \;\middle|\; \begin{array}{l} 2x_1 + x_2 \leq 6 \\ x_1 + r_{11}^{-} - r_{11}^{+} = 2 \\ x_1 + r_{12}^{-} - r_{12}^{+} = 4 \\ x_2 + r_{21}^{-} - r_{21}^{+} = 1 \\ x_2 + r_{22}^{-} - r_{22}^{+} = 8 \\ x_j \geq 0,\ r_{jk}^{-} \geq 0,\ r_{jk}^{+} \geq 0, j = 1,2, k = 1,2 \\ r_{jk}^{-} \cdot r_{jk}^{+} = 0, j = 1,2, k = 1,2 \end{array} \right\} \quad (14)$$

The model is convex optimization model, non-linear constraints

$r_{jk}^{-} \cdot r_{jk}^{+} = 0, j = 1,2, k = 1,2$

can be ignored. Initial and final solutions are in following simplex tables:

		-1 x_1	0 r_{11}^-	-1 r_{11}^+	0 r_{12}^-	-2 r_{12}^+	4 x_2	0 r_{21}^-	-2 r_{21}^+	0 r_{22}^-	-3 r_{22}^+	0 d_1	b
0	d_1	1	0	0	0	0	1	0	0	0	0	1	6
0	r_{11}^-	1	1	-1	0	0	0	0	0	0	0	0	2
0	r_{12}^-	1	0	0	1	-1	0	0	0	0	0	0	4
0	r_{21}^-	0	0	0	0	0	1	1	-1	0	0	0	1
0	r_{22}^-	0	0	0	0	0	1	0	0	1	-1	0	8
$z_j - c_j$		1	0	1	0	2	-4	0	2	0	3	0	0
-2	r_{21}^+	1	0	0	0	0	0	-1	1	0	0	1	5
0	r_{11}^-	1	1	-1	0	0	0	0	0	0	0	0	2
0	r_{12}^-	1	0	0	1	-1	0	0	0	0	0	0	4
4	x_2	1	0	0	0	0	1	0	0	0	0	1	6
0	r_{22}^-	-1	0	0	0	0	0	0	0	1	-1	-1	2
$z_j - c_j$		3	0	1	0	2	0	2	0	0	3	2	14

Let's analyze feasibility range of non-basic variable x_1. Sets of basic decision variables X and slack variables D are
X = {4},
D = {}.
Using (6), the feasibility range of non-basic variable x_1 is determined as follows

$$0 \le x_1 \le \frac{6}{1}$$

The feasible sub-optimal solution with $x_1 = 3$ is calculated as

$$\begin{vmatrix} r_{21}^{+} \\ r_{11}^{-} \\ r_{12}^{-} \\ x_2 \\ r_{22}^{-} \end{vmatrix} \begin{pmatrix} 5 \\ 2 \\ 4 \\ 6 \\ 2 \end{pmatrix} - 3 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 1 \\ 3 \\ 5 \end{pmatrix}$$

$r_{11}^{-} < 0$, so relationships (8) and (9) have to be used. We obtain
$r_{11}^{+} = |r_{11}^{-}| = 1$
and
$r_{11}^{-} = 0$
The variable values in sub-optimal solution are
$x_1 = 3$ $x_2 = 3$
$r_{11}^{+} = 1$ $r_{21}^{+} = 2$
$r_{12}^{-} = 1$ $r_{22}^{-} = 5$
all other variables values are equal to zero.

Let's analyze stability of solution against changes of the RHS component b_1. The lower bound $\underline{\lambda}$ will be set first. Using the suggested procedure, the lower bound is set as follows

$$\underline{\lambda} = \max\left(\frac{-5}{1}; \frac{-6}{1}\right) = -5.$$

When $\underline{\lambda} < -5$, the distance variable r_{21}^{+} leaves the basis. Testing the dual feasibility of new basic solution we determine that the distance variable r_{21}^{-} is the entering variable. The solution structure stays the same after one step of dual simplex method. Thus, $\underline{\lambda} = -6$, because when $\underline{\lambda} < -6$, the variable x_2 is the leaving variable and structure of the solution is surely changed.

The upper bound $\overline{\lambda}$ will be determined as

$$\overline{\lambda} = \frac{-2}{-1} = 2 .$$

This is the real upper bound for b_1 constant change. When $\overline{\lambda} > 2$ than the basis is replaced. The variable r_{22}^{-} is the leaving variable and the variable d_1 is the entering variable. The structure of solution is changed.

5 Conclusion

There was described a method for finding a sub-optimal solution, feasibility interval of non-basic variables and RHS in the article. It was compared with standard method for sensitivity analysis in LP models. Conclusion of the comparation is, that the use of new method is not so much complicated, and brings reasonable solutions.
Principles of the suggested method can be also used for cost coefficients sensitivity analysis. Its specification and description is out of the range of this article.

References

[1] Houška, M.: Modified Simplex Algorithm for Solving Non-Convex PWLP Models. In Proceedings of Matematical methods in economics, PEF ČZU v Praze, 2003

[2] Houška, M.: The Method of Cost Evaluated Distance Variables for Solving of Some Non-Convex Piecewise Linear Programming Models. In Proccedings of the 3rd Ph.D. seminar, PEF ČZU v Praze, Praha, 2003

[3] Houška, M., Brožová, H.: The Method of Cost Evaluated Distance Variables for Solving Piecewise Linear Programming Models. In: Proceedings of the 20th International Conference MME, Ostrava, 2002

[4] Jensen, P. A., Bard, J.F.: Operations Research Models and Methods. Wiley, Canada, 2002, 700 p.

[5] Keha, A. B., de Farias, I. R., Nemhauser, G. L.: Models for Representing Piecewise Linear Cost Functions. Operations Research Letters 32, 44-48, 2004.

[6] Hulgeri, A., Sudarshan, S.: Parametric Query Optimization for Linear and Piecewise Linear Cost Functions. Proceedings of the 28th VLDB Conference, Hong Kong, 2002

Modelling the Equilibrium on External Market Case of Slovakia

Jaroslav Husár
Katedra operačného výskumu a ekonometrie EU Bratislava, EU Bratislava, Dolnozemská 1b, Bratislava
husar@euba.sk

Abstract

The Slovak economy suffers from serious problems in the field of export and import, international trade. So today's research agenda is highly oriented on this problem. There are noteworthy differences in the effectiveness of the foreign sector performance between Slovakia and many developed countries. To get satisfactory economic performance, we see as one of the important problem to use tools that will help to create the strategy for the phenomenon captured under the popular heading globalisation. Empirical techniques based on model are the field that has become the interest of our research. The paper will present the model based on the principles of Mundell-Fleming. We maintain that one of the principles might be the problem of reaching the equilibrium in foreign sector. But what determines the equilibrium. The paper goal is to get the answer.

Keywords

Foreign trade, modelling, fixed exchange rate, flexible exchange rate, multipiers

Introduction

Recent economic performance of Slovakia is not better than in the historical context from 1993. It concerns particularly the foreign trade of Slovakia. We had a difficult economic time over the last several years. Comparison with other countries of Central Europe shows that we need a better economic engine. The sluggish nature of the recovery was particularly evident in two fields: foreign trade and labour market. To get a clear picture of the foreign sector development we prepared the statistics on GDP and foreign trade in Table 1.

Table 1: Gross domestic product and foreign trade, mld. SKK, current prices

Indicator	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Gross dom. prod.	**406.6**	**486.1**	**568.9**	**628.6**	**708.6**	**775.0**	**835.7**	**908.8**	**989.3**	**1096.4**
For. trade, turnover	477.0	552.1	643.1	739.5	862.5	992.5	1056.2	1326.9	1548.3	1639.3
Export	228.3	287.2	326.4	334.7	397.4	456.8	510.0	652.4	732.3	781.4
Import	248.7	264.9	316.7	404.8	465.1	535.7	546.2	674.5	816.0	857.9
Trade balance	**-20.4**	**22.3**	**9.7**	**-69.9**	**-67.7**	**-78.8**	**-36.2**	**-22.1**	**-83.7**	**-76.5**

Trade balance (deficit) is not improving. More indicative, we think, is the recalculation of the foreign trade performance in percentages that is given in Table 2. In the year 2003 the economy foreign trade expansion was **183.0** %, export performance was **77.2** % and import burden was **78.5** %. Government administration has applied fixed exchange rate regime up to 1998 and than shifted to floating exchange rate regime. Expectations of the policy makers did not come true. We consider the **optimalisation of the instruments** of the monetary policy, fiscal policy and trade policy as a way for the management of the Slovak economy and foreign trade problems. We do not want analyse the monetary and fiscal policies in the context of events, both historical and recent, that shaped them. In the research project we concentrate on the model building for the improvement of the foreign trade performance. We believe it can help us to understand the overall economy functioning.

Table 2: Foreign trade dependence of the Slovak economy, % in current prices

Indicator	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
Foreign trade expansion	**117.3**	**113.6**	**113.0**	**117.6**	**121.7**	**128.1**	**126.4**	**146.0**	**156.5**	**149.5**
Exp. performance	56.1	59.1	57.4	53.2	56.1	58.9	61.0	71.8	74.0	71.3
Import burden	61.2	54.5	55.7	64.4	65.6	69.1	65.4	74.2	82.5	78.2

1. New approach to the foreign trade performance modelling

We want to devote to examination of the operation of monetary policy in Slovakia that works thorough the financial markets by, as it is known, way of a portfolio-balance mechanism. To construct a model that helps to analyze monetary policy, we first of all have to make precise the meaning of the money and various types of assets that may be considered. The real life situation is more sophisticated then we need it for our model construction. Some details are not required. We need a brief discussion of the determinants of the demands for these assets. In any case we need to discuss the supply of the various financial assets, with particular reference to the supply of money, and examine the asset market equilibrium conditions and hence the determinants of the domestic interest rate, the exchange rate (this is the subject of other paper), and the net capital outflow.

The characteristics and types of financial assets

We know that in real economy there are many types of financial assets. There is no need in this paper to make precise all the types of the financial assets in Slovakia. Not to complicate the model we consider (we concentrate on) only some of them. To make any analysis of the financial sector in Slovakia and in the same time the model construction tractable a number of simplifying assumptions regarding the number of assets we accepted. After the long and tedious discussion the asset choice was restricted to just three: domestic money, domestic bonds and foreign bonds. In principle we had no problems with the money as an asset generally accepted in the final settlement for purchases. A supply of money can be thought of as the stock of such assets outstanding at a point in time.

As far as a bond is concerned we accepted that it is an imperfectly liquid asset. There were no questions. To convert it into cash it has to be sold on a market. There is a risk of making capital loss if the selling price is below the price at which the bond was purchased. We know that the price of a bond on a market is the present value of the income stream to which the bond is a claim (formula is given in the NBS regulations). In principle, bond yields are inversely related to bond price. Again, there are two types of bonds: those that are redeemable at a fixed future date and those which are perpetuities. In an open economy context Slovak residents (domestic) may choose to hold their domestic bonds, denoted B, issued by domestic government, or foreign currency bonds issued by the foreign government, denoted B^f, which are held by both foreign and domestic residents. You see that our model construction is getting more and more complicated. For the model purpose there are two important differences between the way in which domestic bonds and foreign bonds are treated in the model. The first is that domestic bonds are assumed to be non-traded while foreign bonds are traded. Hence the domestic bond stock is all held by domestic residents while the stock of foreign currency-denominated bonds, B^f, is held partly by domestic residents, B^f_d, and partly by foreign residents, B^f_f. The second difference between domestic and foreign bonds is that foreign bonds are denominated in foreign currency. Hence exchange rate movements will affect the domestic value of foreign currency bonds (a depreciation of the domestic currency will raise the domestic value of foreign-currency-denominated bonds and opposite!!). There are two risks – associated with the illiquidity and with the movements of exchange rate.

So we decided to exclude all other financial assets from our analysis (or are assumed to move in proportion with included assets!!).

On this base we are in the position to try to formulate the mathematical expression to be included in the model. Therefore, in aggregate the financial wealth constraint for the domestic non-bank private sector is given by:

$$\frac{A}{P} = \frac{M^s}{P} + \frac{B^s}{P} + \frac{EB_d^f}{P} \qquad \text{or} \quad 1 = \frac{M^s}{A} + \frac{B^s}{A} + \frac{EB_d^f}{A} \tag{1}$$

where A denotes the total stock of nominal non-bank private sector wealth measured in home currency unit (SKK), M^s is the money stock held by the domestic non-bank private sector, B^s is the total stock of domestic government debt held by the domestic private sector, B^f_d is the stock of foreign currency debt held by the domestic non-bank private sector, and E is the exchange rate measured as the domestic price of foreign currency to convert the foreign-currency-denominated bond stock into domestic currency units. P is the domestic price level, used to convert nominal values into real values.

The essence of our model is that the basic choice confronting domestic residents is how to allocate their financial wealth between these competing assets. The problem is (in order to understand this process) that we need to specify the specific demand functions for each of these assets (must be considered).

Demand functions for financial assets

We consider that demand functions for money and for domestic and foreign-currency bonds are all rather similar. Money, as we know is a unique asset to facilitate transactions. Money is also demanded as an asset because it serves as riskless asset and store of value. Money is fundamentally different to other assets. Thus if individuals believe that bond prices are going to fall, they will hold their wealth in cash, because in that way they can protect the value of their wealth. High bond prices are associated with low interest rates. Hence a fall in interest rates increases the demand for money as an asset. We consider (as the other authors do) that both of the components of the demand for money, the asset and the transactions demand, should be expressed as a demand for real money balances, $M^d/P = m^d$. According to these arguments the demand for real money balances may be summarized as:

$$\frac{M^d}{\overline{P}} = \lambda_0 + \lambda_1 Y - \lambda_2 i - \lambda_3 (\bar{i}^* + x) + \lambda_4 \frac{\overline{A}}{\overline{P}} \tag{2}$$

After our economic analysis we have agreed upon that eq. (2) says that the demand for money depends positively on both the level of real income, reflecting the transactions demand for money, and on the stock of financial wealth, reflecting portfolio size. This is not quite common formulation of the demand for money function. The returns on domestic bonds and foreign bonds, are i and $i^* + x$, respectively, are negatively related to the demand for real money balances, where x denotes the expected depreciation of the domestic currency.

Next our step was to analyze the bonds behaviour. The demands for foreign and domestic bonds, like the demand for money, also depend on the level of real income, although the direction of the *effect* is counter-intuitive. The crucial point is that if stock of non-bank private sector wealth is assumed to be given, then the rise in income that raises the demand for money through the transaction motive must necessitate a fall in the demand for other assets. Therefore a rise in income will reduce the joint demand for foreign and domestic bonds. For simplicity it is (in our attitude assumed that the demand for both domestic and foreign bonds fall as income rises (although this needs not to be the case).

Next we want to stress is that the demands for bonds are also assumed to depend *positively* on their own returns and *negatively* on the returns of the competing asset. This is to assume that assets are *gross substitutes*. After our analysis we say that the real demand for domestic and foreign bonds can be summarised as:

$$\frac{B^d}{P} = \phi_0 - \phi_1 Y + \phi_2 i - \phi_3 (\bar{i}^* + \bar{x}) + \phi_4 \frac{\overline{A}}{\overline{P}} \tag{3}$$

$$\frac{F^d}{P} = \varsigma_0 - \varsigma_1 Y - \varsigma_2 i + \varsigma_3 (\bar{i}^* + \bar{x}) + \varsigma_4 \frac{\overline{A}}{\overline{P}} \tag{4}$$

In these demand relationships it is important to remember that the foreign rate of interest, the expected depreciation of the exchange rate, the stock of financial wealth and the price level is given.

Asset supplies

Our framework needs more to be said about the assets. To move from the demand for each of the assets to the respective asset market equilibrium some assumptions need to be made about the supply of each of three assets, subject to the wealth constraint, given in (1).

What do we know? Domestic bonds are a liability of the domestic government and are issued to *finance the budget deficit*. Or more clearly, by purchasing government bonds (or securities) the banking sector or private sector efficiently lends to the government. In this context, in our model we consider the supply of domestic bonds to be wholly under control of the authorities and hence can be regarded as an exogenous variable. The supply of foreign currency bonds, on the other hand, are most easily thought of as issued by a foreign government and hence a policy variable for that government. Since the Slovakia is, as it is known, is a small country relative to the Europe (world), the supply of foreign bonds for domestic residents to by can be assumed to be infinite at given rate of interest ($\bar{i}^*$).

Next step in our model was to clarify the supply of money. Again, it is more complicated as it is treated here for our purposes. We have accepted two most important functions of NBS. The NBS compelled to provide main banking services, especially that of providing credit to the government. The second function of the central bank is to hold the nations foreign currency and gold reserves. So if the government needs to raise several million of crowns without raising taxation, the government simply approaches the NBS with its T-bills. It means that NBS have increased assets (lending to government) and issued notes and coins as high powered money (liabilities). Our purpose in Table 1 was to show that the consolidated balance sheet the combined positions of the banking system as a whole, which ultimately gives two identical measures of the quantity of money. From the right hand side of the balance we have $M_s = H_p + D_p$. By definition this definition must also be equal to the monetary assets of the banking system that are: $M_s = D + R$.

2. The core of the model of policy multipliers with perfect capital mobility

As we have already said. In Slovak Republic we have experienced two economic policies as far as the external market is concerned. Up to 1998 we used fixed exchange rates and then the flexible exchange rates.

After the discussions we started with the developing of a model (Mundell-Fleming type) that is focusing on the IS curve, LM curve, and the FE (foreign exchange equilibrium).[1] The three principal equations of the constructed model are

$$Y = \frac{\delta - \beta_1 \bar{i} + \overline{G} + \delta_1 \overline{Y}^* + (\delta_2 + \mu_2)(EP^*/P)}{1 - \alpha_1(1-\tau_1) + \mu_1} \tag{5}$$

$$\frac{M_s}{P} = \lambda_0 + \lambda_1 Y - \lambda_2 i - \lambda_3 i^* \tag{6}$$

$$i - i^* \tag{7}$$

The equation (5) represents the IS curve, the equation (6) the LM curve and the equation (7) is the bonds markets equation when there is a perfect capital mobility and static expectations. Now we are in the situation to study the economic fiscal and monetary policies. It requires substituting the equation (7) into equations (5) and (6). To study our economic policy in the last decade, we consider as a principal aspect to derive the multipliers. As we know, we have to do it for the two stages of our economic development. For the period started from 1993 up to 1998 and the following period.

[1] See the authors' paper in MME 2004, Brno, 2004.

Multipliers for the fixed exchange rates

In this stage of our research we concentrate our work on developing a mathematical tool that can fully describe the principal relationships. With fixed rates the money supply is endogenous, so we solve the model for income and the money supply. Differentiating the IS and LM functions and arranging in matrix form we have:

$$\begin{bmatrix} k & 0 \\ -\lambda & 1 \end{bmatrix}\begin{bmatrix} \partial Y \\ \partial M \end{bmatrix}=\begin{bmatrix} -\beta_1\partial_{i^*}+\partial G+\delta_1\partial Y^*+(\delta_2+\mu_2)\partial(EP^*/P) \\ +(M/P^2)\partial P-(\lambda_2+\lambda_3)\partial i^* \end{bmatrix} \tag{8}$$

Where k = 1 - $\alpha_1(1-\tau) + \mu_1$. Inverting left and 2 x 2 matrix, assuming that P = 1, gives:

$$\begin{bmatrix} \partial Y \\ \partial M \end{bmatrix}=\frac{1}{k}\begin{bmatrix} 1 & 0 \\ \lambda_1 & k \end{bmatrix}\times\begin{bmatrix} -\beta_1\partial_{i^*}+\partial G+\delta_1\partial Y^*+(\delta_2+\mu_2)\partial(EP^*/P) \\ +(M/P^2)\partial P-(\lambda_2+\lambda_3)\partial i^* \end{bmatrix} \tag{9}$$

There are three types of policy multipliers. They are as follows:

$$\frac{\partial Y}{\partial G}=\frac{1}{k}\rangle 0; \qquad \frac{\partial Y}{\partial i^*}=\frac{-\beta_1}{k}\langle 0; \qquad \frac{\partial Y}{\partial Y^*}=\frac{\delta_1}{k}\rangle 0; \text{ and } \qquad \frac{\partial Y}{\partial E}=\frac{(\delta_2+\mu_2}{k}\rangle 0;$$

We assume that interest rate is constant. Now we want to know the aggregate demand curve. This can be derived from the top equation from (9). It can be solved for $\partial P/\partial Y$ which is:

$$\frac{\partial P}{\partial Y}=\frac{1-\alpha_1(1-\tau_1)+\mu_1}{\delta_2+\mu_2}\langle 0;$$

We came to an interesting result. We see that the slope of AD curve is related to government expenditure multiplier. This logical deduction could not have been done without the mathematics, or our model.

Multipliers for the flexible exchange rates

Currently we have flexible exchange rates. So our next analysis is concerned with the modification of the model. Under flexible exchange rates the money supply becomes exogenous and the exchange rate endogenous, so the system is now:

$$\begin{bmatrix} k & -(\delta_2+\mu_2 \\ -\lambda_1 & 0 \end{bmatrix}\begin{bmatrix} \partial Y \\ \partial E \end{bmatrix}=\begin{bmatrix} -\beta_1\partial_{i^*}+\partial G+\delta_1\partial Y^*+(\delta_2+\mu_2)\partial(EP^*/P) \\ -\partial M+(M/P^2)\partial P-(\lambda_2+\lambda_3)\partial i^* \end{bmatrix} \tag{10}$$

Again, with the inversion of the left hand 2 x 2 matrix we get:

$$\times\begin{bmatrix} -\beta_1\partial_{i^*}+\partial G+\delta_1\partial Y^*+(\delta_2+\mu_2)\partial(EP^*/P) \\ -\partial M+(M/P^2)\partial P-(\lambda_2+\lambda_3)\partial i^* \end{bmatrix}$$

$$\begin{bmatrix} \partial Y \\ \partial E \end{bmatrix}=\frac{1}{-(\delta_2+\mu_2)\lambda_1}\begin{bmatrix} 0 & (\delta_2+\mu_2) \\ \lambda_1 & k \end{bmatrix}\times\begin{bmatrix} -\beta_1\partial_{i^*}+\partial G+\delta_1\partial Y^*+(\delta_2+\mu_2)\partial(EP^*/P) \\ -\partial M+(M/P^2)\partial P-(\lambda_2+\lambda_3)\partial i^* \end{bmatrix} \tag{11}$$

What are the policy multipliers? We have to go back to derivatives:

$$\frac{\partial Y}{\partial M}=\frac{1}{\lambda_1}\rangle 0; \text{ and } \frac{\partial Y}{\partial i^*}=\frac{\lambda_2+\lambda_3}{\lambda_1}\rangle 0;$$

What is the behaviour of the AD curve? The AD curve with flexible exchange rates is given by the top raw from equation (11), that is:

$$\frac{\partial P}{\partial Y} = -\lambda_1 \langle 0 \, .$$

This is the essence of our model that we have been developing for the foreign trade management based on the coordination of the three policies: monetary, fiscal and trade.

Conclusion

The efforts were concentrated on the theoretical approach as far the foreign trade problems is concerned in Slovakia. The reason was the sluggish nature of the recovery in the field of foreign trade. We consider the good economic policy making as the first task of government administration. In this paper we presented the core of the model in expressions from (5) to (11). We understand the movements of the multipliers of the fixed and flexible exchange rates.

References

[1] Husár, J., Aplikovaná makroekonómia. Sprint, Bratislava, 2003

[2] Husár, J., Macroeconomic model for the analysis of the Slovak economy with regard to international sector. *Mathematical Methods in Economics*, Praha 2004

[3] Hušek, R., Pelikán, J., Aplikovaná ekonometrie, Professional Publishing, Praha, 2004

[4] Pánková, V., Capital mobility in developing economies. Panel data comparison. *Mathematical Methods in Economics*, Praha 2004

[5] Rublíková, E., Ivaničová, Z., Modelling exchange rate SKK/CZK. *Mathematical Methods in Economics*, Praha 2004

Some Problems of VAR Models of the Monetary Transmission Mechanism

Roman Hušek[1]
University of Economics, Department of Econometrics, 4. W. Churchill Sq., CZ – 130 67, Prague 3
(email: husek@vse.cz)

Abstract

The prospect of adopting euro currency in the new EU member states has boosted a number of studies focusing on estimation, functioning and comparison of monetary transmission mechanism in an international environment. Such analysis is often based on the theory of optimum currency area by Mundell (1961) and usually uses macroeconomic vector autoregression models (VAR). Generally, VAR methodology is focused on autonomous shocks, and their pattern of influencing the system is described by the impulse response functions and variance decomposition. In this paper, we discuss the possibilities encompassed in the VAR approach and at the same time we stress its known barriers. This task if of considerable importance, as various barriers (limitations) of the VAR approach are often ignored in many studies, which frequently leads researchers into inaccurate and often completely incorrect results and conclusions.

Keywords

VAR model, monetary transmission mechanism, impulse response function.

JEL Classification

C 320, E 520, F 410

1 Introduction

The admission of ten new member states into EU during the year 2004, their foreseen ingress into the ERM2 regime and their prospects of adopting a common euro currency have given rise to a number of studies focusing on the functioning of a monetary transmission mechanism (MTM). Such studies have been published both locally and in foreign publications. The studies are mainly based on Mundell's theory of **optimum currency area** (Mundell, 1961) and they focus mainly on the international comparison of effects of monetary policy on individual economies with different macroeconomic cycles, different financial structure and different readiness or degree of ability to adopt a common monetary policy (e.g. Demertzis et al., 1997 or Formánek and Hušek, 2004).

Vector autoregression (VAR) models are the most common tool for econometric analysis and quantification of exogenous shocks arising from monetary policy and for international comparison of functioning of transmission mechanisms in different economies. In this paper, we shall therefore discuss in detail the possibilities encompassed in the VAR approach and at the same time we shall stress its barriers. Some barriers or limitations of the VAR approach are frequently ignored in many studies and this may lead to inaccurate and often completely incorrect results and conclusions.

2 VAR analysis of the transmission mechanism

VAR models used in the MTM analysis differ from structural models by the purpose for which they are built and estimated. We use VAR models when we aim to analyse the MTM in order to assess

[1] This paper is a part of the research project of the Grant Agency of the Czech Republic No. 402/03/1299.

the reaction of macroeconomic variables within the system to a monetary policy measure (shock). VAR models may not be applied in order to provide the central bank (CB) with an optimal policy rule.

1.1. Theoretical background for VAR models

VAR models of the monetary transmission mechanism share common basic features with linear structural models of MTM and may be expressed in matrix form as

$$\mathbf{A}\begin{bmatrix}\mathbf{x}_t\\ \mathbf{z}_t\end{bmatrix} = \mathbf{C}(L)\begin{bmatrix}\mathbf{x}_{t-1}\\ \mathbf{z}_{t-1}\end{bmatrix} + \mathbf{B}\begin{bmatrix}\mathbf{v}_{xt}\\ \mathbf{v}_{zt}\end{bmatrix} \quad , \tag{1}$$

where

- **x** is a vector of stationary target variables (e.g. GDP and inflation),
- **z** – vector of stationary control (instrumentary) variables (e.g. interest rates and monetary aggregates),
- $\mathbf{C}(L)$ – matrix of polynoms of the lagged operator L of definite order,
- **v** – vector of structural random elements (white noise shocks) in equations of target and control variables, with identical covariance matrix $\mathbf{\Lambda}$,
- $\mathbf{A}$,$\mathbf{B}$ – matrixes of parameters.

Unlike structural simultaneous models, VAR models are not based on an assumption of exogeneity of monetary policy tools. Hence, with VAR models, all variables are considered to be endogenous.

Structural VAR model (1) not containing a level (constant) may be transformed (solving by the endogenous variables) into a reduced form with no feedback action among the non-lagged endogenous variables. Such a reduced form is commonly described as the **standard form** of a VAR model. Therefore, for a regular **A** we may write

$$\begin{bmatrix}\mathbf{x}_t\\ \mathbf{z}_t\end{bmatrix} = \mathbf{A}^{-1}\mathbf{C}(L)\begin{bmatrix}\mathbf{x}_{t-1}\\ \mathbf{z}_{t-1}\end{bmatrix} + \begin{bmatrix}\mathbf{u}_{xt}\\ \mathbf{u}_{zt}\end{bmatrix} \quad , \tag{2}$$

where

- **u** is a vector of normally distributed random elements of the reduced form - in equations of the target and control variables; it has a general covariation matrix $\mathbf{\Sigma}$ of order n.

The elements $\mathbf{u}_{xt}$ and $\mathbf{u}_{zt}$ are the so called **pure innovations** (shocks) that represent autonomous changes in vectors $\mathbf{x}_t$ and $\mathbf{z}_t$. Hence, the following relation holds for vectors **v** and **u**:

$$\mathbf{A}\begin{bmatrix}\mathbf{u}_{xt}\\ \mathbf{u}_{zt}\end{bmatrix} = \mathbf{B}\begin{bmatrix}\mathbf{v}_{xt}\\ \mathbf{v}_{zt}\end{bmatrix} ,$$

which may be re-written as

$$\mathbf{u}_t = \mathbf{A}^{-1}\mathbf{B}\mathbf{v}_t \quad . \tag{3}$$

By definition, the relationship between covariation matrixes $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ (identity matrix) is

$$E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{A}^{-1}\mathbf{B}E(\mathbf{v}_t\mathbf{v}_t')\mathbf{B}'\mathbf{A}^{-1} \quad . \tag{4}$$

The parameters of the standard form (2) may be directly estimated by ordinary least squares (OLS) method, with the resulting estimates being consistent and asymptotically profuse. If the estimates of the standard form (2) are set in to the equation (4), we may re-write (4) as

$$\text{est.}\mathbf{\Sigma} = \text{est.}\mathbf{A}^{-1}\ \text{est.}\mathbf{B\Lambda}\ \text{est.}\mathbf{B}'\ \text{est.}\mathbf{A}^{-1} \ . \tag{5}$$

2.2 Identification problem

VAR model (2) with no levels, with n equations and endogenous variables and with a common maximum lag p in all equations has a total of pn^2 parameters in the matrix **A**. It is therefore obvious, that the model is over-parameterized. Hence, it is necessary to omit some of the variables before the estimation – but not at the expense of loosing important information. The reduction of variables is further complicated by the fact that lagged explanatory variables tend to be strongly collinear – which makes it impossible to use standard tests for statistical significance in order to select non-significant variables to eliminate.

The estimated matrix **Σ** is symmetric and it contains only $(n^2+n)/2$ individual elements, which is also the maximum number of identifiable parameters in matrixes **A** and **B**. Therefore, in order to reach identification, we have to impose such identifying (null) restrictions on the parameters of those two matrixes, so that the number of their parameters does not exceed the maximum of $(n^2+n)/2$. Similarly to structural models of simultaneous equations, with VAR models we make distinction between exact identification and over-identification.

The matrix **A** contains $n^2 - n$ unknown elements (all **A** diagonals are equal to 1) and with the diagonal matrix **B** we also have n unknown values of variance $\sigma^2_{u\,it}$ for the total of n^2 unknown values of the structural model. It is therefore clear that the **exact identification** of n^2 unknown and $(n^2+n)/2$ known independent elements of the estimated matrix **Σ** may be reached by imposing $n^2-[(n^2+n)/2] = (n^2-n)/2$ additional (null) restrictions on the parameters in matrixes **A** and **B** of the structural model. This may be done in various ways. We shall briefly mention the most widely used approaches to VAR model identification.

The exact identification is usually obtained by Choleski decomposition (factorization) of matrixes **A** and **B** (Sims, 1980 or Hamilton, 1994), which for both matrixes may be expressed as

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_{11} & 0 & 0 & \dots & 0 \\ 0 & b_{22} & 0 & \dots & 0 \\ 0 & 0 & b_{33} & \dots & 0 \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & b_{nn} \end{bmatrix}. \tag{6}$$

The results of identification of structural shocks $\mathbf{v}_t$, using the estimated vector of pure innovations (shocks) $\mathbf{u}_t$ performed by Choleski decomposition, are dependent on the ordering of variables of the structural VAR model. The actual ordering and the corresponding structural shocks selection are often performed on an ad hoc basis. As the matrix **A** has a lower-triangular shape, endogenous variables may be hierarchically ordered according to the theoretical (expected) structure of the modelled system. The triangular shape of matrix A therefore means that only one-way causalities exist among the variables. Over the last ten years, other identifying short-run restrictions are being used. Such restrictions have not a triangular (recursive) structure and they lead to a general (non-triangular) matrix **A**, which allows us to assume feedback interactions among some endogenous variables of the model.

The imposing of $(n^2-n)/2$ restrictions of the parameters is not a generally sufficient condition for reaching the exact identification of the model. The procedure for exact identification gets more complicated when **non-linear** restrictions are present in the model (Enders, 1995).

If, corresponding to economic theory, there are more than $(n^2-n)/2$ identifying restriction in any model, the system becomes **over-identified**, which may result in non-linear parameter restrictions. Nevertheless, sophisticated software products (EViews, Pc Give) allow us to test the significance of over-identifying restrictions (including non-linear restrictions), e.g. using the χ^2 statistics..

Blanchard and Quah (1989) have proposed an alternative method of identification of structural VAR models. Assuming stationarity of variables, this method is based on applying long term null restrictions on cumulative impulse response functions (IRF), i.e. on the reactions of levels of individual variables to different exogenous shocks – e.g. monetary policy. If null restrictions are applied, the long term reaction of levels of stationary variables to specified autonomous shocks converges to zero. The application of Blanchard and Quah (BQ) decomposition in MTM analysis is presented by Enders (1995) or Hušek and Formánek (2005) and other authors.

2.3 Analysis of impulse response functions

VAR models allow us to identify the deviations from a monetary policy rule chosen by the CB. However, this identification is not based on changing the systematic elements of monetary policy, i.e. using changes in matrixes **A** – it is based on influences imposed by structural **exogenous shocks** that do not have effect on the parameters in matrixes A and **B**. Therefore, autonomous shocks do not affect reactions of monetary authorities. The MTM analysis performed using VAR methodology may be simplified by the assumption that structural shocks are **linear combinations** of residuals of the estimated reduced form of the VAR model.

In order to derive and interpret IR functions representing the dynamic effects of exogenous shocks influencing individual endogenous variables, we shall write a structural VAR model for n variables as

$$\mathbf{A}_0\mathbf{y}_t = 1\sum_{i=1}^{p}\mathbf{A}_i\mathbf{y}_{t-i} + \mathbf{B}\mathbf{v}_t \ , \tag{7}$$

where **y** is the vector of n stationary endogenous variables.

We may rewrite (7) as

$$\left[\mathbf{A}_0 - \mathbf{A}(L)\right]\mathbf{y}_t = \mathbf{B}\mathbf{v}_t \ , \tag{8}$$

where $\mathbf{A}(\mathrm{L}) = \sum_{i=1}^{p}\mathbf{A}_i L^i$.

Assuming the existence of inversion $[\mathbf{A}_0 - \mathbf{A}(L)\ ]^{-1} = \mathbf{C}(L)$ we may multiply the equation (8) by a matrix $\mathbf{C}(L)$ and re-write (e.g. Hamilton, 1994) the standard VAR model in the form of **vector moving averages** (**VMA**)

$$\mathbf{y}_t = \mathbf{C}(L)\mathbf{v}_t \, . \tag{9}$$

For a sufficiently large s, the equation (9) may be expressed as

$$\mathbf{y}_t = \sum_{i=1}^{s}\mathbf{C}_i\mathbf{v}_{t-i} = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + \ldots + \mathbf{C}_s\mathbf{v}_{t-s} \ , \tag{10}$$

which leads to $\mathbf{C}_0 = \mathbf{A}^{-1}\mathbf{B}$.

Therefore, matrix $\mathbf{C}_s$ may be interpreted as

$$\mathbf{C}_s = \frac{\partial \mathbf{y}_{t+s}}{\partial \mathbf{v}_t} \ , \tag{11}$$

so its ij^{th} element represents the impact of unitary shock (innovation) of j-th variable of the system at time t (v_{jt}) on the value of i-th variable at time $t+s$ ($y_{i,t+s}$), with all other shocks being constant. By simulation procedure for the changing values of s we obtain a function describing reactions (responses) of arbitrary i-th variable to exogenous shock (impulse) in j-th variable.

VAR model in the VMA form (10) is therefore an appropriate tool for identification of relations among time sequences of arbitrary pairs of variables, as coefficients of the matrix **C** are nothing but

dynamic multiplicators. The graphic representations of all elements of matrix $\mathbf{C}_s$ in *i*-th row of *j*-th column, i.e. marginal values

$$\frac{\partial y_{i,t+s}}{\partial v_{jt}} \tag{12}$$

as functions of the lag length *s*, are the IRF. Prior to IRF identification, all shocks of the vector v_t have to be transformed (e.g. using Choleski factorization) into **orthogonal** variables (innovations) that are pair wise uncorrelated and have unitary variance.

Another advantage of a VAR model in the VMA form is the possibility of **variance decomposition** of the forecast errors $\mathbf{e}_s$ of the vector $\mathbf{y}_s$, i.e. for *s* periods. The vector of forecast errors for a prediction horizon of *s* periods may be re-written as

$$\mathbf{e}_s = (\mathbf{y}_{t+s} - E_t\mathbf{y}_{t+s}) = \mathbf{C}_0\mathbf{v}_t + \mathbf{C}_1\mathbf{v}_{t-1} + ... + \mathbf{C}_s\mathbf{v}_{t-s} \ ,$$

hence

$$\text{var}(\mathbf{e}_s) = \mathbf{C}_0\mathbf{\Lambda}\mathbf{C}_1' + ... + \mathbf{C}_s\mathbf{\Lambda}\mathbf{C}_s' \ . \tag{13}$$

Therefore , we may use (13) to identify the relative influence of variances of individual structural shocks (orthogonal shocks) on the total variance – as the pair wise shock covariances have zero value.

For relatively small samples, estimation of large number of parameters of the VAR model usually leads to statistically insignificant estimates and IRFs. Their quality may be improved by e.g. using apriori information, i.e. by using Bayesian VAR models of MTM with changing coefficients.

3 Problems and barriers of VAR models of MTM

Although VAR models of MTM have many advantages over structural models of simultaneous equations – especially in the field of analysis of effects of monetary policy – it is necessary to mention here their drawbacks and limitations in comparison with other econometric models.

a) Due to their inherent a-theoretical character, VAR models are not particularly useful for theoretical analysis and subsequent choosing of monetary policy tools.

b) If strong emphasis is put on the maximum consistence between data and the VAR model, we may end up with spurious relationships or even with ambiguous interpretation of estimated parameters.

c) For small samples, VAR models are over parameterized, so that degrees of freedom are low and the estimates of parameters and IRFs are not statistically significant.

d) Although all variables of the VAR model should be stationary, some authors suggest not to use differentiation of variables in order to stationarize them, as information concerning long term relationships among variables is lost from the model by such approach.

e) Although VAR models are generally presented as ultimately unrestricted and encompassing all uniformly lagged variables as regressors, in some cases it is necessary to define individual equations of the VAR model with different lag lengths for specific variables – and therefore such model has the character of restricted vector autoregression. To set the optimum lag length in such situation, it is not convenient to use e.g. block F test and LR test, which require normality of random elements in the equations (usually not met in many financial variables). Instead, it is preferable to use **information criteria**, proposed by Akaike, Schwartz or Hannan and Quinn – they are incorporated in many software products (EViews, Pc Give).

f) Block F tests, i.e. causality tests allow us to identify variables in VAR models that have a statistically significant influence on future values of all variables resolved by the model. However, block F test cannot explain the sign of the estimated parameters, nor the length of prevalence of the individual shock influences. This may be determined only by analysis of the IRF and by variance decomposition.

g) Monetary theory provides general guidelines for the ordering of variables in VAR models of MTM, but for specific tasks the ordering may change and we have to bear in mind that the change of ordering may alter the results obtained from the VAR model.

h) Both IRFs and variance decomposition often are not suitable for exact interpretation, as their confidence intervals are usually very broad.

i) While assessing the application of VAR models of MTM for study of effects of monetary policy in the Czech Republic, it is important to keep in mind that regardless of many applications focused on the study of effects of monetary policy in open economies, VAR models are less successful in open economies than in closed economies.

4 Conclusions

The application of vector autoregression on MTM shows that VARs are a suitable tool for analysing the reactions of macroeconomic variables on impulses provided by actual monetary policy. It is substantial to identify monetary policy actions of the CB using restrictions that are independent of other (competing) models of MTM - whereas at the same time, monetary policy tools are considered to be endogenous variables. VAR methodology is focused on autonomous shocks, and their pattern of influencing the system is described by the IR functions and variance decomposition.

Unlike studies focused on the analysis of effects of monetary shocks in closed economies, the application of VAR models of MTM in open economies is problematic in the sense of difficult identification of structural monetary shocks. A significant problem of VAR methodology used for MTM analysis – not only for open economies – is the generally low statistical significance of the results, therefore in practice it is very difficult to reach goals, theoretically stipulated for the analysis.

5 References

[1] Blanchard, O.J.- Quah, D.: The dynamic effects of aggregate demand and supply disturbance. American economic review 79 (1989), 656-673.

[2] Demertzis, M.- Hallet, H.A.- Rummel, O.: Does *a core periphery regime make Europe into an o optimal currency area?* In: Welfens, P.J.J.: European monetary union, Heidelberg 1997.

[3] Enders, W.: *Apllied econometric time series*. J. Willey, N. York 1995.

[4] Formánek, T.- Hušek, R.: *Estimation of expected costs of introducing Euro to CR*. In: Mathematical methods in economics 2004, Brno 2004, 109-114.

[5] Hamilton, J.: *Time series analysis*. Princeton University Press, Princeton 1994.

[6] Hušek, R,- Formánek, T.: *Estimation of the CR sacrifice ratio for the transition period*. Prague Economic Papers,1,2005, 51-63.

[7] Mundell, R.A.: *A theory of optimum currency areas*. American economic review 53 (1961), 717-725.

[8] Sims, C.A: *Macroeconomics and reality*. Econometrica 48 (1980), 1-48.

Extremes of Stochastic Processes

Jana Husová
Agricultural Faculty, University of South Bohemia
Studentská 13, České Budějovice, Czech Republic

Abstract

Many economic problems are studied as stochastic processes. It may be important to know the behaviour of their extremes values. In the paper a weakly convergent sequence of stochastic processes is supposed. A weak convergence of processes, which are obtained as extremes from original processes, is studied. Characteristics of index collections of such processes are established.

Keywords

Stochastic processes, weak convergence, extremes

1 Introduction

Let us consider that we are able estimate some parameters and we know that such estimates converge in distribution to real values of these parameters. The parameters are indexed by an index set T. The aim of this paper is to look more closely at conditions under which we can obtain the same results also for parameters derived as extremes over subsets of the index set T from primary parameters.

The distribution of random variables extremes is studied in many books and papers. We don't study such extremes of random variables collections, we suppose that we know the distribution of such extremes and our aim is to study behaviour of such extremes as a process.

Now, we introduce the notation. We have space of bounded continuous real-valued functions on some measure space (T, μ) ($T \subset \mathbf{R}^m$ for some $m \in \mathbf{N}$ and μ is Lebesgue's measure), we denote it briefly by $C_b(T)$, with L_2-metric. Let us recall L_2-metric for functions:

$$\rho(f,g)_2 = \left(\int_T (f-g)^2 d\mu \right)^{\frac{1}{2}}.$$

In whole article we will use weak convergence or convergence in distribution, the definition can be found in [4, p. 2,17]. We will use this convergence

on different spaces. If we have convergence of n-dimensional distribution then we use the definition on space $\mathbf{R}^n$, for convergence of whole process we use space $C_b(T)$ or some subspace of $C_b(T)$.

We will consider a sequence $(X_n(t), t \in T, n \in \mathbf{N})$ of stochastic processes with trajectories in $C_b(T)$ such that $X_n(t) \xrightarrow{\mathcal{D}} X(t)$ in the space $C_b(T)$. Our aim is to find collections of subsets of set T, $\mathcal{A} \subset \mathcal{P}(T)$ ($\mathcal{P}(T)$ is collection of all subsets of T), and derive conditions under which sequence of stochastic processes $(Y_n(A), A \in \mathcal{A})$ defined by

$$Y_n(A) = \sup_{t \in A} X_n(t) \tag{1}$$

converges in distribution. Supreme processes $(Y_n(A), A \in \mathcal{A})$ are real-valued functions defined on $\mathcal{A}$.

2 Results

It is worth pointing that for space of bounded functions on compact space equipped with a supreme metric the result is very easy since the function of supreme

$$\begin{aligned} f : l^{+\infty}(T) &\rightarrow l^{+\infty}(\mathcal{A}) && (2) \\ x &\mapsto y : y(A) = \sup_{t \in A} x(t) && (3) \end{aligned}$$

is continuous on this space. So if we know distribution of finite dimensional projections we have convergence of whole processes in distribution, for more details we refer the reader to [2].

In case of L_2-metric there is necessary the assumption of continuity of process trajectories. Without such assumption there is no good result. But it isn't enough to suppose continuous trajectories, see following example.

Example: Let us have processes on space $((0,1], \mathcal{B}([0,1], \lambda)$ such that for all ω:

$$\begin{aligned} X_n(t) &= 1 - nt \quad \forall t \in \left(0, \frac{1}{n}\right], \\ &= 0 \quad \forall t \in \left[\frac{1}{n}, 1\right]. \end{aligned}$$

Then these processes are deterministic and they converge to the function $X(t) = 0$ for all $t \in (0,1]$. So we have the convergence in distribution, too. But $Y_n((0,1/4]) = 1$ *a.s.* for all n and $Y((0,1/4]) = 0$ *a.s.* It is easily seen

that if we take this sequence and a sequence of $Z_n(t) = 0\ , t \in (0,1]$ so the mixed sequence of functions converges but the mixed sequence of suprema doesn't.

$\triangle$

At first we will suppose processes with trajectories such that

$$|X(s) - X(t)| \leq K\rho(s,t)^p \quad a.s. \tag{4}$$

for some $p > 0$, where $\rho(\cdot,\cdot)$ is euclidian metric on T.

So, in fact, we consider that processes have $a.s.$ trajectories in a space of bounded Hölder functions with parameter p and eqquipped with L_2-metric, we denote it briefly by $(C_b^{0,p}(T), \rho)$. In such case the one dimensional projection is continuous function. So convergence of finite dimensional distributions of processes is necessary condition for convergence of whole processes.

We consider convergence in distribution of processes X_n to some process X. Due to continuity of projections and continuity of function (3) for fixed A we have convergence of one dimensional distribution of processes $(Y_n(A), A \in \mathcal{A})$ for all $\mathcal{A} \subset \mathcal{P}(T)$. The space of real-valued functions on $\mathcal{A}$ is equipped with a supreme metric and we can write

$$||y_1 - y_2||_s = \sup_{A\in\mathcal{A}} |y_1(A) - y_2(A)| \leq \sup_{t\in\cup\mathcal{A}} |x_1(t) - x_2(t)|, \tag{5}$$

where $\cup\mathcal{A} = \{x, \exists A \in \mathcal{A} : x \in A\}$. It follows easily that for every $\varepsilon > 0$ and for any $x_1, x_2 :\ \rho(x_1, x_2) \leq \delta$ we can find δ such that there is satisfied $\sup_{t\in\cup\mathcal{A}} |x_1(t) - x_2(t)| \leq \varepsilon$. So we get following theorem.

Theorem 1 *For every $\mathcal{A} \subset \mathcal{P}(T)$ the function $f : (C_b^{0,p}(T), \rho) \to (l^{+\infty}(\mathcal{A}), ||\cdot||_s)$ such that*

$$f(x(t)) := (\sup_{t\in A} x(t)) \tag{6}$$

is continuous. Then convergence in distribution of processes $(X_n(t), t \in T)$ implies convergence in distribution of processes $(Y_n(A), A \in \mathcal{A})$ defined by (1).

Proof: The continuity of function follows from inequality (5) and following remark. So convergence in distribution is straightforward from continuous mapping theorem (see [1, Th. 5.1, p. 30]).

$\triangle$

Now we turn to arbitrary continuous functions.

At first recall that the space of continuous functions isn't compact in L_2-metric. Some convergent sequences haven't limits in space of continuous functions. For example we can take sequence of functions from Example and define functions on whole interval $[0, 1]$. Then the limit function isn't continuous.

The other problem is that in case of $(C_b(T), ||\cdot||_2)$ there isn't necessary for convergence in distribution of processes convergence in distribution of finite dimensional distributions since one dimensional projection isn't continuous mapping in this space. But in case when finite dimensional distributions don't converge, the convergence of suprema over some subsets has no sense.

If we want to have convergence in distribution of process in space of bounded functions on $\mathcal{A}$ with supreme metric, $(l^{+\infty}(\mathcal{A}), ||\cdot||_s)$, we need at first convergence in distribution of one dimensional projection to a tight and measurable limit $Y(A)$ ($Y_n(A)$ aren't ordinary measurable) and moreover asymptotical tightness of whole process (see [4, ch. 1.5]).

The convergence of one dimensional marginals, $Y_n(A) \xrightarrow{\mathcal{D}} Y(A)$, for some $A \subset T$ is studied in another books and papers. We remark only that it isn't sufficient to suppose convergence in distribution of finite dimensional marginals of processes (X_n). (It is sufficient only for $A \subset T$ containing finite many elements.)

If we consider convergence in distribution of whole process and moreover convergence of every finite distributions then we have convergence in distribution of supremum over every finite $A \subset \mathcal{P}(T)$. Since there is convergence in distribution of vectors $(X_n(t_i), t_i \in A) \xrightarrow{\mathcal{D}} (X(t_i), t_i \in A)$ we obtain convergence in distribution of maximum, i.e. $Y_n(A) \xrightarrow{\mathcal{D}} Y(A)$, where A contains finite many of $t \in T$.

If we want to have convergence in distribution over infinite subset we need some another assumptions.

Let us suppose convergence in distribution of one dimensional marginals of processes Y_n indexed by some collection $\mathcal{A}$ to measurable and tight limits. And the question is if we also have convergence in distribution of whole process indexed by this collection $\mathcal{A} \subset \mathcal{P}(T)$. According to [4, Th. 1.5.4. and Th. 1.5.7, p. 37], we need to find semimetric $\rho^{\mathcal{A}}$ on $\mathcal{A}$ such that space $(\mathcal{A}, \rho^{\mathcal{A}})$ is totally bounded and the sequence (Y_n) is asymptotically uniformly $\rho^{\mathcal{A}}$-equicontinuous in probability. It means that for every $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that

$$\limsup_n P^* \left(\sup_{\rho^{\mathcal{A}}(A,B)<\delta} |Y_n(A) - Y_n(B)| > \varepsilon \right) < \eta, \tag{7}$$

where P^* denote outer measure.

We can define several semimetrics on collections of subsets of T but there is a problem with assumption (7). So we recall only Hausdorff semimetric on subsets which is given by:

$$\rho^{\mathcal{A}}(A,B) := \max(\sup_{s\in A}\inf_{t\in A\cap B}\rho(s,t), \sup_{s\in B}\inf_{t\in A\cap B}\rho(s,t)). \tag{8}$$

For this semimetric we get

$$\sup_{A,B:\ \rho^{\mathcal{A}}(A,B)<\delta} |Y_n(A) - Y_n(B)| \le w_{\cup\mathcal{A}}(X_n,\delta) \tag{9}$$

where $w_{\cup\mathcal{A}}(\cdot,\cdot)$ is modul of continuity (see [1, p. 54]) taken over $\cup\mathcal{A}$. So convergence of finite dimensional distributions, total boundness of $\mathcal{A}$ under $\rho^{\mathcal{A}}$ and asymptotical uniform ρ-equicontinuity of processes X_n in probability on $\cup\mathcal{A}$ is sufficient for convergence in distribution of Y_n as process.

The characterization of collections $\mathcal{A} \subset \mathcal{P}(T)$ to be totally bounded under Hasdorff semimetric is given by following theorem.

Theorem 2 *The collection $\mathcal{A} \subset \mathcal{P}(T)$ is totally bounded if for every ε : $0 < \varepsilon < 1$ there exists $B_\varepsilon \subset T$ such that $\mu(B_\varepsilon) = r_\varepsilon$ where $r_\varepsilon \to 0$ as $\varepsilon \to 0$ and*

$$\mathcal{A}\backslash B_\varepsilon := \{A\backslash B_\varepsilon; A \in \mathcal{A}\} \tag{10}$$

is VC-class.

For more details about VC-classes see [3, ch. 2] or [4, ch. 2.6.].

Proof: The collection is totally bounded if it has for all ε finite covering number, we denote it by $N(\varepsilon,\mathcal{A})$. It is easily seen that

$$N(\varepsilon + r_\varepsilon, \mathcal{A}) \le N(\varepsilon, \mathcal{A}\backslash B_\varepsilon)$$

so the implication follows. △

There is a possibility to show the converse half of this theorem if we take $\mathcal{A} := \{\bar{A}, A \in \mathcal{A}\}$. There is no change in distribution of suprema if we replace A by $\bar{A}$ since assumption of continuous trajectories of $X_n(t)$, recall that $\rho^{\mathcal{A}}(A,\bar{A}) = 0$.

In conclusion of this paper we remark that for arbitrary processes with bounded continuous trajectories (we suppose convergence in distribution on space with L_2-metric) there is very difficult to obtain some good results for

convergence of their suprema as process. We need convergence in distribution of processes $Y_n(A)$ for all $A \in \mathcal{A}$, asymptotical uniform equicontinuity in probability of original processes on $\cup\mathcal{A}$ and total boundness of collection $\mathcal{A}$ under some semimetric, the best result we get for Hausdorff semimetric.

References

[1] Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley, New York.

[2] Husová, J.: Slabá konvergence suprema náhodných procesů. *Sborník prací 13. letní školy JČMF ROBUST 2004.* JČMF, Praha, 2004, p. 133 - 139.

[3] Pollard, D. (1984): *Convergence of Stochastic Processes*. Springer, New York.

[4] van der Vaart, A., W., Wellner, J., A. (1996): *Weak Convergence and Empirical Processes*. Springer-Verlag, New York.

New Criteria for Stochastic DEA

Petr Chovanec[1]

Charles University Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, 186 75 Prague, Czech Republic
e-mail:chovanec@matfyz.cz

Abstract

By its nature, Data Envelopment Analysis (DEA) leaves no room for uncertainty in data such as measurement errors. To improve this fact, we consider α-stochastic efficiency concept, and we relate this problem to the stochastic programming problem. Probability inequalities are employed for introducing new criteria, and two special cases for normal and for general distribution are discussed. The strengths of new criteria are illustrated with a numerical example.

Keywords

DEA, Stochastic DEA, stochastic programming, technical efficiency

1 Introduction

Data Envelopment Analysis (DEA) is the methodology for evaluating the relative productive efficiency of decision making units (DMUs) that produce multiple-outputs using multiple-inputs. DEA was proposed for the first time in 1957 by Farrell; nonetheless, the wide usage of this method begun with its generalization and the linear programming formulation which is due to Charnes et al. (for more information on the DEA history see, e.g. [2]). The main advantages of this approach is that DEA is completely non-parametric, i.e. no explicit form of production function is necessary. Therefore, there are a host of applications in such a distant areas beginning with banking industry, through public economics to agricultural economics. However, the DEA approach does not count for any uncertainty or noise in data, and this has been often cause for criticism.

There were several attempts to improve this disadvantage. In 1993, Land et al. ([3]) have taken the Banker et al. model ([1]) and transform it in a such way that every single constraint must hold only with probability $1 - \alpha$. This approach have used separated chance constraints, on the opposite, the concept of Huang and Li ([5]) uses the joint probability of all constraints together, and this probability should exceed $1 - \alpha$.

In the following section we would introduce DEA more deeply, after that we show stochastic extensions, and we sum up previous attacks. Later we state theorem and with its help we introduce new criteria for efficiency in stochastic DEA. At the end, we demonstrate the abilities of our criteria on a numerical example.

[1]This work was supported by the Grant Agency of the Czech Republic under grants 402/03/H057, 402/04/1294 and 402/05/0115.

2 Data Envelopment Analysis

We assume that there exists multifunction $F : \mathbb{R}^n_{0,+} \rightrightarrows \mathbb{R}^m_{0,+}$ such that

- $F(\mathbf{0}) = \mathbf{0}$
- gph $F := \{(\mathbf{x}, \mathbf{y}) | \mathbf{y} \in F(\mathbf{x})\}$ is closed and convex
- $\mathbf{x}_1 < \mathbf{x}_2 \Rightarrow F(\mathbf{x}_1) \subsetneqq F(\mathbf{x}_2)$.

where the inequality symbol "$<$" denotes non-strict inequality in every index and strict in at least one . We will use this notation in this paper.

Definition 1 *We call $n+m$-tuple $(\mathbf{x}, \mathbf{y}) \in$ gph F efficient if there exists no $(\mathbf{x}^*, \mathbf{y}^*) \in$ gph F such that $\mathbf{x}^* \leq \mathbf{x}$, $\mathbf{y}^* \geq \mathbf{y}$ with at least one strict inequality.*

Unfortunately, the multifunction F is usualy unknown and there are only finitely many observations of gph F; hence, we have to define relative concepts.

Definition 2 *$n+m$-tuple $(\mathbf{x}, \mathbf{y})$ is said to be relatively efficient based on available evidence if the other $n+m$-tuples do not show that there exists $(\mathbf{x}^*, \mathbf{y}^*)$ for which $\mathbf{x}^* \leq \mathbf{x}$, $\mathbf{y}^* \geq \mathbf{y}$ with at least one strict inequality holds.*

From now on, we assume that we have only finite number of observations. We would also call gph F the production possibility set, $\mathbf{x}$ inputs, $\mathbf{y}$ outputs and $n+m$-tuple $(\mathbf{x}, \mathbf{y})$ DMU[2].

In 1984, Banker et al. ([1]) suggested the following linear program to determine relative efficiency of k-th unit among J units[3]

$$\max_{\theta, \boldsymbol{\lambda}} \ \theta \tag{1}$$

$$\begin{aligned}
\text{subject to} \quad & \sum_{j=1}^{J} \lambda_j y_{ij} \geq \theta y_{ik}, && \forall i = 1, \ldots, m \\
& \sum_{j=1}^{J} \lambda_j x_{ij} \leq x_{ik}, && \forall i = 1, \ldots, n \\
& \sum_{j=1}^{J} \lambda_j = 1 \qquad \lambda_j \geq 0, && \forall j = 1, \ldots, J,
\end{aligned}$$

where x_{ij} is i-th input of j-th DMU and analogously y_{ij} is i-th output of j-th DMU. As one of the points is used the origin $(\mathbf{0}, \mathbf{0})$ to ensure convexity properties of the graph.

They have also shown that k-th DMU is efficient iff the optimal value of (1) is equal to one. Hence, the question of efficiency of given unit can be collapsed to finding such $\boldsymbol{\lambda}$ that (1) has feasible solution with additional constraint $\theta \leq 1$.

3 Stochastic DEA

As we can see, standard DEA model is not well-suited to deal with randomness; one observation with random term could cause relative inefficiency of efficient units. First attempt to improve

[2]In other words, we have DMUs which produce output $\mathbf{y}$ from input $\mathbf{x}$. If there exist such DMU that produce from $\mathbf{x}$ output $\mathbf{y}$, then $(\mathbf{x}, \mathbf{y})$ is in production possibility set. Under our assumptions, this production possibility set possesses free shutdown, free disposal, convexity and monotocity properties.

[3]In this case, we assume variable returns to scale and output-oriented model, but there exist a lot of DEA models with respect to both returns to scale and orientation. For another types of DEA model, see, e.g. [2].

this fact was done by Land et al. ([3]). They have replaced the first two constraints in the model (1) with its probabilistic analogies:

$$\mathrm{P}\left(\sum_{j=1}^{J}\lambda_j y_{ij} \geq \theta y_{ik}\right) \geq 1-\alpha, \qquad \forall i = 1,\ldots,m$$

$$\mathrm{P}\left(\sum_{j=1}^{J}\lambda_j x_{ij} \leq x_{ik}\right) \geq 1-\alpha, \qquad \forall i = 1,\ldots,n$$

Although this model counts for randomness, it still does not count for correlations. Huang and Li ([5]) have utilized this fact and they have proposed the following concept.

Definition 3 *([5], definition 2) DMU_k is called α-stochastically efficient if for every $\boldsymbol{\lambda}$ satisfying $\sum_{i=1}^{J}\lambda_i = 1$ and $\boldsymbol{\lambda} \geq 0$*

$$\mathrm{P}\left(\begin{array}{l}\sum_{j=1}^{J}\lambda_j y_{ij} \geq y_{ik}, \forall i = 1,\ldots,m \\ \sum_{j=1}^{J}\lambda_j x_{ij} \leq x_{ik}, \forall i = 1,\ldots,n\end{array}\right) \leq \alpha \tag{2}$$

holds with at least one strict inequality inside the probability.

In addition to the Land's model, this concept solves the problem of correlation between inputs and outputs. On the other hand, it brings many theoretical problems. In many cases, we are not able to compute it exactly, even there are cases when this probability as a function of $\boldsymbol{\lambda}$ is not convex. These disadvantages make the usage of this concept complicated. But if we look on this concept more precisely, we can see that this efficiency can be related to the stochastic programming with joint probability constraint and random technology matrix (for more information see, e.g. [6]). This efficiency is only the question of non-emptyness of the set induced by $\sum_{i=1}^{J}\lambda_i = 1$, $\boldsymbol{\lambda} \geq 0$, and the opposite inequality in (2). Because there are only few results in this area, we would not address this question directly, but only indirectly through probability inequalities in the section 4.

In [5], the authors have suggested sufficient criteria for DMU to be α-stochastically inefficient. The following theorem is only the restated Theorem 1 from [5].

Theorem 1 *If the optimal value of (3) is greater than zero, then DMU_k is not α-stochastically efficient.*

$$\max_{\boldsymbol{\lambda}} \mathrm{P}\left(\sum_{i=1}^{m}(y_{ik} - \sum_{j=1}^{J}\lambda_j y_{ij}) + \sum_{i=1}^{n}(\sum_{j=1}^{J}\lambda_j x_{ij} - x_{ik}) < 0\right) - \alpha \tag{3}$$

$$\begin{aligned} \textit{subject to} \quad & \mathrm{P}(\sum_{j=1}^{J}\lambda_j y_{ij} > y_{ik}) > 1-\varepsilon, && \forall i = 1,\ldots,m \\ & \mathrm{P}(\sum_{j=1}^{J}\lambda_j x_{ij} < x_{ik}) > 1-\varepsilon, && \forall i = 1,\ldots,n \\ & \sum_{j=1}^{J}\lambda_j = 1, && \boldsymbol{\lambda} \geq 0. \end{aligned}$$

where ε they call "non-Archimedean" infinitesimal quantity, for more details see [5].

Proof can be found in [5] and is straightforward; for practical reasons, the authors also suggest ε, such that $\Phi^{-1}(\varepsilon) = 4$, small enough[4]. Although this theorem gives us the nice criterion for inefficiency, it can easily fail.

[4] Φ^{-1} means the quantile of the normal distribution $N(0,1)$.

Example 1 *Suppose that the only random elements are i-th inputs. Then* $\mathrm{P}(\sum_{j=1}^{J}\lambda_j x_{ij} > x_{ik})$ *can be down to* $1-\alpha$ *to* DMU_k *be* α*-stochastically efficient, but the criterion* (3) *would work only if this term would be more than* $1-\varepsilon$*.*

Even though this is the very easy example, it can be extended to much more complicated cases. Moreover, it reveals the main difficulty in this approach – the theorem 1 impose very strict constraints, and, therefore, the set of solutions is relatively small. This disadvantage lead us to introducing new criteria in the following section. To this purpose, we would develop mathematical notation, and introduce mathematical tools necessary to prove this criteria.

4 New criteria

Since (2) can be rewritten as $\mathrm{P}\{\{\sum_{j=1}^{J}\lambda_j y_{1j} \geq y_{1k}\}\cap\cdots\cap\{\sum_{j=1}^{J}\lambda_j y_{mj} \geq y_{mk}\}\cap\{\sum_{j=1}^{J}\lambda_j x_{1j} \leq x_{1k}\}\ldots\{\sum_{j=1}^{J}\lambda_j x_{nj} \leq x_{nk}\}\}$, and because De Morgan rule $\mathrm{P}(A_1\cap\cdots\cap A_n) = 1-\mathrm{P}(\bar{A}_1\cup\cdots\cup\bar{A}_n)$ holds, we can apply certain inequalities on probability of union (e.g. well-known Bonferonni bound). In the following paragraphs we will introduce probability inequalities stated in [6] and we will modify them for our purpose.

Firstly, we would define

$$S_k(\boldsymbol{\lambda}) = \sum_{1\leq i_1<\cdots<i_k\leq n+m} \mathrm{P}(A_{i_1}(\boldsymbol{\lambda})\cap\cdots\cap A_{i_k}(\boldsymbol{\lambda}))$$
$$\bar{S}_k(\boldsymbol{\lambda}) = \sum_{1\leq i_1<\cdots<i_k\leq n+m} \mathrm{P}(\bar{A}_{i_1}(\boldsymbol{\lambda})\cap\cdots\cap \bar{A}_{i_k}(\boldsymbol{\lambda}))$$

where A_{i_j}'s are events and denotes

$$A_{i_k}(\boldsymbol{\lambda}) = \{\sum_{j=1}^{J}\lambda_j y_{i_k j} \geq y_{i_k k}\} \qquad \text{if } i_k \leq m$$
$$A_{i_k}(\boldsymbol{\lambda}) = \{\sum_{j=1}^{J}\lambda_j x_{(i_k-m+1)j} \leq x_{(i_k-m+1)k}\} \qquad \text{otherwise.}$$

and $\bar{A}$ denotes the opposite event. With this notation, we can restate the binomial moment bounds from [6, pp. 315-324]

$$\mathrm{P}(A_1(\boldsymbol{\lambda})\cap\cdots\cap A_{n+m}(\boldsymbol{\lambda})) \geq S_1(\boldsymbol{\lambda}) - (n+m-1) \tag{4}$$
$$\mathrm{P}(A_1(\boldsymbol{\lambda})\cap\cdots\cap A_{n+m}(\boldsymbol{\lambda})) \geq 1 - \bar{S}_1(\boldsymbol{\lambda}) + \frac{2}{n+m}\bar{S}_2(\boldsymbol{\lambda}) \tag{5}$$
$$\mathrm{P}(A_1(\boldsymbol{\lambda})\cap\cdots\cap A_{n+m}(\boldsymbol{\lambda})) \leq 1 - \bar{S}_1(\boldsymbol{\lambda}) + \bar{S}_2(\boldsymbol{\lambda}) \tag{6}$$
$$\mathrm{P}(A_1(\boldsymbol{\lambda})\cap\cdots\cap A_{n+m}(\boldsymbol{\lambda})) \leq 1 - \frac{2}{i(\boldsymbol{\lambda})+1}\bar{S}_1(\boldsymbol{\lambda}) + \frac{2}{i(\boldsymbol{\lambda})(i(\boldsymbol{\lambda})+1)}\bar{S}_2(\boldsymbol{\lambda}) \tag{7}$$

where $i(\boldsymbol{\lambda}) = 1 + \left\lfloor \frac{2\bar{S}_2(\boldsymbol{\lambda})}{\bar{S}_1(\boldsymbol{\lambda})} \right\rfloor$[5].

Because (2) must be smaller than α for every $\boldsymbol{\lambda}$ to DMU_k be α-stochastically efficient, if we find any $\boldsymbol{\lambda}$ such that this do not hold, we know that DMU_k is not efficient. On contrary, if we know that for every feasible $\boldsymbol{\lambda}$ the upper bound is smaller than α then DMU_k is surely efficient. This lead us to formulating the following theorems

[5] In [6], there are also inequalities using higher binomial moments. Also these inequalities can be applied to our case, and theorems 2 and 3 would hold analogously. However, this has only small practical effect because usually only the first two moments are known.

Theorem 2 *If there exists* $\boldsymbol{\lambda}$ *such that RHS of* (4) *or* (5) *exceed* α *then* DMU_k *is not efficient.*

Theorem 3 *If the maximal value of RHS of* (6) *or* (7) *wrt.* $\boldsymbol{\lambda}$ *is smaller than* α *then* DMU_k *is efficient.*

These theorems has been already proved in previous discussion. Theorem 2 seems to be really strong; however, it has a hidden caveat – $\boldsymbol{\lambda}$. The finding of this $\boldsymbol{\lambda}$ could be difficult, because the rhs's of (4) and (5) do not need to be concave functions on convex set. However, we propose two ways of finding it. The first one is suitable only for normally distributed variables, but the second is general.

The following theorem gives us a clue how to search for optimal $\boldsymbol{\lambda}$. The conditions imposed in the problem 8 ensure that the problem (4) is the problem of convex programming.

Theorem 4 *If* $(x_{i1}, \ldots, x_{iJ})^T$, $\forall i = 1, \ldots, n$ *and* $(y_{i1}, \ldots, y_{iJ})^T$, $\forall i = 1, \ldots, m$ *has an* n*-variate normal distribution (can have different ones, could have either non-degenerate or degenerate) then the following program is a well-defined problem of convex programming.*

$$\max_{\boldsymbol{\lambda}} S_1(\boldsymbol{\lambda}) - (n + m - 1) \tag{8}$$

$$\begin{aligned} \textit{subject to} \quad & \mathrm{P}(\sum_{j=1}^{J} \lambda_j x_{ij} \leq x_{ik}) \geq \frac{1}{2} \\ & \mathrm{P}(\sum_{j=1}^{J} \lambda_j y_{ij} \geq y_{ik}) \leq \frac{1}{2} \end{aligned}$$

Proof: In this proof we utilize the idea of Kataoka ([4]). Let us denote by $\bar{x}_{ij}$ the mean of x_{ij} and by C_i the covariance matrix of $(x_{i1}, \ldots, x_{iJ})$; also $\boldsymbol{\lambda}_k = (\lambda_1, \ldots, \lambda_k - 1, \ldots, \lambda_J)^T$. Without loss of generality, we will work with probability p and later we will switch it for $\frac{1}{2}$. Let $\boldsymbol{\lambda}_k^T C_{ik} \boldsymbol{\lambda}_k \neq 0$

$$p \geq \mathrm{P}(\sum_{j=1}^{J} \lambda_j x_{ij} \leq x_{ik}) = \mathrm{P}(\sum_{j=1}^{J} \lambda_j x_{ij} - x_{ik} \leq 0) =$$

$$\mathrm{P}\left(\frac{\sum_{j=1}^{J} \lambda_j x_{ij} - x_{ik} - \sum_{j=1}^{J} \lambda_j \bar{x}_{ij} + \bar{x}_{ik}}{\sqrt{\boldsymbol{\lambda}_k^T C_i \boldsymbol{\lambda}_k}} \leq \frac{-\sum_{j=1}^{J} \lambda_j \bar{x}_{ij} + \bar{x}_{ik}}{\sqrt{\boldsymbol{\lambda}_k^T C_{ik} \boldsymbol{\lambda}_k}} \right) = \Phi\left(\frac{-\sum_{j=1}^{J} \lambda_j \bar{x}_{ij} + \bar{x}_{ik}}{\sqrt{\boldsymbol{\lambda}_k^T C_{ik} \boldsymbol{\lambda}_k}} \right)$$

This is equivalent to

$$\Phi(p) \sqrt{\boldsymbol{\lambda}_k^T C_i \boldsymbol{\lambda}_k} + \sum_{j=1}^{J} \lambda_j \bar{x}_{ij} - \bar{x}_{ik} \geq 0. \tag{9}$$

If $\boldsymbol{\lambda}_k^T C_{ik} \boldsymbol{\lambda}_k = 0$ then $\sum_{j=1}^{J} \lambda_j x_{ij} - x_{ik}$ is almost surely constant; therefore, (9) would also hold.

As C_i is a positive semidefinite matrix, function $\sqrt{\boldsymbol{\lambda}_k^T C_i \boldsymbol{\lambda}_k}$ is convex function. Since $p \geq \frac{1}{2}$, $\Phi(p)$ is positive, and (9) is a concave function. In addition, the hypograph of concave function is convex. Therefore $\mathrm{P}(\sum_{j=1}^{J} \lambda_j x_{ij} \leq x_{ik})$ is a concave function on the convex set induced by constraints.

This idea can be applied to $\mathrm{P}(\sum_{j=1}^{J} \lambda_j y_{ij} \geq y_{ik}) \leq \frac{1}{2}$ too, but with minor changes. Because we have $1 - \Phi(\cdot)$ instead of $\Phi(\cdot)$, it is necessary to change the inequality in the constraints of (8) to get the concave function and the convex induced set. Thus, the problem (8) is a well-defined problem of convex programming. □

Back to our original problem, theorem 2 works for any $\boldsymbol{\lambda}$. Hence, we can also use the $\boldsymbol{\lambda}$ that was the solution of standard DEA model without any stochasticity (e.g. (1)). The main advantage of this approach is that we do not need any assumption on distribution. Although this approach could seem to be quite strange, as we will see in the numerical example, it works.

5 Numerical example

For illustration of criteria introduced in this article, we have used one-input one-output production function with three DMUs all with additive random term. First DMU_1 would be $(1,1) + (0, \{\frac{-1}{3}, 0, \frac{1}{3}\})$ with equal probabilities, DMU_2 is $(2, \frac{4}{3})$ $+(0, \{\frac{-1}{3}, 0, \frac{1}{3}\})+(\{\frac{-1}{3}, 0, \frac{1}{3}\}, 0)$ both with equal probabilities; therefore, nine possibilities all with probability $\frac{1}{9}$. DMU_3 is $(1.5, .8) + (0, \{-.2, 0, .2\})$ also with equal probabilities. The DMU_0 is artificial point $(0,0)$. We can easily see that DMU_3 is not α-stochastically efficient with $\alpha = 5\%$. We suppose that we know the distribution random terms, but we do not know the means. Our observations, therefore, are our estimators of mean. We have made five sets of observations from random generator[6] and in the following text we will analyze efficiency of DMU_3 with both Huang-Li and our criteria.

First part of Huang-Li criteria, the existence of such $\boldsymbol{\lambda}$ in our case is $\lambda_0 + \lambda_1 + \lambda_2 = 1$ (the positive weights corresponding to respective DMUs) and

$$\lambda_1 + (\tilde{x}_2 + \frac{1}{3})\lambda_2 < 1.5; \quad \lambda_1(\tilde{y}_1 - \frac{1}{3}) + \lambda_2(\tilde{y}_2 - \frac{1}{3}) > \tilde{y}_3.$$

By $\tilde{x}_i$ and $\tilde{y}_i$ we denote observations of DMU_i's input or output respectively.

draw	$\tilde{x}_1$	$\tilde{y}_1$	$\tilde{x}_2$	$\tilde{y}_2$	$\tilde{x}_3$	$\tilde{y}_3$	HL criteria	New criteria
1	1	1	2	$\frac{4}{3}$	1.5	3	fail	(4),DEA
2	1	$\frac{2}{3}$	2	$\frac{4}{3}$	1.5	3	fail	(4),DEA
3	1	$\frac{2}{3}$	$\frac{5}{3}$	$\frac{4}{3}$	1.5	2	ineff.	(4),DEA
4	1	$\frac{2}{3}$	$\frac{7}{3}$	$\frac{5}{3}$	1.5	1	ineff.	(4),DEA
5	1	$\frac{2}{3}$	2	$\frac{5}{3}$	1.5	3	fail	(4),DEA

As we can see, Huang-Li criterion has find inefficiency only in two cases, in the others it has failed. On contrary, our criterion based on (4) with weights given by the standard DEA has succeeded in all cases.

6 Conclusion

DEA is widely used in economics for its advantages; in this work, we have introduced its stochastic extensions. We have summarized previous attacks, and we have paid special attention to α-stochastic efficiency concept. We have analyzed this concept and found the connection with stochastic programming with joint probability constraint with random technology matrix. Unfortunately, there are few results in this part of stochastic programming, but there exists approximating techniques. Moreover, we have found weaknesses in the criterion by Huang and Li ([5], sufficient condition for DMU to be inefficient), and bearing both these weaknesses and approximating techniques in mind, we have introduced new criteria for DMU to be inefficient. As we can see in the numerical example, our new criteria can be applied also in cases where Huang-Li criterion fails.

We hope that in the future, criteria like Huang-Li and ours can be applied for constructing the technical efficiency measure in stochastic environment, and we aim our research in this area.

[6]Full results, the GAMS code, data and random numbers can be downloaded from http://chovanec.matfyz.cz/data/mme05.txt

References

[1] Banker, R. D., Charnes, A. and Cooper W. W.: *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis.* Management Science, 1984, vol. 30, pp. 1078-1092.

[2] Cooper, W.W., Seiford, L.M. and Zhu, J.: *Handbook on Data Envelopment Analysis.* Springer, Boston, 2004.

[3] Land, K.C., Lovell, C.A.K, and Thore, S.: *Chance-constrained Data Envelopment Analysis*, Managerial and Decision Economics, 1993, vol. 14, pp. 541–554.

[4] Kataoka, S.: *A Stochastic Programming Model*, Econometrica, 1963, pp. 181–196.

[5] Huang, Z. and Li S. X.: *Stochastic DEA Models with Different Types of Input-Output Disturbances.* Journal of Productivity Analysis, 2001, vol. 15, no. 2, pp. 95–113.

[6] Prékopa, A.: *Probabilistic Programming*, In: Handbooks in Operations Research and Management Science 10, Stochastic Programming (Ruszczyński, A., Shapiro, A., eds.). Elsevier, Amsterdam, 2003.

Exchange Rate and Monetary Policy in Slovak Economy

Ivaničová Zlatica, Chocholatá Michaela

University of Economics, Faculty of Economic Informatics
Department of Operations Research and Econometrics
Dolnozemská cesta 1/b, Bratislava

Abstract

During the last four years the Slovak crown has surprisingly appreciated against the Euro (11,8 %). On the basis of analytical approaches like the international parities conditions it is possible to say, that these conditions don't hold. The current inflow of the capital to the Slovak economy influences strongly the balance of payments and the Slovak currency is being appreciated. Exchange rate policy of the National bank of Slovakia through interventions influences the development of the Slovak crown only for a very short period.
Our presentation is focused on the analysis of the exchange rate market using the idea of the Mundell-Fleming model which was modified for the Slovak economy for the years 1999-2004.

Keywords

Mundell-Fleming model, exchange rate, foreign exchange reserves, BRIBOR, EURIBOR, balance of payments, current account, base money

1. Introduction

The gradual appreciation of the Slovak crown in spite of many interventions of the National Bank of Slovakia (NBS) oriented against this development causes the discussion focused on the rightfulness of this development. If we don't want to admit the strong volatility of exchange rates and if we want to search economic rules that influence the development of exchange rate, we have to extend the analysis in a following way:

- analysis of the relationship between the equilibrium in the money (M2) and foreign exchange market, and
- especially the analysis of the relationship between exchange rate and current account, balance of payments, BRIBOR 3M and EURIBOR 3M and foreign exchange reserves.

2. Modification of the selected equations of Mundell-Fleming model

In the analysis we used the basic theoretical premises of the Mundell-Fleming model, namely two equations were modified for our research.

- Modification of the equilibrium condition on the money market goes from following equation:

$$Md(Y,i) = Ms(m,Z),$$

where Y gross national product,
i interest rate,
Md money demand,
Ms money supply,
m money multiplier,
Z volume of base money.

More detailed characteristic of the above mentioned equation is as follows: the base money created through domestic credit (DU) granting by the National Bank of Slovakia and through purchase of foreign exchange by the NBS is:

$$Z = DU + DR = (DU + DR)_{-1} + \Delta DU + \Delta DR$$

where $(DU + DR)_{-1}$ volume of base money in previous period,
ΔDU change in sources of domestic credit,
ΔDR change of foreign exchange reserves.

The change of foreign exchange reserves will be reflected in surplus or deficit of the balance of payments in non-banking sector, i.e. $\Delta DR = X$. Market transactions based on sale or purchase of domestic debt instruments by NBS is used as a tool of monetary policy. Let the change of market transactions is defined as $\Delta DU = U$. If it holds that $Ms = m(DU + DR)_{-1} = mZ$, than considering the above mentioned relations it is possible to form the equation for defining the equilibrium on the money market:

$$Md(Y,i) = m(Z_{-1} + U + X). \qquad (1)$$

The model equation characterizing the open economy in Mundell-Fleming model is defined through the balance of payments. This equation includes international transactions connected with the export and import of goods and services, inflow and outflow of the capital, purchase and sale of the foreign currency. The following equation (2) can be also defined as the equation characterising the equilibrium on the foreign exchange market:

$$X = In(e) - Pt(Y,e) + K(Y,i), \qquad (2)$$

where X surplus or deficit of the international balance of payments,
K value of the net non-banking capital flow,
In a *Pt* receipts and payments on the current account of the balance of payments .
If it holds that *X=0*, than the equation (2) defines the equilibrium on the foreign exchange market.

3. Analysis of the SKK/EUR exchange rate

On the basis of the equation (1) we estimated the relation between M2, M2 in previous period, U and X. The analysis was done for period 1999Q1 – 2004Q4. The best model chosen from 10 attempts is following:

$$M2_t = -31{,}7466 + 1{,}147295 M2_{t-1} - 4{,}55 X_t + 0{,}829692 U_t$$

(0,0592) (0,0000) (0,0002) (0,0056) $R^2 = 0{,}986$; DW = 1,75

where the significance levels of the estimated parameters are in parenthesis.
On the present, the sharp inflow of the capital to Slovakia influences the growth of foreign exchange reserves. In the relations of the base money, foreign exchange reserves grow and the domestic credit granting by the central bank should decrease. Simultaneously it comes to depreciation of the Slovak crown and the result is, the increase of exchange rate. This relation will be apparent by estimation of the equation (2).
With the help of equation (2) we defined the influence of the foreign exchange reserves, current account and balance of payments on the exchange rate. As an other variable we also used the interest rate differential between BRIBOR 3M and EURIBOR 3M.
The 1. estimated equation is as follows

$$e_t = 45{,}7207 - 0{,}2648 X_{t-1} - 0{,}3829 BRER_t + 0{,}0219 BPGDP_t$$

(0,0000) (0,0000) (0,055) (0,078) $R^2 = 0{,}7297$; DW=1,76,

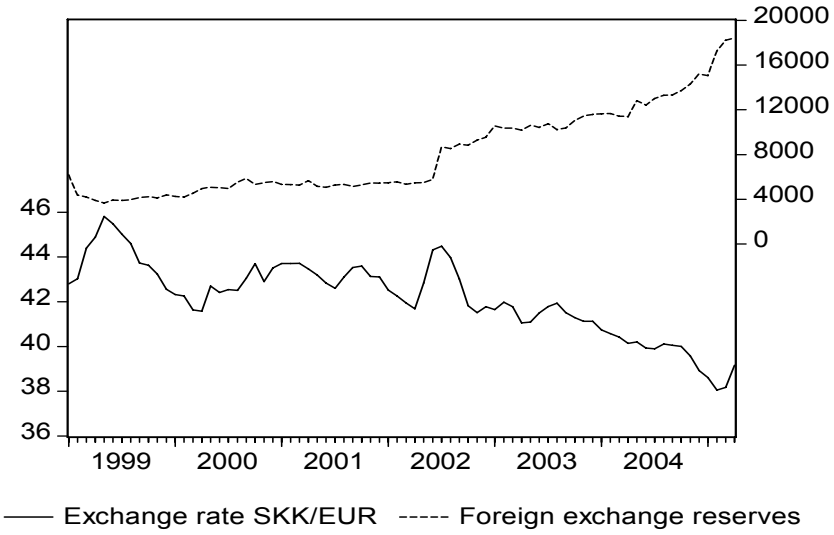
where

$$BRER = BRIBOR\ 3M - EURIBOR\ 3M$$

$$BPGDP = BP / GDP\ ;\ BP - \text{ saldo of total balance of payments, } GDP - \text{ gross domestic product}$$

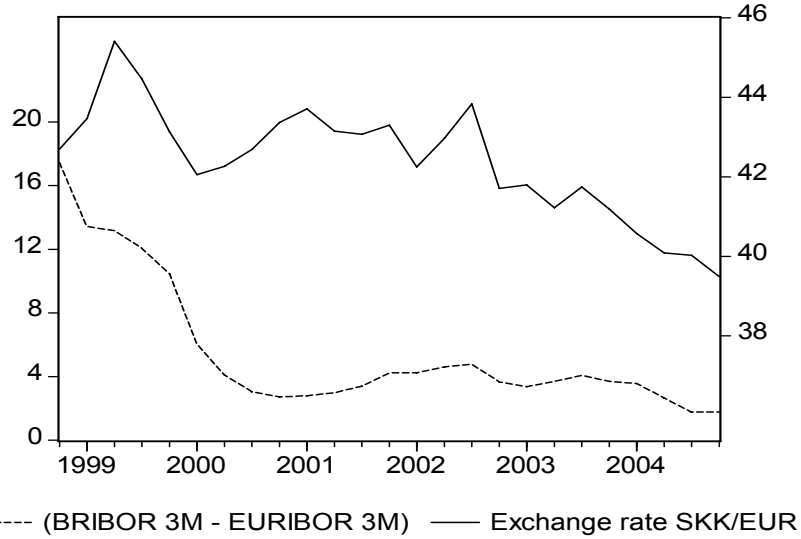
Growth of the foreign exchange reserves pushes down the exchange rate, that means that the Slovak crown appreciates (see Figure 1).

Figure 1



The same relations hold for relations exchange rate – interest rate (the interest rate of EMU countries is embodied) – see Figure 2.

Figure 2



Positive development of the balance of payments has a positive influence on the currency, i.e. on the decline of the exchange rate (see Figure 3).
Because the Slovak economy is very open, we analysed also the influence of the current account on the development of the exchange rate.

The 2. estimated equation is as follows:

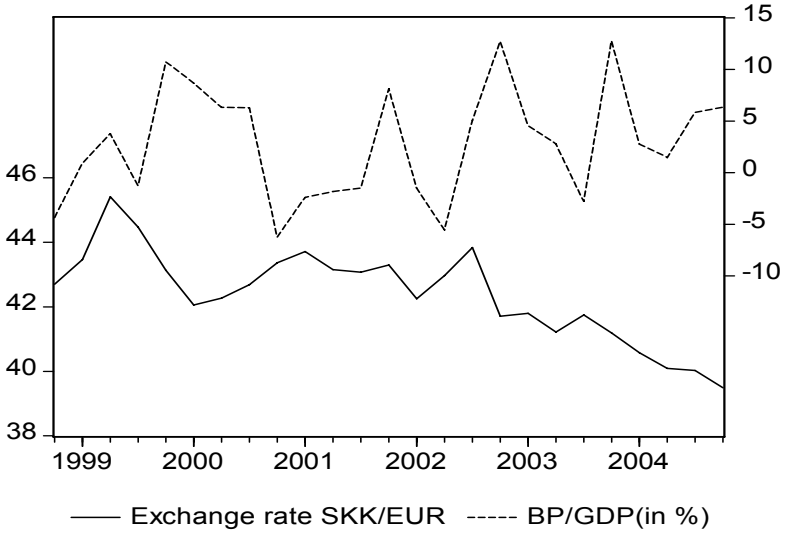
$$e_t = 43{,}64885 - 0{,}21014 X_{t-1} - 0{,}071444 CAGDP_t$$

(0,0000) (0,0008) (0,0577) $R^2 = 0{,}698$; DW=2,04

where

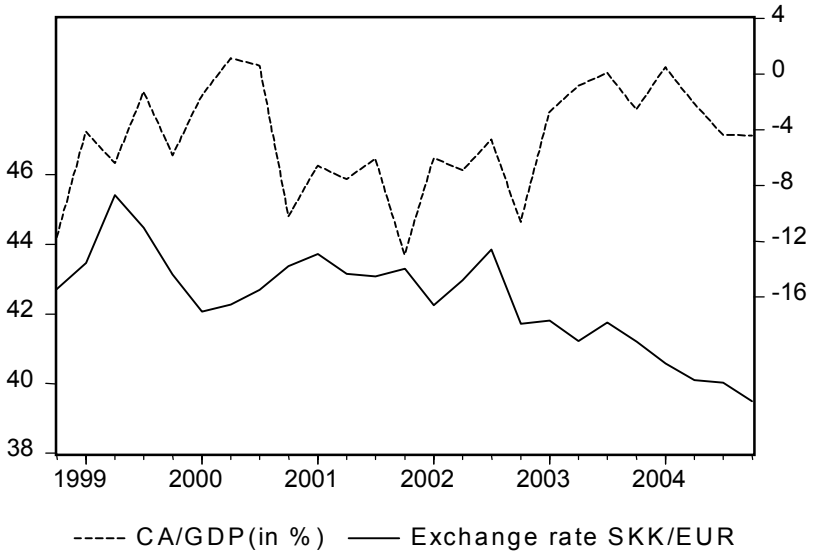
$$CAGDP = CA / GDP; \; CA - current \; account$$

Figure 3



The hypothesis of negative influence of the growth of the foreign exchange reserves is fulfilled; growth of the net export pushes down the exchange rate – see Figure 4.

Figure 4



References

[1] Gandolfo G.: International Economics II, Springer Verlag, Berlin Heidelberg, 1987

[2] Husár J.: Teoretické východiská konštrukcie modelu so zahraničným obchodom pre SR a ČR, Zborník výskumných prác, KOVE, FHI, EU Bratislava, 2003

[3] Husár J.:Makroekonómia, KART, Bratislava 1998

[4] Chocholatá M.: Overenie kurzovej politiky v rámci Marshall – Lernerovej podmienky pre Slovenskú republiku. Zborník z 1. Medzinárodného seminára doktorandov, Praha, 2004

[5] Ivaničová Z.: Co integration of the Exchange Rate and Balance of Trade. Ex post analyse for Slovak and Czech Republic, Journal Economics and Informatics, Faculty of Economic Informatics, University of Economics, Bratislava, 2/2004

[6] Visser H.: A Guide to International Monetary Economics, Exchange Rate Theories, Systems and Policies. Edward Elgar Publishing, Inc. UK, 2000

A MS Excel based support system for data envelopment analysis models[1]

Josef Jablonský

Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67 Praha 3
jablon@vse.cz, http://nb.vse.cz/~jablon

Abstract

Data envelopment analysis (DEA) is a tool for evaluation of efficiency and performance of decision making units. DEA models are based on the definition of efficiency as the ratio of the sum of outputs produced by the unit divided by the sum of inputs spent in the production process. The standard data envelopment analysis models split the units into inefficient and efficient ones without further possibility to classify the efficient units. Super-efficiency DEA models are the extension of the standard DEA models that make it possible to classify the efficient units. The paper informs about an original add-in MS Excel application that offers a simple tool for solving basic DEA models including super-efficiency models.

Keywords

data envelopment analysis, efficiency, super-efficiency, MS Excel

1 Introduction

Data envelopment analysis (DEA) models are used as a tool for evaluation of efficiency, performance or productivity of homogenous decision making units, i.e. units that produce several identical or equivalent effects. These effects can be denoted as the outputs of the decision making units. We consider required or positive outputs of the unit, i.e. their higher values lead (supposing that other characteristics are unchanged) to higher performance of the unit. For producing the outputs the decision making units spend several inputs that are usually minimised, i.e. their lower values lead to higher performance of the unit. Supposing the simplest case in evaluation of performance – one input and one output – the performance of the units can be simply expressed as the ratio

$$\frac{\text{output}}{\text{input}}.$$

In such a case we can receive many different financial characteristics with data that can be taken from financial statements of the evaluated unit. These simple ratio characteristics does not correspond each other. That is why for evaluation of overall efficiency of the decision making unit it is necessary to take into account simultaneously several inputs and outputs.

Let us consider the set of homogenous units $U_1, U_2, \ldots, U_n$ that is described by r outputs and m inputs. Let us denote $\mathbf{X} = \{x_{ij}, i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n\}$ the matrix of inputs and $\mathbf{Y} = \{y_{ij}, \; i = 1, 2, \ldots, r, \; j = 1, 2, \ldots, n\}$ the matrix of outputs. In general, the measure of efficiency of the unit U_q can be expressed as

[1] The research is supported by the Grant Agency of Czech Republic – grant no. 402/03/1360.

$$\frac{\text{weighted sum of inputs}}{\text{weighted sum of inputs}} = \frac{\sum_i u_i y_{iq}}{\sum_j v_j x_{jq}},$$

where v_j, j = 1, 2, ..., m are weights assigned to the j-th input and u_i, i = 1, 2, ..., r are weights of the i-th output.

DEA models suppose that there is a production possibility set consisting of all possible (feasible) combinations of inputs and outputs. Production possibility set is defined by efficient frontier. The units with the combination of inputs and outputs lying on the efficient frontier are marked as efficient because there does not exists a unit with better (lower) inputs and better (higher) outputs simultaneously.

2 Data envelopment analysis models

The evaluation of efficiency of the unit U_q by a DEA model consists in maximisation of its efficiency score under the constraints that the efficiency scores of all other units cannot be greater than 1 (100 %). The weights of all inputs and outputs have to be greater than zero in order to include all the characteristics into the model. Such a model can be formulated as follows:

$$\begin{aligned}
&\text{maximise} && \frac{\sum_i u_i y_{iq}}{\sum_j v_j x_{jq}} \\
&\text{subject to} && \frac{\sum_i u_i y_{ik}}{\sum_j v_j x_{jk}} \le 1, && k = 1,2,...,n, \\
& && u_i \ge \varepsilon, && i = 1,2,\ldots, r, \\
& && v_j \ge \varepsilon, && j = 1,2,\ldots, m.
\end{aligned}$$

This problem can be simply transformed into a standard linear programming problem. Its matrix formulation is as follows:

$$\begin{aligned}
&\text{maximise} && f = \mathbf{u^T y_q} \\
&\text{subject to} && \mathbf{v^T x_q} = 1, \\
& && \mathbf{u^T Y} - \mathbf{v^T X} \le 0, \\
& && \mathbf{u} \ge \varepsilon, \mathbf{v} \ge \varepsilon,
\end{aligned} \qquad (1)$$

where $\mathbf{x_q}$ ($\mathbf{y_q}$) is the vector of inputs (outputs) of the unit U_q. The model (1) if often known as primal CCR (Charnes, Cooper, Rhodes) model. From the computational point of view can be more efficient to work with the dual model:

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

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\boldsymbol{\lambda} \ge 0$, is vector of weights assigned to evaluated units, $\mathbf{s}^+$ a $\mathbf{s}^-$ are vectors of positive and negative slacks in input and output constraints, ε is an infinitesimal constant and θ is a scalar variable expressing the reduction rate of inputs in order to reach the efficient frontier. The model (2) can be explained as follows. In evaluation of efficiency of the unit U_q the model tries to find

a virtual unit described by its inputs $\mathbf{X\lambda}$ and outputs $\mathbf{Y\lambda}$ given by a linear combination of inputs and outputs of other units of the set. The virtual inputs and outputs are better (not worse) than inputs and outputs of the evaluated unit, i.e. it holds $\mathbf{X\lambda} \leq \mathbf{x_q}$ and $\mathbf{Y\lambda} \geq \mathbf{y_q}$. The unit U_q is efficient if the virtual unit is identical with the evaluated unit, i.e. $\mathbf{X\lambda} = \mathbf{x_q}$ a $\mathbf{Y\lambda} = \mathbf{y_q}$. It can be found out by solving the model (2) with the following results:

1. The optimum value of the variable θ equals to 1.
2. The optimum values of all slacks s^+ and s^- equals to zero.

The unit U_q is CCR efficient if the optimum value of the objective function z* of the model (2) equals to 1. Otherwise the unit is not efficient. The optimum value of the objective function z* is denoted as the efficiency score of the evaluated unit U_q. The lower value of the efficiency score indicates lower efficiency of the evaluated unit comparing to other units of the set. The main part of the efficiency score is the variable θ. If the unit is not efficient this variable is usually lower than 1 and it expresses the reduction rate of inputs in order to reach the efficiency frontier.

The models (1) and (2) are input oriented models, i.e. they try to find out how to improve the inputs of the evaluated unit in order to reach the efficiency frontier. Similarly the output oriented model can be formulated as follows:

$$\begin{aligned}
&\text{maximise} && g = \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{i=1}^{r} s_i^+\right), && \\
&\text{subject to} && \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq}, && i = 1, 2, ..., m, \qquad (3)\\
& && \sum_{j=1}^{n} \lambda_j y_{ij} - s_i^+ = \phi y_{iq}, && i = 1, 2, ..., r, \\
& && \lambda_j \geq 0,\ s_i^+ \geq 0,\ s_i^- \geq 0. &&
\end{aligned}$$

The results of the model (3) can be explained analogously to the previous one. The unit U_q is efficient if the optimum value of the objective function g* = 1. The unit is not efficient if the optimum value g* is greater than 1. The variable ϕ expresses the expansion rate of outputs in order to reach the efficient frontier.

The presented models suppose constant returns to scale (CRS) – if a unit with its input and outputs ($\mathbf{x}$, $\mathbf{y}$) is efficient then a unit with inputs and outputs ($\alpha\mathbf{x}$, $\alpha\mathbf{y}$), $\alpha > 0$, is also efficient. The appropriate models with assumption of variable (VRS), non decreasing (NDRS) or non increasing (NIRS) returns to scale can be obtained by adding a new constraint to the models (2) and (3) which limits the sum of the weights λ as presented in the following list:

Input oriented models

$$\begin{aligned}
&\text{min} && z = \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{i=1}^{r} s_i^+\right), \\
&\text{st} && \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq}, \\
& && \sum_{j=1}^{n} \lambda_j y_{ij} - s_i^+ = y_{iq}, \\
& && \lambda_j \geq 0,\ s_i^+ \geq 0,\ s_i^- \geq 0, \\
&\text{CRS} && \textstyle\sum_{j=1}^{n} \lambda_j - \text{free},
\end{aligned}$$

Output oriented models

$$\begin{aligned}
&\text{max} && g = \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{i=1}^{r} s_i^+\right), \\
&\text{st} && \sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{iq} \\
& && \sum_{j=1}^{n} \lambda_j y_{ij} - s_i^+ = \phi y_{iq} \\
& && \lambda_j \geq 0,\ s_i^+ \geq 0,\ s_i^- \geq 0, \\
&\text{CRS} && \textstyle\sum_{j=1}^{n} \lambda_j - \text{free},
\end{aligned}$$

VRS	$\sum_{j=1}^{n} \lambda_j = 1,$	VRS	$\sum_{j=1}^{n} \lambda_j = 1,$
NDRS	$\sum_{j=1}^{n} \lambda_j < 1,$	NDRS	$\sum_{j=1}^{n} \lambda_j < 1,$
NIRS	$\sum_{j=1}^{n} \lambda_j > 1,$	NIRS	$\sum_{j=1}^{n} \lambda_j > 1.$

The DEA models can identify not only the relative efficiency of the decision making units belonging to the set of units but they can specify the target values of inputs $\mathbf{x}'_q$ and outputs $\mathbf{y}'_q$ for inefficient units. These target values identify the virtual units lying on the efficient frontier. They can be computed as follows:

$$\mathbf{x}'_q = \mathbf{X}\boldsymbol{\lambda}^*,$$
$$\mathbf{y}'_q = \mathbf{Y}\boldsymbol{\lambda}^*,$$

where $\boldsymbol{\lambda}^*$ is the vector of optimum values of weights of the above formulated models.

The efficiency score in basic DEA models is limited to unit. Depending on the selection of the DEA model used in the evaluation and on the relation between the number of units on the one hand and the number of inputs and outputs on the other hand, the number of efficient units can be relatively high. That is why several definitions of super-efficiency were formulated in order to classify the efficient units. In the super-efficiency models the efficiency score of inefficient units remains unchanged (lower than 1 in input oriented models) and the efficiency score of efficient units is higher than 1. In this way the model makes it possible to classify the efficient units – it can be one of the very important outputs of the analysis. All the super-efficiency models are based on removing of the evaluated efficient unit from the set of units. This removal changes the efficient frontier. Then the super-efficiency score is measured as a distance between new efficient frontier and the inputs and outputs of the evaluated units. First super-efficiency model was Andersen and Petersen model [1]. This model adds to models (2) and (3) a new constraint $\lambda_q = 0$ (the weight of the evaluated unit is fixed to zero). This constraint can be added to any of the eight models formulated in the above presented list of DEA models. Nevertheless, it is possible to show that the feasible (and optimum) solution of the super-efficiency models exists always just under constant returns to scale assumptions.

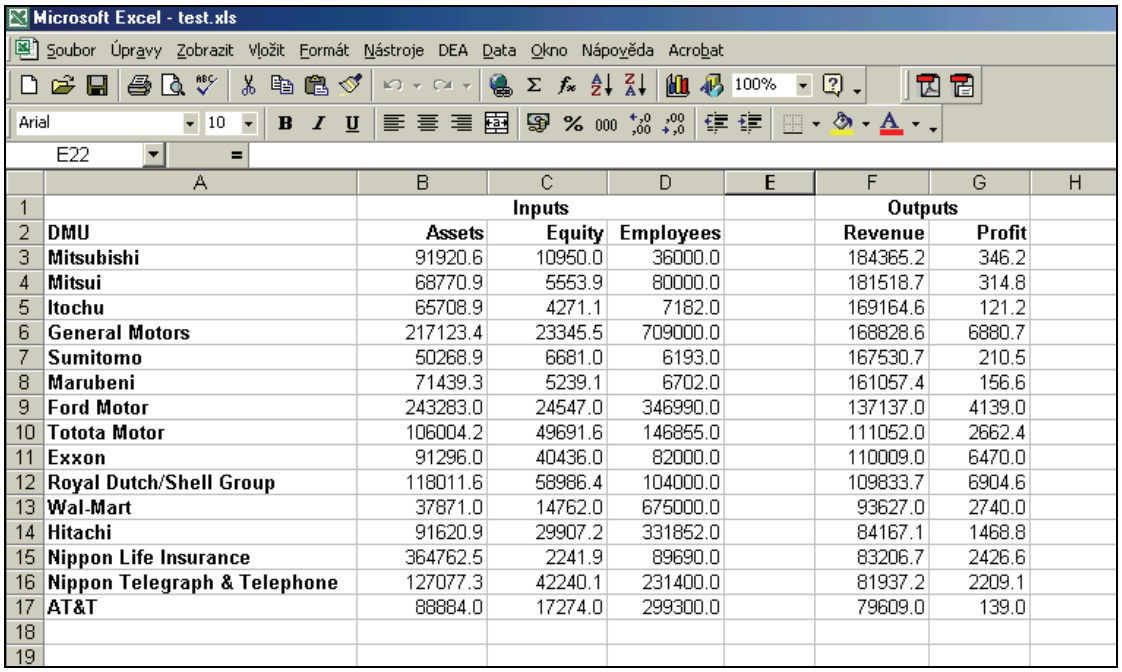
3 Computer implementation

The mathematical formulation of the DEA models shows that they are standard linear programming problems. The efficiency score for any of the decision making unit of the set of units is computed by solving one linear optimisation problem with (n+m+r+1) variables and (m+r) constraints. Even for a higher number of units (n) it is a low-sized LP problem that can be solved without problems by any of the optimisation systems. In order to receive the efficiency score for all the units of the set the optimisation problem of the mentioned size has to be solved n-times. Even though, the problems with approximately one hundred units can be solved by means of current optimisation systems in several seconds.

Several specialised software products for DEA models are available on the current software market. Nevertheless, they are often too expensive and hardly available for decision makers. The DEA models can be solved by any optimisation solvers. The decision makers need a user-friendly environment that can call an optimisation engine, connected with the module for presentation of results. In the past we experimented with modelling system LINGO that contains, except modelling environment, a powerful optimisation solver. Unfortunately the disadvantage is its low availability for frequent users. That is why we decided to build an add-in application in MS Excel environment that works with internal MS Excel solver. In this way the system can be used on any computers with MS Excels spreadsheet, i.e. on almost all computers. In this section we describe in brief how to work with this system.

After the application DEAExcel.xla is once installed the new item „DEA“ is added to the main menu of MS Excel. This new menu contains just two items – *Settings* and *Model selection*. The *Settings* item contains possibility to specify several parameters of the system:

- *Language* – one of the three available language versions of the system,
- *Tolerance* – a constant with initial value 10^{-6} which is used for testing of zero variables values (MS Excel solver returns often values very close to zero instead of zeroes),
- *Title* – a text which is displayed in a heading part of the output of results,
- *Epsilon* – a value of infinitesimal constant – an initial value is 10^{-8},
- *Normalisation of input data* – a switch (on/off) with an initial value “on” which specifies whether the normalisation of input data is done (a transformation to a comparable scale)



	A	B	C	D	E	F	G	H
1		Inputs				Outputs		
2	DMU	Assets	Equity	Employees		Revenue	Profit	
3	Mitsubishi	91920.6	10950.0	36000.0		184365.2	346.2	
4	Mitsui	68770.9	5553.9	80000.0		181518.7	314.8	
5	Itochu	65708.9	4271.1	7182.0		169164.6	121.2	
6	General Motors	217123.4	23345.5	709000.0		168828.6	6880.7	
7	Sumitomo	50268.9	6681.0	6193.0		167530.7	210.5	
8	Marubeni	71439.3	5239.1	6702.0		161057.4	156.6	
9	Ford Motor	243283.0	24547.0	346990.0		137137.0	4139.0	
10	Totota Motor	106004.2	49691.6	146855.0		111052.0	2662.4	
11	Exxon	91296.0	40436.0	82000.0		110009.0	6470.0	
12	Royal Dutch/Shell Group	118011.6	58986.4	104000.0		109833.7	6904.6	
13	Wal-Mart	37871.0	14762.0	675000.0		93627.0	2740.0	
14	Hitachi	91620.9	29907.2	331852.0		84167.1	1468.8	
15	Nippon Life Insurance	364762.5	2241.9	89690.0		83206.7	2426.6	
16	Nippon Telegraph & Telephone	127077.3	42240.1	231400.0		81937.2	2209.1	
17	AT&T	88884.0	17274.0	299300.0		79609.0	139.0	
18								
19								

Figure 1: A spreadsheet with input data set for DEA analysis.

Input data set for DEA analysis can be prepared in any sheet of a MS Excel file. As an illustrative example we use the analysis of efficiency of 15 firms taken from [5]. In the example there are three inputs (total assets, equity and the number of employees) and two outputs (revenues and profit). A possible lay-out of the input data set is presented on Figure 1. After the input data set is prepared and the parameters are set up it is possible to continue by the model selection – Figure 2, where except others the ranges with data sets and the model types are specified.

The dialog window *Model selection* consists of the following items:

- *DMU's labels* – a range with labels of the evaluated units (not obligatory – when it is not specified the system works with default labels DMU1, DMU2,...),
- *Inputs/outputs labels* – ranges with labels of the inputs and outputs (not obligatory – when it is not specified the system works with default labels INP1, INP2,...and OUT1, OUT2...),
- *Matrix of inputs and outputs* – two continuous ranges containing the matrix of inputs and the matrix of outputs – in our example it is the range B3:D17 for the inputs and F3:G17 for the outputs,
- *Model orientation* – one of the two choices – input- or output-oriented model,
- *Returns to scale* – one of the four choices – CRS, VRS, NIRS, NDRS,
- *Super efficiency* – a switch which sets up the selection of the super-efficiency model,

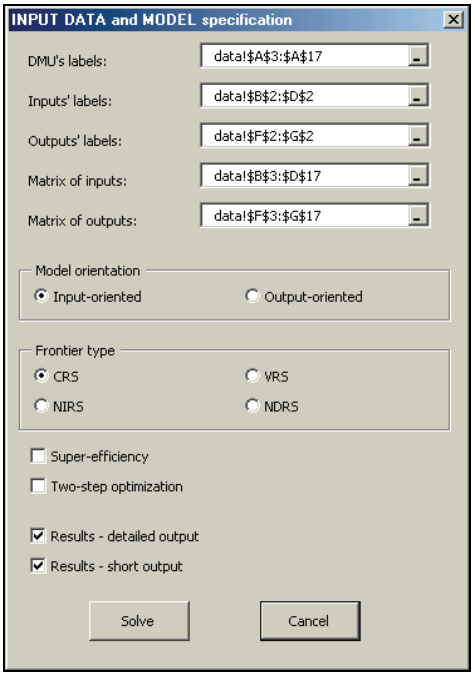


Figure 2: Input data and model selection.

- *Optimisation in two steps* – a switch which specifies whether the optimisation will be realised in one or two steps (the first step is the optimisation of the reduction variable θ or the expansion variable ϕ and the second one is the maximisation of the slack variables s^+ a s^-,
- *Detailed/brief output of results* – two choices which switches on/off a brief and/or detailed output of results – for both the outputs the system generates a single sheet with output information.

The brief output information sheet for our example is presented on Figure 3. This sheet contains first of all the following information:

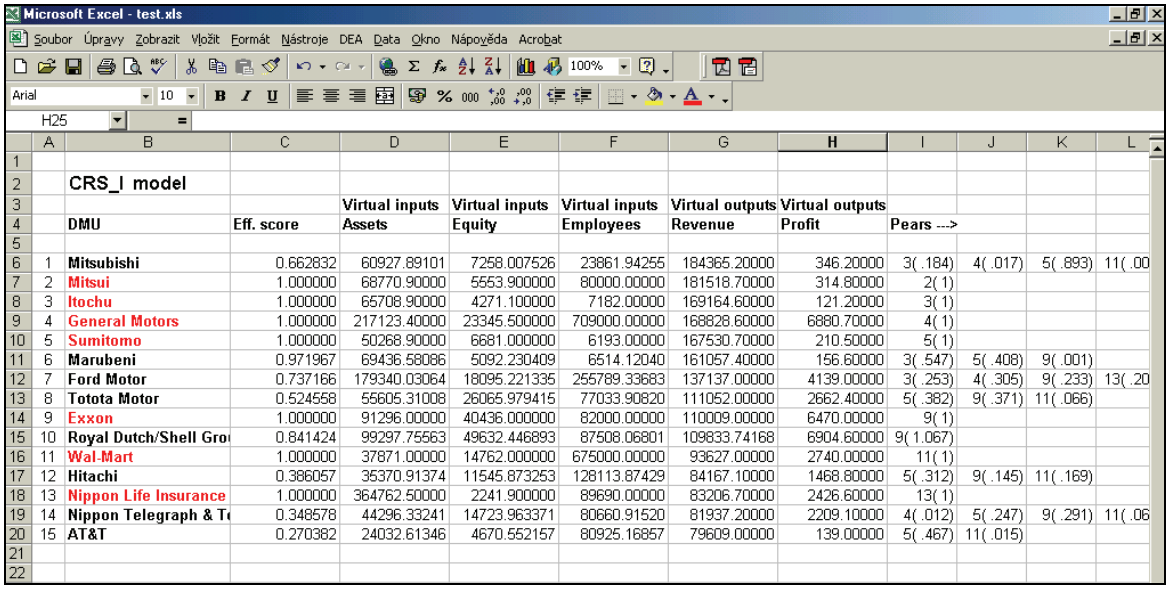
- specification of the DEA model used in the analysis (CRS_I is input oriented model with constant returns to scale)
- DMU labels and the efficiency scores computed by the model (the efficient units are marked by red colour),
- values of virtual inputs and outputs (target values for reaching the efficient frontier),
- non-zero weights of the units (their linear combination of units using these weights gives the virtual inputs and outputs).

Except this information the detailed output sheet contains the optimum values of slack variables and for comparison the original values of inputs and outputs.

4 Conclusions

The availability of powerful software products is one of the assumptions for wider utilisation of any modelling tools. MS Excel offers a user-friendly environment that can be simply extended by specialised modelling applications which can be used by any interested professionals. That is why there are often developed simple but powerful enough MS Excel based applications for different categories of models and such applications are quite popular. Of course there are usually available specialised stand-alone software products which can offer more tools and better user comfort but they are usually too expensive. We tried to develop the MS Excel add-in applications for solving basic DEA models including super-efficiency models that were described in this paper. This application can be downloaded from download section of the web page http://nb.vse.cz/~jablon and tested by any professionals. The application can solve problems up-to 250 decision making units and 20 inputs and

the same number of outputs. This size is sufficient for most of the real-world problems. The author will be grateful for any future comments to this application.



Microsoft Excel - test.xls

Soubor Úpravy Zobrazit Vložit Formát Nástroje DEA Data Okno Nápověda Acrobat

Arial 10 100%

H25 =

CRS_I model

		DMU	Eff. score	Virtual inputs Assets	Virtual inputs Equity	Virtual inputs Employees	Virtual outputs Revenue	Virtual outputs Profit	Pears --->			
6	1	Mitsubishi	0.662832	60927.89101	7258.007526	23861.94255	184365.20000	346.20000	3(.184)	4(.017)	5(.893)	11(.00
7	2	Mitsui	1.000000	68770.90000	5553.900000	80000.00000	181518.70000	314.80000	2(1)			
8	3	Itochu	1.000000	65708.90000	4271.100000	7182.00000	169164.60000	121.20000	3(1)			
9	4	General Motors	1.000000	217123.40000	23345.500000	709000.00000	168828.60000	6880.70000	4(1)			
10	5	Sumitomo	1.000000	50268.90000	6681.000000	6193.00000	167530.70000	210.50000	5(1)			
11	6	Marubeni	0.971967	69436.58086	5092.230409	6514.12040	161057.40000	156.60000	3(.547)	5(.408)	9(.001)	
12	7	Ford Motor	0.737166	179340.03064	18095.221335	255789.33683	137137.00000	4139.00000	3(.253)	4(.305)	9(.233)	13(.20
13	8	Totota Motor	0.524558	55605.31008	26065.979415	77033.90820	111052.00000	2662.40000	5(.382)	9(.371)	11(.066)	
14	9	Exxon	1.000000	91296.00000	40436.000000	82000.00000	110009.00000	6470.00000	9(1)			
15	10	Royal Dutch/Shell Gro	0.841424	99297.75563	49632.446893	87508.06801	109833.74168	6904.60000	9(1.067)			
16	11	Wal-Mart	1.000000	37871.00000	14762.000000	675000.00000	93627.00000	2740.00000	11(1)			
17	12	Hitachi	0.386057	35370.91374	11545.873253	128113.87429	84167.10000	1468.80000	5(.312)	9(.145)	11(.169)	
18	13	Nippon Life Insurance	1.000000	364762.50000	2241.900000	89690.00000	83206.70000	2426.60000	13(1)			
19	14	Nippon Telegraph & T	0.348578	44296.33241	14723.963371	80660.91520	81937.20000	2209.10000	4(.012)	5(.247)	9(.291)	11(.06
20	15	AT&T	0.270382	24032.61346	4670.552157	80925.16857	79609.00000	139.00000	5(.467)	11(.015)		

Figure 3: DEA analysis – the brief output information sheet.

Reference

[1] Andersen,P., Petersen,N.C.: A procedure for ranking efficient units in data envelopment analysis. Management Science 39 (1993), pp.1261-1264.

[2] Cooper, W.W., Seiford, L.M, Tone, K.: Data Envelopment Analysis. Kluwer Publ., Boston 2000.

[3] Jablonský,J.: Modely hodnocení efektivnosti produkčních jednotek (Models for evaluation of efficiency of production units). Politická ekonomie, 2 (2004), pp.206-220.

[4] Jablonský,J., Dlouhý,M.: Modely hodnocení efektivnosti produkčních jednotek (Models for evaluation of efficiency of production units) Professional Publishing, Praha 2004.

[5] Zhu,J. Quantitative Models for Performance Evaluation and Benchmarking. Kluwer Publ., Boston 2003.

Quantile Regression: An Application to the Wages in the Czech Republic

Jana Kalčevová
Department of Econometrics
Faculty of Informatics and Statistics, University of Economics
W. Churchill Sq. 4, 130 67, Prague 3
kalcevov@vse.cz

Abstract

The analysis of wages is focused on the dependency of wages on various factors. For that analysis we usually use wage regression and assume the relationship between logarithm of wages and schooling or experiences. In the work with a large data file the problem occurs – we detect heteroscedasticity often. The estimations of regression coefficients by OLS are unbiased and consistent however they are inefficient. Therefore we use the other methods which assume heteroscedasticity. One of them could be robust regression, the other one could be quantile regression. In this paper quantile regression is introduced and an application on wages in ČR is presented.

Keywords

Analysis of wages, Quantile regression, Heteroscedasticity

1 Introduction

In 1978 Roger Koenker and Gilbert Bassett Jr. published the article Regression quantiles [3] that was one of the first papers discussed this issue. In 1982 the same authors introduced a new class of tests for heteroscedasticity in linear models based on quantile statistics [4] and proved that the new tests are robust, in contrast to classical methods based on least-squares residuals. In this time the quantile regression method is useful diagnostic tool in economic papers for wages analyses because it extracts better information than standard mean regression, especially in the case of heteroscedasticity.

In 2004 Pedro S. Martins and Pedro T. Pereira used this method and published the paper [5] when they showed the impact of education on wages. They used the Mincerian form of human capital earning function and the quantile regression introduced in [3] applied on data from 16 countries. As one can see in their paper the highest returns to education have Portugal (12.6%) and Austria (9.7%) but it is appropriate to note that Portugal has the lowest level of average schooling (6.5, the Czech Republic had 11.7% in 1996). Vice versa the lowest returns to education have Sweden (4.1%) and Norway (6.0%).

2 Quantile regression

Let y_i be a value of dependent variable for individual i and x_i be a vector of factors those may affect variable y. Consider the linear regression model in form:

$$y_i = x_i\beta + u_i \tag{1}$$

where u_i is an error term and β is a vector of regression parameters.

The standard econometric method used to estimate parameters of model (1) is OLS (Ordinary Least Squares), which implies the minimization of

$$\psi_{OLS} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 = \sum_{i=1}^{n}(y_i - x_i'\beta)^2 \tag{2}$$

where $y_i = \beta' x_i + u_i \; \forall i$. This method is based on some assumptions presented for example in [2]. Formulation (2) together with assumptions imply that the fitted curve is the prediction of the mean of dependent variable y_i. When the variance of u_i is non-constant OLS is not robust. In such case we can compute with the White correction for heteroscedasticity (e.g. in [8]) or study an alternative method that minimizes the sum of the absolute residuals rather than the sum of the squared residuals. The Least Absolute Deviation (LAD) method minimizes:

$$\psi_{LAD} = \sum_{i=1}^{n}|y_i - \widehat{y_i}| = \sum_{i=1}^{n}|y_i - x_i'\beta| = \sum_{i=1}^{n}(y_i - x_i'\beta)\ \mathrm{sign}(y_i - x_i'\beta) \tag{3}$$

where $\mathrm{sign}(G) = 1$ for $G \geq 0$ and $\mathrm{sign}(G) = -1$ for $G < 0$. In this case we speak about median regression that is the special case of quantile regression. In the median regression we leave 50% of observations above the median

value and remaining 50% below. Now we would like to fit curves that would leave different percentages of the observations above and below the fitted curve. For this step we can use weights for positive (θ) and negative ($1-\theta$) errors in equation (3).

Assume the conditional quantile function of y_i given x_i:

$$Q_\theta(y_i|x_i) = x_i'\beta_\theta,\ y_i = x_i'\beta_\theta + u_{\theta i} \tag{4}$$

The relation (4) implies $u_{\theta i} = y_i - x_i'\beta_\theta$ from where it follows that $Q_\theta(u_{\theta i}|x_i) = 0$. The θth quantile regression method (QRM) looks for vector β_θ that minimizes:

$$\psi_{QRM,\theta} = \theta \sum_{i|y_i \geq x_i'\beta_\theta} |y_i - x_i'\beta_\theta| + (1-\theta) \sum_{i|y_i < x_i'\beta_\theta} |y_i - x_i'\beta_\theta| \tag{5}$$

Note useful features of QRM. The minimization problem in (5) has a linear programming representation, which makes estimation easy. QRM can be used to characterize the entire conditional distribution of y given x by looking at different values of θ. Quantiles are equivariant to monotone transformations and are robust. Median regression estimators are weakly biased however can be more efficient to mean regression estimators when error term is non-normal. And finally, the flexible structure of the quantile regression model is able to detect some forms of heteroscedasticity. The detailed description of QRM properties is in [1, 4, 10].

3 Analysis of Wages

In this paper data provided by The Czech Statistical Office are processed. Data are from Microcensus 1996 and 2002 and had to be transform before analysis.

Let $\ln E_i$ be the natural logarithm of hourly wages for individual i and X_i be a vector of factors those may affect earnings. The standard model used to earning analysis is based on the human capital earning function developed by Jacob Mincer in 1974 [6]. It has the Mincerian form:

$$\ln E_i = \varphi(X_i) + u_i \tag{6}$$

where u_i is a random independent identically distributed error term that reflects unobserved characteristics. The vector X_i usually includes a measure of schooling or education, a measure of experiences and some other factors such as abilities, gender, race, occupation, region etc.

From [7, 9] we can see that function $\varphi(X_i)$ has usually form:

$$\varphi(X_i) = \beta_0 + \beta_1 S_i + \beta_2 Exp_i + \beta_3 Exp_i^2 \tag{7}$$

where S_i is number of the schooling years, Exp_i is the level of experiences (replaced by potential experiences defined as age minus years of schooling minus 6) and Exp_i^2 is the square of the level of experiences put in equation due to effect of decreasing age-earning profile for a given level of experiences).

The mathematical background applied to the introduced model with wages can be written as:

$$\ln E_i = x_i'\beta_\theta + u_{\theta i},\ Q_\theta(\ln E_i|x_i) = x_i'\beta_\theta \tag{8}$$

where x_i is the vector of exogenous variables and β_θ is the vector of parameters, $Q_\theta(\ln E_i|x_i)$ is the θth conditional quantile of $\ln E_i$ given x_i and the θth regression quantile is defined as the solution to the problem:

$$\begin{aligned} \min_{\beta_\theta \in R^k} \psi_{QRM,\theta} \quad = \quad & \min_{\beta_\theta \in R^k} \{\theta \sum_{i|\ln E_i \geq x_i'\beta_\theta} |\ln E_i - x_i'\beta_\theta| \\ & +(1-\theta) \sum_{i|\ln E_i < x_i'\beta_\theta} |\ln E_i - x_i'\beta_\theta|\} \end{aligned} \tag{9}$$

where $0 < \theta < 1$. The formula (9) should be shortly written as:

$$\min_{\beta_\theta \in R^k} \psi_{QRM,\theta} = \min_{\beta_\theta \in R^k} \sum_i \Omega_\theta(\ln E_i - x_i'\beta_\theta) \tag{10}$$

where $\Omega_\theta(G)$ is the function defined as $\Omega_\theta(G) = \theta G$ for $G \geq 0$ and $\Omega_\theta(G) = (\theta - 1)G$ for $G < 0$. This problem does not have an explicit form but it can be solved by standard linear programming methods[1].

Note that for setting $\theta = 0.5$ in the formula (10) will be received the formula for the median regression.

In the next part we can see the results of QRM for the analyzed data. The quantile regression was applied on whole data file from Microcensus 1996 and also from 2002 restricted on working population. In the model (7) was included gender in addition and the acquired results are in the table 1.

As one can see for both robust regression and quantile regression the return to education was increasing from 1996 till 2002 and return to practice was decreasing. It means education affected level of wage in 2002 more than

[1]This work was supported by the Grant Agency of the Czech Republic – grant no. 402/03/1360.

	1996		2002	
method	RR	QR(0.5)	RR	QR(0.5)
S_i	0.0882276	0.0833524	0.0915482	0.0892246
Std.Err.	0.0011373	0.0012902	0.0023637	0.0022417
Exp_i	0.0239918	0.0239769	0.0124077	0.0142063
Std.Err.	0.0007483	0.0008391	0.0017446	0.0016564
Exp_i^2	-0.0004482	-0.0004475	-0.0002509	-0.0002656
Std.Err.	0.0000176	0.0000189	0.0000383	0.0000356
G_i	0.2532109	0.2637491	0.2307276	0.2574624
Std.Err.	0.0044134	0.0055249	0.0089201	0.009595
const	2.491879	2.53532	2.997754	2.983074
Std.Err.	0.0150954	0.0176438	0.0335744	0.0334769

Table 1: Results for robust regression (RR) and quantile regression (QR)

six years ago. The level of constant is higher in 2002 because the minimal wage (and average wage too) was growing up. From the results we can see that regression parameter for gender decreased from 1996. It means that the impact of gender declines and wages for females lead up to males.

4 Conclussion

In [5] the authors analysed the impact of education on wages in sixteen countries. They used data from various sources such as microcensus come mainly from 1995 and processed information for fully worked men in given countries. In comparison (after analysis based on the same restrictions – men working 35 hours or more per week) the Czech return to education at the first decile was 7.13% in 1996 (and 7.11% in 2002) and at the ninth decile it was 9.79% (and in 2002 11.19% for comparison). By using OLS (or better robust regression) the Czech return to education was about 8.20% (in 2002 8.79%). In Table 2 in [5] the authors presented the corresponding values for the analysed countries. Compared with them the Czech return to education in 1996 has the eighth rank between Germany and Spain. The unweighted mean is 6.5% at the first decile, 9.1% at ninth decile and 7.9% for OLS estimation of return to education. As one can see, the Czech Republic has all values higher than average.

The quantile regression is very relevant tool for estimation of regression coefficients, mainly due to assumption of heteroscedasticity. It is the reason why the quantile regression is a standard part of the major statistical

computing softwares, including R, Splus, Stata and SAS, as well as OLS or robust regression.

References

[1] BUCHINSKY, M.: The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. Journal of Applied Econometrics, Vol. 13, p. 1-30, 1998.

[2] HUŠEK, R.: Ekonometrická analýza. Ekopress, Praha, 1999.

[3] KOENKER, R., BASSETT, G.: Regression Quantiles. Econometrica, Vol. 46, p. 33-50, 1978.

[4] KOENKER, R., BASSETT, G.: Robust Test for Heteroscedasticity Based on Regression Quantiles. Econometrica, Vol. 50, p. 43-61 1982.

[5] MARTINS, P.S., PEREIRA, P.T.: Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries. Labour Economics 11, p.355-371, 2004.

[6] MINCER, J.: Schooling, Experience and Earnings. NBER, New York, 1974.

[7] POLACHEK, S., SIEBERT, S.: The Economics of Earnings. Cambridge University Press, Cambridge 1993.

[8] WHITE, H.: A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. Econometrica, Vol. 48, p. 817-838, 1980.

[9] WILLIS, R.: "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," in Ashenfelter, O. and Layard, R.: Handbook of Labor Economics. North-Holland, Amsterdam 1986.

[10] YU, K., VAN KERM, P., ZHANG, J.: Bayesian Quantile Regression: An Application to the Wage Distribution in 1990s Britain. IRISS Working Paper Series, No. 2004-10, Luxembourgh, 2004.

On Stability of Stochastic Programming Problems with Linear Recourse

Vlasta Kaňková

Department of Econometrics
Institute of Information Theory and Automation, AS ČR
Pod Vodárenskou věží 4, 182 08 Praha 8
kankova @utia.cas.cz

Abstract

[1] Analyzing (generally nonlinear) stochastic programming problems with recourse we can see that this type of the problems is a composition of two (inner and outer) optimization problems. A solution of the outer problem depends on the "underlying" probability measure while a solution of the inner problem depends on the solution of the outer problem and on the random element realization. Consequently (in the case of the optimal solution of the outer problem) the optimal value and the solution set of the inner problem depend also on the "underlying" probability measure. The aim of the contribution is to investigate this dependence.

Keywords

Stochastic programming problems with linear recourse, stability, Lipschitz function, strongly convex function, Wasserstein metric

1 Introduction

Let $b := b(\omega)$, $q := q(\omega)$ be m and n_1–dimensional random vectors; $T := T(\omega), (m \times n)$, $m \leq n$ be a random matrix defined on a probability space $(\Omega, \mathcal{S}, P)$; $W(m \times n_1)$ be a deterministic matrix. We denote by ξ, s–dimensional random vector (generally with $s \leq n + n_1 + mn$) determined by $\xi := \xi(\omega) = [b(\omega), q(\omega), T(\omega)]$ and by F^{ξ}, $P_{F^{\xi}}$ the distribution function and the probability measure corresponding to the random vector ξ. Let, moreover, $g_0(x, z)$ be a function defined on $R^n \times R^s$; $C \subset R^n$ be a nonempty, closed set. Symbols x, $y := y(x, z)$ denote n–dimensional decision vector and n_1–dimensional decision vector function defined on $R^n \times R^s$. (R^n denotes the n–dimensional Euclidean space.)

We consider stochastic programming problem with linear recourse in the form.

Find

$$\varphi(F^{\xi}) = \min_{x \in C} \mathsf{E}_{F^{\xi}} \{ g_0(x, \xi) + \min_{\{y \in R^{n_1} : Wy = b - Tx,\, y \geq 0\}} q^{'} y \}. \tag{1}$$

$q^{'}$ denotes a transposition of the vector q; $\mathsf{E}_{F^{\xi}}$ denotes the operator of the mathematical expectation corresponding to the distribution function F^{ξ}.

Evidently the problem (1) is a composition of the outer and inner problems. The solution $x(F^{\xi})$ of the outer problem depends on the probability measure $P_{F^{\xi}}$, the solution $y(x, \xi)$ of the inner problem depends on the solution x of the outer problem and on the realization of the random element ξ. Consequently the solution of the inner problem depends (in the case of the optimal solution $x(F^{\xi})$ of the outer problem) also on the probability measure $P_{F^{\xi}}$. In details,

[1]This research was supported by the grant agency of the Czech Republic under Grants 402/04/1294, 402/05/0115 and the Grant Agency of AS CR under Grant A 7075202.

if the underlying probability measure P_{F^ξ} is replaced by another P_{G^ξ}, then the solution $x(F^\xi)$ and the solution $x(G^\xi)$ can be different and consequently the solutions of the inner problem $y(x(F^\xi), \xi)$, $y(x(G^\xi), \xi)$ can be also different. The aim of this contribution is to investigate the stability of the inner problem (considered with respect to the probability measure space equipped with a suitable metric). To this end we denote

$$\begin{aligned} Q(x, \xi) &= \min_{\{y \in R^{n_1} : Wy = b - Tx,\, y \geq 0\}} q^{'} y, \\ \mathcal{Y}(x, \xi) &= \{y \in R^{n_1} : Q(x, \xi) = q^{'} y;\, Wy = b - Tx,\, y \geq 0\}, \\ Q_{F^\xi}(x) &= \mathsf{E}_{F^\xi}\{g_0(x, \xi) + \min_{\{y \in R^{n_1} : Wy = b - Tx,\, y \geq 0\}} q^{'} y\}, \\ \mathcal{X}(F^\xi) &= \{x \in C : Q_{F^\xi}(x) = \varphi(F^\xi)\}. \end{aligned} \tag{2}$$

It follows from the relations (2) that $Q(x, \xi)$, $\mathcal{Y}(x, \xi)$ depend on the vector x only through Tx. In particular, the following implication holds for every ξ.

$$x, x^{'} \in C, \quad Tx = Tx^{'} \Rightarrow Q(x, \xi) = Q(x^{'}, \xi) \quad \text{and simultaneously } \mathcal{Y}(x, \xi) = \mathcal{Y}(x^{'}, \xi). \tag{3}$$

2 Motivation

An original motivation for the stochastic models with recourse has been to compensate unfulfilled constraints with random elements in optimization problems with a random factor. This approach has surely more possibilities than the approach by a penalty function. However, it can happen a case in which the inner problem can be relevant to the outer problem. We introduce an example of economic problem leading to the maximization; we employ to it the former symbols.

Example. We consider "underlying" problem of a production planning in which quantity b of raw materials is random. In the case of a deterministic technology matrix T and the cost vector c we obtain (under some additional assumptions) the "underlying" linear programming problem.

Find

$$\max\{c^{'} x |\, Tx \leq b,\ x \geq 0\}.$$

Let us suppose a situation in which the unutilized raw materials can be employed for a next production and, moreover, this second production can be organized by a relative independent manager. Evidently, the aim of the second manager is to maximize profit from this additional production. We can suppose that the additional problem can be considered in the form.

Find

$$\max_{\{y \in R^{n_1} : Wy = b - Tx,\, y \geq 0\}} q^{'} y.$$

Supposing that the main manager has also a profit from the inner problem, his or her decision is determined by the problem.

Find

$$\varphi(F^\xi) = \max_{x \in C} \mathsf{E}_{F^\xi}\{c^{'} x + K^* \max_{\{y \in R^{n_1} : Wy = b - Tx,\, y \geq 0\}} q^{'} y\} \quad \text{for some} \quad K^* > 0. \tag{4}$$

Under the assumptions of a complete information about P_{F^ξ} the decision of the second manager is determined by the problem.

Find

$$\max_{\{y \in R^{n_1} : Wy = b - Tx(F^\xi),\, y \geq 0\}} q^{'} y. \tag{5}$$

Evidently, the decision, the optimal value as well as the other properties of the inner problem depends on the "underlying" probability measure P_{F^ξ}. Consequently, the investigation of the stability of the last problem with respect to the "underlying" probability measure space is a reasonable problem. (In the general model we set $K^* = 1$.)

3 Some Definitions and Auxiliary Assertions

To recall the first assertion we define the Wasserstein metric $d_{W_1}(\cdot, \cdot)$. To this end let $\mathcal{P}(R^s)$ denote the set of all (Borel) probability measures in R^s. If $\mathcal{M}_1(R^s) = \{\nu \in \mathcal{P}(R^s) : \int\limits_{R^s} \|z\|\nu(dz) < \infty\}$ and $\mathcal{D}(\nu, \mu)$ denotes the set of measures in $\mathcal{P}(R^s \times R^s)$ whose marginal measures are ν, μ, then

$$d_{W_1}(\nu, \mu) = \inf\{\int\limits_{R^s \times R^s} \|z - \bar{z}\|\kappa(dz \times d\bar{z}) : \kappa \in \mathcal{D}(\nu, \mu)\}, \quad \nu, \mu \in \mathcal{M}_1(R^s).$$

(For more details see e.g. [5].)

Proposition 1. [2], [8] Let $X \subset R^n$ be a nonempty compact set, $P_{F^\xi} \in \mathcal{M}_1(R^s)$. Let, moreover, $g(x, z)$ be a function defined on $R^n \times R^s$ such that

1. $g(x, z)$ is uniformly continuous on $X \times R^s$,
2. for every $x \in X$, $g(x, z)$ is a Lipschitz function of $z \in R^s$ with the Lipschitz constant L not depending on $x \in X$,

then for every $P_{G^\xi} \in \mathcal{M}_1(R^s)$ the following inequalities are valid

a. $|\mathsf{E}_{F^\xi} g(x, \xi) - \mathsf{E}_{G^\xi} g(x, \xi)| \leq L d_{W_1}(P_{F^\xi}, P_{G^\xi})$ for every $x \in X$,

b. $|\bar{\varphi}(F^\xi) - \bar{\varphi}(G^\xi)| \leq L d_{W_1}(P_{F^\xi}, P_{G^\xi})$, where $\bar{\varphi}(F^\xi) = \inf\{\mathsf{E}_{F^\xi} g(x, \xi) | x \in X\}$.

If, moreover, X is a convex set and, simultaneously, $g(x, z)$ is a strongly convex (with a parameter $\rho > 0$) function on X, then also

$$\|\bar{x}(F^\xi) - \bar{x}(G^\xi)\|^2 \leq \frac{16}{\varrho} L d_{W_1}(P_{F^\xi}, P_{G^\xi}), \quad \text{where} \quad \bar{x}(F^\xi) = \arg\min\{\mathsf{E}_{F^\xi} g(x, \xi) | x \in X\}. \quad (6)$$

(For the definition of strongly convex functions see e.g. [1]).

To the next auxiliary assertions we introduce the following system of the assumptions.

A.1 W is a complete recourse matrix,

A.2 q is a deterministic vector and, moreover, there exists $u \in R^m$ such that $u'W \leq q$,

A.3 $P_{F^\xi} \in \mathcal{M}_1(R^s)$,

A.4 P_{F^ξ} is absolutely continuous with respect to the Lebesgue measure on R^s,

A.5 for any simplicial cone $\mathcal{K} \subset R^m$, the function $P_{F^\xi}[u + \mathcal{K}]$ is locally Lipschitzian on R^m.

For the definition of simplicial cone or equivalently for the definition of a conical hull of m linearly independent vectors in R^m see e.g. [1], for the definition of the complete recourse matrix see e.g. [4] and for the definition of local Lipschitzian function see e.g [7].

Remark. The conditions, under which A.5 are fulfilled, are introduced e.g. in [8].

First we formulate (for us auxiliary) assertions from the linear and stochastic programming.

Lemma 1. If A.1 and A.2 are fulfilled, $C \subset R^n$ is a nonempty, convex, compact set, q, T are deterministic, then

1. for $z \in R^s$, $Q(x, z)$ is a piecewise linear, convex, continuous and Lipschitz function of x on C (consequently also of Tx on TC) with the Lipschitz constant not depending on $z \in R^s$,
2. for every $x \in R^n$, $Q(x, z)$ is a convex and Lipschitz function of $z \in R^s$ with the Lipschitz constant not depending on x (consequently also on Tx),
3. if, moreover, A.3 is fulfilled, $g_0(x, z)$ is a uniformly continuous on $C \times R^s$, Lipschitz function of z with the Lipschitz constant $L > 0$, then for $P_{G^\xi} \in \mathcal{M}_1(R^s)$

   a. $|Q_{F^\xi}(x) - Q_{G^\xi}(x)| \leq L d_{W_1}(P_{F^\xi}, P_{G^\xi})$ for every $x \in C$,
   b. $|\varphi(F^\xi) - \varphi(G^\xi)| \leq L d_{W_1}(P_{F^\xi}, P_{G^\xi})$.

Proof. First, let $\varepsilon > 0$ be arbitrary. It follows from [4], that for every $z \in R^s$ (under the assumptions) $Q(x, z)$ is a piecewise linear and convex real–valued function on R^n. Consequently, $Q(x, z)$ is a bounded function on $C(\varepsilon)$. The continuity and the Lipschitz property of $Q(x, z)$ as a function of x on compact set C follows from the properties of the convex functions (for details see e.g. [6]). The independence of the corresponding Lipschitz constant on z follows from the Linear programming theory. (The symbol $C(\varepsilon)$ denotes ε neighbourhood of the set X.) The assertion 2 follows from the theory of linear programming (for more details see e.g. [4]) and from the theory of convex functions. The assertion 3 follows from the assertion 2 and Proposition 1.

To the next, we denote by W_i, (m, m), $i = 1, \ldots, r$ nonsingular submatrices of W and by q_i a vector of the corresponding components of q. Furthermore, we denote by $\bar{d}_i$, D_i the $n - m$–dimensional vector and matrix $(m, n - m)$ that are given by components and columns not belonging to q_i and W_i, $i = 1, 2, \ldots, r$. We introduce the following assumption.

A.6. There exists $\Omega'' \subset \Omega$, $P_{F^\xi}\{\Omega''\} = 1$ such that if $W_i^{-1}u \geq 0$ (componentwise) $i \in \{1, \ldots, r'\}$ for some $u = b(\omega) - Tx$, $\omega \in \Omega''$, $x \in C$, then $d_i' - q_i' W_i^{-1} D_i \geq 0$ (componentwise) just for one $i \in \{1, \ldots, r'\}$.

Proposition 2. Let A.1, A.2, A.4 and A.6 be fulfilled, $x \in C$, q, T be deterministic. Then there exist $\Omega' \subset \Omega$, $P\{\Omega'\} = 1$ and constants L, L', $d := d(x, \omega) > 0$, $\omega \in \Omega'$ such that

1. for $\omega \in \Omega'$, $\mathcal{Y}(x, \xi(\omega))$ is a singleton; we denote its value by $y(x, \xi(\omega))$,
2. $\|y(x, \xi(\omega)) - y(x', \xi(\omega))\| \leq L\|Tx - Tx'\|$ for every $x' \in C$, $\|x - x'\| \leq d$, $\omega \in \Omega'$,
3. $\|y(x, \xi(\omega)) - y(x', \xi(\omega))\| \leq L'\|x - x'\|$ for every $x' \in C$, $\|x - x'\| \leq d$, $\omega \in \Omega'$.

($\|\cdot\| := \|\cdot\|_n$ denotes the Euclidean norm in R^n.)

Proof. Let $x \in C$ be arbitrary. It follows from the theory of the linear programming that (under A.4, A.6) there exist $\Omega_i' := \Omega_i'(x) \subset \Omega$, $i = 1, \ldots, r'$, $r' \leq r$, $\Omega' = \bigcup_{i=1}^{r} \Omega_i'$, $P\{\Omega'\} = 1$ and constants L, L', $d := d(x, \omega) > 0$ such that

1. the solution set $\mathcal{Y}(x, \xi(\omega))$ is a singleton determined by the relation

   $$y(x, \xi(\omega)) = W_i^{-1}[b(\omega) - Tx] > 0 \quad \text{componentwise,} \quad \omega \in \Omega_i' \quad \text{for some } i \in \{1, 2 \ldots, r'\},$$

2. for $x' \in C$, $\|x - x'\| \leq d$, $\mathcal{Y}(x', \xi(\omega))$ is also a singleton determined by the same submatrix W_i and, moreover,

   $$\|y(x, \xi(\omega)) - y(x', \xi(\omega))\| = \|W_i^{-1}[b(\omega) - Tx] - W_i^{-1}[b(\omega) - Tx']\| \leq L\,\|Tx - Tx'\|.$$

Now, it is easy to see that the assertions of Proposition 2 are valid.

Proposition 3. [8] Let C be a polyhedral set, q, T be deterministic, A.1, A.2 and A.3 be fulfilled. If there exists an n–dimensional deterministic vector c such that $g_0(x, z) = c^{'}x$ independent on $z \in R^s$, then there exists constants $L^* > 0$, $\delta^* > 0$ such that

$$\Delta[\mathcal{X}(F^{\xi}), \mathcal{X}(G^{\xi})] \leq L^*[d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})]^{\frac{1}{2}},$$

whenever $P_{G^{\xi}} \in \mathcal{M}_1(R^s)$, $d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}}) \leq \delta^*$. (The symbol $\Delta[\cdot, \cdot]$ denotes the Hausdorff distance of closed subsets of R^n; for the definition see e.g. [1]).

To recall the next assertion we set v := Tx. Let

$$f_i(v) := P_{F^{\xi}}\{z \in R^s : W_i^{-1}z \geq W_i^{-1}v\}, \quad d_i = -q_i^{'}W_i, \ i = 1, \ldots, r^{'}$$

for W_i that corresponds to the optimal base for some $z \in R^s$.

Definition. [3]. If $f : R^s \to R^{\bar{s}}$ is a locally Lipschitzian function of $v \in R^s$, then generalized directional derivative $\nabla f(v; u)$ is defined as the set consisting of all points $t \in R^{\bar{s}}$ which a limit of the points

$$t^k = (\lambda_k)^{-1}(f(v^k + \lambda_k u) - f(v^k)), \quad \text{where } v^k \to v \quad \text{and } \lambda_k \downarrow 0\,(k \to \infty).$$

Proposition 4. [8] Let q and T be deterministic, A.1, A.2, A.3, A.4 and A.5 be fulfilled, v := Tx. Let $V \subset R^s$ be convex, compact set and

$$\emptyset = \ker D \bigcap (lin\{1\})^{\perp} \bigcap \times_{i=1}^{l} \nabla f_i(v, u) \quad \text{for all } v \in V, \ u \in R^s, \ \|u\| = 1, \tag{7}$$

then $\mathsf{E}_{F^{\xi}} Q(\xi, x)$ is a strongly convex function of $v \in V$.

Cases, when the relation (7) is fulfilled can be found in [8]. ($\ker D$ denotes the kernel of the linear mapping given by the matrix $D(s, r^{'})$ whose columns are the vectors $d_1, \ldots, d_{r'}$, $(lin\{1\})^{\perp}$ is the orthogonal complement of the linear subspace spanned by the vector $\mathbf{1} \in R^r$whose components are identically 1 and $\times_{i=1}^{l} \nabla f_i(v, u)$ is the cartesian product.)

Lemma 2. Let the assumptions of Proposition 4 be fulfilled, C be a convex, compact set. If

1. $g_0(x, z)$ is uniformly continuous on $C \times R^s$ and, moreover, for every $z \in R^s$ a convex function on C,
2. for every $x \in C$, $g_0(x, z)$ is a Lipschitz function of $z \in R^s$ with the Lipschitz constant not depending on $x \in C$,

then there exist constants K, $K^{'} > 0$ such that

a. $\|Tx(F^{\xi}) - Tx\|^2 \leq K|\varphi(F^{\xi}) - Q_{F^{\xi}}(x)|$ for every $x \in C$, $x(F^{\xi}) \in \mathcal{X}(F^{\xi})$,

b. $\Delta[T\mathcal{X}(F^{\xi}), T\mathcal{X}(G^{\xi})]^2 \leq K^{'} d_{W_1}(P_{F^{\xi}}, P_{G^{\xi}})$ for every $P_{G^{\xi}} \in \mathcal{M}_1(R^s)$.

Proof. Let $x^1, x^2 \in C$, $\lambda \in (0, 1)$ be arbitrary, $x(\lambda) = \lambda x^1 + (1 - \lambda)x^2$, $x(F^{\xi}) \in \mathcal{X}(F^{\xi})$. It follows from Lemma 1 and from Proposition 4 that there exists $\rho > 0$ such that

$$\mathsf{E}_{F^{\xi}}\{g_0(x(\lambda), \xi) + Q(\,x(\lambda), \xi)\} \leq$$

$$\lambda \mathsf{E}_{F^{\xi}}[g_0(x^1, \xi) + Q(x^1, \xi)] + (1 - \lambda)\mathsf{E}_{F^{\xi}}[g_0(x^2, \xi) + Q(x^2, \xi)] - \lambda(1 - \lambda)\rho\|Tx^1 - Tx^2\|^2.$$

Setting $x^1 := x$, $x^2 := x(F^\xi)$ in the last inequality we can obtain that

$$\lambda(1-\lambda)\rho\|Tx - Tx(F^\xi)\|^2 \le \lambda[\mathsf{E}_{F^\xi}(g_0(x,\xi)+Q(x,\xi)) - \mathsf{E}_{F^\xi}(g_0(x(F^\xi),\xi)+Q(x(F^\xi),\xi))]+$$
$$[\mathsf{E}_{F^\xi}(g_0(x(F^\xi),\xi)+Q(x(F^\xi),\xi)) - \mathsf{E}_{F^\xi}(g_0(x(\lambda),\xi)+Q(x(\lambda),\xi))].$$

Since $\mathsf{E}_{F^\xi}(g_0(x(F^\xi),\xi)+Q(x(F^\xi),\xi)) - \mathsf{E}_{F^\xi}(g_0(x(\lambda),\xi)+Q(x(\lambda),\xi)) \le 0$, we can see that the assertion a of Lemma 2 follows from Proposition 1.

To prove the assertion b, let $x(F^\xi) \in \mathcal{X}(F^\xi)$ be arbitrary. We obtain from the triangular inequality, Proposition 1, the relation (3) and the assertion a, the existence of a constant $K' = 2K$ such that for $x(G^\xi) \in \mathcal{X}(G^\xi)$

$$\|Tx(F^\xi) - Tx(G^\xi)\|^2 \le K|Q_{F^\xi}(x(F^\xi)) - Q_{F^\xi}(x(G^\xi))| \le$$
$$K\{|Q_{F^\xi}(x(F^\xi)) - Q_{G^\xi}(x(G^\xi))| + |Q_{G^\xi}(x(G^\xi)) - Q_{F^\xi}(x(G^\xi))|\} \le k' d_{W_1}(P_{F^\xi}, P_{G^\xi}).$$

4 Main Results

In this section we introduce a few systems of the assumptions under which the optimal solution and optimal value of the inner problem are "stable" with respect to the underlying probability measure space equipped with the Wasserstein metric.

Theorem 1. Let $x(F^\xi) \in \mathcal{X}(F^\xi)$, q, T be deterministic, C be a nonempty, convex, compact set. If

1. the assumptions of Propositions 2 and 3 are fulfilled, then there exist constants $L_Q^1 > 0$, $d_Q^1 > 0$ such that for $P_{G^\xi} \in \mathcal{M}_1(R^s)$, $d_{W_1}(P_{F^\xi}, P_{G^\xi}) \le d_Q^1$ there exists $x(G^\xi) \in \mathcal{X}(G^\xi)$ fulfilling the relation
$$|\,Q(x(F^\xi),\xi) - Q(x(G^\xi),\xi)\,| \le L_Q^1[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}}, \quad \xi \in R^s, \tag{8}$$
2. the assumptions of Lemma 2 are fulfilled, then there exists a constants $L_Q^2 > 0$ such that
$$|Q(x(F^\xi),\xi) - Q(x(G^\xi),\xi)| \le L_Q^2[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{for every} \quad P_{G^\xi} \in \mathcal{M}_1(R^s), \quad \xi \in R^s, \tag{9}$$
3. A.1, A.2, A.3 and A.4 be fulfilled, $g(x,z) := g_0(x,z)$ fulfils a complete system of the assumptions of Proposition 1, then there exists a constants $L_Q^3 > 0$ such that
$$|Q(x(F^\xi),\xi) - Q(x(G^\xi),\xi)| \le L_Q^3[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{for every} \quad P_{G^\xi} \in \mathcal{M}_1(R^s), \quad \xi \in R^s. \tag{10}$$

Proof. First we consider the case 1. To this end let $x(F^\xi) \in \mathcal{X}(F^\xi)$ be arbitrary. First, it follows from Proposition 3 that there exists $L^* > 0$, $\delta^* > 0$ such that

$$\Delta[\mathcal{X}(F^\xi), \mathcal{X}(G^\xi)] \le L^*[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{whenever} \quad P_{G^\xi} \in \mathcal{M}_1(R^s),\, d_{W_1}(P_{F^\xi}, P_{G^\xi}) \le \delta^*. \tag{11}$$

However since it follows from Lemma 1 that there exists a constant $L' > 0$ such that

$$|Q(x,\xi) - Q(x',\xi)| \le L'\|x - x'\| \quad \text{for every } x,\, x' \in C,\, \xi \in R^s,$$

we can see that there exists $x(G^\xi) \in \mathcal{X}(G^\xi)$ such that

$$|Q(x(F^\xi),\xi) - Q(x(G^\xi),\xi)| \le L'\|x(F^\xi) - x(G^\xi)\| \le L'L^*[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}},$$

whenever $P_{G^\xi} \in \mathcal{M}_1(R^s)$, $d_{W_1}(P_{F^\xi}, P_{G^\xi}) \leq \delta^*$. Consequently setting $L_Q^1 = L^{'} L^*$ we obtain the assertion in the case 1.

To prove the case 2. First, it follows from Lemma 2 that there exists $K^{'}$ such that

$$\Delta[T\mathcal{X}(F^\xi), T\mathcal{X}(G^\xi)] \leq K^{'}[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{whenever} \quad P_{G^\xi} \in \mathcal{M}_1(R^s). \tag{12}$$

However since it follows from Lemma 1 that there exists a constant $L^{'} > 0$ such that

$$|Q(x, \xi) - Q(x^{'}, \xi)| \leq L^{'}\|Tx - Tx^{'}\| \quad \text{for every } x, x^{'} \in C, \ \xi \in R^s$$

we can (according to the relation (3)) see that there exists $x(G^\xi) \in \mathcal{X}(G^\xi)$ such that

$$|Q(x(F^\xi), \xi) - Q(x(G^\xi), \xi)| \leq L^{'}K^{'}[W_1(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{whenever} \quad P_{G^\xi} \in \mathcal{M}_1(R^s).$$

Consequently, setting $L_Q^2 = L^{'}K^{'}$ we obtain the assertion 1.

It remains to consider the case 3. First, it follows from the assumptions, Lemma 1 and from the properties of the convex functions that $g_0(x, z) + Q(x, z)$, $z \in R^s$ is a strongly convex function on C. Consequently, $\mathcal{X}(F^\xi)$, $\mathcal{X}(G^\xi)$ are singletons and according to the assertion of Proposition 1 for $P_{G^\xi} \in \mathcal{M}(R^s)$ there exist constants $L^{'} > 0$ such that

$$\|x(F^\xi) - x(G^\xi)\|^2 \leq L^{'} d_{W_1}(P_{F^\xi}, P_{G^\xi}).$$

Now, surely, we can see that there exists a constant $L_Q^3 > 0$ fulfilling the relation (10).

Corollary. Let $x(F^\xi) \in \mathcal{X}(F^\xi)$, q, T be deterministic, C be a nonempty, convex, compact set. If

1. the assumptions of the case 1 of Theorem 1 is fulfilled, then there exists $\bar{L}_Q^1$, $\bar{d}_Q^1 > 0$, $x(G^\xi) \in \mathcal{X}(G^\xi)$ such that

   $$|\mathsf{E}_{F^\xi} Q(x(F^\xi), \xi) - \mathsf{E}_{F^\xi} Q(x(G^\xi), \xi)| \leq \bar{L}_Q^1[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \text{ whenever } d_{W_1}(P_{F^\xi}, P_{G^\xi}) \leq \bar{d}_Q^1,$$

2. at least one of the system of the assumptions of the cases 2 or 3 of Theorem 1 are fulfilled, then there exists $\bar{L}_Q^1$, $\bar{d}_Q^1 > 0$, $x(G^\xi) \in \mathcal{X}(G^\xi)$ such that

   $$|\mathsf{E}_{F^\xi} Q(x(F^\xi), \xi) - \mathsf{E}_{F^\xi} Q(x(G^\xi), \xi)| \leq \bar{L}_Q^1[d_{W_1}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{for every} \quad P_{G^\xi} \in \mathcal{M}_1(R^s).$$

Proof. The assertion of Corollary 1 follows from the assertion 1 of Theorem 1.

Theorem 2. Let $x(F^\xi) \in \mathcal{X}(F^\xi)$, q, T be deterministic, C be a nonempty, convex, compact set. If at least one of the system of the assumptions of the cases 1, 2, or 3 of Theorem 1 is fulfilled, then there exists $\Omega^{'} \subset \Omega$, $P_{F^\xi}(\Omega^{'}) = 1$ such that for $\omega \in \Omega^{'}$

a. $\mathcal{Y}(x(F^\xi), \xi(\omega))$ is a singleton $y(x(P_{G^\xi}), \xi(\omega)) = y(x(F^\xi), \xi(\omega))$,

b. $y(x(F^\xi), \xi(\omega)) = y(x(P_{F^\xi}), \xi(\omega))$ is in the point $x(F^\xi)$ a continuous function in the space $\mathcal{M}_1(R^s)$ equipped with the Wasserstein metric d_{W_1}.

Proof. First, we consider the case of the assumptions of the case of Theorem 1. It follows from Proposition 2 that there exists $\Omega^{'}$, $P_{F^\xi}(\Omega^{'}) = 1$ such that for $\omega \in \Omega^{'}$ the solution set of the inner problem is a singleton determined by

$$y(x(F^\xi), \xi(\omega)) = W_i^{-1}[b(\omega) - Tx(F^\xi)] \quad \text{for some } i \in \{1, \ldots, r\}. \tag{13}$$

Furthermore, according to the assertion 3 of Proposition 2 there exists $d := d(x(F^\xi), \omega) > 0$ such that for every $x \in C$ fulfilling the relation $\|x(F^\xi) - x\| \leq d$, the solution of the inner problem is also singleton determined by

$$y(x, \xi) = W_i^{-1}[b(\omega) - Tx] \tag{14}$$

and, moreover, there exists a constant L such that

$$\|y(x, \xi(\omega)) - y(x(F^\xi), \xi(\omega))\| \leq L\|x - x(F^\xi)\| \quad \text{for every} \quad x \in C,\ \|x - x(F^\xi)\| \leq d,\ \omega \in \Omega^{'}.$$

Employing the assertion of Proposition 3 we can see that there exist constant L^*, $\delta^* > 0$ and $x(G^\xi) \in \mathcal{X}(G^\xi)$ such that

$$\|x(F^\xi) - x(G^\xi)\| \leq L^*[d_{W_i}(P_{F^\xi}, P_{G^\xi})]^{\frac{1}{2}} \quad \text{whenever} \quad d_{W_1}(P_{F^\xi}, P_{G^\xi}) \leq d^*.$$

Consequently, setting $d^1_{\mathcal{Y}} = \min\{\delta^*, d\}$, $L^1_{\mathcal{Y}} = LL^*$ we have finished the proof.

Since the proofs under the assumptions of the cases 2, 3 is very similar we can them omit.

5 Conclusion

In the contribution we have investigated the stability of the inner problem in (1) considered with respect to the "underlying" probability measure. To this end, we assume that the measure is known to the manager of the inner (see e.g. analysis of the special case leading to the relations (4) and (5). However, it happen surely often that empirical data are known only. Consequently, to study the case of empirical estimates seems to be reasonable also. However the investigation in this direction is over the possibilities of this paper.

References

[1] J-B. Hiriart-Urruty and C. Lemaréchal: Fundamentals of Convex Analysis. Springer, Berlin 2001.

[2] M. Houda: Stability and Estimates in Stochastic Programming (Spesial cases) (in Czech). Diploma Work. Faculty of Mathematics and Physics, Charles University.

[3] B. Kummer: Lischitzian Inverse Functions, Directional Derivates and Applications in $C^{1,1}$ Optimization. J. Optimization Theory and Applications *70* (1991), 561–582.

[4] J. Mayer: Stochastic Linear Programming Algorithms (A Comparison Based on a Model Management System) Gordon and Breach Science Publishers, Amsterdam 1998.

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

[6] R. T Rockafellar: Convex Analysis. Princeton University Press, New Yersey 1970.

[7] R. T Rockafellar and R. J-B. Wets: Variational Analysis: Springer, Berlin 1998.

[8] W. Römisch and R. Schulz: Stability of Solutions for Stochastic Programs with Complete Recourse. Math. Oper. Res. *18* (1993), 590–609.

Optimal Range for the iid Test Based on Integration across the Correlation Integral

Evžen Kočenda and Ľuboš Briatka

CERGE-EI, Prague; CERGE-EI, Charles University and Academy of Sciences of the Czech Republic, P.O.Box 882, Politických vězňů 7, 111 21 Prague, Czech Republic; tel. (420) 224005149, fax (420) 224227143; e-mail: evzen.kocenda@cerge-ei.cz and lubos.briatka@cerge-ei.cz

Abstract

This paper builds on Kočenda (2001) and extends it in three ways. First, new intervals of the proximity parameter ε (over which the correlation integral is calculated) are specified. For these ε-ranges new critical values for various lengths of the data sets are introduced and through Monte Carlo studies it is shown that within new ε-ranges the test is even more powerful than within the original ε-range. The range that maximizes the power of the test is suggested as the optimal range. Second, an extensive comparison with existing results of the controlled competition of Barnett et al. (1997) as well as broad power tests on various nonlinear and chaotic data is provided. Test performance with real (exchange rate) data is provided as well. The results of the comparison strongly favor our robust procedure and confirm the ability of the test in finding nonlinear dependencies as well its function as a specification test. Finally, new user-friendly and fast software is introduced.

Keywords

chaos, nonlinear dynamics, correlation integral, Monte Carlo, single-blind competition, power tests, high-frequency economic and financial data

1 Introduction

Growing interest as well as practical needs of researchers in testing for nonlinearity and chaos in high-frequency economic and financial data has prompted the design of various methods to accomplish the task. Among them, a well known BDS test was devised by Brock, Dechert, Scheinkman and LeBaron (1996) as a non-parametric method of testing for nonlinear patterns in time series. The method is based on the correlation integral described by Grassberger and Procaccia (1983) and is unique in its ability to detect nonlinearities independent of linear dependencies in the data. The null hypothesis is that data in a time series are independently and identically distributed (iid); an alternative is not specified. In order to conduct the BDS test, two free variables (embedding dimension m and proximity parameter ε) must be chosen ex ante, with limited guidance from statistical theory; thus it is likely that inappropriate values may be chosen. Kočenda (2001) has suggested an alternative test which, through integrating across the correlation integral, avoids arbitrary selection of the proximity parameter ε and allows for running the test across an empirically endorsed set of embedding dimensions m.

This paper builds on Kočenda (2001) and increases the operational ability of the alternative test in three ways. First, we improve the choice of the interval of the proximity parameter ε over which the correlation integral is calculated and suggest the optimal range that maximizes the power of the test. We also bring necessary sets of critical values for various lengths of data and provide an analysis of their sensitivity with respect to the choice of proximity parameter range.

Second, we compare the test with existing results of the single-blind controlled competition of Barnett et al. (1997) and in addition perform power tests on various chaotic (nonlinear) and noisy chaotic data. The results strongly attest to our robust procedure. The performance of the test surpasses that of other tests for non-linear dependencies. The improvement allows for new economic insights especially when the procedure is used on financial data as a specification test.

Third, new compact software to run the test, as well as allowing for associated simulations, is introduced as

freeware.

2 Theoretical Backgrounds and Motivation

The test of Kočenda (2001) suggests considering an OLS-estimate of the correlation dimension (defined presently) over a range of ε-values, and is thus closer in spirit to the original correlation dimension than is the BDS test (for full details see the original paper).[1] The test rests upon the concept of the correlation integral, developed by Grassberger and Procaccia (1983). Formally, let $\{x_t\}$ be a scalar time series of the size T generated randomly according to a density function f. Form m-dimensional vectors, called m-histories, $x_t^m = (x_t, x_{t+1}, \ldots, x_{t+m-1})$. The sample correlation integral (or correlation sum) at embedding dimension m is computed as

$$C_{m,T}(\varepsilon) = 2\sum_{t=1}^{T_{m-1}} \sum_{s=t+1}^{T_m} I_\varepsilon(x_t^m, x_s^m)/(T_m(T_m - 1)), \qquad (2.1a)$$

where $T_m = T - m + 1$, and $I_\varepsilon(x_t^m, x_s^m)$ is an indicator function of the event $\left\| x_t^m - x_s^m \right\| = \max_{i=0,1,\ldots m-1} \left| x_{t+i} - x_{s+i} \right| < \varepsilon$. Further, the correlation integral at embedding dimension m is defined as

$$C_m(\varepsilon) = \lim_{T\to\infty} C_{m,T}(\varepsilon), \qquad (2.1b)$$

Thus, the sample correlation integral measures the fraction of pairs that lie within the tolerance distance ε for the particular embedding dimension m. If, in the limit of large T and small ε, the correlation integral scales as ε^D, then the exponent D defines the correlation dimension as

$$D = \lim_{\varepsilon\to 0} \lim_{T\to\infty} \frac{\ln C_{m,T}(\varepsilon)}{\ln \varepsilon}. \qquad (2.2)$$

The alternative statistic uses a number of tolerance distances chosen from a specific range for each particular embedding dimension by calculating the slope of the log of the correlation integral versus the log of the proximity parameter over a broad range of values of the proximity parameter. The estimates of the correlation dimension, slope coefficients β_m, can be estimated as

$$\beta_m = \frac{\sum_{\varepsilon} \left(\ln(\varepsilon) - \overline{\ln(\varepsilon)}\right) \cdot \left(\ln(C_{m,T}(\varepsilon)) - \overline{\ln(C_{m,T}(\varepsilon))}\right)}{\sum_{\varepsilon} \left(\ln(\varepsilon) - \overline{\ln(\varepsilon)}\right)^2}, \qquad (2.3a)$$

which equals to calculating the slope coefficient β_m from the least squares regression

$$\ln\left(C_{m,T}(\varepsilon_i)\right) = \alpha_m + \beta_m \ln(\varepsilon_i) + u_i; \quad i = 1, \ldots, n \qquad (2.3b)$$

where $\ln(\varepsilon)$ is the logarithm of proximity parameter (tolerance distance), $\ln(C_{m,T}(\varepsilon))$ is the logarithm of the sample correlation integral (correlation sum), m is the embedding dimension, and the variables with a bar denote the mean of their counterparts without a bar.

3 Proximity Parameter Range and Critical Values

Kočenda (2001) performed a Monte Carlo study with 10,000 replications of the distribution of the β_m statistic under the null hypothesis of iid data. Critical values were tabulated for data-length of 500, 1000, and 2500 observations allowing for nine embedding dimensions m (2-10). The range of proximity parameter ε, for which the critical values were generated, extends over the specific interval: n = 41 proximity parameters ε ranging over the interval (0.25σ, 1.00σ) in proportionally equal increments (σ being standard deviation of a

[1] It is worthwhile noting that originally an important reason to develop the BDS test was that point estimates of the correlation dimension were very unstable across values of ε.

sample).

The above original interval is chosen sensibly to allow for hidden patterns corresponding to very narrow tolerance distances. However, a single ε-range prevents observing whether and how sensitive the tabulated critical values are to a choice of different ε-range. Further, the originally chosen ε-range does not need to be an optimal range; e.g. the range that, when used, maximizes the power of the test. We elaborate on this issue in the following sections.

The issue of the ε-range choice is complicated by the fact that we cannot theoretically derive a correct range of proximity parameters. This is due to the fact that the behavior of the β_m statistic within an ε-range is closely related to the composition of the analyzed data. Therefore it is possible that one ε-range is more appropriate for some kind of data and a different ε-range for another one. For this reason, we select two additional ranges of proximity parameters to study whether and how the critical values behave.

To repeat, Kočenda (2001) provides critical values for proximity parameter in a range of (0.25σ, 1.00σ); this is our first (original) interval. This range is the most discriminating range of the proximity parameters ε, which means that it takes into account primarily very small tolerance distances among the data in a sample.

The second interval is motivated by the results of Monte Carlo studies performed by Hsieh and LeBaron (1988), who found that the power and size of the BDS test is maximized when proximity parameter ε is chosen between 0.5 and 1.5 of the standard deviation of the sample. For this reason we have chosen to calculate values of correlation integral for (0.50σ, 1.50σ) range of proximity parameters.

The third interval represents the broadest range of sensible proximity parameter values that are used in the empirical literature. Since we concur with Kanzler (1999), who shows that the asymptotic normality of the BDS test depends on the correct choice of proximity parameter, we employ the broad interval of proximity parameters within (0.25σ, 2.00σ) in order to avoid omitting its possible correct values when computing the statistic of the Kočenda test.

4 Performance, Power of the Test and Optimal Range

4.1 Performance in Competition

In order to compare performance of the Kočenda test with other tests for nonlinearity and chaos we have exploited the study of Barnett et al. (1997) who performed a well-known single-blind controlled competition to compare the power of five highly regarded tests for nonlinearity or chaos against various alternatives. The data used in this competition were simulated by five different generating specifications at two sample sizes. The "small" sample size contains 380 and "large" sample size 2000 observations. The samples were generated by the following five models:

1. Fully deterministic, chaotic Feigenbaum recursion (FEIG) in the form of the logistic equation:

$$y_t = 3.57 y_{t-1}\left(1 - y_{t-1}\right), \tag{4.1}$$

where the initial condition was set at $y_0 = 0.7$;

2. A generalized autoregressive conditional heteroscedasticity model (GARCH) of the form:

$$y_t = h_t^{1/2} u_t, \tag{4.2}$$

where h_t is defined by $h_t = 1 + 0.1 y_{t-1}^2 + 0.8 h_{t-1}$, with $h_0 = 1$ and $y_0 = 0$;

3. A nonlinear moving average model (NLMA) of the form:

$$y_t = u_t + 0.8 u_{t-1} u_{t-2}; \tag{4.3}$$

4. An autoregressive conditional heteroscedasticity model (ARCH) of the form:

$$y_t = \left(1 + 0.5 y_{t-1}^2\right)^{1/2} u_t, \tag{4.4}$$

with the value of the initial observation set at $y_0 = 0$;

5. An autoregressive moving average model (ARMA) of the form:

$$y_t = 0.8 y_{t-1} + 0.15 y_{t-2} + u_t + 0.3 u_{t-1}, \tag{4.5}$$

with $y_0 = 1$ and $y_1 = 0.7$.

To summarize: our results show that the test performance is very satisfactory since with a correctly

selected ε-range it performs equally or better than the tests included in the competition performed by Barnett et al. (1997). Further, the time needed to run the test with our new software is negligible.

4.2 Power Tests

To conduct the power tests we pursued the following strategy. For each of the five models described in specifications (4.1-4.5), and used by Barnett et al. (1997), we have generated 1000 samples of data. The data were generated in three sizes of 500, 1000, and 2500 observations. We have used all samples to perform a battery of power tests.

Across all proximity ranges as well as sample sizes the test always accurately rejects the null hypothesis for FEIG and ARMA processes. For the remaining processes the power of the test uniformly improves with the sample size as one would expect. GARCH, ARCH, and NLMA processes pose some challenge to the test at sample size of 500 observations; the power of the test dramatically improves for ARCH and NLMA processes when samples of 1000 and 2500 observations are used. An interesting picture emerges when the power of the test is compared among the three ranges of the proximity parameter. The power is lowest for interval $(0.25\sigma, 1.0\sigma)$ but improves radically for intervals $(0.5\sigma, 1.5\sigma)$ and $(0.25\sigma, 2.0\sigma)$. For these two intervals the power of the test is extremely high, above 90% in most cases even for the small sample size of 500 observations; an exception is the power in the GARCH process. The power of the test is highest for both wide intervals and large sample size (2500 observations) in all five processes; it is near 100% of correct rejections for the GARCH process and precisely 100% for remaining processes.

4.3 Sensitivity Analysis and Optimal Range

4.3.1 Optimal range

In the previous section we have shown which of the chosen ranges performs better than the others. We now analyze whether the power of the test improves while the given range is expanded or contracted. Since the range $(0.50\sigma, 1.50\sigma)$ is a subset of the range $(0.25\sigma, 2.00\sigma)$ and the range $(0.50\sigma, 1.50\sigma)$ came out as the best option for the lower embedding dimensions (core of the applied research), such an approach can be considered as a robustness check of the range.

The sensitivity analysis of the power of the test against nonlinear non-iid processes was pursued with the following strategy. First, we have expanded the range $(0.50\sigma, 1.50\sigma)$ by small amounts (0.10σ) from the lower and upper bound separately. Expansion reached 0.30σ and 2.10σ at the lower and upper bound, respectively. Second, we have expanded the range $(0.50\sigma, 1.50\sigma)$ in a similar fashion from the lower and upper bound at the same time. Third, we have contracted the range $(0.50\sigma, 1.50\sigma)$ from one end separately and later from both ends at the same time as well. Finally, we have contracted the range $(0.50\sigma, 1.50\sigma)$ from one end and expanded it from the opposite end. This effectively meant, shifting the range. The upper limit we tested reached 3.00σ, the lower limit was 0.25σ.

The results of the above exercise can be summarized and generalized in this way: by contracting the range the test was losing its power while by expanding the range the power was increasing; such an increase in power was limited, though. Beyond the range $(0.60\sigma, 1.90\sigma)$ the power ceased to increase or its increase was negligible, depending on a specific embedding dimension. Hence, we consider the range $(0.60\sigma, 1.90\sigma)$ of the distance ε to be the optimal range for conducting the test.

4.3.2Competition and power

Additionally, we have again used the samples of the data that were used for the blind competition of Barnett et al. (1997) and run the Kočenda test with the new optimal range on them. The test rejected the null hypothesis of iid for the whole set of data and rejected the iid hypothesis with higher statistical significance than when performed with the three earlier ranges. The results also show superior performance of the test with respect to other tests for nonlinearity. Further, similarly to the power tests that compare performance under different ranges, we report the results of the power tests that were performed with the optimal range $(0.60\sigma, 1.90\sigma)$; results show superior performance of the test over other ranges compared in Section 4.2.

5 Performance of the Test with Real Data

As a complementary analysis to the performance of the test with chaotic artificial data we have also performed analysis with real data. For the comparative purpose as well as that of replication accuracy we used the data on exchange rates employed in the studies by Kugler and Lenz (1990, 1993) and Brock, Hsieh and LeBaron (1993); for the sake of consistency we use the same notation as in those studies. These data samples were used earlier by Kočenda (2001) and Belaire-Franch (2003), which allows some comparison.

5.1 Analysis of ARCH corrected weekly exchange rates

Kugler and Lenz (1990) analyzed nonlinear dependence of weekly exchange rate changes for four currencies against the US dollar from 1979 to 1989 (575 observations, the rate of change of the log exchange rate $x_t = \Delta \log S_t$). The data were corrected to account for the present ARCH process by transformation into the ARCH corrected rate of changes in the form

$$\Delta \log S_t^h = \Delta \log S_t \Bigg/ \left(\hat{\alpha}_0 + \sum_{\tau=1}^{6} \hat{\alpha}_\tau \Delta \log S_{t-\tau}^2 \right)^{0.5} \qquad (5.1)$$

where α-coefficients were obtained by OLS regression of $(\Delta \log S_t)^2$ on constant and six lagged variables.

We have replicated the original study of Kugler and Lenz (1990) with the same results as Kočenda (2001) or Belaire-Franch (2003); the null hypothesis is rejected for the French Franc (FRF), Japanese Yen (JPY) and the Deutsche Mark (DEM). The Swiss Franc (CHF) is the only currency where the null of iid cannot be rejected.

5.2 Analysis of daily exchange rates

Brock, Hsieh, and LeBaron (1993) analyzed the daily closing bids for five major currencies in U.S. Dollars: Swiss Franc (CHF), Canadian Dollar (CAD), Deutsche Mark (DEM), British Pound (GBP), and Japanese Yen (JPY) during the period from January 2, 1974 to December 30, 1983 (2,510 observations). The specification of the model resulted in the mean equation

$$r_t = \beta_0 + \sum_{i=1}^{j} \beta_i r_{i-1} + \beta_M D_{M,t} + \beta_T D_{T,t} + \beta_W D_{W,t} + \beta_R D_{R,t} + \beta_H D_H + u_t \qquad (5.2a)$$

where, $u_t | \Omega_{t-1} \sim D(0, h_t)$, and variance equation

$$h_t = \phi_0 + \psi u_{t-1}^2 + \phi h_{t-1} + \phi_M D_{M,t} + \phi_T D_{T,t} + \phi_W D_{W,t} + \phi_R D_{R,t} + \phi_H D_H \qquad (5.2b)$$

where r_t is the rate of change of the nominal exchange rate at time *t*, $D_{M,t}$, $D_{T,t}$, $D_{W,t}$, and $D_{R,t}$, are dummy variables for days in a week (Monday, Tuesday, Wednesday, and Thursday) and D_H is the number of holidays between two successive trading days excluding week-ends. The order of the AR process was found to be *j = 6*, *5*, *6*, and *0* respectively for CHF, CAD, DEM, and GBP.

After estimation, the Kočenda test is run on standardized residuals $z_t = u_t / h_t^{\frac{1}{2}}$. The results are rejections for all four time series.

5.3 Analysis of weekly exchange rates

Kugler and Lenz (1993) analyzed the non-linear dependence of exchange rate changes for ten currencies against the US dollar. The sample period is from 1979 to 1989 (575 observations, the rate of change of the log exchange rate $x_t = \Delta \log S_t$). In order to check whether the detected dependence can be attributed solely to an ARCH process, the authors estimated the following GARCH-M model

$$\Delta \log S_t = \beta_0 + \sum_{\tau=1}^{3} \beta_\tau \Delta \log S_{t-\tau} + \beta_4 \sqrt{h_t} + \eta_t \qquad (5.3)$$

$$h_t = \alpha_0 + \alpha_1 \eta_{t-1}^2 + \alpha_2 h_{t-1} \qquad \eta_t = \varepsilon_t \sqrt{h_t}$$

After estimation the fitted residuals $\hat{\varepsilon}_t = \eta_t / \sqrt{h_t}$ are subjected to the test. The results of the test with the new interval (0.60σ, 1.90σ) provide even stronger results than in Kočenda (2001). The null hypothesis of iid is rejected in all cases but one: only the Swiss Franc is still shown to be without any nonlinear dependency.

6 Conclusions

In this paper we extend the test of Kočenda (2001) and deliver two improvements to the operational ability of the procedure.

First, we improve the choice of the range of the proximity parameter ε over which the correlation integral is calculated, tabulate new sets of critical values for various lengths of data, and provide a sensitivity check to the robustness of critical values with respect to the choice of range of proximity parameters. We perform a series of power tests and suggest the range that maximizes the power of the test. Unless assumptions of a particular research project dictate otherwise, the interval (0.60σ, 1.90σ) for the ε-range choice should be used as a template option.

Second, we compare the test with existing results of the single-blind controlled competition of Barnett et al. (1997) and additionally perform power tests on various chaotic (nonlinear) and noisy chaotic data. The results strongly attest to our robust procedure. The improvement allows for new economic insights especially when the procedure is used on financial data as a specification test. Comparison with real exchange rate data is provided as well.

As a final contribution, we introduce a new compact program to run the test as well as simulations. The software is very fast, user-friendly, and may be downloaded from http://home.cerge-ei.cz/kocenda/software/ as freeware, subject to appropriate citation.

References

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J.A., Kaplan, D.T., Jensen, M.J., 1997. A single-blind controlled competition among tests for nonlinearity and chaos. *Journal of Econometrics*, vol. 82, pp. 157-192.

Belaire-Franch, J., 2003. A Note on Resampling the integration Across the Correlation Integral with Alternative Ranges. *Econometric Reviews*, 22(4), 337-349.

Brock, W., 1988. Nonlinearity and Complex Dynamics in Economics and Finance in Anderson, P. W., Arrow, K. J., and Pines, D., editors, 1988, *The Economy as an Evolving Complex System*, Santa Fe Institute Studies in the Sciences of Complexity, v. 5, Addison-Wesley Publishing Company

Brock, W., Dechert, W., Scheinkman, J., LeBaron, B., 1996. A Test for Independence Based on the Correlation Dimension. *Econometric Reviews*, 15, 197-235.

Brock, W. A., Hsieh, D. A., and LeBaron, B., 1993. *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, MIT Press, Cambridge, Massachusetts, Third edition.

Grassberger, P., and Procaccia, I., 1983. Measuring the Strangeness of Strange Attractors. *Physica 9D*, 189-208.

Hsieh, D. A., 1993. Implications of Nonlinear Dynamics for Financial Risk Management. *Journal of Financial and Quantitative Analysis*, 28(1), 41-64.

Hsieh, D. A., and LeBaron, B., 1988. The BDS Test. Later published as the Chapter 2, in Brock, W. A., Hsieh, D. A., and LeBaron, B., 1993. Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence, MIT Press, Cambridge, Massachusetts, Third edition.

Kanzler, L., 1999. Very Fast and Correctly Sized Estimation of the BDS Statistic. Manuscript. Christ Church and Department of Economics, University of Oxford

Kočenda, E., 2001. An Alternative to the BDS Test: Integration across the Correlation Integral. *Econometric*

Reviews, 20(3), 337-351.

Kočenda, E., and Briatka, Ľ, 2004. Advancing the iid Test Based on Integration across Correlation Integral: Ranges, Competition, and Power, *CERGE-EI Working Paper* No. 235.

Kugler, P., and Lenz, C., 1990. Sind Wechselkursfluktuationen zufullig oder chaotisch? (Are Exchange Rate Fluctuations Random or Chaotic?). *Schweizerische Zeitschrift fhr Volkswirtschaft und Statistik*, 126(2), 113-128.

Kugler, P., and Lenz, C., 1993. Chaos, ARCH, and the Foreign Exchange Market: Empirical Results from Weekly Data. *Rivista Internazionale di Scienze Economiche e Commerciali*, 40(2), 127-140.

Stability and Lyapunov Exponents in Keynesian and Classical Macroeconomic Models

Jan Kodera, Karel Sladký, Miloslav Vošvrda

Department of Econometrics
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

Abstract

In this article we compare dynamical properties of Keynesian and Classical macroeconomic models. We start with an extended dynamical IS-LM neoclassical model generating behaviour of the real product, interest rate, expected inflation and the price level over time. Limiting behaviour, stability, existence of limit cycles and other specific features of these models will be compared.

Keywords: Macroeconomic models, Keynesian and Classical model, nonlinear differential equations, linearization, asymptotical stability, Gershgorin's circles, Lyapunov exponents.

JEL: C00,E12,E13

1 Macroeconomic Models

In this article we study dynamic macroeconomic systems employing the IS (i.e. Investment–Saving) and LM (i.e. Liquidity–Money) models. This approach enables to enclose also the money market and moreover, introducing a Phillips curve enables to extend the dynamic models also to labour market (e.g. Fischer [1], Sargent [7], Sargent [8], Tobin [10]).

Let (in continuous time $t \geq 0$) $Y(t)$, $S(\cdot,\cdot)$ and $I(\cdot,\cdot)$ be denote the real product, savings and real investments of the considered economy respectively. Recall that for the nominal interest $R(t)$ it holds $R(t) = r(t) + \pi^e(t)$ where $r(t)$ is the real rate of interest and $\pi^e(t)$ the expected inflation, in contrast to the inflation $\pi(t)$. The dynamics of the IS model is then given by the following differential equation (see e.g. Takayama [9]) $\dot{Y} = \alpha\{I(Y(t), r(t)) - S(Y(t), r(t))\}$ or on taking logarithms by

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = \alpha\{i(y(t), r(t)) - s(y(t), r(t))\} \tag{1}$$

where $y(t) = \ln Y(t)$, and $i(\cdot,\cdot) = \frac{I(\cdot,\cdot)}{Y(\cdot,\cdot)}$, $s(\cdot,\cdot) = \frac{S(\cdot,\cdot)}{Y(\cdot,\cdot)}$, is the so-called propensity to invest, to save respectively. Observe that for an equilibrium point $Y(t) \equiv Y^\star$, $y(t) \equiv y^\star$, $r(t) \equiv r^\star$ we have $I(Y^\star, r^\star) = S(Y^\star, r^\star)$ or $i(y^\star, r^\star) = s(y^\star, r^\star)$.

Denoting by $p(t)$ the price level at time t, the dynamics of the money market is described by the following differential equation

$$\frac{\mathrm{d}r(t)}{\mathrm{d}t} = \beta\left\{\ell(y(t), R(t)) - \ln\frac{M^s}{p(t)}\right\} = \beta\{\ell(y(t), r(t) + \pi^e(t)) - (m^s - \bar{p}(t))\} \tag{2}$$

where $\ell(y(t), R(t)) = \ln(L(Y(t), R(t))$, $m^s = \ln M^s$, $\bar{p}(t) = \ln p(t)$; $L(\cdot,\cdot)$ and M^s is reserved for demand for money and money supply respectively. In (1), (2) α, β are positive constants signifying the speed of adjustment of the respective market.

To obtain a complete dynamic model of the economy we need to include equations for expected inflation $\pi^e(t)$ and the price level $p(t)$. According to Tobin [11], for $\pi^e(t)$ the following adaptive equation is valid

$$\frac{\mathrm{d}\pi^e(t)}{\mathrm{d}t} = \gamma\left[\pi(t) - \pi^e(t)\right] \tag{3}$$

where γ is the coefficient of adaptation and $\pi(t)$ is the inflation. Recalling that $\pi(t) = \frac{\dot{p}(t)}{p(t)} = \frac{\mathrm{d}}{\mathrm{d}t}\bar{p}(t)$, from (3) we immediately get

$$\frac{\mathrm{d}\pi^e(t)}{\mathrm{d}t} = \gamma\left[\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}(t) - \pi^e(t)\right]. \tag{4}$$

For what follows we need to express $\frac{\mathrm{d}}{\mathrm{d}t}\bar{p}(t)$. To this end we shall assume that the development of the price level $p(t)$ over time is in accordance with changes of the so-called cost function $C(y(t))$. In particular, the well-known condition of profit maximization $p(t) - \frac{\mathrm{d}C(y)}{\mathrm{d}y} = 0$ is the base for the following adjustment formula for $p(t)$, where δ is a constant:

$$\frac{\mathrm{d}\bar{p}(t)}{\mathrm{d}t} = \delta\left(\frac{\mathrm{d}C(y)}{\mathrm{d}y} - e^{\bar{p}(t)}\right) \tag{5}$$

In fact, the above formula is in accordance with the traditional theory of perfectly competitive firms (see e.g. [5]) and as such is interpreted in many treatises on monetary and price dynamics (cf. e.g. [3]).

In what follows we shall use shorthand notations only, i.e., we replace $\frac{\mathrm{d}\bar{p}(t)}{\mathrm{d}t}$ by $\dot{\bar{p}}$, similarly for the time derivatives $\dot{y}$, $\dot{r}$, $\dot{\pi}^e$, and $\frac{\mathrm{d}C(y)}{\mathrm{d}y}$ is replaced by $C'(y)$. Moreover, we shall often omit the argument t. Hence, (cf. (1), (2), (4), (5)) using such a model the system describing an economy from the Keynesian point of view has the following form:

$$\left.\begin{array}{rcl} \dot{y} &=& \alpha[i(y,r) - s(y,r)], \\ \dot{r} &=& \beta[\ell(y, r+\pi^e) - (m^s - \bar{p})], \\ \dot{\pi}^e &=& \gamma[\dot{\bar{p}} - \pi^e], \\ \dot{\bar{p}} &=& \delta[C'(y) - e^{\bar{p}}], \end{array}\right\} \tag{6}$$

where $i(y,r)$, $s(y,r)$, $\ell(y, r+\pi^e)$ and $C(y)$ are real investment, real savings, real money demand and cost functions respectively, depending on production y, rate of interest r, (expected) inflation π^e and the price level p.

Classical models that describe (commodity) price level, interest rate, production and expected inflation dynamics have similar structure of right hand sides (RHS) of differential equations, but left hand sides (LSD) are permuted as follows:

$$\left.\begin{array}{rcl} \dot{r} &=& \alpha[i(y,r) - s(y,r)], \\ -\dot{\bar{p}} &=& \beta[\ell(y, r+\pi^e) - (m^s - \bar{p})], \\ \dot{\pi}^e &=& \gamma[\dot{\bar{p}} - \pi^e]. \end{array}\right\} \tag{7}$$

Since for classical models the real product $y(t)$ is assumed to be constant, in (7) we ignore the equation $-\dot{y} = \delta[C'(y) - e^{\bar{p}}]$.

Just introduced models are the base for establishment of macroeconomic models of price and monetary dynamics. Recall that the vector $\mathbf{x}^* = (y^*, r^*, \pi^{e*}, \bar{p}^*)$ whose elements are obtained as a solution of the following set of equations:

$$\left.\begin{array}{lcl} i(y,r) & = & s(y,r), \\ \ell(y, r+\pi^e) & = & m^s - \bar{p}, \\ e^{\bar{p}} & = & C'(y), \end{array}\right\} \tag{8}$$

is the equilibrium point both of the Keynesian model given by the set of equations (6) and Classical models given by the set of equation (7). This equilibrium point is said to be (asymptotically) locally stable if every solution of the considered system, starting sufficiently close to $\mathbf{x}^*$ converges to $\mathbf{x}^*$ as $t \to \infty$. Similarly, $\mathbf{x}^*$ is said to be (asymptotically) globally stable if every solution regardless the starting point converges to $\mathbf{x}^*$. It is well known (cf. e.g. [4] or [9]) that an equilibrium point (and also a stable point) of the system need not exist, hence the system is unstable. Recall that having found equilibrium points, the system need not converge to some or any of the equilibrium points (in the latter case the system is unstable). Furthermore, if the considered system is unstable and nonlinear, then the system can also exhibit limit cycles (i.e. its trajectory remains in a bounded region) or even chaotic behavior. In words, in contrast to above phenomena, stability is equivalent to monotone or oscillating convergence toward the equilibrium point.

To identify a chaotic behavior of a macroeconomic model, it is plausible to compare dynamical behaviour of the macroeconomic model by an exponential divergence of nearby trajectories measured by the so-called Lyapunov exponents. The most important is the maximal Lyapunov exponent, negative for stable models, positive for unstable models and infinite for the chaotic behavior – for details see Lorenz [6].

2 Approximation and Linearization of the Models

To find an analytical form of the output $y(t) = \ln Y(t)$, interest rate $r(t)$, expected inflation $\pi^e(t)$ and the price level $p(t)$ we need to assume that the functions $i(\cdot,\cdot)$, $s(\cdot,\cdot)$, $C(\cdot)$ are of a specific analytical form. As usual, the functions $s(\cdot,\cdot)$ as well as demand for money $\ell(y,R)$ can be well approximated by linear functions, whereas it is necessary to approximate $i(\cdot,\cdot)$ and sometimes also $C(\cdot)$ by suitable nonlinear functions. In what follows, we assume that savings $S(Y(t), r(t))$ can be well approximated by the following expression

$$S(Y(t), r(t)) = Y(t) \cdot [s_0 + s_1 \cdot y(t) + s_2 \cdot r(t)] \quad \text{with} \quad s_0 < 0, \text{ and } s_1, s_2 > 0. \tag{9}$$

Hence the propensity to save $s(\cdot,\cdot) = S(\cdot,\cdot)/Y(\cdot)$ can be written as

$$s(Y(t), r(t)) \stackrel{\text{def}}{=} \bar{s}(y(t), r(t)) = s_0 + s_1 \cdot y(t) + s_2 \cdot r(t) \tag{10}$$

Similarly, the demand for money is described by the traditional Keynesian demand-for-money function being in the following form

$$\ell(y(t), R(t)) = \ell_0 + \ell_1 y(t) - \ell_2 R(t) - \ell_3 \pi^e(t) = \ell_0 + \ell_1 y(t) - \ell_2 [r(t) + \pi^e(t)] - \ell_3 \pi^e(t), \tag{11}$$

where the parameters $\ell_i > 0$, $i = 0, 1, 2, 3$ are given. On the other hand, it is convenient to assume that the propensity to invest $i(y(t), r(t))$ is a product of $\frac{1}{r(t)+1}$ and the so-called logistic function. Hence the propensity to invest is assumed to be given analytically as

$$i(y(t), r(t)) = \frac{1}{r(t)+1} \cdot \frac{k}{1 + be^{-ay(t)}} \tag{12}$$

where the parameters $k, a > 0$ and b is an arbitrary real number. Similarly, we shall assume that the cost function $C(\cdot)$ is also a logistic function given analytically as

$$C(y(t)) = \frac{h}{1 + de^{-cy(t)}} \tag{13}$$

where the parameters $h, c > 0$ and d is an arbitrary real number. Hence

$$\frac{\mathrm{d}C(y)}{\mathrm{d}y} = \frac{h}{(1 + de^{-cy})^2}(-cd\, y) \tag{14}$$

and we can assume that the "central" part of $C(y(t))$ can be well approximated by a linear function

$$C(y(t)) = d_0 + d_1 y(t) \tag{15}$$

Since $\pi^{e\star} = 0$ to calculate the values $y^\star$, $r^\star$, $p^\star$, on inserting (10), (11), (12) and (13) into (8) we have

$$\frac{1}{r^\star + 1} \cdot \frac{k}{1 + be^{-ay^\star}} = s_0 + s_1 y^\star + s_2 r^\star \tag{16}$$

$$\ell_0 + \ell_1 y^\star - \ell_2 r^\star = m^s - \bar{p}^\star \tag{17}$$

$$\bar{p}^\star = -\ln d_1 \stackrel{\text{def}}{=} -\bar{d}_1 \tag{18}$$

In virtue of (18) from (16), (17) the equilibrium values $y^\star$, $r^\star$ can be found as a solution to

$$\frac{k}{1 + be^{-ay^\star}} = (s_0 + s_1 y^\star + s_2 r^\star)(1 + r^\star) \tag{19}$$

$$r^\star = \frac{1}{\ell_2}(\ell_0 - (m^s + \bar{d}_1) + \ell_1 y^\star) \Longleftrightarrow y^\star = \frac{1}{\ell_1}\left((m^s + \bar{d}_1) - \ell_0 + \ell_2 r^\star\right) \tag{20}$$

From (19), (20) we get

$$\left[s_0 + s_2\left(\frac{\ell_0}{\ell_2} - \frac{m^s + \bar{d}_1}{\ell_2}\right) + \left(s_1 + s_2\frac{\ell_1}{\ell_2}\right) y^\star\right] \cdot \left[1 + \frac{\ell_0}{\ell_2} - \frac{m^s + \bar{d}_1}{\ell_2} + \frac{\ell_1}{\ell_2} y^\star\right] = k \cdot \frac{1}{1 + be^{-ay^\star}} \tag{21}$$

Hence finding the solution to (21) and inserting this value into (20) we immediately get the pair of equilibrium points $y^\star$, $r^\star$. We can observe that:

1. The RHS of (21) is the so-called logistic function (an increasing function having an inflection point at $y = \frac{1}{a}\ln b$ that is convex in the interval $(0, \frac{1}{a}\ln b)$ and concave in $(\frac{1}{a}\ln b, \infty)$);

2. The LHS of (21) is a quadratic function (in fact, for real-life models this function differs only slightly from a straight line).

Hence there exist at most three, in real models usually only one, pair(s) of equilibrium points $y^\star$, $r^\star$ for $y \geq 0$. More insight in the properties of the equilibrium points, especially with respect to the stability, can be obtained by linearization around the neighborhood of the equilibrium point $(y^\star, r^\star, \pi^{e\star}, \bar{p}^\star)$ with $\pi^{e\star} = 0$. To check stability of the linearized model, (i.e., that all eigenvalues of the matrix of the linearized system have negative real parts), let us recall that all eigenvalues of the matrix lay in the union of the Gershgorin's circles. The centers of circles are diagonal elements of the matrix and the radius is equal to the minimum of row of column sums of the absolute values of the corresponding off-diagonal elements. For details see e.g. Fiedler [2].

3 Stability and Speed of Adjustment

3.1 Keynesian Model

In particular, on employing (16), (17), (18) for the Keynesian model we have:

$$\begin{bmatrix} \frac{\mathrm{d}(y(t)-y^\star)}{\mathrm{d}t} \\ \frac{\mathrm{d}(r(t)-r^\star)}{\mathrm{d}t} \\ \frac{\mathrm{d}(\pi^e(t))}{\mathrm{d}t} \\ \frac{\mathrm{d}(p(t)-p^\star)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} \alpha(D_y - s_1) & \alpha(D_r - s_2) & 0 & 0 \\ \beta\ell_1 & -\beta\ell_2 & -\beta(\ell_2+\ell_3) & \beta \\ 0 & 0 & -\gamma & \gamma d_1 \\ 0 & 0 & 0 & -\delta d_1 \end{bmatrix} \begin{bmatrix} y(t) - y^\star \\ r(t) - r^\star \\ \pi^e(t) \\ p(t) - p^\star \end{bmatrix} \tag{22}$$

where

$$D_y = \frac{1}{1+r}\bigg|_{r=r^\star} \cdot \frac{\partial}{\partial y}\frac{k}{1+be^{-ay(t)}}\bigg|_{y=y^\star}, \quad D_r = \frac{\partial}{\partial r}\frac{1}{1+r}\bigg|_{r=r^\star} \cdot \frac{k}{1+be^{-ay(t)}}\bigg|_{y=y^\star}.$$

and

$$k = \left[s_0 + s_2\left(\frac{\ell_0}{\ell_2} - \frac{m^s + \bar{d}_1}{\ell_2}\right) + \left(s_1 + s_2\frac{\ell_1}{\ell_2}\right)y^\star\right] \cdot \left[1 + \frac{\ell_0}{\ell_2} - \frac{m^s + \bar{d}_1}{d_1\ell_2} + \frac{\ell_1}{\ell_2}y^\star\right] \cdot \left[1 + be^{-ay^\star}\right].$$

To verify if the obtained equilibrium point is stable, we shall have a look at the eigenvalues of the matrix

$$\boldsymbol{A} = \begin{bmatrix} \alpha(D_y - s_1) & \alpha(D_r - s_2) & 0 & 0 \\ \beta\ell_1 & -\beta\ell_2 & -\beta(\ell_2+\ell_3) & \beta \\ 0 & 0 & -\gamma & \gamma d_1 \\ 0 & 0 & 0 & -\delta d_1 \end{bmatrix} \tag{23}$$

Employing the "nearly" upper triangular structure of the matrix $\boldsymbol{A}$ we can immediately conclude that the eigenvalues λ_1, λ_2, λ_3, λ_4 of $\boldsymbol{A}$ are equal to δd_1, γ and the remaining two eigenvalues λ_3, λ_4 can be calculated as the two eigenvalues of the matrix

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \alpha(D_y - s_1) & \alpha(D_r - s_2) \\ \beta\ell_1 & -\beta\ell_2 \end{bmatrix} \tag{24}$$

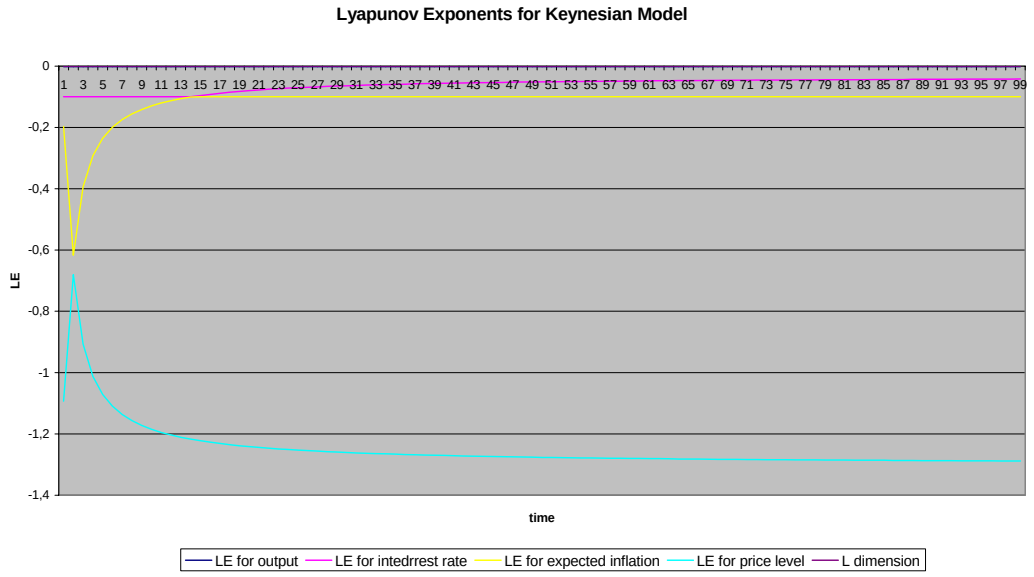


Figure 1:

In particular, if the following two equations Keynesian model

$$\begin{bmatrix} \frac{\mathrm{d}(y(t)-y^\star)}{\mathrm{d}t} \\ \frac{\mathrm{d}(r(t)-r^\star)}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} \alpha(D_y - s_1) & \alpha(D_r - s_2) \\ \beta\ell_1 & -\beta\ell_2 \end{bmatrix} \begin{bmatrix} y(t) - y^\star \\ r(t) - r^\star \end{bmatrix} \tag{25}$$

is stable, then also our extended Keynesian model given by (22) must be stable. Obviously, the eigenvalues of $\tilde{A}$ are (the symbols $\operatorname{tr}\tilde{A}$ and $\det\tilde{A}$ are reserved for trace and determinant of $\tilde{A}$

$$\lambda_{3,4} = \frac{1}{2}\left(\operatorname{tr}\tilde{A} \pm \sqrt{(\operatorname{tr}\tilde{A})^2 - 4\det\tilde{A}}\right)$$

and $\det\tilde{A}$ must be positive in order to exclude the possibility of a saddle point. For the asymptotic stability $\operatorname{Re}\lambda_{3,4} < 0$, hence if $\operatorname{tr}\tilde{A} = \alpha(D_y - s_1) - \beta\ell_2 < 0$ both (23) and 24) are stable, in case that $\alpha(D_y - s_1) > \beta\ell_2$ the equilibrium is not asymptotically stable and the limit cycle occurs. In particular, sufficient conditions for the stability of the matrix $\boldsymbol{A}$ of the considered four-equation Keynesian model are $D_y - s_1 < 0$ along with $D_y - s_1 > D_r - s_2$, $\ell_1 < \ell_2$ or $D_y - s_1 > \frac{\beta}{\alpha}\cdot\ell_1$, $D_r - s_2 > \frac{\beta}{\alpha}\cdot\ell_2$. An interesting case is when eigenvalues of $\tilde{A}$ are purely imaginary, i.e. if $\alpha(D_y - s_1) = \beta\ell_0$.

Lyapunov exponents for the considered four-equation Keynesian model with the following values of parameters $\alpha = 20$, $\beta = 1$, $\gamma = 0.1$, $\delta = 0.02$, $a = 0.1, b = 1.5, s_0 = -0.16, s_1 = 0.07, s_2 = 0.016, l_0 = 0.25, l_1 = 0.4, l_2 = -0.06, l_3 = -0.06, d = -1, d_2 = 0.3, m^s = 0.65, k = 0.4$ are presented in Figure 1.

The Lyapunov dimension of the Keynesian model attractor is equal 0. It means that practically all eigenvalues of the Keynesian model attractor are negative. Thus the Keynesian model is not chaotic macroeconomic system.

3.2 Classical Model

In particular, on employing (16), (17), (18) for the Classical model we have:

$$\begin{bmatrix} \frac{\mathrm{d}(r(t)-r^\star)}{\mathrm{d}t} \\ \frac{\mathrm{d}(p(t)-p^\star)}{\mathrm{d}t} \\ \frac{\mathrm{d}(\pi^e(t))}{\mathrm{d}t} \end{bmatrix} = \begin{bmatrix} \alpha(D_r - s_2) & 0 & 0 \\ \beta\ell_2 & -\beta & \beta(\ell_2+\ell_3) \\ 0 & \gamma d_1 & -\gamma \\ 0 & \delta d_1 & 0 \end{bmatrix} \begin{bmatrix} r(t)-r^\star \\ p(t)-p^\star \\ \pi^e(t) \end{bmatrix} \tag{26}$$

where D_r and k take on the same values as in Section 3.1.

Lyapunov exponents for the classical model with the following values of parameters $\alpha = 200, \beta = 0.2, \gamma = \delta = 1, a = 0.1, b = 1.5, s_0 = -0.16, s_1 = 0.07, s_2 = 0.016, l_0 = 0.25, l_1 = 0.4, l_2 = -0.06, l_3 = -0.06, d = -1, d_2 = 0.3, m^s = 0.65, k = 0.4, y = 4.5$ are presented in Figure 2.

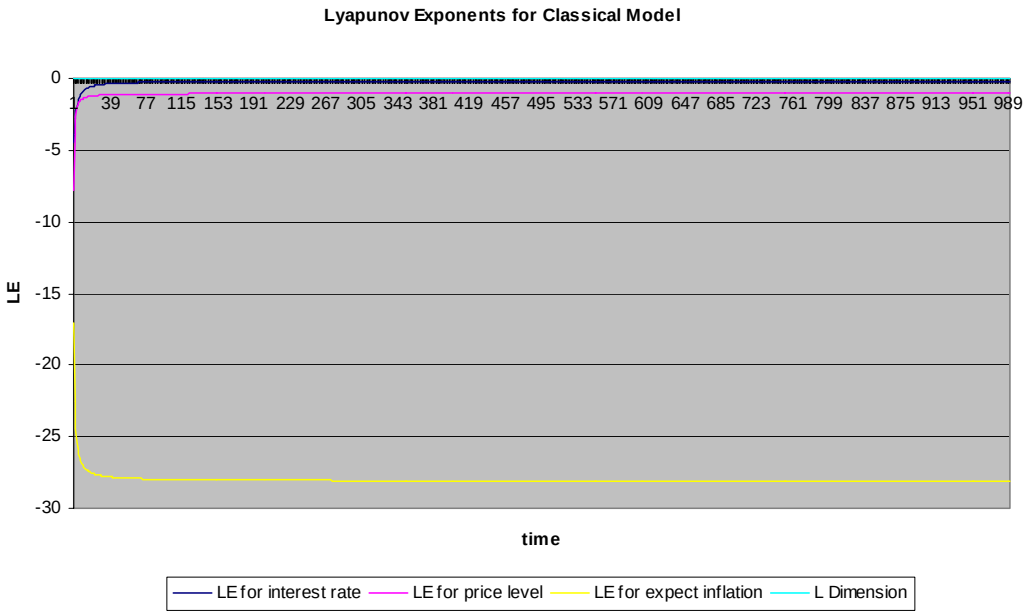


Figure 2:

The Lyapunov dimension for the classical model attractor is equal to 0 also. It means that practically all eigenvalues of the classical model attractor are negative. Thus the classical model is not chaotic macroeconomic system.

Acknowledgement

This research was supported by the Grant Agency of the Academy of Sciences of the Czech Republic under Grant A 7075202 and by the Grant Agency of the Czech Republic under Grant 402/03/H057.

References

[1] S. Fischer: Keynes–Wicksell and neoclassical models of money and growth. American Economic Review 62 (1972), 880–890.

[2] M. Fiedler: Speciální matice a jejich použití v numerické matematice. SNTL, Praha 1981.

[3] P. Flaschel, R. Franke, and W. Semmler: Dynamic Macroeconomics. The MIT Press, Cambridge MA, 1997.

[4] J. Guckenheimer and P. Holmes: Non-Linear Oscillation, Dynamical Systems and Bifurcations of Vector Fields. Springer–Verlag, New York 1986.

[5] D. Laider and S. Estrin: Introduction to Microeconomic. Philip Allan, London 1989.

[6] H.-W. Lorenz: Nonlinear Dynamical Economics and Chaotic Motion. Second edition, Springer-Verlag, Berlin 1993.

[7] T. Sargent: Interest rate and prices in the long run: a study of the Gibson paradox. Journal of Money, Credit, and Banking 5 (1973), 385–449.

[8] T. Sargent: Macroeconomic Theory. Second edition. Academic Press, New York 1987.

[9] A. Takayama: Analytical Methods in Economics. Harvester Wheatsheaf, Hertfordshire 1994.

[10] J. Tobin: Money and economic growth. Econometrica 33 (1965), 671–684.

[11] J. Tobin: Keynesian models of recession and repression. American Economic Reviews 65 (1975), 195–202.

Stochastic dominance and CVaR in portfolio selection problem

Miloš Kopa [1]
Charles University Prague, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovska 83, 186 75 Prague, Czech Republic
e-mail:*kopa@karlin.mff.cuni.cz*

Abstract
This paper describes second-order stochastic dominance rules concerning portfolio selection problem. We consider discrete probabilistic distributions of returns of the assets using the scenario approach. As was shown in [4], the conditional value-at-risk corresponds to second-order stochastic dominance. Using this property, the necessary and sufficient condition for a non-dominated portfolio relative to all possible portfolios created from a set of assets is derived.

Keywords
stochastic dominance, CVaR, portfolio selection problem

1 Introduction

The portfolio selection problem may be regarded as a two-step procedure. Firstly, an efficient set among all available portfolios is chosen and then the risk preferences of decision maker to this set are applied. This paper deals with the first step. The second section recalls the basic ideas and results of stochastic dominance approach. A portfolio is non-dominated in the considered set of assets if there exists no convex combination of the assets which dominates the portfolio.

The risk preferences of decision maker can be described by a von Neumann-Morgenstern utility function. More details about utility functions and characterizations of utility functions can be found in [2]. Applying conditional value-at-risk (CVaR) is another way how to express the risk attitude of decision makers. If the yields or losses of assets in the portfolio are described by discrete probabilistic distributions then CVaR can be computed as a solution of linear programming problem. This property will be used in the sequel.

In section 3, CVaR in context of stochastic dominance is analyzed.

In section 4, necessary and sufficient condition for a non-dominated portfolio in the considered set of assets is proved.

2 Stochastic dominance

For two random variables X_1 and X_2 with respective cumulative probability distributions functions $F_1(x)$, $F_2(x)$ we say that X_1 *dominates* X_2 *by first degree stochastic dominance*: $X_1 \succeq_{FSD} X_2$ if

$$\mathbb{E}_{F_1} u(x) - \mathbb{E}_{F_2} u(x) \geq 0$$

for every utility function u, i.e. for every continuous nondecreasing function u, such that these expected values exist. Let us denote by U_1 the set of all such functions. We say that X_1 *dominates* X_2 *by second degree stochastic dominance*: $X_1 \succeq_{SSD} X_2$ if

$$\mathbb{E}_{F_1} u(x) - \mathbb{E}_{F_2} u(x) \geq 0$$

[1]This work was supported by the Grant Agency of the Czech Republic (grant 201/05/H007)

for every $u \in U_2$ where $U_2 \subset U_1$ denotes the set of all concave utility functions such that these expected values exist.

For the development of the third or higher degree of stochastic dominance see [3], [7] and [8]. Set

$$F_i^{(2)}(t) = \int_{-\infty}^{t} F_i(x)dx \qquad i = 1, 2.$$

The following necessary and sufficient conditions for stochastic dominance were proved in [1].

Lemma 2.1:

Let $F_1(x)$ and $F_2(x)$ be cumulative distribution functions of X_1 and X_2. Then

$$\begin{aligned} X_1 \succeq_{FSD} X_2 \quad &\Leftrightarrow \quad F_1(x) \leq F_2(x) \quad \forall x \in \mathbb{R} \\ X_1 \succeq_{SSD} X_2 \quad &\Leftrightarrow \quad F_1^{(2)}(t) \leq F_2^{(2)}(t) \quad \forall t \in \mathbb{R}. \end{aligned}$$

Definition 2.2:

For a set $A = \{X_1, X_2, ..., X_n\}$ of random variables, a random variable $\overline{X}$ is called SSD-non-dominated (FSD-non-dominated) in A if there exists no

$$\mathbf{c} \in C = \{\mathbf{c} \in \mathbb{R}^n : \sum_{j=1}^{n} c_j = 1, \quad c_j \geq 0, j = 1, 2, \ldots, n\}$$

such that

$$\sum_{j=1}^{n} c_j X_j \succeq_{SSD} \overline{X} \quad (\sum_{j=1}^{n} c_j X_j \succeq_{FSD} \overline{X}).$$

Otherwise, $\overline{X}$ is called SSD-dominated (FSD-dominated) in A.

If we modify the set C to the set

$$C' = \{\mathbf{c} \in \mathbb{R}^n : c_j = 1, c_i = 0, \forall i \neq j\}$$

we obtain the same definition of stochastic dominance in the set of random variables as in [4]. The reason why C is used in Definition 2.2 instead of C' is that the aim of portfolio selection problem is to find the optimal convex combinations of assets in A.

Consider now the quantile model of stochastic dominance [4]. The *first quantile function* $F_X^{(-1)}$ corresponding to a real random variable X is defined as the left continuous inverse of F_X:

$$F_X^{(-1)}(v) = \min\{u : F_X(u) \geq v\}. \tag{1}$$

Directly from Lemma 2.1 we can see that

$$X_1 \succeq_{FSD} X_2 \quad \Leftrightarrow \quad F_1^{(-1)}(p) \leq F_2^{(-1)}(p) \quad \forall p \in (0, 1\rangle.$$

The *second quantile function* $F_X^{(-2)}$ is defined as

$$\begin{aligned} F_X^{(-2)}(p) &= \int_{-\infty}^{p} F_X^{(-1)}(t)dt \quad \text{for} \;\; 0 < p \leq 1 \\ &= 0 \qquad\qquad\qquad\quad \text{for} \;\; p = 0 \\ &= +\infty \qquad\qquad\qquad\quad \text{otherwise.} \end{aligned}$$

The function $F_X^{(-2)}$ is convex and it is well defined for any random variable X satisfying the condition $\mathbb{E}|X| < \infty$. For the proof of the following basic properties of the second quantile function and more details about dual stochastic dominance see [4].

Theorem 2.3:

For every random variable X with $\mathbb{E}|X| < \infty$ we have:

$$\begin{aligned} F_X^{(-2)}(p) &= \sup_{\nu}\{\nu p - \mathbb{E}\max(\nu - X, 0)\} \\ X_1 \succeq_{SSD} X_2 &\Leftrightarrow \frac{F_1^{(-2)}(p)}{p} \geq \frac{F_2^{(-2)}(p)}{p} \quad \forall p \in \langle 0, 1\rangle. \end{aligned}$$

3 CVaR for scenario approach

Let Y be a random loss variable corresponding to an yield described by random variable X, i.e. $Y = -X$. We assume that $\mathbb{E}|Y| < \infty$. For a fixed level α, the *value-at-risk* VaR is defined as the α-quantile:

$$\mathrm{VaR}_\alpha(Y) = F_Y^{(-1)}(\alpha)$$

where $F_Y^{(-1)}$ is given by (1). We follow [5] in defining *conditional value-at-risk* CVaR as the solution of the optimization problem

$$\mathrm{CVaR}_\alpha(Y) = \min_{a\in\mathbb{R}}\{a + \frac{1}{1-\alpha}\mathbb{E}[Y - a]^+\} \tag{2}$$

where $[x]^+ = \max(x, 0)$.

If we use in Theorem 2.3 $-Y$ and $1-\alpha$ instead of X and p, respectively, we can directly see from the definition of CVaR that:

$$\begin{aligned} \frac{F_X^{(-2)}(p)}{p} &= \sup_{\nu}\{\nu - \frac{1}{p}\mathbb{E}\max(\nu - X, 0)\} \\ &= -\inf_{\nu}\{-\nu + \frac{1}{p}\mathbb{E}\max(\nu - X, 0)\} \\ &= -\inf_{a}\{a + \frac{1}{1-\alpha}\mathbb{E}\max(Y - a, 0)\} \\ &= -\mathrm{CVaR}_\alpha(Y). \end{aligned}$$

Therefore Theorem 2.3 leads to the following result.

Lemma 3.1:

Let $Y_i = -X_i$ and $\mathbb{E}|X_i| < \infty$ for $i = 1, 2$. Then

$$X_1 \succeq_{SSD} X_2 \Leftrightarrow \mathrm{CVaR}_\alpha(Y_1) \leq \mathrm{CVaR}_\alpha(Y_2) \quad \forall \alpha \in \langle 0, 1\rangle.$$

In this paper we limit our attention to scenario approach, i.e. from now on we will assume that Y is a discrete random variable which takes values $y^i, i = 1, ..., N$ with equal probabilities. Then (2) can be rewritten as a linear programming problem. Moreover CVaR_α(Y) can be calculated using the following formula:

$$\mathrm{CVaR}_\alpha(Y) = \frac{1}{N}\sum_{y^i > \mathrm{VaR}_\alpha(Y)} y^i.$$

Finally, the assumptions of Theorem 2.3 and Lemma 3.1 are fulfilled. For more details we refer to [5].

Let $y^{[k]}$ be the k-th smallest element among $y^1, y^2, ..., y^N$, i.e.
$y^{[1]} \leq y^{[2]} \leq \ldots \leq y^{[N]}$. In context of stochastic dominance the description of CVaR_α(Y) as a function of α will be more useful. The following result is a special case of more general property proved in [6].

Lemma 3.2:

If $\alpha \in \langle \frac{k}{N}, \frac{k+1}{N} \rangle$ and $\alpha \neq 1$ then

$$\mathrm{CVaR}_\alpha(Y) = y^{[k+1]} + \frac{1}{(1-\alpha)N} \sum_{i=k+1}^{N} (y^{[i]} - y^{[k+1]}) \tag{3}$$

for k = 0,1,...,N-1 and $\mathrm{CVaR}_1(Y) = y^{[N]}$.

Proof:

Consider a random variable Y which takes values $y^i, i = 1, ..., N$ with probabilities $p_1, p_2, ..., p_N$. For a chosen α let j_α be defined such that

$$\alpha \in \langle \sum_{j=1}^{j_\alpha - 1} p_j, \sum_{j=1}^{j_\alpha} p_j \rangle.$$

Then the following formula was proved in [6]:

$$\mathrm{CVaR}_\alpha(Y) = \frac{1}{1-\alpha} \left[\left(\sum_{j=1}^{j_\alpha} p_j - \alpha \right) y^{[j_\alpha]} + \sum_{j=j_\alpha+1}^{N} p_j y^{[j]} \right].$$

Since $p_i = 1/N$, $i = 1, ..., N$ we set: $j_\alpha = k + 1$ and the lemma follows.

Combining Lemma 3.1 with Lemma 3.2 we obtain the necessary and sufficient conditions of the second-order stochastic dominance. This conditions can be more easily verified than conditions in Lemma 2.1, Theorem 2.3 or Lemma 3.1.

Theorem 3.3:

Let $Y_1 = -X_1$ and $Y_2 = -X_2$ be discrete random variables which take values y_1^i and y_2^i, $i = 1, ..., N$ with equal probabilities. Then

$$X_1 \succeq_{SSD} X_2 \quad \Leftrightarrow \quad \mathrm{CVaR}_\alpha(Y_1) \leq \mathrm{CVaR}_\alpha(Y_2) \quad \forall \alpha \in \{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}\}.$$

Proof:

Let $\alpha_k = k/N$, $k = 0, 1, ..., N-2$. Lemma 3.2 implies: $\mathrm{CVaR}_{\beta_1}(Y_i) = \mathrm{CVaR}_{\beta_2}(Y_i)$, $i = 1, 2$ for all $\beta_1, \beta_2 \in \langle \frac{N-1}{N}, 1 \rangle$. Thus it suffices to show that if

$$\mathrm{CVaR}_{\alpha_k}(Y_1) \leq \mathrm{CVaR}_{\alpha_k}(Y_2) \tag{4}$$

and

$$\mathrm{CVaR}_{\alpha_{k+1}}(Y_1) \leq \mathrm{CVaR}_{\alpha_{k+1}}(Y_2) \tag{5}$$

then $\mathrm{CVaR}_\alpha(Y_1) \leq \mathrm{CVaR}_\alpha(Y_2)$ for all $\alpha \in \langle \alpha_k, \alpha_{k+1} \rangle$. To obtain a contradiction, suppose that (4) and (5) holds and there exists $\overline{\alpha} \in \langle \alpha_k, \alpha_{k+1} \rangle$ such that $\mathrm{CVaR}_{\overline{\alpha}}(Y_1) > \mathrm{CVaR}_{\overline{\alpha}}(Y_2)$. From continuity of CVaR in α there exists $\alpha^1 \in \langle \alpha_k, \alpha_{k+1} \rangle$ and $\alpha^2 \in \langle \alpha_k, \alpha_{k+1} \rangle$, $\alpha^1 \neq \alpha^2$ such that

$$\mathrm{CVaR}_{\alpha^1}(Y_1) = \mathrm{CVaR}_{\alpha^1}(Y_2) \tag{6}$$

$$\mathrm{CVaR}_{\alpha^2}(Y_1) = \mathrm{CVaR}_{\alpha^2}(Y_2). \tag{7}$$

Substituting (3) into (6) and (7) we conclude that $\alpha^1 = \alpha^2$, contrary to $\alpha^1 \neq \alpha^2$ and the proof is complete.

4 Stochastic dominance in a set of random variables

Consider m assets represented by a random vector of yields $\mathbf{X} = (X_1, X_2, \ldots, X_m)'$ which takes vectors $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N$ with equal probabilities where $\mathbf{x}^i = (x_1^i, x_2^i, \ldots, x_m^i)'$. Let $\mathbf{Y} = -\mathbf{X}$ be a random vector of losses. Consider a portfolio represented by a discrete random variable V_B describing the random yields of the portfolio where $B = \{Z_1, Z_2, \ldots, Z_M\}$ denotes the set of assets from which the portfolio is constructed. Let V_B takes the values v^i, $i = 1, ..., N$ with equal probabilities. Let $W_B = -V_B$ be a random variable of loss of the portfolio. Let $A = \{X_1, X_2, \ldots, X_m\}$. We derive the sufficient and necessary condition for second-order stochastic dominance of V_B in the set A.

Theorem 4.1:

Let $\alpha_k = k/N, \quad k = 0, 1, \ldots, N-1$. Let

$$D^* = \min_{D, t_j, b_k} D \tag{8}$$

$$\begin{aligned} \text{s.t.} \quad b_k + \frac{1}{1-\alpha_k}\mathbb{E}\max(\mathbf{t}'\mathbf{Y} - b_k, 0) - \text{CVaR}_{\alpha_k}(W_B) &\leq D, \quad k = 0, 1, \ldots, N-1 \\ \sum_{j=1}^{m} t_j &= 1 \\ t_j &\geq 0, \quad j = 1, 2, \ldots, m. \end{aligned}$$

If $D^* \leq 0$ then V_B is SSD-dominated in A. If $D^* > 0$ then V_B is SSD-non-dominated in A.

Proof:

Let $\mathbf{t}^*$,b_k^*, $k = 0, 1, \ldots, N-1$ be an optimal solution of (8). If $D^* \leq 0$ then

$$b_k^* + \frac{1}{1-\alpha_k}\mathbb{E}\max(\mathbf{t}^{*'}\mathbf{Y} - b_k^*, 0) \leq \text{CVaR}_{\alpha_k}(W_B) \quad \forall \alpha_k \in \left\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}\right\} \tag{9}$$

Since from the definition of CVaR we have

$$\text{CVaR}_{\alpha_k}(\mathbf{t}^{*'}\mathbf{Y}) = \min_{b_k}\left\{b_k + \frac{1}{1-\alpha_k}\mathbb{E}\max(\mathbf{t}^{*'}\mathbf{Y} - b_k, 0)\right\}$$

we conclude from (9) that

$$\text{CVaR}_{\alpha_k}(\mathbf{t}^{*'}\mathbf{Y}) \leq \text{CVaR}_{\alpha_k}(W_B)$$

and applying Theorem 3.3 we can conclude that $\sum_{j=1}^{m} t_j X_j$ dominates V_B. If $D^* > 0$ then such k exists that

$$b_k + \frac{1}{1-\alpha_k}\mathbb{E}\max(\mathbf{t}'\mathbf{Y} - b_k, 0) - \text{CVaR}_{\alpha_k}(W_B) \geq D > 0$$

for all feasible solutions of (8). Therefore

$$\begin{aligned} \text{CVaR}_{\alpha_k}(\mathbf{t}'\mathbf{Y}) &= \min_{b_k}\left\{b_k + \frac{1}{1-\alpha_k}\mathbb{E}\max(\mathbf{t}'\mathbf{Y} - b_k, 0)\right\} \\ &\geq D + \text{CVaR}_{\alpha_k}(W_B) \\ &> \text{CVaR}_{\alpha_k}(W_B) \end{aligned}$$

for any choice of $\mathbf{t}$. Since we found k such that

$$\text{CVaR}_{\alpha_k}(\mathbf{t}'\mathbf{Y}) > \text{CVaR}_{\alpha_k}(W_B)$$

for all $\mathbf{t}$ such that $\mathbf{1}'\mathbf{t} = 1, t_j \geq 0, j = 1, 2, \ldots, m$, the necessary and sufficient condition of SSD-dominance in Theorem 3.3 is not fulfilled for any convex combination of the assets, which completes the proof.

Non-linear problem (8) has $N+m+1$ constraints and $N+m$ variables. Following [5],[6], it can be rewritten as a linear programming problem with N^2+N+1 constraints and $N(N+1)+m+1$ variables:

$$D^* = \min_{D, t_j, b_k, w_k^i} D \tag{10}$$

$$\begin{aligned}
\text{s.t.} \quad b_k + \frac{1}{(1-\frac{k-1}{N})N} \sum_{i=1}^{N} w_k^i - \text{CVaR}_{\frac{k-1}{N}}(W_B) &\leq D, && k = 1, \ldots, N \\
w_k^i &\geq \mathbf{t}'\mathbf{y}^i - b_k, && i, k = 1, \ldots, N \\
w_k^i &\geq 0, && i, k = 1, \ldots, N \\
\sum_{j=1}^{m} t_j = 1, \quad t_j &\geq 0, && j = 1, 2, \ldots, m.
\end{aligned}$$

Using (10) instead of (8) in Theorem 4.1 we obtain a linear programming criterion for SSD-dominated and SSD-non-dominated portfolio in A.

References

[1] G. Hanoch, H. Levy: The efficiency analysis of choices involving risk, *Review of Economic Studies* **36** (1969), 335–346.

[2] M. Kopa: Utility functions and portfolio selection problem, *Bulletin of the Czech Econometric Society* **19** (2003), 87–108.

[3] H. Levy: Stochastic dominance and expected utility: Survey and analysis, *Management Science* **38**, 4 (1992), 555–593.

[4] W. Ogryczak, A. Ruszczyński: Dual stochastic dominance and related mean-risk models, *SIAM J. Optim.* **13** (2002), 60–78.

[5] G. Ch. Pflug: Some remarks on the value-at-risk and the conditional value-at-risk, *in Probabilistic Constrained Optimization: Methodology and Applications,* S. Uryasev, ed., Kluwer Academic Publishers, Norwell, MA, 2000, pp. 278–287.

[6] S.Uryasev, R. T. Rockafellar: Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* **26** (2002), 1443–1471.

[7] G. A. Whitmore: Third degree stochastic dominance, *American Economic Review* **60** (1970), 457–459.

[8] G. A. Whitmore: Stochastic dominance for the class of completely monotonic utility functions, *in Studies in the Economics of Uncertainty* Thomas B. Fomby and Tae Kun Seo, eds., Springer Verlag, New York, 1970, 77–88.

Tree Approach to the Time Bounded Transportation Problem

Petr Kučera
Czech University of Agriculture in Prague
Faculty of Economics and Management
Department of Operational and Systems Analysis
kucera@pef.czu.cz

Abstract

The time bounded transportation problem (TBTP) is related to the vehicle routing problem (VRP). The main difference is that the routes are paths (not cycles), i.e. vehicles do not return to the central city. Costs are given for straight routes between each pair of cities and represent time necessary for going through. Each path must not exceed a given time limit. The sum of time for all routes is to be minimized.

The method suggested here is based on combining approaches of the Mayer method for the VRP and the Christofides method for the traveling salesman problem (TSP). The Mayer method, choosing cities for single routes, properly constructs a spanning tree for all cities on each route and these trees can be utilized in the first step of the Christofides method. In addition, the Christofides method achieves the best approximation ratio among all methods for the TSP.

This method is tested on several instances and compared with the saving method, which is considered to be one of the best methods for the TBTP.

Keywords

Time bounded transportation problem, vehicle routing problem, traveling salesman problem, Mayer method, Christofides method

1 Introduction

The time bounded transportation problem (TBTP) is defined as follows: One central city and other n (ordinary) cities are given and for each pair of cities a cost is given, representing time necessary for going through the straight route between them. The cost matrix is supposed to be symmetric. The goal is to find a set of paths so that each of them has one of its endpoints in the central city, its length does not exceed a given time limit and each city except for the central one lies on exactly one of the paths.

This problem has many practical instances, e.g. transportation newspapers from the printers' to shops, grocery products (dumplings etc.) from the manufactory to restaurants, daily reports from affiliated branches to the headquarters etc. Each vehicle is required to visit all the cities on its route until a given time, but we do not mind how it gets from the end back to the start of its route to realize it next time.

Nevertheless, TBTP has been studied relatively little. It belongs to the NP-hard problems, for which there is no efficient algorithm finding their theoretical optimum. It is related to the vehicle routing problem (VRP), where routes are cycles instead of paths, and so to the traveling salesman problem (TSP), i.e. the task to construct one cyclic route containing all the cities, too. Thus heuristics (approximation methods) for the TBTP can be derived from the methods for the VRP and the TSP.

Let us introduce some notation. The central city will be indexed by 0 and the other cities by numbers from 1 to n. The cost matrix will be denoted by $\mathbf{C}$ (and so single costs c_{ij}, $i, j = 0, \ldots, n$).

2 Tree Approach

For the VRP, more preciously for its special case where capacities (demands) of cities and vehicles are specified (so called multipletours traveling salesman problem, cf. e.g. [2]), the Mayer method is often used. It only separates cities into groups so that each group contains cities on a route for one vehicle

and then some method for the TSP must be used to determine the order of the cities on the single cycles. The algorithm of the Mayer method is here:

Mayer Method:

1) Choose from the cities, which have not been put onto any route, such a city i so that c_{0i} is maximum possible (i.e. the remotest city from the central one) as the first city of the currently created route.
2) Let P denote the set of cities already being in the currently created route and Q the set of cities, which have not been put onto any route except for the central one.
   If $Q = \emptyset$
   then stop
   else find the minimum c_{ij} so that $i \in P$ and $j \in Q$.
3) If it is possible to add the city j to the route regarding to the capacities of the cities and the vehicle
   then do it and go to 2) to search for other cities for this route
   else go to 1) to start creating a new route.

In fact, groups of the cities for single routes are not the only output of the Mayer method. Actually, this method works by the same way as the Prim (Jarník) algorithm for the minimum spanning tree. Thus the edges determined in the step 2) form the tree spanning all the cities on the route with the minimum total cost.

Let us remind the well-known Christofides method for the TSP, which uses the minimum spanning tree, too:

Christofides Method:

1) Construct the minimum cost tree T spanning all the cities.
2) Find the minimum cost perfect matching M of the cities with odd degree in T. Now $T \cup M$ forms a Eulerian graph.
3) Find a Eulerian walk in this Eulerian graph.
4) Create a cyclic route (Hamiltonian cycle) by preserving the first occurrence of each city in the Eulerian walk and deleting all other occurrences.

These two methods fit together to design a method for the TBTP:

Tree Approach for the TBTP:

1) Choose from the cities, which have not been put onto any route, such a city i so that c_{0i} is maximum possible (i.e. the remotest city from the central one) as the first city of the currently created route. Set T to the graph consisting of the only vertex i.
2) Let P denote the set of non-central cities in the currently created route and Q the set of non-central cities, which have not been put onto any route.
   If $Q = \emptyset$
   then stop
   else find the minimum c_{ij} so that $i \in P$ and $j \in Q$.
3) Add the vertex j and the edge $\{i, j\}$ to T. Derive T_0 from T by adding the vertex 0 and the edge $\{0, k\}$ such that $k \in P \cup \{j\}$ and c_{0k} is minimum possible. (Note that both T and T_0 are trees).
4) Find the minimum cost "nearly perfect" matching M of all such cities of T that have odd degree in T_0. Obviously, there is odd number of such cities, so exactly one of them, say l, is to be isolated in this matching.
5) Now there are exactly two vertices of odd degree in $T \cup M$: l and 0. Find a walk from l to 0 containing all the edges of $T \cup M$.
6) Create a path by preserving the first occurrence of each city in the walk and deleting all other occurrences.
7) If the total cost (the time for going through) of this path does not exceed the time limit given in the input of this task
   then declare the path from the last executing of the step 6) to be the currently created route and go to 2) to search for other cities for this route
   else declare the path obtained *before* the last executing of the steps 2) to 6) to be a finished route and go to 1) to start creating a new path.

Let us remark that all paths are supposed to contain at least two non-central cities in this formulation of the algorithm.

3 Other Methods

The only reference to another approach to the TBTP, which author has managed to find, is [3] (even only in Czech). It is based on the modification of the savings method for the TSP (cf. [1]). As it is not the main subject of this contribution, let us describe it only briefly. One of the cities, let it have an index 0, is chosen and for all pairs of cities the savings $s_{ij} = c_{i0}+c_{0j}-c_{ij}$ are computed. Then the edges with the maximum possible saving are consecutively added to the solution so that they could form (after adding others) a cyclic route. In the end the city 0 is joined.

The modification for the TBTP in [3] is not completely specified. It is clear, that the central city is to be the city 0 from the above description. But the savings method performs a global search in the whole network and so the route may be disconnected during its construction. If we want to obtain several routes of the solution of the TBTP at once by this way, a question arises what to do when two connected components are to be joined but when it has been done, the cost (time) exceeds the given limit. In the tests in the next chapter we will concentrate on three following possibilities how to manage with this complication:

Version 1: If this situation appears, both this components are declared to be finished and are joined by their closer end vertex to the central city. The same is done when the attempt to add an edge to a single route causes the limit excess.

Version 2: In such cases the edge is not considered and the next edge is processed.

Version 3: Paths are constructed one by one using the local search: First the edge with the greatest possible saving is chosen. Then other edges are consecutively connected to the endpoints of the created path so that their savings are maximum possible and by connecting the central city the time limit is not exceeded, until no such edge exists.

4 Test Computations and Their Results

For testing the same randomly generated cases, which were used in [2] for the VRP, were taken. Let us remind that the central city was located to the middle of the attended region where there were 12 other cities. Of course, the capacities were not considered and the costs represented time. The time limit for routes was set to 250.

Seven of these cases were solved by the tree approach and comparatively by all three versions of the savings method. The reasons why all the cases from [2] were not computed are explained in the next chapter. The results are summarized in the Table 1 in the percentage form, where 100 p.c. is the best result, and in the Table 2 the order of results, which single methods achieve for each test case, is mentioned (1 = the best, 4 = the worst).

	Tree Approach	Savings Method V.1	Savings Method V.2	Savings Method V.3
Case 1	109.9%	114.6%	100.0%	106.2%
Case 2	108.5%	100.0%	100.0%	117.9%
Case 3	103.3%	114.2%	100.0%	100.0%
Case 4	100.0%	114.0%	107.0%	107.0%
Case 5	119.0%	100.0%	100.0%	108.3%
Case 6	105.2%	100.0%	102.8%	100.0%
Case 7	112.1%	100.0%	100.0%	100.0%

Table 1: Test Cases Results in Percentage Form

	Tree Approach	Savings Method V.1	Savings Method V.2	Savings Method V.3
Case 1	3	4	1	2
Case 2	3	1	1	4
Case 3	3	4	1	1
Case 4	1	4	2	2
Case 5	4	1	1	3
Case 6	4	1	2	1
Case 7	4	1	1	1

Table 2: Order of the Test Cases Results

5 When is the Tree Approach Better than the Savings Method?

The tree approach achieved different success at single cases in comparison to the savings method. In the case 4 it gave the best result while in the cases 5 and 7 it completely failed. Thus some properties, which would have influence on the quality of the tree approach solution, were searched for. It was found that there is an important dependence of the results of the tree approach on the ratio between the farthest and the nearest city from the central one among cities lying on the convex hull of the set of all cities. Big values of this ratio properly indicate that the central city does not lie near the middle of the attended region. In the case 4 this ratio was 2.10, in the cases 5 and 7 it was 1.26 and 1.32, respectively, and e.g. in the cases 1, 2 and 3 between 1.65 and 1.8. In the cases 8, 9 and 10 this ratio was small (in the range from 1.2 to 1.65) and so there was not a big hope of a good result of the tree approach. Thus these cases were not computed.

6 Conclusions

The tree approach shows to be suitable for cases where the central city is far from the middle of the attended region. In the opposite case the savings method gives better results (its "right" form is obviously the version 2 from the chapter 4). So the existence of the tree approach extends possibilities how to find a good solution for some instances of the TBTP. There is a question how the tree approach will behave at some different instances, namely larger ones with a greater number of cities, but it can be expected that the both tree approach and savings method will achieve their good results at the similar types of instances as at the observations here.

References

[1] Clarke, G., Wright, J.W.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points, Operations Research, 12, p. 568-581, 1964

[2] Kučera, P., Dömeová, L.: Solution of the Multipletours Traveling Salesman Problem with Vehicle Capacity Preference, Mathematical Methods in Economics, Prague, 2003

[3] Pelikán, J.: Časově omezený rozvozní problém, Firma a konkurenční prostředí, Brno, 2005 (to appear)

Long Memory in Volatility or Parameter Inconstancy? The Case of Prague Stock Exchange

Alexandr Kuchynka

Dept. of Statistics and Operations Research
Faculty of Economics
University of West Bohemia in Pilsen
Email: alexk@kso.zcu.cz

Abstract

The paper summarizes some stylized facts in financial returns, in particular so called long-range dependence and IGARCH effect. The long range dependence demonstrates itself by a very slow decay of the sample autocorrelation function of absolute and squared returns, especially at larger lags. However, it has been argued elsewhere that these effects can be explained by nonstationarity due to shifts in unconditional variance. One possibility how changes of unconditional variance in GARCH model can occur is to allow for time-varying parameters. In this paper, we verify whether the above mentioned stylized facts can be observed in Prague stock index PX 50 and we perform the test of parameter constancy of the estimated GARCH model.

Keywords

GARCH models, long-range dependence, IGARCH effect, structural breaks, PX 50

1 Stylized facts and a possible explanation

In accordance with other authors (cf. [11]), we say that log returns of stock indices, exchange rate etc. display the *long-range dependence (LRD) effect*, if the sample autocorrelation functions (ACF) of absolute and squared returns have the following features: they are all positive, decay relatively fast for the first few lags, and at larger lags are low but decay very slowly. Moreover, we can often observe that the autocorrelations of absolute returns exceed those of squared returns (so called Taylor effect). These phenomena have been extensively documented in the literature (for instance, [4] show the slow decay of the sample ACF of powers of absolute daily log returns in the Standard and Poors series for years 1928 – 1990).

It has been shown that the popular GARCH(1,1) model given by

$$\varepsilon_t = h_t^{1/2}\xi_t$$
$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$$

where $\{\xi_t\}$ is i.i.d. with zero mean and unit variance, is not flexible enough to capture these stylized facts (see [10], among others). One of the reasons is that GARCH(1,1) process is strongly mixing with geometrical rate which implies that it „forgets“ its past quickly. The

same applies to the processes of its absolute values and squares which should also exhibit short memory property, i.e. their sample ACF should vanish quite quickly. This can be related to the so called *IGARCH effect* whose name comes from the observation that the fitting of GARCH(1,1) models on log returns very often results in obtaining estimates such that $\alpha+\beta$ is close to one. This is true in particular for longer samples, in shorter sub samples the sum of both coefficients is found to be substantially smaller than one (see [1] and the references therein). Nevertheless, it has been argued that both LRD and IGARCH effects can be easily generated by nonstationarity due to shifts in unconditional variance rather than by the genuine long-memory in the data. Note that in weakly stationary GARCH(1,1) models the change of the unconditional variance corresponds to change in its parameters, the variance being given by $\omega/(1-\alpha-\beta)$.

The fact that the structural break in unconditional variance can be easily mistaken for long memory has been extensively studied by Mikosch and Starica [11]. Their work was anticipated by Lamoureux and Lastrapes [8] who were among the first ones who noted the relationship between breaks in unconditional volatility and estimated volatility persistence. However, the latter did not provide any theoretical explanation of this phenomenon nor they offered systematic testing procedure able to detect such breaks. Some theoretical perspective can be found in Granger and Hyung [6] and Diebold and Inoue [3]. During the last ten years, several tests of structural stability in GARCH models have been proposed in the literature. In this paper, we will focus on Lundbergh and Teräsvirta (2002) (hereafter LT) and modified Inclan and Tiao (1994) (hereafter IT) tests.

LT test of parameter constancy of GARCH model was derived as a part of a unified framework aimed to check the adequacy of an estimated GARCH model. As the alternative a model with smoothly changing parameters is considered:

$$h_t=\left[\omega_1+\alpha_1\varepsilon_{t-1}^2+\beta_1 h_{t-1}\right]\left[1-F(t)\right]+\left[\omega_2+\alpha_2\varepsilon_{t-1}^2+\beta_2 h_{t-1}\right]F(t)$$

where $F(t)=\dfrac{1}{1+\exp(-\gamma t)},\gamma>0$ is the logistic transition function. Under the null $H_0:\omega_1=\omega_2,\alpha_1=\alpha_2,\beta_1=\beta_2$, the parameter γ of the transition function is not identified. Nonetheless, this can be circumvented by replacing the transition function by a first-order Taylor approximation. This yields the following auxiliary model

$$h_t=\delta_1+\delta_2\varepsilon_{t-1}^2+\delta_3 h_{t-1}+\delta_4 t+\delta_5\varepsilon_{t-1}^2 t+\delta_6 h_{t-1}t$$

Now the null becomes $H_0':\delta_4=\delta_5=\delta_6=0$ which can be tested by means of the Lagrange multiplier test. In practice, the test is carried out using an artificial regression (for more details, see [9]). Of course, it is possible to test the constancy of a subset of parameters only.

On the other side, the test suggested by Inclan and Tiao [7] is a CUSUM-type test. Define the cumulative sum of squares $C_k=\sum_{t=1}^{k}a_t^2, k=1,...,T$ and the corresponding centered and normalized variable $D_k=C_k/C_T-k/T$. The proposal of Inclan and Tiao was to employ the statistic $IT=\sup_k\left|\sqrt{T/2}D_k\right|$. Under the assumption that a_t are normally, *iid* random variables with zero mean and constant variance, the asymptotic distribution of this test is given by

$$IT\xrightarrow{\mathcal{L}}\sup_r\left|W^*(r)\right|$$

where $W^*(r) \equiv W(r) - rW(1)$ is a Brownian bridge and *W(r)* is a standard Wiener process. Whereas the assumption of normality is not crucial and can be easily relaxed, the independence requirement turns out to be a more serious issue. If the test is applied to conditionally heteroskedastic processes like GARCH, it suffers from severe size distorsions which precludes its application to financial time series. Therefore, there is a need for its modification which would cover this case of dependence in the data. Fortunately, Sansó et al. [12] proposed the modified IT statistics $IT_m = \sup_k \left| T^{-1/2} \hat{\omega}_4^{-1/2} D_k \right|$ where $\hat{\omega}_4$ is some consistent estimator of the long-run fourth moment of a_t. If multiple structural breaks are supposed to occur in the analysed series, the use of the iterated cumulative sum of squares (ICSS) algorithm is recommended (see [7] for a detailed description).

2 Data

In this paper, we analyse daily log-returns of Prague Stock Exchange index PX 50. The series covers the period between January 6,1995 and January 3, 2005 and contains 2490 observations (see Figure 1)[1]. The descriptive statistics are summarized in Table 1.

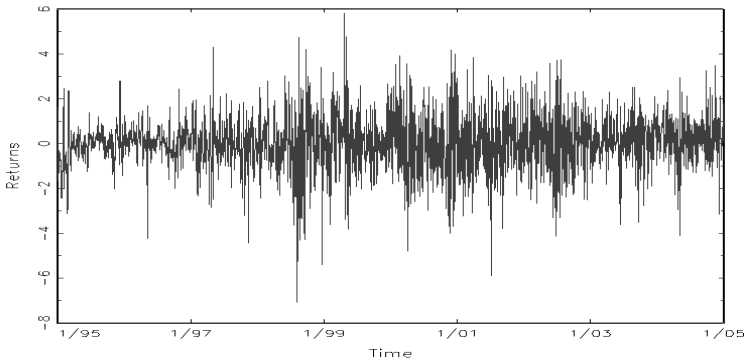


Figure 1. Plot of the log-returns of PX 50 index

Mean	*Minimum*	*Maximum*	*Median*	*St. dev.*	*Skewness*	*Kurtosis*
0.0246	*-7.0772*	*5.8200*	*0.0248*	*1.2055*	*-0.1867*	*5.0360*

Table 1. Summary statistics for daily returns (2490 observations)

It is interesting to examine autocorrelations functions (ACF) of raw, absolute and squared returns (see Figure 2). It is clearly visible that returns are correlated very little whereas ACF for absolute returns remains positive even for large lags. Note that Taylor effect also seems to be present in the data.

[1] The dataset was obtained from the website of the Prague Stock Exchange and from Datastream. The analysis of the series has been carried out using GAUSS 6.0 (the purchase of the software was funded by the grant of FRVŠ nr. 1306Ab/2004). To a large extent, we built on the code written by Dick van Dijk and Andreu Sansó.

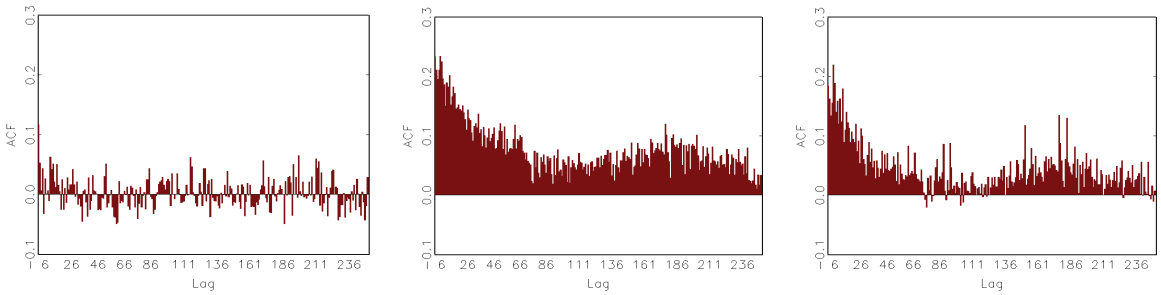


Figure 2. Sample autocorrelation functions for the PX-50 returns, absolute returns and squared returns, respectively, calculated from the whole sample.

In the next step, an AR(1)-GARCH(1,1) model was estimated by quasi maximum likelihood method. The results are not reported here in order to save space but they are available upon request. However, it should be mentioned that the estimate of the sum $\alpha + \beta$ equals 0.9892 which suggests highly persistent GARCH process. In order to check whether this could be caused by nonstationarity in unconditional variance, we performed both LT and modified IT tests. The results are reported in Table 2 and 3, respectively. The LT test suggests a change in the intercept which implies a shift in the unconditional variance; however, it is silent about the location of such a break (or breaks). This is supported by the result of modified IT test which detected four breaks.

test	*LM statistic*	*p-value*
intercept	*8.23*	*0.004*
ARCH parameter	*0.10*	*0.758*
all parameters	*10.27*	*0.016*

Table 2. Results of Lundbergh-Teräsvirta test of parameter constancy (χ^2 form)

observation	*52*	*504*	*857*	*1886*
date	*29.3.1995*	*4.2.1997*	*1.7.1998*	*5.8.2002*

Table 3. Location of breaks in unconditional variance as indicated by modified Inclán-Tiao test

If we split the whole sample into five subperiods as suggested by the results of the modified IT test, we find out that they differ quite substantially. For instance, the second period can be characterized as a tranquil one followed by a burst in volatility in the years 1997 and 1998. Finally, in the last period we witness a decrease in volatility. Moreover, it seems that unconditional mean is time-dependent as well. The descriptive statistics for sub samples can be found in Table 4.

sub sample	*nr. of observations*	*mean*	*minimum*	*maximum*	*st.deviation*
06/01/1995 – 28/03/1995	*51*	*-0.5658*	*-3.1260*	*2.4586*	*1.2868*
29/03/1995 – 03/02/1997	*452*	*0.0629*	*-4.2253*	*2.7923*	*0.6228*
04/02/1997 – 30/06/1998	*353*	*-0.0553*	*-4.4366*	*4.3134*	*1.0227*
01/07/1998 – 02/08/2002	*1029*	*-0.0039*	*-7.0772*	*5.8200*	*1.5076*
04/08/2002 – 03/01/2005	*605*	*0.1407*	*-4.0957*	*3.4783*	*1.0174*

Table 4. Descriptive statistics for subperiods

Time-varying character of volatility can also be confirmed by a visual inspection of the evaluation-weighted estimate for volatility obtained by using one-sided estimator with normal kernel. The result is shown in Figure 3. The picture reflects slowly growing volatility between the years 1995 and 1998 and the consequent sudden rise. On the contrary, the last period beginning at the end of 2002 seems to be more tranquil.

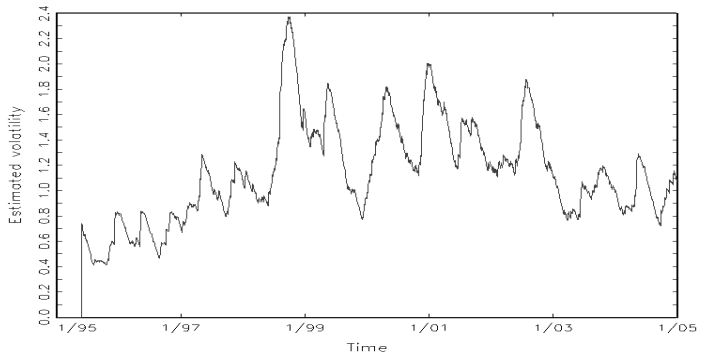


Figure 3. Estimated time-varying standard deviation using one-sided smoothing

It is interesting to examine ACF of absolute returns for individual subperiods separately. In Figure 4 we see that the ACF die out quite quickly. This suggests that the so called „long-memory effect“ characterized by slow decay of ACF and observed in the whole sample is spurious.

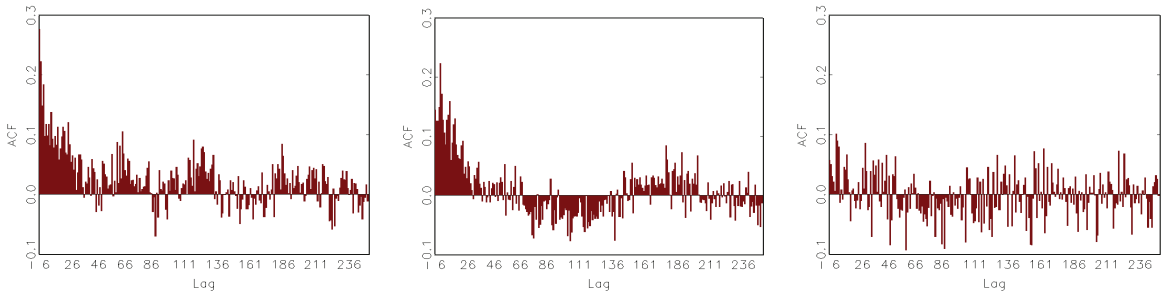


Figure 4. Sample autocorrelation functions of the PX-50 absolute returns, calculated for three non-overlapping sub samples (observations 1-856,857-1885,1886-2490)

3 Conclusions

There exists a growing amount of literature trying to assess the adequacy of standard GARCH models. Due to the fact that their performance is often considered unsatisfactory, two approaches can be taken: the first one consists in „augmenting“ the baseline GARCH model with more sophisticated features like more complex nonlinearity or fractional integration which should allow for a richer dynamics and which are supposed to be relevant to explain empirically observed stylized facts. However, there is another explanation why GARCH models possibly do not do their job well and this stresses the role of structural breaks in the data. In fact, both approaches have distinct implications for forecasting: a genuine long memory (generated typically by fractional integration) implies that even observations in the distant past have influence on the present and therefore are still relevant for forecasting whereas the presence of structural breaks suggests exactly the opposite. For this reason, it is very likely that this problem will keep the academic community busy in the future.

References

[1] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31, 307-327.

[2] Bollerslev, T., R.Y. Chou, and K.F. Kroner (1992), ARCH modelling in finance. A review of the theory and empirical evidence, *Journal of Econometrics* 52, 5-59.

[3] Diebold, F. X. and A. Inoue (2001), Long Memory and Structural Change, *Journal of Econometrics*, 105, 131-159.

[4] Ding, Z., C.W.J. Granger, and R.F. Engle (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83-106.

[5] Granger, C.W.J. and Z. Ding (1995), Some properties of absolute returns. An alternative measure of risk, *Annales d´économie et de statistique* 40, 67-92.

[6] Granger, C.W.J. and N. Hyung (1999), Occasional structural breaks and long-memory, *Discussion paper 99-14*, University of California, San Diego.

[7] Inclán, C. and G.C. Tiao (1994), Use of Cumulative Sums of Squares for Retrospective Detection of Changes in Variance, *Journal of the American Statistic Association* 89, 913-923.

[8] Lamoureux, C.G. and W.D. Lastrapes (1990), Persistence in Variance, Structural Change and the GARCH Model, *Journal of Business and Economic Statistics* 8, 225-234.

[9] Lundbergh, S. and T. Teräsvirta (2002), Evaluating GARCH Models, *Journal of Econometrics*, 110, 417-435.

[10] Malmsten, H., and T. Teräsvirta (2004), Stylized facts of financial time series and three popular models of volatility, *SSE/EFI Working Paper Series in Economics and Finance 563*, Stockholm School of Economics.

[11] Mikosch, T. and C. Stărică (2004), Nonstationarities in Financial Time Series, the Long-Range Dependence, and the IGARCH Effects, *Review of Economics and Statistics* 86, 378-390.

[12] Sansó, A., V. Arragó, and J.L. Carrion (2004), Testing for Change in the Unconditional Variance of Financial Time Series, *Revista de Economía Financiera* 4, 32-53.

Modification of the EOQ Model for the Annual Constant Demand Situation and the Possibility of its Usage in a Supply Chain[1]

Martina Kuncová
University of Economics, Prague

Abstract

Supply chain management is the latest concept in the attempt to optimize the entire process of manufacturing the product, right down to the point of its consumption, and all the stages in-between. Inventory is an important supply chain driver because changing inventory policies can dramatically alter the supply chain's efficiency and responsiveness. The basic EOQ model and formula has been created to solve the problem of the optimal order quantity via the inventory cost minimization. Demand is expected to be fixed and for the model it is necessary to know the inventory ordering and holding cost. Because of the fact that the annual fixed demand may or may not be fixed for the shorter period (month, week, working day) and the curves of the fixed and variable inventory cost need not be smooth (depending on the order quantity), the new model, playing upon the EOQ model and its presumptions, is presented. Owing to the complexity of the new model, the procedure of using the MS Excel software for obtaining the best solution is demonstrated as well as the possibility of using this model in a small supply chain.

Keywords

Inventory optimization, EOQ model, fixed demand, inventory ordering cost, inventory holding cost, supply chain, network economy

1 The Basic Facts about a Supply Chain

Supply chain can be characterized as a net that consists of suppliers, manufacturing centres, warehouses, distribution centres, and retail outlets, as well as raw materials, work-in-process inventory, and finished products that flow between the facilities. Because of the fact that each subject can be a part of various chains, it is possible also to talk about the network economy, where the small supply chain, that we may observe, is only one of its segments. All the subjects of the selected supply chain are connected by three flows in both ways – downstream (to the customer) or upstream (to the supplier). These three flows are: information flow, materials flow and financial flow (Fig. 1).

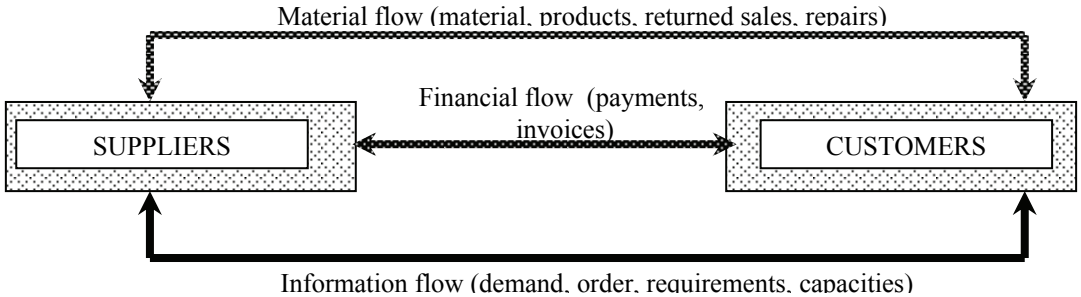


Fig 1: The three basic flows in a supply chain

[1] This paper was supported by Grant No. 402/05/0148 from the Grant Agency of the Czech Republic

2 Inventory Control

The inventories are usually held because of a mismatch between supply and demand. Inventory optimisation and inventory control belong to the most important fields in business companies. In case no appropriate software is available, managers must estimate the order quantity and order time themselves, mostly on the basis of their individual knowledge of customers' behaviour. Nowadays, when it is necessary to be very quick and perfect in reaction to the customer's demand, this approach seems to be insufficient. Of course the best way is to use some special software, but corresponding cost of it may be quite significant and additionally, the installation of such systems may require radical changes in company's structure. It also can take a year or more to install and to get familiarized with it. Even big firms should consider buying the expensive software specialized just on the partial problems in company's operations. However, as it will be shown later, using standard spreadsheets (e.g. MS Excel) is suitable access to solve the company's problems. There are at least two possibilities how to manage the inventory – analytically, using mathematical model, or apply a simulation theories and methods on this problem.

2.1 Retailer's inventory control for the fixed demand

The wide spread mathematical method for finding the optimal order quantity is via the EOQ formula. The term EOQ – Economic Order Quantity – has been used for a long time. It can be instrumental in finding out how many products should be ordered to minimize the inventory costs. These costs include the inventory holding (or carrying) costs and the ordering costs. The inventory holding costs tell us how much money it costs to hold one unit of the product in stock for one period (usually one year). The ordering costs indicate the sum of money we have to pay for one transfer of the ordered products from the supplier to our company. The basic EOQ formula is

$$q^* = \sqrt{\frac{2Qc_2}{c_1}}$$

where: q^* = the optimal order quantity (EOQ) for minimization of the inventory costs

Q = constant demand in units per one period

c_1 = the inventory carrying costs (per one unit per one period)

c_2 = the costs per order

Using this strategy, the minimal total inventory costs per year are:

$$TC^* = \sqrt{2 \cdot Q \cdot c_2 \cdot c_1}$$

In this case we suppose that the demand and the costs are constant through the year and that there are no lead times (or they are very small and fixed).

One of the main problems when using this formula is that the formula supposes the ordering costs to go down smoothly in dependence on the number of products ordered. But in reality it need not be true. If these costs rise only when an order is made, then the curve is not so smooth (Fig.2).

Comparing these two situations we can find out that the EOQ formula overestimates the inventory costs. Let's show it on the example.

Suppose the annual demand is 260 units, the inventory ordering costs are 120 crowns and the inventory holding costs are 18 crowns per unit per year. Following the EOQ formula, we obtain these results:

$$q = \sqrt{\frac{2 \cdot Q \cdot c_2}{c_1}} = \sqrt{\frac{2 \cdot 260 \cdot 120}{18}} \approx 59 \qquad TC^* = \sqrt{2 \cdot Q \cdot c_2 \cdot c_1} = \sqrt{2 \cdot 260 \cdot 120 \cdot 18} = 1059{,}8$$

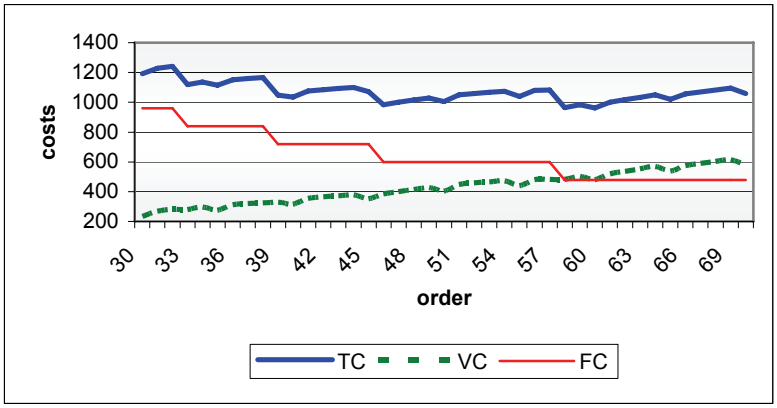


Fig 2: Total, fixed and variable costs for various orders

For the same situation it is possible to use MS Excel. But now it is necessary to choose the right period for monitoring the inventory. The period should correspond with the demand, the fluctuation of inventory and also with the length of potential lead time. We can prove that the annual inventory costs can be dependent of the period length.

If we choose a month as a significant interval, it is possible to find out the average demand per period:

$\overline{Q_{i,k}} = \frac{Q}{k}$, where k is the number of intervals during the year (here k=12), so here $\overline{Q}_{i,k} = 21{,}667$.

As you can see, there might be a problem with non-integer demand. If we suppose real demand to be nearly the same in every period, we can also define that it may oscillate between lower bound $Q_{i,k}^{\ l}$ and upper bound $Q_{i,k}^{\ u}$, where $Q_{i,k}^{\ l}$ is the nearest smaller integer number and $Q_{i,k}^{\ u}$ is the nearest higher integer number. Then we have to calculate the probability of these demands to achieve the given average demand:

$$\overline{Q_{i,k}} = p_l \cdot Q_{i,k}^{\ l} + p_u \cdot Q_{i,k}^{\ u}$$

where p_l is the probability of the $Q_{i,k}^{\ l}$ demand occurence and p_u is the probability of the other possibility $Q_{i,k}^{\ u}$. The sum of these probabilities must be equal to 1, so according to the formulas mentioned above we can specify the probabilities:

$$\overline{Q_{i,k}} = p_l \cdot Q_{i,k}^{\ l} + (1-p_l) \cdot Q_{i,k}^{\ u} = p_l (Q_{i,k}^{\ l} - Q_{i,k}^{\ u}) + Q_{i,k}^{\ u}$$

$$p_l = \frac{\overline{Q_{i,k}} - Q_{i,k}^{\ u}}{Q_{i,k}^{\ l} - Q_{i,k}^{\ u}}$$

$$p_u = 1 - p_l$$

In the test case the probabilities are:

$$p_l = \frac{\overline{Q_{i,k}} - Q_{i,k}^{\ u}}{Q_{i,k}^{\ l} - Q_{i,k}^{\ u}} = \frac{21{,}667 - 22}{21 - 22} = 0{,}333 = 1/3$$

$$p_u = 1 - p_l = 1 - 1/3 = 2/3$$

To obtain the estimation of the real demand we can use a simulation. If we know the probabilities of the two possible cases, we generate demand in MS Excel via the functions IF and RAND(), where the function RAND() generates a random variable between zero and one. The form of the function is[2]:

$$=\text{IF}(\text{RAND}()<p_l;O_{i,k}^{\ l};Q_{i,k}^{\ u}),$$

so for the test case :

$$=\text{IF}(\text{RAND}()<1/3;21;22)$$

Total annual inventory costs are made by the total annual fixed costs and total annual variable costs. Fixed costs rise only when an order is made, so we can create a binary variable b_i that is equal to one

[2] In Czech version of MS Excel is the function IF called KDYŽ and the function RAND() calles NÁHČÍSLO()

only when ordering a given number of goods. Variable costs are dependent on the inventory level at the end of the period. This level is determined by the previous level, actual order and actual demand. It is clear that all the levels depend on the start inventory for a given year. According to these assumptions we can define a mathematical model (for the retailer to minimize the inventory costs) as follows:

Minimize

$$TC_r = \sum_{i=1}^{k} TC_i = \sum_{i=1}^{k} (FC_i + VC_i) = \sum_{i=1}^{k} b_i c_2 + \sum_{i=1}^{k} Z_i \cdot \frac{c_1}{k} = \sum_{i=1}^{k} b_i c_2 + \sum_{i=1}^{k} (Z_{i-1} + b_i \cdot q_i - Q_i) \cdot \frac{c_1}{k}$$

Conditions:

$$Z_i - Z_{i-1} - b_i \cdot q_i + Q_i = 0$$

$$-Z_{i-1} - b_i < 0$$

$q_i \geq 0$ and q_i is an integer number,

$Z_i \geq 0$ and Z_i is and integer number,

$b_i \geq 0$ a $b_i \in \{0;1\}$ i = 1, 2, …, k

This model is too complicate to solve it via MS Excel solver. But it is possible to use the EOQ model result, the optimal order q, generate demand for all periods and then use data tables to count the costs for another quantity. For the test case, q is equal to 59 and the order is made only when the inventory level in the previous period is lower than the reorder point r. If the lead time is less or equal to one period, then the reorder point r can be set to the value $Q_{i,k}^{u}$. So the function that was made for signalizing an order is:

$$=IF(Z_{i-1}<r;q^*;0)$$

for the test case: $= IF(Z_{i-1}<22;59;0)$

Using all the functions mentioned above we can obtain various results (because of the small variability of demand). One of the simulation experiments with all necessary facts are in the Table 1.

Table 1: Annual fixed, variable and total costs for q=59 and monthly inventory monitoring (in case of the start inventory level is equal to 30 units)

	RETAILER				
	demand	**order**	**inventory**	**FC_i**	**VC_i**
month	**Q_i**	**q**	**Z_i**		
1	22	0	8	0	12
2	21	59	46	120	69
3	22	0	24	0	36
4	22	0	2	0	3
5	21	59	40	120	60
6	22	0	18	0	27
7	22	59	55	120	82,5
8	22	0	33	0	49,5
9	22	0	11	0	16,5
10	22	59	48	120	72
11	21	0	27	0	40,5
12	21	0	6	0	9
Total	260	236	318	**480**	**477**
Average			26,5		
			TC_r=	**957**	

As you can see, a strategy of ordering the EOQ amount can give us lower costs than expected (other simulations experiments gave the similar results). A radical change can occur when there is no inventory at the beginning of the year. In this situation simulation showed that the average costs are around 1150 crowns (more than expected). So the initiatory inventory level has a big impact on the costs.
Data tables can show that the given amount must not be the best way of ordering. In case there is no inventory, then the best strategy is to order 66 units because the total annual costs are 913 on average (whereas for the q=59 they are around 1150 crowns per year). For the case of 30 units in inventory at the beginning the best strategy could be ordering 58 units with the average annual inventory costs 920 crowns.
So for the retailer to find out the best order there exist a possible way how to do it: in case of known inventory holding and ordering costs, it is possible to create a table in MS Excel (using the result of EOQ for the first q) with interdependencies described above and with the given start inventory level. Then create a data table with tested ordering quantities and through various simulation experiments find the best quantity to order.

2.2 Supplier's inventory control for the fixed demand

A supplier is in worse position. To be able to find the best strategy, it is necessary to know whether the supplier can influence the decision of the retailer or not and if he has all the information about demand and ordering strategy or not.
If the managers have all the information from the retailer and must take the retailer's order as a fact, they are able to count the period when the retailer orders a how much it will be. Then one of the possibilities is to order one period before (if we suppose the lead time to be smaller or equal to one period) and pass the whole order to the retailer, so the inventory holding costs are minimal and the fixed (ordering) costs are the same as for the retailer. Therefore the annual inventory costs of the supplier are lower than for the retailer (if we suppose the supplier to be exposed to pay the same inventory ordering and holding costs). For our test case it is possible to try various orders of the retailer and supplier and count the costs – everything is in the Table 2:

Table 2 – Total annual inventory costs of the retailer, supplier and the chain when various order strategies are used (the initial inventory level is 30 for the retailer and 0 for the supplier)

Retailer's order	Supplier's order	Avg. TC retailer	TC supplier	TC chain	Optimum
59	59	960	480	1440	Order - EOQ formula
59	118	960	362,5	1322,5	-
63	63	940	480	1420	-
63	126	940	370,85	1310,85	-
58	58	**920**	480	1400	Retailer
58	118	**920**	370,85	**1290,85**	Retailer, Supply Chain
260	260	2420	**120**	2540	Supplier

The results show us that using the EOQ formula to find the best ordering strategy does not lead to the optimal solution neither for the retailer nor for the supplier (and nor for the supply chain). If the supplier could influence the retailer's order, than for it the best strategy is to order once a year (the order is equal to the year demand) and than put everything to the retailer. But this is not a good solution, because retailer's costs are too high. If the retailer uses data tables to find the best strategy, it orders 58 units with the minimum average inventory costs 920 crowns. Provided the supplier only copies this order, it also does not lead to the best solution for him (not even for the supply chain). But if the supplier tries to find its best strategy (including the retailer's one), then it is possible to find an optimum for the whole chain.

2.3 Best solution for the supply chain

To find a best solution for the whole chain it is necessary to know all the information about demand, costs, orders, about the strategy of the other members of the chain and also to respect all the conditions of the members.

In case retailer and suppliers know all the costs connected with inventory and in case that the demand is nearly fixed, it is possible to find a best strategy, that is optimal for the whole chain and is also good for the individual members. As was described above, data tables and various functions in MS Excel can help with this problem. But the main condition is sharing all the information through the chain and the possibility to cooperate.

3 Conclusion

Managing the supply chain is nowadays seen as a one big and hard task because it is a very complex network of various companies and individuals with different problems, needs and possibilities. Sometimes the only achievable goal is to optimize a part of the network – but it is still better than do nothing. In case of looking for the best ordering strategy to reach the minimum inventory costs (provided the inventory holding and ordering costs are known) when annual demand is stable but demand in shorter periods is slightly variable, we cannot recommend to use an EOQ formula. If the firms can not use special software to find a best solution, then using MS Excel, its functions and data tables is highly recommended. Via this tool it is possible to simulate demand, calculate costs and search for the favourable decisions that could lead to the supply chain optimum.

References

[1] Chopra, S., Meindl, P.: Supply Chain Management: Strategy, Planning and Operation. Prentice Hall Inc., New Jersey 2001

[2] Donovan, R.M.: Inventory Efficiency in the Supply Chain. http://www.rmdonovan.com/pdf/perfor3.pdf

[3] Donovan, R.M.: Reduce Inventories and Improve Business Performance. http://www.rmdonovan.com/pdf/perform_97_7.pdf

[4] Fiala, P.: Modelování a analýza produkčních systémů. Professional Publishing, Praha 2002

[5] Fiala, P.: Modelování dodavatelských řetězců. Professional Publishing, Praha 2005

[6] Hausman W.H.: *Information Distortion and Collaboration: The Bullwhip Effect,* http://www.supplychain.com/

[7] Jablonský, J.: Operační výzkum. Professional Publishing, Praha 2002

[8] Simchi-Levi D., Kaminsky P., Simchi-Levi E.: *Designing and Managing the Supply Chain*. USA, McGraw-Hill 2000

[9] Ter-Manuelianc, A.: Matematické modely řízení zásob (Mathematical Models of Inventory Control – in Czech). Institut řízení, Prague 1980

[10] Waller, D.L.: Operations Management: A Supply Chain Approach. International Thomson Business Press, London, UK, 1999

Author:
Ing. Martina Kuncova
Department of Econometrics
University of Economics
W.Churchill Sq. 4
130 67 Prague 3
kuncovam@vse.cz

Some Notes to Black-Scholes Equation

Ladislav Lukáš
University of West Bohemia, Faculty of Economics,
Department of Statistics and Operations Research,
Husova 11, 30614 Plzeň
lukasl@kso.zcu.cz

Abstract

Paper concerns with well-known fair-pricing model of European options described by Black-Scholes equation. The notes discuss various links between diffusion model and B-S pricing model in order to broaden understanding of the B-S equation and its solution.

Keywords

Pricing model, Black-Scholes equation, diffusion model, standard heat flow equation.

1 Introduction

For certain stochastic systems, there is an interesting connection between statistics and dynamics. This one occurs in a rather evident way when discussing the Black-Scholes equation. Well, an option is a contract that gives the buyer the right, but not the obligation, to buy or sell a given asset at a future time for a fixed price, called the strike price. American options can be exercised at any time up to and including the day of expiration of the option. On the contrary, the European options can only be exercised on the day of expiration of the option.

Issuing of options as financial instruments intends to protect investors from the stock market randomness. Since the early seventies, the option market became rapidly to be very successive in its development. The theoretical studies of options were directed in finding fair and presumably riskless prices of these instruments. The turning point in those studies is the fundamental work of Black and Scholes [1], no doubt. Since that time, their method has been proven to be very useful for investors trading in various option markets. Therefore, the B-S equation is still most popular in quantitative finance analysis.

The important requirements under which a theoretical model of fair option price has been developed are generally known, and they include

i) absence of arbitrage opportunities, i.e., identical cash flows have the identical values,
ii) efficient market hypothesis, i.e., the market incorporates instantaneously any information concerning future market evolution,
iii) existence of a unique riskless strategy for a portfolio in a complete market,
iv) the underlying asset value follows a geometric Brownian motion,
v) the stock representing underlying asset does not pay a dividend during the life of the option.

The standard derivation of the B-S equation is based upon the assumption iv), i.e., the geometric Brownian motion for the underlying asset is described by the stochastic differential equation

$$\mathrm{d}s = \mu s \mathrm{d}t + \sigma s \mathrm{d}W, \tag{1}$$

where t is the time, s denotes the asset value, and μs and $\sigma^2 s^2$ are drift and variance of the random walk, respectively. Since $\mathrm{d}W$ is the increment of a Wiener process with probability distribution of $N(0,\mathrm{d}t)$, i.e., a normal distribution with zero average and variance $\mathrm{d}t$, the resulting trajectory will be continuous. This assumption also maintains that the asset value obeys a lognormal distribution, because Eq.(1) is equivalent to $\mathrm{d}s/s = \mu \mathrm{d}t + \sigma \mathrm{d}W$. However, crucial step of standard derivation of the B-S equation is an application of the Ito's formula.

At present, there is a huge literature both theoretical and empirical ones devoted to various problems concerning option pricing, e.g. see [2],[3],[6],[7],[9]-[12]. However, the very basic result assumes a risk neutral environment, which enables expressing the mean values of European call/put options at expiration time T as follows

$$\mathrm{E}[\max(s(T)-x,\,0)] \quad , \quad \mathrm{E}[\max(x-s(T),\,0)] \,, \tag{2}$$

where $s(T)$ is the asset value at the expiration time T of the option, and x is the option strike price.

The option price problem is to find a fair price for the option today. When assuming $T-t$ time period to expiration, the solution of the B-S equation as an expectation value of the final price at expiration brought at present value by risk-free interest rate r gives the corresponding E-call/-put options the following expressions

$$C(t) = \exp(-r(T-t))\,\mathrm{E}[\max(s(T)-x,\,0)] \quad , \quad P(t) = \exp(-r(T-t))\,\mathrm{E}[\max(x-s(T),\,0)] \,, \tag{3}$$

where $C(t)$ and $P(t)$ denote the fair prices avoiding an arbitrage of the E-call and E-put options at the time t, respectively.

In the simplest case of a European option, the B-S equation can be solved explicitly in order to obtain the resulting well-known E-call and E-put option price formulae. They reads as follows

$$C(t) = s(T)N(d_1) - x\,\exp(-r(T-t))\,N(d_2) \quad , \quad P(t) = x\,\exp(-r(T-t))\,N(-d_2) - s(T)N(-d_1) \,, \tag{4}$$

where $N(d_1)$, $N(d_2)$ are values of distribution function of the standard normal distribution $N[0,1]$, such as $N(a) = (2\pi)^{-1/2} \int_{-\infty}^{a} \exp(-\xi^2/2)\mathrm{d}\xi$, at the arguments $d_1 = (\ln(s(T)/x) + (r+0.5\sigma^2)(T-t))/(\sigma\,\mathrm{sqrt}(T-t))$, and $d_2 = d_1 - \sigma\,\mathrm{sqrt}(T-t)$, and finally, σ denotes volatility.

In general, the B-S equation is nothing else but a diffusion equation, i.e., a partial differential equation of parabolic type, with one-dimensional state variable, i.e., fair option price. However, solutions of the diffusion equation with other initial and/or boundary conditions given by the option contract terms can be obtained by numerical procedures.

This article follows up with [8].

2 Difussion equation

In order to pick up the links between diffusion and B-S equation we remind briefly the corresponding physical background, see [4],[5]. Let assume line random walk of a particle at discrete time steps $k\tau$, k=1,2,..., realizing by each time step τ a space shift of specified length h either to the right with given probability p, or to the left with complementary probability $q = 1-p$, when undergoing independent random impacts. The Bernoulli scheme gives a probability $f(x,t)$ of the particle to be at the position x after running a random walk during time period $t = n\tau$, when starting from $x = 0$ at $t = 0$

$$f(x,t) = C(m,n)p^{\mathrm{m}}\,q^{\mathrm{n-m}} \,, \tag{5}$$

where m and n-m denote number of shifts made to the right and to the left, respectively. Balancing walked shifts in order to get x, one reads $mh - (n-m)h = x$, or $m - (n-m) = x/h$. Whereas, resuming to reach x during next period τ from all possible current positions, one gets a difference equation for $f(x,t)$

$$f(x,\,t+\tau) = p{\cdot}f(x-h,\,t) + q{\cdot}f(x+h,\,t) \,, \tag{6}$$

with evident initial conditions $f(0,0) = 1$, and $f(x,0) = 0$, for any $x \neq 0$. as – as

In order to get well-known partial differential equation of parabolic type describing un-steady diffusion phenomenon in one-dimensional state space, first we subtract the term $f(x,t)$ to both sides of Eq.(6), further assuming that $f(x,t) \in C^{2,1}$ we approximate time and space differences using Taylor expansions with the simplest possible expressions letting both h, $\tau \to 0$, and finally we recast the equation (6) into resulting form as follows

$$f(x,\,t+\tau) - f(x,t) = p\,f(x-h,\,t) + q\,f(x+h,\,t) - f(x,t) = p(\,f(x-h,\,t) - f(x,t)) + q(\,f(x+h,\,t) - f(x,t)) \,,$$

$$f(x,\,t+\tau) - f(x,t) \approx \tau\,\partial f(x,t)/\partial t \,,$$

$$f(x-h,\,t) - f(x,t) \approx -\,h\,\partial f(x,t)/\partial x + h^2/2\;\partial^2(x,t)/\partial x^2 \,,$$

$$f(x+h,\,t) - f(x,t) \approx \;h\,\partial f(x,t)/\partial x + h^2/2\;\partial^2(x,t)/\partial x^2 \;\,,$$

$$\tau\,\partial f(x,t)/\partial t + o(\tau) = -(p-q)\,h\,\partial f(x,t)/\partial x + h^2/2\;\partial^2(x,t)/\partial x^2 + o(h^2)\,,$$

$$f_t = -\,c\,f_x + k\,f_{xx}\,, \tag{7}$$

where as usual f_t , f_x and f_{xx} denote $\partial f(x,t)/\partial t$, $\partial f(x,t)/\partial x$ and $\partial^2(x,t)/\partial x^2$, respectively, and c, k are constants introduced by relations $c = -(p-q)h/\tau$, $k = h^2/2\tau$, and called flow rate and diffusion coefficient, respectively.

In the special case, when $c = 0$, $k = 1$, which holds formally for $p = q = ½$, and $\tau = h^2/2$, the equation (7) takes well-established form of standard heat flow equation

$$f_t = f_{xx}\,. \tag{8}$$

In particular, the condition $\tau < h^2/2$ maintains numerical stability of finite difference approximation of equation (8), where f_t will be approximated by forward differences and f_{xx} by central differences, and τ and h , τ, $h > 0$, represent time and space steps of the mesh.
Note that equation (8) admits a particular solution taking the form $f(x, t) = (1/\sqrt{t})\,\exp(-x^2/4t)$, $t > 0$.

3 Black-Scholes equation

Except standard derivation of B-S equation mentioned above, there are known other alternative ways how to get such equation. Let us discuss one of such approach in more details, following [8] and [11].

Let $v(s,t)$ denotes the price of E-call option related to the asset value s at the time t. Following the assumption i), one can express no change of value in incremental sense between δs and δv at any time as

$$\delta s = \eta\cdot\delta v, \quad \eta = \delta s/\delta v \approx (v_s)^{-1}, \tag{9}$$

where η denotes appropriate volume of options to be sold, and in the infinitesimal setting $\eta \approx 1/(\partial v/\partial s) = (v_s)^{-1}$.

The incremental values of s and v during the time period Δt under the interest rate r are

$$\Delta s = s\cdot r\cdot\Delta t, \quad \Delta v = v\cdot r\cdot\Delta t, \tag{10}$$

Now, combining Eq.(9) and Eq.(10) with accepting $\eta \approx (v_s)^{-1}$, $\delta s = \Delta s - s\cdot r\cdot\Delta t$, $\delta v = \Delta v - v\cdot r\cdot\Delta t$, and after rearranging we arrive at

$$\Delta s - (v_s)^{-1}\cdot\Delta v = (s - (v_s)^{-1}\cdot v)r\cdot\Delta t\,, \tag{11}$$

which actually expresses the condition that the hedge equity $\Delta s - (v_s)^{-1}\cdot\Delta v$ grows as investing the net equity $s - (v_s)^{-1}\cdot v$ at the fixed interest rate r. Further, the equation (10) serves as a starting point for derivation of the B-S equation.

The Taylor expansion for the increment Δv gives

$$\Delta v = v(s+\Delta s,\, t+\Delta t) - v(s,t) = v_t\cdot\Delta t + v_s\cdot\Delta s + (v_{tt}(\Delta t)^2 + 2v_{ts}(\Delta t\cdot\Delta s) + v_{ss}(\Delta s)^2)/2. \tag{12}$$

However, the third term proportional to $(\Delta s)^2$ has to be recasted because of the assumption iv), which also says keeping in mind the Ito's formula applied to Eq.(1) that $(\Delta s)^2 = \sigma^2 s^2\Delta t$. Hence, substituting expression (12) into Eq.(11) gives

$$\Delta s - (v_s)^{-1}\cdot((\,v_t + \sigma^2 s^2/2\cdot v_{ss})\Delta t + v_s\cdot\Delta s + (v_{tt}(\Delta t)^2 + 2v_{ts}(\Delta t\cdot\Delta s))/2 = (s - (v_s)^{-1}\cdot v)r\cdot\Delta t\,,$$

further, cancelling all terms containing Δs because the expected value of Δs is zero in accordance with iv), and dividing by Δt assuming $\Delta t \neq 0$, we get

$$(v_s)^{-1}\cdot((\,v_t + \sigma^2 s^2/2\cdot v_{ss}) = (s - (v_s)^{-1}\cdot v)r\,,$$

which finally after multiplying by v_s, and rearranging gives the B-S equation in the famous form

$$v_t = r\,v - r\,s\,v_s - \sigma^2 s^2/2\cdot v_{ss}, \tag{13}$$

which hardly resembles the equation (8), the well-known canonical form of parabolic partial differential equation with one-dimensional state space.

The boundary condition adjusted to the time of expiration $t = T$ of the E-call option denoting $x = c$ is

$$v(s,T) = \max(s(T)-c,\,0)\,, \tag{14}$$

which stands in correspondence with the first expression of (2).

Now, we shall concern with steps converting the B-S equation (13) to the canonical form (8). In order to get such form we have to apply a sequence of mappings.

First, we notice that B-S equation (13) is homogeneous and invariant under the linear map $x \to \gamma \cdot x$, whenever $\gamma > 0$.

Step a) Turn the coefficients to be independent upon the variable s. Standard way to do it, there is an introduction of dimensionless variable s/c, and application of mapping $s \to u$, where the new variable is $u = \ln(s/c)$, or $s = c \cdot \exp(u)$, in exponential form respectively. Hence, we have

$$u_s = (s/c)^{-1}\, c^{-1} = s^{-1}. \tag{15}$$

Instead of $v(s,t)$, we introduce a function $\varphi(u,t)$ defined implicitly by identity $\varphi(u,t) = v(s,t)$, and calculate the corresponding derivatives

$\varphi_s = \varphi_u \cdot u_s = (s)^{-1} \cdot \varphi_u$,

$\varphi_{ss} = (\varphi_u \cdot u_s)_s = \varphi_{us} \cdot u_s + \varphi_u \cdot u_{ss} = \varphi_{uu} \cdot (u_s)^2 + \varphi_u \cdot u_{ss}, = (s)^{-2} \cdot \varphi_{uu} - (s)^{-2} \cdot \varphi_u = (s)^{-2} \cdot (\varphi_{uu} - \varphi_u)$, with $u_{ss} = (s^{-1})_s = s^{-2}$.

Inserting them into Eq.(13), and after rearranging we get an equation with constant coefficients

$$\varphi_t = r \cdot \varphi - (r - \sigma^2/2) \cdot \varphi_u - \sigma^2/2 \cdot \varphi_{uu} = r \cdot \varphi - \alpha \cdot \varphi_u - \beta \cdot \varphi_{uu}\ , \tag{16}$$

where new coefficients α, β have been introduced by $\alpha = r - \sigma^2/2$, $\beta = \sigma^2/2$. The equation (16) already resembles the equation of diffusion (7), except the term $r \cdot \varphi$, which causes the solution $\varphi(u,t)$ to grow proportionate to $\exp(rt)$, as generally known.

Step b) Tame the solution, i.e., remove a growth of order as $\exp(rt)$. Following standard approach, one assumes $\varphi(u,t)$ to be decomposed in a multiplicative way containing a new function $y(u,t)$ defined by identity as follows

$$\varphi(u,t) = \exp(-r(T-t))y(u,t). \tag{17}$$

Calculate the corresponding derivatives and terms

$\varphi_t = \partial(\exp(-r(T-t))y(u,t))/\partial t = \exp(-r(T-t))(r \cdot y + y_t)$,

$r \cdot \varphi = \exp(-r(T-t))r \cdot y$, $\quad \alpha \cdot \varphi_u = \alpha \exp(-r(T-t))y_u$, $\quad \beta \cdot \varphi_{uu} = \beta \exp(-r(T-t))y_{uu}$,

substituting them into Eq.(16), and multiplying the equation by $\exp(r(T-t))$ we get

$$y_t = -\alpha \cdot y_u - \beta \cdot y_{uu}\ , \tag{18}$$

which already stands in correspondence with diffusion equation (7) except signs of the coefficients.

Step c) Rescale the independent variables u, t in order to get unit coefficients instead of α, β using the following substitutions, which introduce new state variable ξ and time scale τ

$$\xi = (\alpha/\beta) \cdot u\ , \quad \tau = (\alpha^2/\beta) \cdot (T\text{-}t). \tag{19}$$

Again, we introduce a new function $\psi(\xi,\tau)$ defined implicitly by identity $\psi(\xi,\tau) = y(u,t)$, and calculate the corresponding derivatives

$y_t = \psi_\tau \cdot \tau_t = -(\alpha^2/\beta) \cdot \psi_\tau$, because $\tau_t = -(\alpha^2/\beta)$,

$y_u = \psi_\xi \cdot \xi_u = (\alpha/\beta) \cdot \psi_\xi$, because $\xi_u = \alpha/\beta$,

$y_{uu} = (\psi_\xi \cdot \xi_u)_u = \psi_{\xi\xi} \cdot (\xi_u)^2 + \psi_\xi \cdot \xi_{uu} = (\alpha^2/\beta^2) \cdot \psi_{\xi\xi}$, because $\xi_{uu} = 0$.

Inserting them into Eq.(18), and after rearranging we get an equation corresponding with standard diffusion equation (7) with $c = -1$, $k = 1$

$$\psi_\tau = \psi_\xi + \psi_{\xi\xi}\ . \tag{20}$$

Step d) Eliminate the term ψ_ξ using simple kinematics substitution, which introduces a co-moving frame with the unit velocity as follows

$$z = \xi + \tau\ , \tag{21}$$

which yields Eq.(20) to turn into a standard heat or diffusion equation with unit diffusion coefficient

$$w_\tau = w_{zz}\ , \tag{22}$$

where new function $w(z, \tau)$ satisfies the following identities $w(z, \tau) = w(\xi + \tau, \tau) = \psi(\xi,\tau)$, keeping in mind that $\psi_t = w_\tau + w_z \cdot z_\tau = w_\tau + w_z$, $\psi_\xi = w_z \cdot z_\xi = w_z$, and $\psi_{\xi\xi} = w_{zz} \cdot (z_\xi)^2 + w_z \cdot z_{\xi\xi} = w_{zz}$.

Transformation of boundary conditions.

The standard temporal boundary condition on the E-call option price given by Eq.(14) at the expiration time $t = T$, i.e., $v(s,T) = s(T) - c$, if $s(T) \geq c$, and $v(s,T) = 0$, if $s(T) < c$, now reads $u \geq 0$ at $\tau = 0$, which

yields $z = \xi$, too. However, the first Eq.(19) says that $u \geq 0$ imply $\xi \geq 0$, if $r \geq \sigma^2/2$ holds. Hence, the boundary conditions prescribed for $w(z,0)$ are following

$$w(z,0) = s(T) - c = c(\exp(u) - 1) = c(\exp(z \cdot \beta/\alpha) - 1), \text{ if } z \geq 0, \quad \text{and } w(z,0) = 0, \text{ if } z < 0. \qquad (23)$$

4 Conclusions

Specific links between the B-S equation and diffusion process has been discussed. First, in order to pick up links between heat and/or diffusion equation and the B-S equation the background of one-dimensional diffusion process has been briefly reminded. Further, the sequence of mappings turning the B-S equation into the standard heat equation has been provided. Except the first one, $u = \ln(s/s(T))$, which introduces the dimensionless quantity and expresses its logarithm, the others are linear substitutions both for rescaling and intrinsic simple kinematics of the solution. Thus any solution of the B-S equation can be tractable as specific compound function being originated by a thermal propagation solution described by standard one-dimensional heat evolution equation. That qualitative structure could be used for building numerical solutions of B-S equation in an alternative way, since the standard heat equation is extremely well investigated both from analytical and numerical points of view, as providing special fast solvers, which deploy its specific feature, i.e., unit diffusion coefficient, into effective difference schemes. Such approach seams to be interesting when a lot of numerical solutions of B-S equation are to be constructed purposing various investment variants analysis.

Acknowledgement: This work was supported by grant No. 402/05/2392 of the Grant Agency of the Czech Republic.

References

[1] Black, F., Scholes, M.: *The pricing of options and corporate liabilities*. J. Political Econom. 81 (3) 1973, pp.637-654

[2] Blaha, Z.S., Jindřichovská, I.: *Opce, swapy, futures deriváty finančního trhu*. (*Options, swaps, futures derivatives of fin. market*.), 2.ed., Management Press, Praha 1997, ISBN 80-85943-29-8

[3] Blake, D.: *Financial Market Analysis.* (*Analýza finančních trhů*), Grada Publishing, Praha, 1995

[4] Drábek, P., Holubová, G.: *Pariální diferenciální rovnice*. (*Partial differential equations*), Vyd. ZČU v Plzni, Plzeň 2001, ISBN 80-7082-766-1

[5] Gneděnko, B.V.: *Kurs teoriji věrojatnostej*. (*Course of theory of probability*), 4.ed., Nauka, Moskva 1965

[6] Jaroš, P., Kodera, J.: *Termínové a opční operace*. (*Term and option operations*), Vydavatelství NAD ZLATO, Praha 1992, ISBN 80-85626-07-1

[7] Jílek, J.: *Termínové a opční obchody*. (*Term and option trades*), GRADA Publishing, Praha 1995, ISBN 80-7169-183-6

[8] Lukáš, L.: *Poznámky k odvození Black-Scholesovy rovnice*. (*Notes to derivation of the Black-Scholes equation*), Econometric Day, Seminar 2004-12-02, ÚTIA ČAV, Praha

[9] Pavlát, Vl.: *Finanční opce*. (*Financial options*), Magnet-Press, Institut pro podporu rozvoje finančního trhu, Praha 1994, ISBN 80-85847-19-1

[10]Pironneau, O., Hecht, F.: *Mesh Adaptation for the Black-Scholes Equations*. ftp://ftp.ann.jussieu.fr/pub/soft/pironneau

[11] Silverman, D.: *Solution of the Black Scholes Equation Using the Green´s Function of the Diffusion Equation*. Google search: „Black-Scholes Equation“ - S.D.(1999), Dept. of Physics and Astronomy, Univ. of California, Irvine, CA

[12] Watkins, T.: *Derivation of the Black_Scholes Equation for Option Value*. http://www2.sjsu.edu/faculty/watkins/blacksch.htm

An Efficient Procedure for 0-1 IP Problem

Otakar Machac and Simona Bohmova
Dept. of Economy and Management of Chemical and Food Industry
The University of Pardubice, CZ-532 10 Pardubice

Abstract

The paper describes an effective procedure of solving so called "Capital Budgeting Problem", with use of spreadsheet tables. The first step of the method consists of the problem formulation and a preoptimality analysis. Next, the algorithm procedure using Microsoft Excel is suggested. This procedure is based on a quick elimination of infeasible combinations and on a reduction of feasible combinations which cannot be the optimal solution. The remaining feasible combinations are analysed in a specifically prepared table finding the optimal solution effectively. Comparison of the suggested procedure with heuristic and exact (using Solver tool in Excel) methods is shown on a numerical example and the advantages of the procedure are discussed.

Keywords

0-1 Integer Programming Problem, Capital Budgeting Problem, Solver, Spreadsheet Tables

1 Introduction

One of the simplest binary integer programming (0-1 IP) models is "Capital Budgeting Problem". This problem can be described as decision-making problem when managers have to choose the set of projects that maximizes total economic contribution and stays within budget and other possible constrains at the same time, for all resources won't cover realization of a whole group of projects. The most important part of the model is that the projects are indivisible (they can be only accepted or refused as a whole) so the decision variables must be binary, where a 1 means a project will be undertaken and a 0 means it is not.

The basic 0-1 Integer Programming Problem can be written as follows:

MAXIMIZE $Z = \sum_{j=1}^{n} c_j x_j$

SUBJECT TO $\sum_{j=1}^{n} a_{ij} x_j \le b_i$ for i = 1, 2 . . . m constraints

$x_j = 0 \vee 1$ for j = 1, 2, . . . n independent projects.

Although 0-1 IP models are basically LP models with added binary variables, they are often more difficult to solve than LP problems, firstly because we cannot use the simplex algorithm. Several solution methods have been suggested by researches and new methods for specialized problems are still being developed. Mainly, we can speak about two different approaches to solve 0-1 IP problems:

- ***exact solution procedures*** - mostly based on a combinatorial approach with important disadvantage in a high laboriousness;
- ***heuristic methods*** - they can lead to a good solution, not necessarily optimal, but can provide useful information on the problem; they are usually simply and clear. The key principle is mostly based on ranking projects according to a particular factor or their combination with following selection of projects till all limits of individual resources are depleted. One of the simplest procedures is shown by Ragsdale [3] recommending to rank projects according to absolute value of contributions and to choose a file of first *k* projects that still fulfil all the constraints.

Some of the methods try to combine both approaches. The group of combinations is reduced using heuristics, followed by combinatorial algorithm that examines selected possible solutions, see [1], [4], etc. Recently, development of spreadsheet techniques has brought new inspirations and possibilities for heuristic and combinatorial solutions methods, see [3] etc. The basic solution procedure of 0-1 IP models is called branch and bound, see [2], [4], [8] etc., and is used by MS Excel's Solver, too.

2 Efficient method for 0-1 IP problem solving with Microsoft Excel

Suggested optimisation method for solving a 0-1 IP problem combines both the heuristic and combinatorial procedures and uses some of the Excel tools. The whole procedure can be summarised in following steps:

1. **Formulation** of 0-1 integer programming problem
2. **Preoptimality analysis**
3. **Selection of primal or dual problem** for the optimisation
4. **Separation** of k-elements subsets for $k = 1, 2 \ldots n$ into 2 groups: k_p subsets containing feasible solutions and k_n subsets with infeasible solutions.
5. **Reduction** of k_p subsets to k_{p+} subsets with possible optimal solutions and to k_{p-} subsets that cannot contain optimal solution.
6. **Complete examination** of all combinations in k_{p+} subsets and determination of the optimal solution.

Problem statement

Next, the whole procedure is described and demonstrated on a following numerical example:

A company is considering 8 mutually independent investment projects. Resources (cash and new employees) required for each investment, limits of both resources and assumed contributions of investments to the company are given in the Table 1.

Tab 1 Entry Data

	A	B	C	D
1	j =	a_{1j}	a_{2j}	c_j
2	3	150	30	180
3	6	220	40	200
4	2	280	54	300
5	1	300	65	300
6	5	320	68	350
7	8	330	63	320
8	7	350	60	280
9	4	450	80	400
10	Sum	2400	460	
11	Limit	1800	360	
12	sum/lim (%)	75,00%	78,26%	

Preoptimality analysis

Preoptimality analysis is very useful for solving the 0-1 IP problems because it reveals some important information on properties of the solution and on a whole effectiveness of the model. As a first step, we suggest to examine relative values of constrains b_i. The number of selected projects is always determined by those constraints that are the strictest in view of the sum of resources required if all projects are realized. In the case of primal problem with constraints: $\sum_{j=1}^{n} a_{ij} x_j \le b_i$, the strictest constraints are those with the smallest ratio $p_i = b_i / \sum_{j=1}^{n} a_{ij}$, for i = 1, 2, .. m constraints and j = 1, 2,... n independent projects. The constraints with the ratio close to 0.5 have the highest influence on

optimality efficiency. On the other hand, the importance (relevance) of the optimizing procedure decreases as the ratio p_i is closer to 1 or 0.

Next, a rank correlation of projects on the basis of project contribution per resource invested, i.e. ratios c_j/a_{ij}, should be examined. The higher a direct correlation among effectiveness of invested resources, the higher effect optimization will bring because projects dominated the others (from all or most points of view) are selected.

Preoptimality analysis reveals that the rank correlation of project efficiency is medium- strength and limits of both sources can be used up to 3/4 of resources required by all projects. Therefore the optimizing procedure seems to be effective; solving the problem can mean a significant contribution comparing to an empirical or a random selection of projects.

Primal or dual problem solving

In some cases, a dual of the original problem is better to solve. The formulation of the dual problem is:

MINIMIZE V = $\sum_{j=1}^{n} c_j x_j$

SUBJECT TO $\sum_{j=1}^{n} a_{ij} x_j \geq d_i$ for i = 1, 2 . . . m constraints

$x_j = 0 \vee 1$ for j = 1, 2, . . . n independent projects

where $d_i = \sum_{j=1}^{n} a_{ij} - b_i$ for i = 1, 2, . . . m constraints.

If the constraints b_i allow using less than 50% of the sum of all resources required by all projects, it would be better to solve the primal problem. On the contrary, if the values of constraints are close to the sum of resources required by all projects (it means we could realize more projects) it would be better to solve the dual problem. In this case, the goal is to eliminate variables that minimally decreased the amount of all possible contributions. Unlike the primal problem, 0-1 variable equals 0 for the undertaken project and 1 for refused one.

Since $p_1 = 75.00\%$ and $p_2 = 78.26\%$, the dual problem is chosen, its formulation is:

minimize $300x_1 + 3000x_2 + 180x_3 + 400x_4 + 350x_5 + 200x_6 + 280x_7 + 320x_8 = z$

subject to: $300x_1 + 280x_2 + 150x_3 + 450x_4 + 320x_5 + 220x_6 + 350x_7 + 330x_8 \geq 600$

$65x_1 + 54x_2 + 30x_3 + 80x_4 + 68x_5 + 40x_6 + 60x_7 + 63x_8 \geq 100$

$x_j = 0 \vee 1$ for j = 1, 2, … 8.

Separation of k-elements subsets into feasible and infeasible solutions

In this step, two boundary combinations for each k = 1, 2, …n elements, with the highest and the lowest value are found. If both boundary combinations for k-elements subset comply with a given constraint, all the rest combinations comply, too. On the contrary, if both boundary combinations do not comply, none of the rest possible k-elements combinations comply either. Finally, if the lowest combination fulfills the conditions and the highest does not (or contrariwise), such a k-elements subset(s) contains some combinations that comply with constraint and other ones that do not.

The separation of k-elements subsets can be efficiently carried out in Excel using option "Data/Sort". Projects are sorted according to constraints either in ascending or descending orders. Examination starts with one-element subsets (k = 1). It is obvious that none of 1-element combination complies with both constraint conditions because even the project with the highest values does not comply either (it is not true that $a_{14} \geq 600$ and it is not true that $a_{24} \geq 100$ either). Next, 2-elements combinations (k = 2) are examined. The total number of 2-elements combinations is 28. Combination of projects number 3 and 6 with the lowest requirements do not comply with both given conditions, whereas combination of projects with the highest requirements, number 7 and 4 or 5 and 4, comply with both conditions already. Hence two elements subsets can contain feasible and thus optimal solution. Next, we will continue with 3-elements combinations, k = 3. Combination of three projects with the lowest values, projects 3, 6 and 2, is a solution feasible for both restricted conditions (150 + 220 + 280) = 650 a (30 + 40 + 54) = 124. If this combination is a feasible solution, all the rest of

possible combinations will be feasible, too. We do not need to examine following combinations for k = 4, 5, 6, 7 and 8 elements as they are all evidently feasible.

Reduction of combinations that cannot contain optimal solution

In this step, feasible combinations are reduced to those with expected optimal solutions. It is useful to prepare new table with projects ranked on the basis of c_j. Since all 3-elements combinations are feasible, the best solution will be found at the bottom of table where the projects are ranked in descending order. The minimal value of objective function for 3-elements combinations has value of 660 for projects combination of 3, 6, 7 ($c_3 + c_6 + c_7 = 660$). Logically, we can exclude possibility that 4 and more elements combinations can have lower value of objective function. In the next step, we must focus on 2-elements combinations containing feasible and infeasible solutions.

Complete examination

As we excluded infeasible combinations and reduced some of feasible solutions, it is obvious that the optimal solution is the former 3-elements combination (projects 3, 6, 7 with V = 660) or one of the 2-elements combinations. Therefore, we have to examine 2-elements combinations completely. Even this step can be solved effectively reducing the number of combinations. For this purpose, following Table is prepared.

Tab 2 Complete Examination of 2-Elements Combinations

	A	B	C	D	E	F
16	j =	j =	sum a_1	sum a_2	sum c_j	optimum
17	4	7	800	140	680	
18	4	8	780	143	720	
19	4	5	770	148	750	
20	4	1	750	145	700	
21	4	2	730	134	700	
22	4	6	670	120	600	
23	4	3	600	110	580	opt
24	7	8	680	123	600	
25	7	5	670	128	630	
26	7	1	650	125	580	opt
27	7	2	630	114	580	opt
28	7	6	570	100	x	
29	7	3	500	90	x	
30	8	5	650	131	670	
31	8	1	630	128	620	
32	8	2	610	117	620	
33	8	6	550	103	x	
34	8	3	480	93	x	

In the first two columns (A; B), 2-elements combinations are prepared. They are made systematically - from the projects with the highest values of a_{1j} to the lowest. In columns C and D, values of total consumption of resources consumed by both projects of particular combination are determined. To do that, following formula was written into cell C17 and copied to all cells of both columns C and D:

=IF($A17=$A$2;B$2;0)+IF($A17=$A$3;B$3;0)+IF($A17=$A$4;B$4;0)+IF($A17=$A$5;B$5;0)+IF($A17=$A$6;B$6;0)+IF($A17=$A$7;B$7;0)+IF($A17=$A$8;B$8;0)+IF($A17=$A$9;B$9;0)+IF($B17=$A$2;B$2;0)+IF($B17=$A$3;B$3;0)+IF($B17=$A$4;B$4;0)+IF($B17=$A$5;B$5;0)+IF($B17=$A$6;B$6;0)+IF($B17=$A$7;B$7;0)+IF($B17=$A$8;B$8;0)+IF($B17=$A$9;B$9;0)

In the formula, first two cells in the row are compared step by step with entry data and using the function IF; the proper amounts of consumption of resources a_1 and a_2, respectively, are calculated in third a fourth cells in the row.

In the next column E, values of objective function for particular combinations are calculated. Formula written into the cells of column E was, in comparison with the former formula, extended by

function IF verifying if particular combination comply with both constraints. Those which do not comply are eliminated with "x". Last column F highlights optimal solution (there can be more than one) – minimum value of objective function for all combinations (Excel's function for finding minimum was used to do that).

As we can see, our dual problem has 3 optimal solutions - combinations of projects 3, 4; 1, 7 and next 2, 7. Objective function has value of 580 for all the three combinations. Since all 1-element combinations are infeasible and minimal value of objective function for 3-elements combinations is 660, all the three combinations are optimal solution of our problem.

Now the dual problem has to be transformed back, to the primal problem. It results from the following table that the dual problem that primal problem has three optimal solutions (to eliminate projects 3;4 or 2;7 or 1;7) - with the same value of objective function (1,750).

Tab 3 Optimal Solutions of Dual and Primal Problems

Dual problem				Primal problem			
Projects	Sum a1j	Sum a2j	Sum cj	projects	Sum a1j	Sum a2j	Sum cj
3; 4	600	110	580	1;2;5;6;7;8	1,800	350	1,750
2; 7	630	114	580	1;3;4;5;6;8	1,770	346	1,750
1; 7	650	125	580	2;3;4;5;6;8	1,750	335	1,750

But, at the same time, we can see that these solutions are not completely the same in point of resource requirements. The worst seems to be solution eliminating projects 3 and 4 with the highest requirements on resources (1800 of a_{1j} and 350 of a_{2j}). The best appears to be third solution with the lowest requirements on resources. It should be pointed out there is no such information if we solve the problem with Excel's Solver.

3 Discussion and Conclusion

Let´s have a look at our example using different solution methods.

If we use Ragsdale's recommendation (see chap. 1) in *ranking projects according contribution values*, we obtain solution where first 5 projects with the highest contribution values (i.e. 4, 5, 8, 2, 1) are undertaken, giving NPV of 1670. The other possible approximate heuristic solution procedure lies in *ranking projects on the basis of ratios* $r_{ij} = c_j/a_{ij}$. For problem with more than one constraint ($i \geq 2$), we can rank projects for each *i* and/or calculate average position using data for all constrains. In our example, as a solution we would select, according r_{1j}, investments 3, 5, 2, 1, 8, and 6 with NPV of 1650. Similarly, ranking investments based on r_{2j}, the selected projects are 2, 3, 4, 5, 6, and 8 with NPV of 1750 (already one of the optimal solutions). If we rank investments on the basis of average position, the same solution as in the case of ranking based on r_{1j} is found.[1]

Of course, our problem can be solved with *MS Excel´s Solver*, too, see [6], [7], etc. To solve IP problems with Solver we must use Solver's binary option (the constraint "decision = bin" indicates that variables are 0-1 variables). The solution of our example using Solver selects projects number 2, 3, 4, 5, 6 and 8 with objective function value of 1,750 mil CZK (identical solution to one we have already obtained using heuristics). Here, we must pay attention to the solver Option dialog that includes a Tolerance field for IP models. The default Tolerance value is 5% and it means that the optimization procedure continues until the solution (its value of objective function) is within 5% of IP optimum value of objective function. A higher Tolerance value can speed up procedure but the solution may not be the optimal one and be even further from the real IP optimum. Tolerance of 0% forces Solver to find the actual IP optimum but is more time-consuming. The former solution founded by Solver uses default Tolerance value of 5%. Setting Tolerance to 4% (and less), Solver founds solution again but chooses projects number 1, 2, 5, 6, 7, 8 to be undertaken with objective function value 1,750 mil CZK. One would think this solution would be better than the one with tolerance value 5% but this is not true because despite this solution has the same objective function value, it uses more resources (1,800 mil CZK and 350 employees). The advantage of Solver can be seen in simplicity of

[1] All these solutions can be found in [ ]

finding the solution but there are disadvantages, too. First, we do not always obtain the true optimal solution and second, Solver does not provide any additional information about the solution.

The main advantage of suggested method is effective connection of heuristic and exact methods together with use of spreadsheet techniques of Excel. The method includes preoptimality analysis and provides more possibilities for postoptimality analysis in which the entry values are changed and influence on feasibility and optimality can be observed. At the same time, the solution procedure is highly effective, particularly when the spreadsheet technique with database and logic functions is used. The presented method is applicable even to more extensive models with more variables where a pure combinatorial solving procedure has its limitations even if more powerful computer is used. Another benefit can be seen in easy interpretation and understandability of the whole procedure. Therefore we believe the suggested method can mean a valuable contribution to the solving of this type of 0-1 integer problems even in practice.

References

[1] Balas, E., Martin, C., H.: Pivot and Complement - A Heuristic for 0-1 Programming. *Management Science,* Vol. 26, No. 1(Jan., 1980), 86-96.

[2] Dennis, T.,L., Dennis L.,B.: Management Science, West Publishing Company, 1991

[3] Ragsdale, C.,T.: Spreadsheet Modelling and Decision Analysis. South-Western College Publishing, 2001

[4] Winston, W. L.: Operations Research: Applications and Algorithms. 2nd edition, PWS-KENT Publishing Company, Boston, 1991, ISBN 0-534-92495-6

[5] Bell, P. C.: Management Science/Operations Research: A Strategic Perspective, South-Western College Publishing, Cincinnati, Ohio, 1999, ISBN 0-324-00466-4

[6] Winston, W. L., Albright, S. C.: Practical Management Science. 2nd edition, Duxbury/ Thomson Learning, 2001, ISBN 0-534-37135-3

[7] More, J. H., Weatherford, L. R., Eppen, Gould, Schmidt: Decision Modeling with Microsoft Excel. 6th edition, Prentice Hall, 2001, ISBN 0-13-017789-X.

[8] Jablonský, J.: Operační výzkum: kvantitativní modely pro ekonomické rozhodování. 1st edition, Professional Publishing, 2002, ISBN 80-86419-23-1

[9] Albright, S.C., Winston, W. L.: Spreadsheet Modeling and Applications - Essentials of Practical Management Science, Thomson Learning-Brooks/Cole, 2005, ISBN 0-534-38032-8

[10] Gros, I.: Kvantitativní metody v manažerském rozhodování. Grada Publishing a.s., 2003, ISBN 80-247-0421-8

[11] Machač, O., Böhmová, S.: An Efficient Spreadsheet Procedure for Solving a 0-1 Integer Programming Problem, Scientific Papers of the University of Pardubice, Ser. A, Faculty of Chemical Technology, 10 (2004), p. 379-391

Fuzzy Time Series Modelling by SCL Learning*

Dušan Marček[1, 2] , Milan Marček[3]
1. Institute of Computer Science, The Silesian University Opava, Czech Republic
2. The Faculty of Management Science and Informatics, University of Žilina, Slovakia
Email: 1. Dusan.Marcek@fpf.slu.cz 2. marcek@fria.utc.sk
3. Kiwa, Ltd., Priemyselná 2, Nitra, e-mail: tech.kiwa@flynet.sk

Abstract

Based on the works [11], [22] a fuzzy time series model is proposed and applied to predict chaotic financial process. The general methodological framework of classical and fuzzy modelling of economic time series is considered. A complete fuzzy time series modelling approach is proposed. To generate fuzzy rules from data, the neural network with Supervised Competitive Learning (SCL)-based product-space clustering is used.

Keywords

Fuzzy time series models, fuzzy bank rules, product-space clustering, fuzzy controllers

1 Introduction

Much of the literature in the field of the fuzzy logic and technology is focused on dynamic processes modelling with linguistic values as its observations (see e.g. [14], [15], [22]). Such a dynamic process is called fuzzy time series. This type of dynamic processes play very important role in making practical applications. Economic and statistical time series analysis is concerned with estimation of relationships among groups of variables, each of which is observed at a number of consecutive points in time. The relationships among these variables may be complicated. In particular, the value of each variable may depend on the values taken by many others in several previous time periods. Very often it is difficult to express exactly these dependencies, or there is not known hypothesis for that. Very frequently, in such cases more sophisticated approaches are considered. These approaches are based on the human experience knowledge and consist of series linguistic expressions each of which takes the form of an ´ if ... then ...´ fuzzy rule, and they are well known under the common name fuzzy controllers. But also, an expert is usually unable linguistically describe the behaviour of economic processes in particular situations. Hence, most recent researches in the fuzzy controllers design for deriving of linguistically interpreted fuzzy rules have been centred on developing automatic methods to build these fuzzy rules using a set of numerical input-output data. For applying of these methods it is supposed that a database describing previous input-output behaviour of a system is available [14]. Majority of these models and data-driven techniques rely on the use Takagi-Sugeno type controllers and fuzzy/non-fuzzy neural networks [6], [8-10], [19], [24], clustering/fuzzy-clustering and genetic algorithm approaches [3], [4], [7], [11], [12], [23], [25].

The goal of this paper is to illustrate that two distinct areas, i.e. fuzzy sets theory and computational networks, may be used to economic time series modelling. We show how to use and how to incorporate both fuzzy sets theory and computational networks to determine the fuzzy relational equations. As an application of proposed method, the estimate of the inflation is carried out in this paper. The characterization of time series is introduced in Section 2. Quantitative modelling methods of time series are presented in Section 3 and 4. They introduce conventional and fuzzy time series modelling and show how to combine neural and fuzzy system to produce fuzzy rules. Concluding remarks are offered in Section 5.

*This work on this contribution was supported by the grants GAČR 402/05/2786 and VEGA 1/2628/05.

2 Conventional and fuzzy time series

Time series models are based on the analysis of chronological sequence of observations on particular variable. Typically, in conventional time series analysis, we assume that the generating mechanism is probabilistic and that the observed values $\{x_1, x_2, ..., x_t, ...\}$ are realisations of stochastic processes $\{X_1, X_2, ..., X_t, ...\}$.

In contrast to the conventional time series, the observations of fuzzy time series are fuzzy sets (the observations of conventional time series are real numbers). Song and Chisson [22] give a thorough treatment of these models. They define a fuzzy time series as follows. Let Z_t, (t = ..., 1, 2, ...), a subset of $\Re$, be the universe of discourse on which fuzzy sets x_t^i, (i = 1, 2, ...) are defined and X_t is the collection of x_t^i, (i = 1, 2, ...). Then X_t, (t = ..., 1, 2, ...) is called a fuzzy time series on Z_t, (t = ..., 1, 2, ...).

3 Quantitative time series modelling methods

In practice, there are many time series in which successive observations are dependent. This dependence can be treated here as an observational relation

$$R_o = \{(y_{t-1}, y_t), (y_{t-2}, y_{t-1}), ...\} \subseteq Y_{t-1} \times Y_t, \quad (1)$$

where Y_t, Y_{t-1} denote the variables and y_t, y_{t-1}, ... denote the observed values of Y_t and Y_{t-1} respectively.

In most real economic processes it is assumed that there exists a functional structure between Y_{t-1} and Y_t, i.e.

$$f: \ Y_{t-1} \rightarrow Y_t \quad (2)$$

belonging to a pre-specified class of mappings [5]. In practice many real models of this functional structure are represented by linear relation

$$y_t = f(y_{t-1}, \phi_1, \varepsilon_t) = \phi_1 y_{t-1} + \varepsilon_t \quad (3)$$

where ϕ_1 is the parameter of this linear relation, ε_t is a random error or noise component that is drawn from a stable probability distribution with zero mean and constant variance.

To determine the model (3) statistical methods are used such that function (3) satisfies some optimality criterion in fitting the observed data R_o.

In the case of the fuzzy time series, the fuzzy relational equations can be employed as the models. Analogously to the conventional time series models, it is assumed that the observation at the time t accumulates the information of the observation at the previous times, i.e. there exists a fuzzy relation such that [22]

$$y_t^j = y_{t-1}^i \circ R_{ij}(t, t - 1), \quad (4)$$

where $y_t^j \in Y_t$, $y_{t-1}^i \in Y_{t-1}$, $i \in I$, $j \in J$, I and J are indices sets for Y_t and Y_{t-1} respectively, " $\circ$ " is the sign for the *max-min* composition, $R_{ij}(t, t - 1)$ is the fuzzy relation among the observations at t and t -1 times. Then Y_t is said to be caused by Y_{t-1} only, i.e.

$$y_{t-1}^i \rightarrow y_t^j \quad (5)$$

or equivalently

$$Y_t \rightarrow Y_{t-1} \quad (6)$$

and

$$Y_t = Y_{t-1} \circ R(t, t - 1), \quad (7)$$

where $R(t, t-1)$ denotes the overall relation between Y_t and Y_{t-1}. In the fuzzy relational equation (7) the overall relation $R(t, t-1)$ is calculated as the union of fuzzy relations $R_{ij}(t, t-1)$,i.e. $R(t, t-1) = \bigcup_{i,j} R_{ij}(t, t-1)$, where "$\bigcup$" is the union operator. In the following, we will use Mamdani´s method [13] to determine these relations. For simplicity, in the following discussion, we can also express y_{t-1}^i and y_t^j as the values of membership functions for fuzzy sets y_{t-1}^i and y_t^j respectively. Since the Eq. (4) is equivalent to the linguistic conditional statement

$$\text{"if } y_{t-1}^i \text{ then } y_t^j \text{ "}, \tag{8}$$

we have $R_{ij}(t, t-1) = y_{t-1}^i \times y_t^j$, where "×" is the Cartesian product and therefore

$$R(t, t-1) = \max_{i,j} \{ \min(y_{t-1}^i, y_t^j) \}. \tag{9}$$

Referring to the above definition by Song and Chisson of the fuzzy time series, in fuzzy time series model Y_t, Y_{t-1} can be understood as linguistic variables and y_t^j, y_{t-1}^i as the possible linguistic values of Y_t, Y_{t-1} respectively.

Equation (7) is called a first-order model of the fuzzy time series of Y_t with lag $p = 1$. This first order model can be extended to the p-th order model. See [22] for details.

4 Determination of fuzzy relations by neural networks

All the above fuzzy time series models can be determined if in particular models the fuzzy relations are known. Since finding the exact solution of fuzzy relations is generally very difficult and in practice unrealistic, hence, more sophisticated approaches are considered very frequently.

In a fuzzy system, a powerful tool for generating fuzzy rules purely from data is neural network. Neural networks can adaptively generate the fuzzy rules in a fuzzy system by SCL-based product-space clustering technique [11]. Next, in a numerical example, we will illustrate and show how to obtain fuzzy rules using the fuzzy sets theory and neural networks.

Let us consider a simple example. The data set used in this example (the 514 monthly inflation rates in the U.S.). A graph oh historical values of inflation are presented in Fig. 1. To build a forecast model the sample period for analysis y_1, ..., y_{344} was defined. The following statistical model was specified

$$y_t = \xi + \phi_1 y_{t-1} + \varepsilon_t, \tag{10}$$

where the variable y_t is explained by only on its previous values, and ε_t is a white noise disturbance term. Using Levinson-Durbin algorithm [2], [18] the model (10) is statistically fitted as

$$\hat{y}_t = -0{,}1248 y_{t-1} \tag{11}$$

At this stage, we will only give some outlines to model a fuzzy time series in a fuzzy environment. The fuzzy time series modelling procedure consists of an implementation of several steps, usually as follows:

1. Define the input-output variables and the universes of discourse.
2. Define (collect) linguistic values and fuzzy sets on the universes of discourse.
3. Define (find) fuzzy relations (fuzzy rules).
4. Apply the input to the model and compute the output.
5. Defuzzify the output of the model.

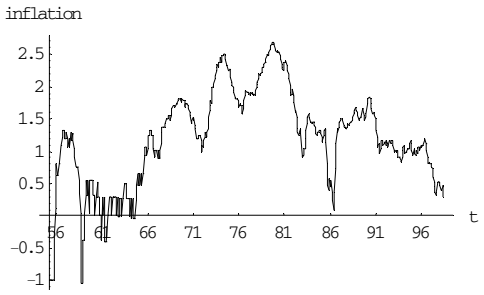


Figure 1: Natural logarithm of monthly inflation from February 1956 to November 1998.

From Step 1 to Step 2, the input data are fuzzified, in Step 3, analogously to the conventional model (10), the fuzzy time series model, i.e. the fuzzy relational model is created. Steps 4, 5 are considered as an application of the model (i.e., analysis of economic structures and the forecasting). Below, we will discuss these steps and apply them to the inflation time series at a more detailed level.

Firstly, in the fuzzification block, we specified input and output variables. The input variable x_{t-1} is the lagged first difference of inflation values $\{y_t\}$ and is calculated as $x_{t-1} = y_{t-1} - y_{t-2}$, $t = 3, 4, \ldots$. The output variable x_t is the first difference of inflation values $\{y_t\}$ and it is calculated as $x_t = y_t - y_{t-1}$, $t = 2, 3, \ldots$. The variable ranges are as follows - 0,75 $\leq x_t$, $x_{t-1} \leq$ 0,75 . These ranges define the universe of discourse within which the data of x_{t-1} and x_t are, and on which the fuzzy sets have to be, specified. The universes of discourse were partitioned into the seven intervals.

Next, we specified the fuzzy-set values of the input and output fuzzy variables. The fuzzy sets numerically represented linguistic terms. Each fuzzy variable assumed seven fuzzy-set values as follows: NL: Negative Large, NM: Negative Medium, NS: Negative Small, Z: Zero, PS: Positive Small, PM: Positive Medium, PL: Positive Large.

Fuzzy sets contain elements with degrees of membership. Fuzzy membership functions can have different shapes. The triangular membership functions were chosen. Fig. 2 shows membership function graphs of the fuzzy sets above.

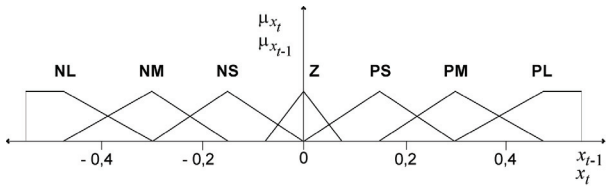


Figure 2: Fuzzy membership functions for each linguistic fuzzy-set value

The input and output spaces were partitioned into the seven fuzzy sets. From membership function graphs μ_{t-1}, μ_t in Fig. 2 is shown that the seven intervals [-0,75; -0,375], [-0,375; -0,225], [-0,225 -0,075], [-0,075; 0,075], [0,075; 0,225], [0,225; 0,375], [0,375; 0,75] correspond respectively to NL, NM, NS, Z, PS, PM, PL.

Next, we specified the fuzzy rule base or the bank of fuzzy relations. The appendix describes the neural network which uses the supervised competitive learning to derive fuzzy rules from data. As shown in Fig. 3(b) the bank contains the 5 fuzzy rules. For example the fuzzy rule of the 34th block corresponds to the following fuzzy relation if x_{t-1}^i = PM then x_t^j = PS.

Finally, we determined the output action given the input conditions. We used the Mamdani´s implication [13]. Each fuzzy rule produces the output fuzzy set clipped at the degree of membership determined by the input condition and the fuzzy rule. When the input value, say $x_{t-1}^i = x_{344}^i$, is applied to the model (4), the output value $x_t^j = x_{345}^j$ can be calculated. Following the above principles, we have obtained the predicted fuzzy value for the inflation $x_t = x_{345}^j = 0{,}74933$.

The inflation values in the output x_t^j, t = 345, 346, ... are not very appropriate for a decision support because they are fuzzy sets. To obtain a simple numerical value in the output universe of discourse, a conversion of the fuzzy output is needed. This step is called defuzzification. The simplest defuzzification scheme was used. This scheme selects the value $\hat{x}_t$ as the middle of the maximum membership function. Following this method, we have obtained the predicted value for the $\hat{x}_{345}$ = - 0,15 . The remaining forecasts for ex post forecast period t = 346, 347, ... may be generated similarly.

5 Conclusion

The method may be of real usefulness in practical applications, where the expert usually can not explain linguistically what control actions the process takes or there is no knowledge of the process. In principle a neural network can derive this knowledge from data. In practice this is usually necessary. Although the method has been carried out in the time series modelling field, it is suitable for other applications as data mining systems, information access systems, etc.

References

[1] G.E.P. Box, and G.M. Jenkins (1970) Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA.

[2] P.J Brockwell, and R.A Davis(1987) Time Series: Theory and Methods. Springer-Verlag, New York, 1987.

[3] B. Carse, T.C. Forgarty, and A. Munro (1996) "Evolving fuzzy based controllers using genetic algorithms", Fuzzy Sets and Systems. Vol. 89: 273-293.

[4] M. Delgato, A.F.G. Skarmita, and F. Martin 1997) "A fuzzy clustering-based rapid prototyping for fuzzy rule-based modelling", IEEE Trans. Fuzzy Systems. Vol. 5, No. 2: 223-233.

[5] M. Fedrizzi, M.M. Fedrizzi, W. Ostasiewic (1993) "Towards fuzzy modelling in economics", Fuzzy Sets and Systems, 54": 259-268.

[6] S. Horikawa, T. Furihashi, and Y. Uchikawa (1992) "On fuzzy modelling using fuzzy neural networks with the backpropagation algorithm", IEEE Trans. Neural Networks. Vol. 3, No. 5: 801-806.

[7] J.Q. Chen, Z.G. Xi, and Z.J. Zhang (1998) "A clustering algorithm for fuzzy model identification", Fuzzy Sets and Systems. Vol. 98: 319-329.

[8] J.S.R. Jang, and C.T. Sun (1995) "Neuro-fuzzy modelling and control", in Proceedings of the IEEE. Vol. 83, No. 3, 378-406.

[9] J.S.R Jang (1993) "Adaptive-network based fuzzy inference system", IEEE Trans. Systems. Man and Cybernetics. Vol. 23, No 3: 665-685.

[10] B. Kosko (1992) "Fuzzy system as universal approximators", Proc. IEEE Int. Conf. on Fuzzy Systems. San. Diego, 1153-1162.

[11] B. Kosko (1992) Neural networks and fuzzy systems - a dynamical systems approach to machine intelligence, Prentice-Hall International, Inc.

[12] R. Li, and Y. Zhang (1996) "Fuzzy logic controller based on genetic algorithms", Fuzzy sets and Systems. Vol. 83: 1-10.

[13] E.H. Mamdani (1997) "Application of a fuzzy logic to approximate reasoning using linguistic synthesis" IEEE Trans. Comput. 26: 1182-1191.

[14] D. Marček (2002) Neural networks and fuzzy time series with applications in economics. Silesian University (in Czech).

[15] D. Marček (2000) "Stock Price Forecasting: Autoregressive Modelling and Fuzzy Neural Network", Mathware & Soft Computing, Vol. 7, No. 2-3: 139-148.

[16] J. Mockoř (1997) "Locally fuzzy models",in Proceedings of the Mathematical Methods in Economics. VŠB TU Ostrava, 140-149.

[17] D.C. Mongomery, L.A. Johnston, J.S. Gardiner (1990) Forecasting and Time Series Analysis. McGraw-Hill, Inc.

[18] P.A. Morettin (1984) "The Levinson algorithm and its applications in time series analysis", International Statistical Review, 52: 83-92.

[19] H. Nomura, I. Hayashi, N. Wakami (1992) "A learning method of fuzzy inference rules by descent method" in IEEE Int´1 Conference on Fuzzy systems. San Diago, California, 203-210.

[20] J.J. Saade (2000) "A defuzzification based new algorithm for the design of Mamdani-type fuzzy controllers", Mathware & Soft Computing. Vol. 7, No. 2-3: 159-173.

[21] P. Siarry, and F. Guely (1998) "A genetic algorithm for optimising Tagaki-Sugeno fuzzy rule bases", Fuzzy Sets and Systems. Vol. 99, 1998: 37-47.
[22] Q. Song, B.S. Chisson (1993) "Fuzzy time series and its models", Fuzzy Sets and Systems, 54: 269-277.
[23] M. Sugeno, and T. Yasukawa (1993) "A fuzzy-logic-based approach to quantitative modelling", IEEE Trans. Fuzzy System. Vol. 1, No. 1: 7-31.
[24] H. Takagi, H, and I. Hayashi (1991) "NN driven fuzzy reasoning", Int´1 Journal of Approximate Reasoning. Vol. 5, No. 3: 191-212.
[25] Y.S. Tarng,, Z.M. Yeh, and C.Y. Nian (1996) "Genetic synthesis of fuzzy logic controllers in turning", Fuzzy Sets and Systems, Vol. 83: 301-310.
[26] C.V. Therien (1992) Discrete random signals and statistical signal processing. Prentice-Hall, USA.

Appendix

Generating fuzzy rules by SCL-based product-space clustering

The neural network pictured in Fig. 4 was used to generate structured knowledge of the form „if A, then B" from a set of numerical input-output data. In Section 4 we defined cell edges with the seven intervals of the fuzzy-set values in Fig. 2. The interval - 0,75 $\le x_t$, $x_{t-1} \le$ 0,75 was partitioned into seven non-uniform subintervals that represented the seven fuzzy-set values NL, NM, NS, Z, PS, PM, and PL assumed by fuzzy variables x_{t-1} and x_t. The Cartesian product of these subsets defines 7 × 7 = 49 fuzzy cells in the input-output product space R^2. As mentioned in [10] these fuzzy cells equal fuzzy rules. Thus, there are total 49 possible rules and thus 49 possible fuzzy relations.

We can represent all possible fuzzy rules as 7-by-7 linguistic matrix (see Fig. 4). The idea is to categorise a given set or distribution of input vectors $\mathbf{x}_t = (x_{t-1}, x_t)$, t = 1, 2, ..., 344 into 7 × 7 = 49 classes, and then represent any vector just by the class into which it falls. We used SCL (Supervised Competitive Learning) [10], [14] to train the neural network in Fig. 3. The software was developed at Institute of Computer Science of Faculty of Philosophy and Science, Opava. We used 49 synaptic quantization vectors.

For each random input sample $\mathbf{x}_t = (x_{t-1}, x_t)$, the wining vector $\mathbf{w}_{i'} = (w_{1i'}, w_{2i'})$ was updated by the SCL algorithm according to

$$\left.\begin{aligned} \tilde{w}_{1i'} &\leftarrow \tilde{w}_{1i} + \eta\,(\tilde{x}_{1t} - \tilde{w}_{1i}) \\ \tilde{w}_{2i'} &\leftarrow \tilde{w}_{2i} + \eta\,(\tilde{x}_{2t} - \tilde{w}_{2i}) \end{aligned}\right\} \quad if\ i = i' \ , \qquad \left.\begin{aligned} \tilde{w}_{1i'} &\leftarrow \tilde{w}_{1i} - \eta\,(\tilde{x}_{1t} - \tilde{w}_{1i}) \\ \tilde{w}_{2i'} &\leftarrow \tilde{w}_{2i} - \eta\,(\tilde{x}_{2t} - \tilde{w}_{2i}) \end{aligned}\right\} \quad if\ i \neq i' \ ,$$

where i' is the winning unit defined $\left\|\tilde{\mathbf{w}}_{i'} - \tilde{\mathbf{x}}_t\right\| \le \left\|\tilde{\mathbf{w}}_i - \tilde{\mathbf{x}}_t\right\|$ for all i, and where $\tilde{\mathbf{w}}_i$ and $\tilde{\mathbf{x}}_t$ is a normalized version of $\mathbf{w}_i$ and $\mathbf{x}_t$ respectively, η is the learning coefficient.

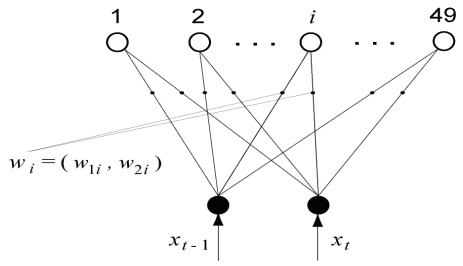


Figure 3: The topology of the network for fuzzy rules generating by SCL-based product-space clustering.

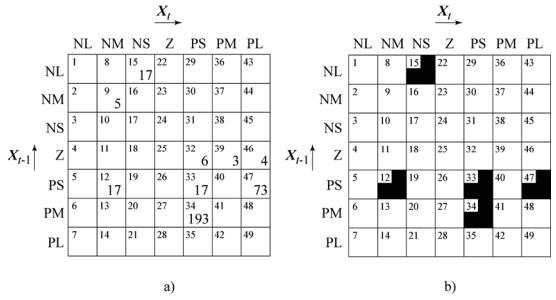


Figure 4: Distribution of input-output data (x_{t-1}, x_t) in the input-output product space $X_{t-1} \times X_t$ (a). Bank of fuzzy rules of the time series modelling system (b).

Supervised Competitive Learning (SCL)-based product-space clustering classified each of the 344 input-output data vectors into 9 of the 49 cells as shown in Fig. 4(a). Fig. 4(b) shows the fuzzy rule bank. We added a rule to the rule bank if the count of input-output vectors in particular cells was larger than the value 0,05N, where N = 344 is number of data pairs (x_{t-1}, x_t), t = 1, 2, ..., N in the input and output series. For example the most frequent rule represents the cell 34. From most to least important (frequent) the fuzzy rules are (PM; PS), (PS; PL), (NL; NS), (PS; PL), and (PS; PS).

Core Equivalence for Economy on Belief

Takashi Matsuhisa

Department of Natural Sciences, Ibaraki National College of Technology
Nakane 866, Hitachinaka-shi, Ibaraki 312-8508, Japan.
E-mail:mathisa@ge.ibaraki-ct.ac.jp

Abstract. We investigate a pure exchange economy under uncertainty with emphasis on the multi-modal logical point of view; the traders are assumed to have a multi-modal logic of belief. We propose a generalized notion of expectations equilibrium for the economy, and we establish the extended core equivalence theorem: The ex-post core for the economy coincides with the set of all its expectations equilibria in belief.
Keywords. Belief, Core equivalence theorem, Exchange economy under uncertainty, Expectations equilibrium in belief, Ex-post core, Multi-modal logic.

1 Introduction

This article relates economies and multi-agent modal logics. The purposes are: First to propose the multi-modal logic, *the logic of belief* **B**, by which the traders use making their decision, secondly to present the notion of *economy for the logic* **B** and the extended notion of equilibrium for the economy, called an *expectation equilibrium in belief*. Finally to establish the core equivalence theorem for the economy with emphasis on modal logical point of view:

Main theorem. *In a pure exchange economy under uncertainty, assume that the traders have the multi-modal logic* **B** *and they are risk averse. Then the ex-post core coincides with the set of all expectations equilibrium allocations in belief.*

In Economic theory, many authors have investigated several notions of equilibrium in an economy under asymmetric information (c.f., the papers cited in Forges et al [5]). E. Einy et al [4] succeeded in extending the core equivalence theorem of Aumann [1] as the equivalence theorem between the ex-post core and the rational expectations equilibria for an economy under asymmetric information.

The serious limitations of the analysis in these researches are its use of the 'partition' structure by which the traders receive information. The partition structure is the Kripke semantics for the modal logic **S5**.[1] The idea has been performed in different settings. This article presents the economy upon the logic of belief, and we extend the core equivalence theorem into the economy.

[1] The logic **S5** is denoted by **KT5** (= **KT45**) in Chellas [3].

This article is organized as follows: In Section 2 we present the multi-agent modal logic **B** and show the finite model property. Section 3 introduces the economy for logic **B** In Section 4 we present the notion of expectations equilibrium, which is a generalized notion of rational expectations equilibrium. We establishes the core equivalence theorem, giving an outline of the proof.

2 Multi-Modal Logics

2.1 Logic of belief

Let T be a set of *traders* and $t \in T$ a trader. The language is founded on as follows: The *sentences* of the language form the least set containing each *atomic* sentence $\mathbf{P}_m (m = 0, 1, 2, \dots)$ and closed under the following operations: Nullary operators for *falsity* $\bot$ and for *truth* $\top$; unary and binary syntactic operations for *negation* $\neg$, *conditionality* $\rightarrow$ and *conjunction* $\wedge$, *disjunction* $\vee$, respectively; two unary operations for *modalities* $\Box_t$, $\Diamond_t$ for $t \in T$. Other such operations are defined in terms of those in usual ways. The intended interpretation of $\Box_t \varphi$ is the sentence that 'trader t believes a sentence φ,' and the sentence $\Diamond_t \varphi$ is interpreted as the sentence that 'a sentence φ is possible for t.'

A *multi-modal logic* L is a set of sentences containing all truth-functional tautologies and closed under substitution and modus ponens. A multi-modal logic L' is an *extension* of L if $L \subseteq L'$. A sentence φ in a modal logic L is a *theorem* of L, written by $\vdash_L \varphi$. Other proof-theoretical notions such as *L-deducibility, L-consistency, L-maximality* are defined in usual ways. (Chellas [3].)

A *normal system* of multi-modal logic L is a multi-modal logic containing the schema $(\mathrm{Df}_\Diamond)$ and is closed under the $2n$ rules $(\mathrm{RK}_\Box)$, $(\mathrm{RK}_\Diamond)$ of inference.

$$(\mathrm{Df}_\Diamond) \quad \Diamond_t \varphi \longleftrightarrow \neg \Box_t \neg \varphi; \qquad (\mathrm{RK}_\Box) \quad \frac{(\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_k) \longrightarrow \psi}{(\Box_t \varphi_1 \wedge \Box_t \varphi_2 \wedge \cdots \wedge \Box_t \varphi_k) \longrightarrow \Box_t \psi} \quad (k \geq 0)$$

Definition 1. The *logic of belief* **B** is the minimal normal system of multi-modal logic.

2.2 Belief structure, model and truth

Let Ω be a non-empty set called a *state space* and 2^Ω the field of all subsets of Ω. Each member of 2^Ω is called an *event* and each element of Ω called a *state*. A *belief structure* is a tuple $\langle \Omega, (B_t)_{t \in T}, (P_t)_{t \in T} \rangle$ in which Ω is a state space and $B_t : 2^\Omega \rightarrow 2^\Omega$ is trader t's *belief operator*. The interpretation of the event $B_t E$ is that 't believes E.' P_t is t's *possibility operator* on 2^Ω defined by $P_t E = \Omega \setminus B_t(\Omega \setminus E)$ for every E in 2^Ω. The interpretation of $P_t E$ is that 'E is possible for t.'

A *model on* a belief structure is a tuple $\mathcal{M} = \langle \Omega, (B_t)_{t \in T}, (P_t)_{t \in T}, V \rangle$ in which $\langle \Omega, (B_t)_{t \in T}, (P_t)_{t \in T} \rangle$ is a belief structure and a mapping V assigns either **true** or **false** to every $\omega \in \Omega$ and to every atomic sentence $\mathbf{P}_m$. The model $\mathcal{M}$ is called *finite* if Ω is a finite set.

Definition 2. By $\models^{\mathcal{M}}_{\omega} \varphi$, we mean that a sentence φ is *true* at a state ω in a model $\mathcal{M}$. *Truth at a state* ω in $\mathcal{M}$ is defined by the inductive way as follows:

1. $\models^{\mathcal{M}}_{\omega} \mathbf{P}_m$ if and only if $V(\omega, \mathbf{P}_m) = \mathbf{true}$, for $m = 0, 1, 2, \ldots$;
2. $\models^{\mathcal{M}}_{\omega} \top$, and not $\models^{\mathcal{M}}_{\omega} \bot$;
3. $\models^{\mathcal{M}}_{\omega} \neg\varphi$ if and only if not $\models^{\mathcal{M}}_{\omega} \varphi$;
4. $\models^{\mathcal{M}}_{\omega} \varphi \longrightarrow \psi$ if and only if $\models^{\mathcal{M}}_{\omega} \varphi$ implies $\models^{\mathcal{M}}_{\omega} \psi$;
5. $\models^{\mathcal{M}}_{\omega} \varphi \wedge \psi$ if and only if $\models^{\mathcal{M}}_{\omega} \varphi$ and $\models^{\mathcal{M}}_{\omega} \psi$;
6. $\models^{\mathcal{M}}_{\omega} \varphi \vee \psi$ if and only if $\models^{\mathcal{M}}_{\omega} \varphi$ or $\models^{\mathcal{M}}_{\omega} \psi$;
7. $\models^{\mathcal{M}}_{\omega} \Box_t\varphi$ if and only if $\omega \in B_t(||\varphi||^{\mathcal{M}})$, for $t \in T$;
8. $\models^{\mathcal{M}}_{\omega} \Diamond_t\varphi$ if and only if $\omega \in P_t(||\varphi||^{\mathcal{M}})$, for $t \in T$:

Where $||\varphi||^{\mathcal{M}}$ denotes the set of all the states in $\mathcal{M}$ at which φ is true; this is called the *truth set of* φ.

We say that a sentence φ is *true in the model* $\mathcal{M}$ and write $\models^{\mathcal{M}} \varphi$ if $\models^{\mathcal{M}}_{\omega} \varphi$ for every state ω in $\mathcal{M}$. A sentence is said to be *valid in* a belief structure if it is true in every model on the belief structure. Let Γ be a set of sentences. We say that $\mathcal{M}$ is a *model for* Γ if every member of Γ is true in $\mathcal{M}$. A belief structure is said to be *for* Γ if every member of Γ is valid in it. Let $\mathbf{C}$ be a class of models on a belief structure. A multi-modal logic L is *sound with respect to* $\mathbf{C}$ if every member of $\mathbf{C}$ is a model for L. It is *complete with respect to* $\mathbf{C}$ if every sentence valid in all members of $\mathbf{C}$ is a theorem of L. A multi-modal logic L is said to have the *finite model property* if it is sound and complete with respect to the class of all finite models in $\mathbf{C}$. We denote by $\mathbf{C}_F$ the class of all finite models in $\mathbf{C}$. We can establish that

Theorem 1. *A normal system of multi-modal logic L has the finite model property; i.e.,* $\vdash_L \varphi$ *if and only if* $\models^{\mathcal{M}} \varphi$ *for all* $\mathcal{M} \in \mathbf{C}_F$. *In particular,* $\mathbf{B}$ *has the finite model property.*

3 Economy for Multi-Modal Logic

3.1 Information structure

We shall give the generalized notion of information partition in the line of Bacharach [2]. First, we note that $P_t(\{\omega\}) = \bigcap_{E \in 2^{\Omega}} \{E \,|\, \omega \in B_t E\}$.

Definition 3. The *associated information structure* $(P_t)_{t\in T}$ with a model on a belief structure $\langle \Omega, (B_t)_{t\in T}, (P_t)_{t\in T}, V\rangle$ for a normal system of multi-modal logic L is the class of t's *associated information functions* $P_t : \Omega \to 2^{\Omega}$ defined by $P_t(\omega) = \bigcap_{E\in 2^{\Omega}} \{E \,|\, \omega \in B_t E\}$. We denote by $\mathrm{Dom}(P_t)$ the set $\{\omega \in \Omega \mid P_t(\omega) \neq \emptyset\}$, called the *domain of* P_t.

Remark 1. M. Bacharach [2] introduces the *strong epistemic model* equivalent to the Kripke semantics for the modal logic **S5**. The strong epistemic model can be interpreted as the belief structure $\langle \Omega, (B_t)_{t\in T}, (P_t)_{t\in T}\rangle$ with B_t satisfying the schemas T, 4, and 5 in Chellas [3].

3.2 Economy on belief

A *pure exchange economy under uncertainty* is a tuple $\langle T, \Sigma, \mu, \Omega, \boldsymbol{e}, (U_t)_{t\in T}, (\pi_t)_{t\in T}\rangle$ consisting of the following structure and interpretations: There are l commodities in each state of the state space Ω , and it is assumed that Ω is *finite* and that the consumption set of trader t is $\mathbf{R}^l_+$; (T, Σ, μ) is the measure space of the traders, Σ is a σ-field of subsets of T whose elements are called *coalitions*, and μ is a measure on Σ; $\boldsymbol{e} : T \times \Omega \to \mathbf{R}^l_+$ is t's *initial endowment* such that $\boldsymbol{e}(\cdot, \omega)$ is μ-measurable for each $\omega \in \Omega$; $U_t : \mathbf{R}^l_+ \times \Omega \to \mathbb{R}$ is t's von-Neumann and Morgenstern utility function; π_t is a subjective prior on Ω for a trader $t \in T$.

Let L be a normal system of multi-modal logic. A pure exchange *economy for* L is a structure $\mathcal{E}^L = \langle \mathcal{E}, (B_t)_{t\in T}, (P_t)_{t\in T}, V\rangle$, in which $\mathcal{E}$ is a pure exchange economy under uncertainty, and $\langle \Omega, (B_t)_{t\in T}, (P_t)_{t\in T}, V\rangle$ is a finite model on a belief structure for L. By the *domain* of the economy $\mathcal{E}^L$ we mean $\mathrm{Dom}(\mathcal{E}^L) = \cap_{t\in T}\mathrm{Dom}(P_t)$. We always assume that $\mathrm{Dom}(\mathcal{E}^B) \neq \emptyset$.

Definition 4. An *economy on belief* is the pure exchange economy for the logic **B**, denoted by $\mathcal{E}^B$. The economy is called *atomless* if (T, Σ, μ) is a non-atomic measure space.

Remark 2. An economy under asymmetric information can be interpreted as the economy $\mathcal{E}^{S5}$ for the multi-modal logic **S5**, in which the belief structure $\langle \Omega, (B_t)_{t\in T}, (P_t)_{t\in T}, V\rangle$ is given by the strong epistemic model, and that $\mathrm{Dom}(\mathcal{E}^B) = \Omega$.

We denote by $\mathcal{F}_t$ the field of $\mathrm{Dom}(P_t)$ generated by $\{P_t(\omega)|\ \omega \in \Omega\}$ and denote by $\Pi_t(\omega)$ the atom containing $\omega \in \mathrm{Dom}(P_t)$. We denote by $\mathcal{F}$ the join of all $\mathcal{F}_t(t \in T)$ on $\mathrm{Dom}(\mathcal{E}^B)$; i.e. $\mathcal{F} = \vee_{t\in T}\mathcal{F}_t$, and denote by $\{\Pi(\omega)|\ \omega \in \mathrm{Dom}(\mathcal{E}^B)\}$ the set of all atoms $\Pi(\omega)$ containing ω of the field $\mathcal{F} = \vee_{t\in T}\mathcal{F}_t$. We shall often refer to the following conditions for $\mathcal{E}^B$: For every $t \in T$,

A-1 $\sum_{t\in T} \mathbf{e}(t, \omega) > 0$ for each $\omega \in \Omega$.

A-2 $\mathbf{e}(t, \cdot)$ is $\mathcal{F}$-measurable on $\mathrm{Dom}(P_t)$;

A-3 For each $x \in \mathbf{R}^l_+$, the function $U_t(x, \cdot)$ is at least $\mathcal{F}$-measurable on $\mathrm{Dom}(\mathcal{E}^B)$, and the function: $T \times \mathbf{R}^l_+ \to \mathbf{R}, (t, x) \mapsto U_t(x, \omega)$ is $\Sigma \times \mathcal{B}$-measurable where $\mathcal{B}$ is the σ-field of all Borel subsets of $\mathbf{R}^l_+$.

A-4 For each $\omega \in \Omega$, the function $U_t(\cdot, \omega)$ is continuous, strictly increasing and quasi-concave on $\mathbf{R}^l_+$.

4 Core Equivalence Theorem

4.1 Expectations equilibrium in belief.

An *assignment* for an economy $\mathcal{E}^B$ on belief is a mapping $\mathbf{x} : T \times \Omega \to \mathbf{R}^l_+$ such that for each $t \in T$, the function $\mathbf{x}(t, \cdot)$ is at least $\mathcal{F}$-measurable on $\mathrm{Dom}(\mathcal{E}^B)$. We denote by $\mathcal{A}ss(\mathcal{E}^B)$ the set of all assignments for the economy $\mathcal{E}^B$. By an *allocation* for $\mathcal{E}^B$ we mean an assignment $\mathbf{a}$ such that $\mathbf{a}(t, \cdot)$ is $\mathcal{F}$-measurable on $\mathrm{Dom}(\mathcal{E}^B)$ for all $t \in T$ and $\sum_{t\in T} \mathbf{a}(t, \omega) \leq \sum_{t\in T} \mathbf{e}(t, \omega)$ for every $\omega \in \Omega$. We denote by $\mathcal{A}lc(\mathcal{E}^B)$ the set of all allocations for $\mathcal{E}^B$.

We shall introduce the revised notion of trader's expectation of utility in $\mathcal{E}^B$. By t's *ex-ante* expectation we mean $\mathbf{E}_t[U_t(\mathbf{x}(t, \cdot)] := \sum_{\omega\in \mathrm{Dom}(P_t)} U_t(\mathbf{x}(t, \omega), \omega)\pi_t(\omega)$ for each $\mathbf{x} \in \mathcal{A}ss(\mathcal{E}^B)$. The *interim* expectation $\mathbf{E}_t[U_t(\mathbf{x}(t, \cdot)|P_t]$ is defined by $\mathbf{E}_t[U_t(\mathbf{x}(t, \cdot)|P_t](\omega) = \sum_{\xi\in \mathrm{Dom}(P_t)} U_t(\mathbf{x}(t, \xi), \xi)\pi_t(\{\xi\})|P_t(\omega))$ on $\mathrm{Dom}(P_t)$.

A *price system* is a non-zero function $p : \Omega \to \mathbf{R}^l_+$ which is $\mathcal{F}$-measurable on $\mathrm{Dom}(\mathcal{E}^B)$. We denote by $\Delta(p)$ the partition on Ω induced by p, and denote by $\sigma(p)$ the field of Ω generated by $\Delta(p)$. The *budget set* of a trader t at a state ω for a price system p is defined by $B_t(\omega, p) := \{\ x \in \mathbf{R}^l_+ \mid p(\omega) \cdot x \leq p(\omega) \cdot \mathbf{e}(t, \omega)\ \}$. Define the mapping $\Delta(p) \cap P_t : \mathrm{Dom}(P_t) \to 2^\Omega$ by $(\Delta(p) \cap P_t)(\omega) := \Delta(p)(\omega) \cap P_t(\omega)$. We denote by $\mathrm{Dom}(\Delta(p) \cap P_t)$ the set of all states ω in which $\Delta(p)(\omega) \cap P_t(\omega) \neq \emptyset$. Let $\sigma(p) \vee \mathcal{F}_t$ be the smallest σ-field containing both the fields $\sigma(p)$ and $\mathcal{F}_t$.

Definition 5. An *expectations equilibrium in belief* for an economy $\mathcal{E}^B$ on belief is a pair $(p, \mathbf{x})$, in which p is a price system and $\mathbf{x}$ is an assignment for $\mathcal{E}^B$ satisfying the following conditions:

EA1 $\mathbf{x}$ is an allocation for $\mathcal{E}^B$;
EA2 For all $t \in T$ and for every $\omega \in \Omega$, $\mathbf{x}(t, \omega) \in B_t(\omega, p)$;
EA3 For all $t \in T$, if $\mathbf{y}(t, \cdot) : \Omega \to \mathbf{R}^l_+$ is $\mathcal{F}$-measurable on $\mathrm{Dom}(\mathcal{E}^B)$ with $\mathbf{y}(t, \omega) \in B_t(\omega, p)$ for all $\omega \in \Omega$, then $\mathbf{E}_t[U_t(\mathbf{x}(t, \cdot))|\Delta(p)\cap P_t](\omega) \geq \mathbf{E}_t[U_t(\mathbf{y}(t, \cdot))|\Delta(p)\cap P_t](\omega)$ pointwise on $\mathrm{Dom}(\Delta(p) \cap P_t)$;
EA4 For every $\omega \in \mathrm{Dom}(\mathcal{E}^B)$, $\sum_{t\in T} \mathbf{x}(t, \omega) = \sum_{t\in T} \mathbf{e}(t, \omega)$.

The allocation $\mathbf{x}$ in $\mathcal{E}^B$ is called an expectations equilibrium *allocation* in belief for $\mathcal{E}^B$.

We denote by $EA(\mathcal{E}^B)$ the set of all the expectations equilibria of a pure exchange economy $\mathcal{E}^B$, and denote by $\mathcal{A}(\mathcal{E}^B)$ the set of all the expectations equilibrium allocations in belief for the economy.

4.2 Ex-post core.

An assignment $\mathbf{y}$ is called an *ex-post improvement* of a coalition $S \in \Sigma$ on an assignment $\mathbf{x}$ at a state $\omega \in \Omega$ if

Imp1 $\mu(S) > 0$; **Imp2** $\int_S \mathbf{y}(t, \omega)d\mu \leq \int_S \mathbf{e}(t, \omega)d\mu$; and
Imp3 $U_t(\mathbf{y}(t, \omega), \omega) > U_t(\mathbf{x}(t, \omega), \omega)$ for almost all $t \in S$.

Definition 6. An allocation $\mathbf{x}$ is said to be an *ex-post core* allocation of an economy with awareness structure $\mathcal{E}^B$ if there is no coalition having an ex-post improvement on $\mathbf{x}$ at any state $\omega \in \mathrm{Dom}(\mathcal{E}^B)$. The *ex-post core* denoted by $\mathcal{C}^{ExP}(\mathcal{E}^B)$ is the set of all the ex-post core allocations of $\mathcal{E}^B$.

4.3 Proof of main theorem

Now we can explicitly state the main theorem, and we shall sketch the proof.

Theorem 2. *Let $\mathcal{E}^B$ be a pure exchange economy with belief structure satisfying the conditions* **A-1***,* **A-2***,* **A-3** *and* **A-4***. Suppose that the economy is atomless. Then the ex-post core coincides with the set of all expectations equilibrium allocations in belief; i.e., $\mathcal{C}^{ExP}(\mathcal{E}^B) = \mathcal{A}(\mathcal{E}^B)$.*

Let $\mathcal{E}^B(\omega)$ be the economy with complete information $\langle T, \Sigma, \mu, \mathbf{e}(\cdot,\omega), (U_t(\cdot,\omega))_{t\in T}\rangle$ for each $\omega \in \Omega$. We denote by $\mathcal{C}(\mathcal{E}^B(\omega))$ the set of all core allocations for $\mathcal{E}^B(\omega)$, and by $\mathcal{W}(\mathcal{E}^B(\omega))$ the set of all competitive equilibria for $\mathcal{E}^B(\omega)$. We can observe that

Proposition 1. *Notations being the same as above, we obtain*

(i) $\mathcal{C}^{ExP}(\mathcal{E}^B) = \{\mathbf{x} \in \mathcal{A}lc(\mathcal{E}^B) \mid \mathbf{x}(\cdot,\omega)) \in \mathcal{C}(\mathcal{E}^B(\omega))$ *for all* $\omega \in \mathrm{Dom}(\mathcal{E}^B)\}$.

(ii) $\mathcal{A}(\mathcal{E}^B) = \{\mathbf{x} \in \mathcal{A}lc(\mathcal{E}^B) \mid$ *There is a price system* p *such that* $(p(\omega), \mathbf{x}(\cdot,\omega)) \in W(\mathcal{E}^B(\omega))$ *for all* $\omega \in \mathrm{Dom}(\mathcal{E}^B)\}$.

We note by the core equivalence theorem of Aumann ([1]) that $\mathcal{C}(\mathcal{E}^B(\omega)) = W(\mathcal{E}^B(\omega))$ for each $\omega \in \mathrm{Dom}(\mathcal{E}^B)$.

Proof of Theorem 2. Let $\mathbf{x} \in \mathcal{A}(\mathcal{E}^B)$. By Proposition 1 (ii) we obtain that for each $\omega \in \mathrm{Dom}(\mathcal{E}^B)$, $(p(\omega), \mathbf{x}(\cdot,\omega)) \in W(\mathcal{E}^B(\omega))$, and thus it follows from the theorem of Aumann ([1]) that $\mathbf{x}(\cdot,\omega)) \in \mathcal{C}(\mathcal{E}^B(\omega))$ for any $\omega \in \mathrm{Dom}(\mathcal{E}^B)$. It has been verified that $\mathcal{A}(\mathcal{E}^B) \subseteq \mathcal{C}^{ExP}(\mathcal{E}^B)$. The converse shall be shown as follows: Let $\mathbf{x} \in \mathcal{C}^{ExP}(\mathcal{E}^B)$. It follows from Proposition 1 (i) that $\mathbf{x}(\cdot,\omega) \in \mathcal{C}(\mathcal{E}^B(\omega))$ for every $\omega \in \mathrm{Dom}(\mathcal{E}^B)$. By the above theorem of Aumann ([1]) there is a $p^*(\omega) \in \mathbf{R}^l_+$ such that $(p^*(\omega), \mathbf{x}(\cdot,\omega)) \in W(\mathcal{E}^B(\omega))$. Let p be the price system defined by $p(\xi) := p^*(\omega)$ for all $\xi \in \Pi(\omega)$ and $\omega \in \mathrm{Dom}(\mathcal{E}^B)$, $p(\omega) := p^*(\omega)$ for $\omega \notin \mathrm{Dom}(\mathcal{E}^B)$. we obtain that $(p(\omega), \mathbf{x}(\cdot,\omega)) \in W(\mathcal{E}^B(\omega))$ for every $\omega \in \mathrm{Dom}(\mathcal{E}^B)$. By Proposition 1 (ii), we have observed that $\mathcal{C}^{ExP}(\mathcal{E}^B) \subseteq \mathcal{A}(\mathcal{E}^B)$. □

References

1. R. J. Aumann: Markets with a continuum of traders, Econometrica 32 (1964) 39–50.
2. M. O. Bacharach: Some extensions of a claim of Aumann in an axiomatic model of knowledge, Journal of Economic Theory 37 (1985) 167–190.
3. B. F. Chellas: *Modal Logic: An introduction.* Cambridge University Press, Cambridge, London, New York, New Rochelle, Melbourne, Sydney, 1980.
4. E. Einy, D. Moreno, and B. Shitovitz: Rational expectations equilibria and the ex-post core of an economy with asymmetric information, Journal of Mathematical Economics 34 (2000) 527–535.
5. F. Forges, E. Minelli, and R. Vohra: Incentive and core of an exchange economy: a survey, Journal of Mathematical Economics 38 (2002) 1–41.

A Comparison of Two Parametric ROC Curves Estimators in Binormal Model [1]

Jaroslav Michálek[2], Marek Sedlačík[3] and Lucie Doudová[4]

[2]Faculty of Economics and Administration, Masaryk University
Lipová 41a, 602 00 Brno
michalek@econ.muni.cz

[3]Faculty of Economics and Management, University of Defence
Kounicova 65, 602 00 Brno
marek.sedlacik@unob.cz

[4]Faculty of Sciences, Masaryk University
Janáčkovo nám. 2a, 662 95 Brno
ldoudova@centrum.cz

Abstract. The receiver operating characteristic curve $ROC(t)$ plays an important role in many economic branches when it is necessary to classify between two populations. There are several approaches how to estimate it: classical estimator based on the empirical cumulative distributive function, the weighted regression estimator, the estimator based on the best unbiased cdf estimator, non-parametric approaches based on the kernels or bootstrapping.
In the contribution the binormal model will be considered and the attention will be concentrated on the comparison of two parametric methods. The first one is the method of reweighted least squares where weights depend on the response variable. The second one is based on the best unbiased cdf estimator. The comparison of both methods will be demonstrated by simulations and the performance of both methods will be discussed for small sample sizes.
...

Key words : *ROC curve, ODC curve, estimate based on the best unbiased cdf estimate (EBBUCE), generalized least squares method estimate (GLME), adaptive generalized least squares method estimate (AGLME).*

1 The cdf curve estimators

Let $X_1, \ldots, X_m$ be a random sample from the distribution with cumulative distributive function (cdf) F and let F_m be the corresponding empirical (sample) cdf. The empirical cdf $F_m(x)$ is in each point x the unbiased and consistent

[1] Supported by GA ČR under the Grant No. 402/04/1308.

estimator of the cdf $F(x)$. In the case when the cdf F belongs to the known family of distributions it is possible to look for its better estimator. This case will be considered here. Under assumptions that F belongs to the normal family of distributions $N(\mu, \sigma^2)$ the best unbiased estimator of the cdf F will be studied.

Let $X_1, \ldots, X_m$ be a random sample from the normal family $N(\mu, \sigma^2)$ with unknown parametr $\Theta = (\mu, \sigma^2)$. Then for each real x we look for the estimate of the parametric function

$$\gamma(\Theta) = P_\Theta(X_1 \leq x) = F(x, \Theta),$$

where F is the cdf of $N(\mu, \sigma^2)$. Further let $\overline{X}$ be the sample mean and S^2 be the sample variance. Then it is simple to derive following results (see [2]):

1. For any real x the indicator of the event $[X_1 \leq x]$ is an unbiased estimate of $\gamma(\Theta)$.
2. The statistic $\mathbf{U} = (\overline{X}, S^2)$ is a complete and sufficient statistic for Θ.
3. The statistic $Y = \frac{X_1 - \overline{X}}{(m-1)S}\sqrt{m}$ is an ancillary statistic for Θ.
4. $\mathbf{U}$ and Y are independent.
5. The density function of Y is given by

$$\begin{aligned} g(y) &= \frac{1}{\beta(\frac{1}{2}, \frac{m-2}{2})}(1-y^2)^{\frac{m-4}{2}} \quad \text{for} \quad y \in <-1, 1> \\ g(y) &= 0 \quad \text{otherwise} \end{aligned} \tag{1}$$

   and $\beta(p, q)$ in (1) stands for the beta function.
6. For the cdf $F_Y(y)$ of Y holds

$$\begin{aligned} F_Y(y) = \int_{-\infty}^{y} g(t)dt &= 0 \quad \text{for} \quad y \leq -1 \\ &= \frac{1}{2} - \frac{1}{2}\beta_{y^2}(\frac{1}{2}, \frac{m}{2} - 1) \quad \text{for} \quad -1 < y \leq 0 \\ &= \frac{1}{2} + \frac{1}{2}\beta_{y^2}(\frac{1}{2}, \frac{m}{2} - 1) \quad \text{for} \quad 0 < y \leq 1 \\ &= 1 \quad \text{for} \quad y > 1, \end{aligned}$$

   where $\beta_a(p, q) = \int_0^a t^{p-1}(1-t)^{q-1}dt$ stands for incomplete beta function, $a \in \langle 0, 1\rangle, p \geq 0, q \geq 0$.

Now let us put

$$Q = Q(x) = \frac{x - \overline{X}}{(m-1)S}\sqrt{m}.$$

Then using previous results, properties of the conditional expectation and Rao-Blackwell Theorem we obtain (see [2] for details) that for any real x the best

unbiased estimate of $\gamma(\Theta) = F(x, \Theta)$ is the statistic

$$\begin{aligned}\widetilde{F}(x) = \widetilde{\gamma}(\Theta) &= 0 \quad \text{for} \quad Q(x) \leq -1 \\ &= \frac{1}{2} - \frac{1}{2}\beta_{Q^2(x)}(\frac{1}{2}, \frac{m}{2} - 1) \quad \text{for} \quad -1 < Q(x) \leq 0 \\ &= \frac{1}{2} + \frac{1}{2}\beta_{Q^2(x)}(\frac{1}{2}, \frac{m}{2} - 1) \quad \text{for} \quad 0 < Q(x) \leq 1 \\ &= 1 \quad \text{for} \quad Q(x) > 1.\end{aligned}$$

2 The ROC and ODC curves

The receiver operating characteristic (ROC) curve is used for classification between two groups of subjects. One will be called the group with condition and the other will be called the group without condition. The ROC is defined as a plot of probability of incorrect classification of subjects from the group without condition (that have low value of a discrimination score X - lower then given cut-point c) versus the probability of correct classification of subjects from the group with condition across all possible cut-point values c of a discrimination score X. The group without condition is characterized by lower values of X and let $F(x)$ be the cdf of X in the group without condition. On the contrary, the group with condition is characterized by higher values of X and the cdf of X in the group with condition will be denoted by $G(x)$. Then the ROC curve is defined as $ROC(t) = 1 - G(F^{-1}(1-t))$ for $0 < t < 1$ if the inverse F^{-1} exists. Equivalently, reversing the axes, the curve $F(G^{-1}(t))$ for $t \in < 0, 1 >$ is called ordinal dominance curve (ODC). Thus the ODC can be viewed as a plot of $G(c)$ on horizontal axis versus $F(c)$ on vertical axis and the ROC curve can be viewed as a plot of $1 - G(c)$ on the vertical axis versus $1 - F(c)$ on the horizontal axis for all cut-points $c \in (-\infty, \infty)$.

A popular approach to ROC and ODC curves estimates is to assume so-called binormal model. This is a semiparametric approach which postulates that both F and G distributions are normal ones. Without loss of generality it is possible to assume that F is $N(0, 1)$ and G is $N(\mu, \sigma^2)$ where μ and σ^2 are unknown parameters. Thus instead of F the notation Φ will be used for standard normal cdf.

3 The ROC and ODC curves estimates

We consider two independent random samples, the first $X_1 \ldots, X_m$ is from the distribution with cdf F and the second $Y_1 \ldots, Y_n$ is from the distribution with cdf G. The classical approach to ROC and ODC curves estimates is based on empirical cdf's. Thus the estimate of ODC curve can be obtained as a plot of $F_m(c)$ on vertical axis versus $G_n(c)$ on horizontal axis. Similarly the ROC curve estimate can be obtained. In the frame of binormal model, the empirical estimates F_m and G_n can be substituted by the best unbiased estimates $\widetilde{F}_m$ and

$\widetilde{G}_n$ of F and G respectively. This way more suitable estimate for small sample sizes m and n will be obtained. The estimate will be called the estimate based on the best unbiased cdf estimate (EBBUCE).

4 Estimates of the ODC curve using the generalized least squares method

The binormal model will be assumed throughout the paragraph. Thus for $0 \leq t \leq 1$ the ODC curve can be expressed as

$$ODC(t) = F(G^{-1}(t)) = \Phi(\mu + \sigma\Phi^{-1}(t)) \tag{2}$$

where unknown parameters μ and σ are to be estimated.

Further the ODC curve will be considered at k points $0 < t_1 < t_2 < \cdots < t_k < 1$ and let us put

$$\beta_i = ODC(t_i) = F(G^{-1}(t_i)) = \Phi(\mu + \sigma\Phi^{-1}(t_i)) \tag{3}$$

Then as a natural estimator of β_i can be used

$$\widehat{\beta}_i = F_m(G_m^{-1}(t_i)), \text{ for } i = 1, \ldots, k, \tag{4}$$

where F_m and G_n are the empirical cdf's and G_m^{-1} is quantile function a corresponding to G_m. From the covariance structure of Brownian bridges can be derived the asymptotic distribution of $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_k)'$ (see [1] for details). It is possible to show that for fixed $0 < t_1 < \cdots < t_k < 1$ and under the binormal assumption

$$\sqrt{n}(\widehat{\beta} - \beta) \sim^A N(0, \lambda\Sigma_1 + \Sigma_2) \tag{5}$$

and

$$\sqrt{n}(\Phi^{-1}(\widehat{\beta}) - \Phi^{-1}(\beta)) \sim^A N(0, \Sigma), \tag{6}$$

where

- $\sim^A$ denotes the asymptotic distribution when $m \to \infty$ and $n \to \infty$ and $\frac{m}{n} \to \lambda$
- $\Sigma = C[\lambda\Sigma_1 + \Sigma_2]C$ and $C^{-1} = \text{diag}(\ldots, \phi(\mu + \sigma\Phi^{-1}(t_i)), \ldots)$
- $\Sigma_1 = (\sigma_{ij}^1)$ and $\sigma_{ij}^1 = \min\{\beta_i, \beta_j\} - \beta_i\beta_j$
- $\Sigma_2 = A\Sigma_0 A$, where $A = \text{diag}(\ldots, (\sigma\phi(\mu + \sigma\Phi^{-1}(t_i)))/(\phi(\Phi^{-1}(t_i))), \ldots)$ and $\Sigma_0 = (\sigma_{ij}^0)$, $\sigma_{ij}^0 = \min\{t_i, t_j\} - t_i t_j$.

Thus using (3) and (4) the following linear regression model can be considered:

$$\Phi^{-1}(\widehat{\beta}_i) = \mu + \sigma\Phi^{-1}(t_i) + \varepsilon_i, \qquad i = 1, \ldots, k. \tag{7}$$

The error vector, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_k)'$, has the distribution specified in (5) and (6). Since Σ depends on the unknown parameters μ a σ, an iterative procedure is needed to find the estimates of μ a σ (see [1]). The results are as follows:

The ordinary least squares estimates of μ a σ in the linear regression model are given by the formula

$$\begin{pmatrix} \widehat{\mu}_0 \\ \widehat{\sigma}_0 \end{pmatrix} = (M'M)^{-1}M'\Phi^{-1}(\widehat{\beta}), \tag{8}$$

where M is the design matrix given by

$$M' = \begin{pmatrix} 1 & \dots & 1 \\ \Phi^{-1}(t_1) & \dots & \Phi^{-1}(t_k) \end{pmatrix} \tag{9}$$

and the vector $\Phi^{-1}(\widehat{\beta})' = (\Phi^{-1}(\widehat{\beta_1}), \dots, \Phi^{-1}(\widehat{\beta_k}))$.

Substituting $\widehat{\mu}_0$ and $\widehat{\sigma}_0$ for μ and σ in the expression for Σ from the formula (6) and writing $\widehat{\Sigma}$ instead of Σ, the one-step generalized least squares estimator will be obtained by

$$\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} = (M'\widehat{\Sigma}^{-1}M)^{-1}M'\widehat{\Sigma}^{-1}\Phi^{-1}(\widehat{\beta}). \tag{10}$$

The procedure could be further iterated and this way the reweighted regression estimator of the ODC will be obtained. Often one step procedure is quite adequate. For asymptotic distribution of $(\widehat{\mu}, \widehat{\sigma})$ holds

$$\sqrt{n}\begin{pmatrix} \widehat{\mu} - \mu \\ \widehat{\sigma} - \sigma \end{pmatrix} \sim^A N(0, (M'\Sigma^{-1}M)^{-1}).$$

Further we call the ODC estimate based on the parameter estimates (10) the generalized least squares method estimate (GLME).

The transformation Φ^{-1} on the left side of the regression model (7) can lead to the unstable GLME estimates when the sample sizes m and n are moderate. Thus in [1] the adaptive procedure for the selection of points $t_1, \dots, t_k$ has been suggested to improve the estimate. Such improved adaptive GLME estimate will be further referred as the adaptive generalized least squares method estimate (AGLME).

5 A comparison of EBBUCE and AGLME estimators

The ROC estimators has been compared by simulations for small sample sizes. The GLM estimate depends on the selection of starting vector $(t_1, \dots, t_k)$ and this dependency can be reduced using the procedure for the optimal choice of points $(t_1, \dots, t_k)$.

Two binormal situations were simulated, in which F is $N(0, 1)$, G is $N(2, 1)$ and corresponding sample sizes are $m = n = 15, 30$. The comparison of discussed ROC curve estimators with the theoretical ROC curve are displayed in Figure 1.

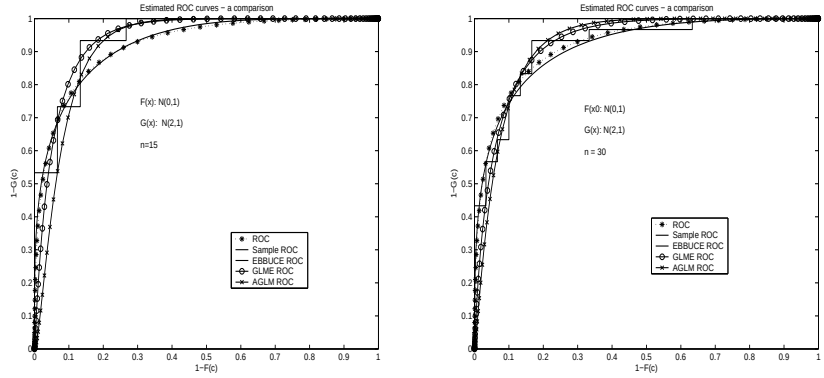


Fig. 1. *Simulations results*

6 Conclusion

For small sample sizes the simulations experiences lead to the conclusions that the EBBUCE estimate is closer to the theoretical ROC curve then AGLME estimate. The AGLME copies more the starting empirical ROC curve.

References

1. F. Hsieh and B. W. Turnbull. Nonparametric and semiparametric estimation of receiver operating characteristic curve. *The Annals of Statistics*, **24** (1996), no. 1, pp. 25–40.
2. J. Michálek and V. Veselý. The ROC and ODC curves estimators by means of the best unbiased CDF estimators. XXIII International Colloquium on the Acquisition Process Management. Proceedings of abstracts and electronic version of contributions on CD-ROM. UD Brno, May 19, 2005, pp. 34.

Analysis of Style Investing and Evaluating Style Analysis of the Mutual Funds

Zdenka Milánová ,
Faculty of Social and Economic Sciences, Institute of Economics, Comenius University.

Abstract

We try to study prices of assets, because some investors classify risky assets into different styles and move funds between these styles depending on their relative performance. We assume that news about one style can affect the prices of other apparently unrelated style, that assets in the same style commove to much while assets in different styles commove too little and high average returns on a style will be associated with common factors for reason unrelated to risk. In the last part we try to apply these conclusions on Slovak financial market.

Key words

Mutual funds, Style Investing, Style Analysis

Classification is the grouping of objects into categories based on some similarity among them [8]. Classification of large number of objects into categories is also in financial markets. The asset classes that investors use in this process are sometimes called investing styles, and the process of allocating funds among styles is known as style investing.

There are several reasons why investors use investing styles:

1. Categorization simplifies problems of choice and allows us to process vast amounts of information reasonably efficiently.
2. Creation of asset categories helps investors evaluate the performance of professional money managers, because style automatically creates a group of managers who pursue that particular style.
3. Style investing simplifies the process of diversification.

Let's create a model of style investing. We consider an economy with 2*n* risky assets in fixed supply, and a risk free asset, cash, in perfectly elastic supply and with zero net return. We model risky asset *i* as a claim to a single liquidating dividend $D_{i,T}$ to be paid at some later time *T*. The eventual dividend equals

$$D_{i,T} = D_{i,0} + \varepsilon_{i,1} + \ldots + \varepsilon_{i,T}, \tag{1}$$

where $\varepsilon_{i,t}$ becomes known at time *t* [5]. We assume

$$\varepsilon_t = \left(\varepsilon_{1,t}, \ldots, \varepsilon_{2n,t}\right)' \sim N\left(0, C_D\right),$$

The price of a share of risky asset *l* at the time *t* is $P_{l,t}$ and the return on the asset between time *t*-1 and time *t* is

$$\Delta P_{l,t} = P_{l,t} - P_{l,t-1}. \tag{2}$$

We build a simple model of style investing. There are just two styles, X and Y, and each risky asset in the economy belongs to one, and only one, of these two styles Risky asset 1 trough *n* are in style X, while *n*+1 through 2*n* are in style Y. For now, we assume that the composition of two styles is the same in every time period.

As a measure of the value of style *X* in time *t*, we use $P_{X,t}$ defined as the average price of a share across all assets in style *X*:

$$P_{X,t} = \frac{1}{n} \sum_{l \in X} P_{l,t} \tag{3}$$

The return on style X between time *t*-1 and time *t* is

$$\Delta P_{X,t} = P_{X,t} - P_{X,t-1}. \tag{4}$$

There are two kinds of investors in our model, „switchers“ and „fundamental traders“. The investment policy of switchers has two distinctive characteristics:

1. Switchers allocate funds at the level of a style
2. How much they allocate to each style depends on that style's past performance relative to other style.

Each period, switchers allocate more funds to styles with better than average performance and finance these additional investments by taking funds away from styles with below average performance.

Switcher's demand [1] for shares of an asset i in style X at time t, is:

$$N_{i,t}^{S} = \frac{N_{X,t}^{S}}{n} = \frac{1}{n}\left[A_X + \sum_{k=1}^{t-1} \theta^{k-1} \left(\frac{\Delta P_{X,t-k} - \Delta P_{Y,t-k}}{2} \right) \right], \quad \text{pre } i = 1,2,\ldots,n, \tag{5}$$

where A_x and θ are constants, with $0 < \theta < 1$.

The second investor type in our model is fundamental trader. They act as arbitrageurs and try to prevent the price of an asset from deviating too far from its expected final dividend. Since they have no constraints on their allocations, they solve

$$\max_{N_t} E_t^F \left(-\exp\left[-\gamma \left(W^F + N_t' \left(P_{t+1} - P_t \right) \right) \right] \right) \tag{6}$$

where $N_t = \left(N_{1,t}, \ldots, N_{2n,t} \right)'$ is a vector of the number of shares allocated to each risky asset, γ governs the degree of risk aversion, E_t^F denote fundamental traders expectations at time t and $P_t = \left(P_{1,t}, \ldots, P_{2n,t} \right)'$. Optimal holdings N_t^F are given:

$$N_t^F = \frac{\left(V_t^F \right)^{-1}}{\gamma} \left(E_t^F \left(P_{t+1} \right) - P_t \right), \tag{7}$$

where

$$V_t^F = \text{var}_t^F \left(P_{t+1} - P_t \right),$$

with the F superscript in V_t^F again denoting a forecast made by fundamental traders.

An **individual stock-level momentum strategy** ranks all stocks on their return in the previous period and buys those stocks that did better than average and sells those that did worse. It can by implemented through

$$N_{i,t} = \frac{1}{2n} \left[\Delta P_{i,t} - \Delta P_{M,t} \right], \qquad i = 1,\ldots,2n. \tag{8}$$

where

$$\Delta P_{M,t} = \frac{1}{2n} \sum_{l=1}^{2n} \Delta P_{l,t}$$

An individual stock-level value strategy buys (sells) those stocks which are trading below (above) fundamental value $P_{i,t}^*$:

$$N_{i,t} = \frac{1}{2n} \left[P_{i,t}^* - P_{i,t} \right], \qquad i = 1,\ldots,2n. \tag{9}$$

A **style-level momentum strategy** buys into styles with good recent performance and avoids styles that have done poorly:

$$N_{i,t} = \frac{1}{2n}\left[\frac{\Delta P_{X,t} - \Delta P_{Y,t}}{2}\right], \quad i \in X,$$
$$N_{j,t} = \frac{1}{2n}\left[\frac{\Delta P_{Y,t} - \Delta P_{X,t}}{2}\right], \quad j \in Y \tag{10}$$

A **style-level value strategy** buys into styles trading below fundamental value and shorts the remaining styles

$$N_{i,t} = \frac{1}{2n}\left[P^*_{X,t} - P_{X,t}\right] \quad i \in X \tag{11}$$
$$N_{j,t} = \frac{1}{2n}\left[P^*_{Y,t} - P_{Y,t}\right] \quad j \in Y.$$

Essentially, in **return based style analysis** a factor model is used to explain fund returns. Return based analysis determines the **mimicking portfolio of mutual funds** or other investment opportunities with positive portfolios weights, i.e., the positively weighted style portfolio that is closest to the mutual fund in a least squares sense
The case when no constraints are imposed on the factor loadings will be referred to as weak style analysis. The case where only the portfolio constraints is imposed will be referred to as semi – strong style analysis and the case where both portfolio and the positivity constraints are imposed will be referred to as strong style analysis [4].

Suppose that *K* factor (mimicking) portfolios with return vector R_t drive the asset returns. In addition, there are *N* mutual funds with return vector r_t, for which we have the linear factor model

$$r_t = a + BR_t + \varepsilon_t \tag{12}$$

where $E\left[\varepsilon_t\right] = E\left[\varepsilon_t R_{i,t}\right] = 0$ for $i = 1,\ldots,K$. In this case $B = C_{rR}C_{RR}^{-1}$ and $a = \mu_r - B\mu_R$, where C_{rR} is a covariance matrix between returns of mutual funds and mimicking, C_{RR} is a covariance matrix of returns of mimicking portfolio and μ is an expected return vector. When using (12) as a factor model , we do not impose any constraints on a and B. If there are no restriction on B, we refer to this as weak style analysis.
If we define a_i as the *i*th element of *a* and b_i as the *i*th row of *B*, then a_i and b_i are the solutions to the problem

$$\min_{\alpha,\beta} E\left[\left(r_{i,t} - \alpha - \beta' R_t\right)^2\right] \tag{13}$$

The vector *b* reflect the fund mimicking positions or the minimum variance hedge position for the mutual fund.
To see the effect of the portfolio constraint $\sum_j \beta_j = 1$, let $\widetilde{a}_i$ and $\widetilde{b}_i$ be the solutions of the problem

$$\min_{\alpha,\beta}\left[\left(r_{i,t} - \alpha - \beta' R_t\right)^2\right],, \tag{14}$$

s.t.

$$\beta' e_K = 1$$

where e_K is K-dimensional vector of ones. Thus, $\widetilde{b}_i$ are the factor exposures which are constrained to sum to one, i.e., they characterize a portfolio. The case where only the portfolio constraints is imposed, will be referred to as semi-strong style analysis . Using standard least squares results, it is straightforward to show that the coefficients $\widetilde{b}_i$ can be written as

$$\tilde{b}_i = b_i + \left(1 - b_i' e_K\right) C_{RR}^{-1} e_K \left(e_K' C_{RR}^{-1} e_K\right)^{-1}. \tag{15}$$

In addition to the portfolio constraints, it is common in style analysis to impose positivity constraints on the estimated factor exposures. The style portfolios $\hat{b}_i$ and the associated intercepts $\hat{a}_i$ are then the solutions to the problem

$$\min_{\alpha,\beta} E\left[\left(r_{i,t} - \alpha - \beta' R_t\right)^2\right], \tag{16}$$

s.t.

$$\beta' e_K = 1,$$
$$\beta \geq 0,$$

We refer to this case as strong style analysis.
The strong style coefficients as given in (16) reflect the positively weighted portfolio of the benchmarks that mimics the mutual fund.

Since we have no detailed information about portfolio of mutual funds in Slovakia, to see consequences of propositions in this article, we are going to create two fund of funds. The first one consists of three bond mutual funds and the second one consists of three stock mutual funds. The first one we call investing style *X* and the second one investing style *Y*. In investing style *X* are these three funds: TAM - Korunový dlhopisový, VÚBAM – Eurofond a VÚBAM – Korunový. Investing style *Y* consists of these funds: TAM - Americký akciový, JTAM - Japan akciový a JTAM - Euro akciový. We have weekly data from 1.1.2002 to 31.1.2003. To get more information see [10].

In Table 1 is a practical application of relation (8). If a return of mutual fund is lower than average market return off all assets $\Delta P_{M,t}$, it will cause sale of some share of this fund. For example mutual fund VÚBAM – Eurofond has in the second week negative return -0,58%, what is less than average market return of all mutual funds and it leads to drain of funds.

Table 1: Individual stock-level momentum strategy in investing style X

Individual stock-level momentum strategy in investing style X							
Week	TAM - Korunový		VÚBAM - Eurofond		VÚBAM - Korunový		$\Delta P_{M,t}$
t	Return	$N_{1,t}$	Return	$N_{2,t}$	Return	$N_{3,t}$	
1	0,38%	0,03%	0,47%	0,04%	0,38%	0,03%	0,21%
2	0,34%	0,11%	-0,58%	-0,04%	0,35%	0,12%	-0,34%

In the Table 2 we can see how works individual stock-level strategy, which buy (sell) assets, which are traded below (above) the fundamental value. Let suppose, that $P_{i,t}$ is a Netto Asset Value and $P_{i,t}^*$ is expected value of the assets. Netto Asset Value of mutual fund VÚBAM-Eurofond is higher than fundamental value 1,0197 and it cause outflow of funds from this mutual fund, because individual stock-level strategy expects decrease of value of share in the next period.

Table 2: Individual stock-level value strategy of investing style X

Individual stock-level value strategy of investing style X						
Week	TAM - Korunový		VÚBAM - Eurofond		VÚBAM - Korunový	
t	NAV	$N_{1,t}$	NAV	$N_{2,t}$	NAV	$N_{3,t}$
1	1,5713	1,21%	1,0444	-0,41%	1,0859	0,56%
2	1,5772	1,11%	1,0493	-0,49%	1,0900	0,50%
	$P^*_{1,t}$	1,6439	$P^*_{2,t}$	1,0197	$P^*_{3,t}$	1,1197

Style-level momentum strategy shifts funds into style with good recent performance and avoids investing styles with poor performance. In the Table 3 there are results of relation (10).In the first period the performance of investing style *X* is better than performance of investing style *Y*, it causes shift of funds from style *Y* to style *X*.

Table 3: Style-level momentum strategy of investing style *X* and *Y*.

Style-level momentum strategy of investing style X and Y				
Week	Return of style X	Demand of X	Return of style Y	Demand of Y
t	$\Delta P_{X,t}$	$N_{i,t}$	$\Delta P_{Y,t}$	$N_{j,t}$
1	0,40%	0,03%	0,02%	-0,03%
2	0,08%	0,07%	-0,71%	-0,07%

In the Table 4, there are results of relation (11). Style – level value strategy buys into styles trading below fundamental value and shorts the remaining styles. Netto Asset Value of investing style *X* is lower than fundamental value of this style and it leads to inflow of funds into this style, because we expect that the performance of style will be better in the future. The inverse situation is in style *Y*.

Table 4: Style-level value strategy of investing style *X* and *Y*.

Style-level value strategy of investing style X and Y				
Week	Investing style *X*		Investing style *Y*	
t	NAV of style *X*	$N_{i,t}$	NAV of style *Y*	$N_{j,t}$
1	1,2339	0,45%	0,9875	-2,28%
2	1,2388	0,37%	0,9877	-2,28%
	$P^*_{X,t}$	1,2611	$P^*_{Y,t}$	0,8508

In the next table are results of strong style analysis (16). We have five mutual funds and we try to create mimicking portfolio, which is a combination of fund of funds VB - Quartett "Sicherheit" and index SAX. The reason is that passively directed mutual funds have lower costs than actively directed mutual funds. When intercept is negative it looks like actively managed mutual funds underrate mimicking portfolio. Negative intercept does not mean that investor should invest into mimicking portfolio. To find out what is better there is computed Sharpe ration for mimicking portfolio and for mutual fund. When there is negative intercept we choose portfolio with higher Sharpe ratio. Mutual funds TAM-Dolárový dlhopisový and TAM - Americký akciový have negative intercept. In this case investor should insert funds into mimicking portfolio with higher Sharpe ratio.

Table 5: Strong analysis of investing style

Mutual fund	Strong analysis of investing style		
	$\widehat{a}$	$\widehat{b}_1$	$\widehat{b}_2$
Sporo - Korunový peňažný	0,0021	0,1556	0,8444
TAM - Dolárový dlhopisový	-0,0010	0,1017	0,8983
VÚBAM – Eurofond	0,0008	0,1529	0,8471
TAM - Americký akciový	-0,0054	0,0527	0,9473
TAM - Korunový akciovo-dlhopisový	0,0022	0,2711	0,7289

Table 6: Sharpe ratio of mimicking portfolio and mutual fund

Mutual fund	Sharpe ratio of mimicking portfolio	Sharpe ratio of mutual fund
Sporo - Korunový peňažný	-0,1289	0,2283
TAM - Dolárový dlhopisový	-0,1571	-0,1772
VÚBAM – Eurofond	-0,1305	-0,0693
TAM - Americký akciový	-0,1779	-0,2498
TAM - Korunový akciovo-dlhopisový	-0,0604	0,1933

The goal of this paper was to show technique of style investing in the space of Slovak financial market. According to limited possibility of this paper the practical illustration was restricted only for few mutual funds and time interval is shorter than in [9], where is illustrated longer time horizon. In this paper is used evaluating style analysis as a tool for decision making between mutual fund and mimicking portfolio which is cheaper in the sense of cost.

References

[1] Barberis, N.–Shleifer, A.: *Style Investing*, University of Chicago, 2001,

[2] Barberis, N.–Shleifer, A.: *Style Investing*, Discussion Paper Number 1908, Harvard University 2000,

[3] Bernstein, R.: *Style Investing*, Wiley, New York, 1995

[4] DeRoon, F.A– Nijman, Th.E.– TerHorst, J.R.: *Evaluating Style Analysis*, Tilburg University, 2000.

[5] Hong, H.–Stein, J.: *A Unified Theory of Underreaction, Momentum Traiding and Overreaction in Asset Markets*, Journal of Finance 54, 2143-2184, 1999

[6] Kohout, P.: *Investiční strategie pro třetí tisiciletí*, Grada, Praha, 2000.

[7] Mlynarovič, V.:*Finančné investovanie–Teórie a aplikácie*, IURA, Bratislava, 1999.

[8] Rosch, E.–Lloyd, B.: *Cognition and Categorization*, New Jersey: Lawrence Erlbaum Associates, (1978)

[9] Practical application: Milánová, Z.: Analýza investičných štýlov podielových fondov, DP, Ekonomická univerzita v Bratislave, 2003.

Alternative Portfolio Selection Models

Vladimír Mlynarovič
Comenius University Bratislava

Abstract

The paper treats alternative portfolio selection models as an extension to the model in the mean – variance space. The basic motivation follows from the facts that recent statistical studies revealed that not all assets follow normal distribution and portfolio management using lower partial risk measures is attracting more attention of practitioners.

At the first part of the paper such lower partial risk measures as lower semi – variance (lower semi – standard deviation), lower semi – absolute deviation, below target risk and conditional value at risk are presented. In the second part of the paper portfolio selection models based on these risk measures are formulated.

Keywords

Lower partial risk measures**,** mean – lower semi-absolute deviation model, mean – conditional value at risk model

1 Introduction

Markowitz mean – variance (MV) portfolio selection model [1] that one can write as a multiple objective problem

$$\text{"max"} \quad \{E(\mathbf{w}), -V(\mathbf{w})\}$$

subject to

$$\mathbf{e}^T\mathbf{w} = 1$$

where

$$E(\mathbf{w}) = \mathbf{E}^T\mathbf{w}$$
$$V(\mathbf{w}) = \mathbf{w}^T\mathbf{C}\mathbf{w}$$

and **E** is the vector of expected returns of assets, **C** is the covariance matrix, **w** is the vector of weights of assets and **e** denotes the vector which elements are equal 1, is still very popular in the practice of fund management. However in the recent years one can observe radical changes in the investment environment. There exist various financial instruments with non-symmetric return distribution such as options and bonds. In addition recent statistical studies revealed that not all common stocks follow normal distribution, which was usually considered to be a valid assumption for these kinds of stocks. It follows that one can not fully rely on MV model and should look for other risk measures than variance.

2 Lower partial risk measures

Portfolio management using lower partial risk (downside risk) is attracting more attention of practitioners in recent years. In this part of the paper a review of such kind of measures is presented. In Zeleny [2] it is shown that almost all commonly used risk measures can be viewed as special cases of the family of three - parameter risk measures. This rich and fairly general group of functions provides the research with an

infinite variety of "risk measures" Ω through different combinations pf appropriate parameters values of c, α and λ:

$$\Omega(c,\alpha,\lambda) = \sum_{r_k \le \lambda} |r_k - c|^{\alpha} p_k, \quad resp. \quad \Omega(c,\alpha,\lambda) = \int_{-\infty}^{\lambda} |r_k - c|^{\alpha} dF(r)$$

where p_k is the probability of kth level r_k of 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 $\alpha > 0$. Possible choices for parameter λ include 8, c, a desired target level return, and some others. Of course it is possible to introduce the "root mean deviation power" and obtain a more homogeneous family of measures.

$$\Omega(c,\alpha,\lambda) = \left[\sum_{r_k \le \lambda} |r_k - c|^{\alpha} p_k \right]^{\frac{1}{\alpha}}, \quad \alpha > 0$$

which are now entirely analogous with so called "L_p metrics", common measures of distance.

Lower Semi-Variance (*Lower Semi-Standard Deviation*)

The most well known lower partial risk the lower semi-variance of Markowitz [1], where $\alpha = 2$, $c = \lambda = E(\mathbf{w})$, then can be written for portfolio $\mathbf{w}$ in the form

$$V_{-}(\mathbf{w}) = \mathrm{E}\left[|R(\mathbf{w}) - E(\mathbf{w})|_{-}^{2} \right]$$

where $R(\mathbf{w})$ is the rate of return of portfolio $\mathbf{w}$ and

$$|u|_{-} = \max\{0, -u\}$$

Lower semi-standard deviation $\sigma_{-}(\mathbf{w})$ is the square root of $V_{-}(\mathbf{w})$. These two measures are more appealing to the practitioners feelings against risk, but they were largely ignored since they are essentially the same as variance when portfolio return follows normal distribution.

Lower Semi-Absolute Deviation

Lower semi-absolute deviation of $R(\mathbf{w})$is defined as follows:

$$W_{-}(\mathbf{w}) = \mathrm{E}\left[|R(\mathbf{w}) - E(\mathbf{w})|_{-} \right]$$

As it is shown in Ogryczak and Ruszczynski [3] this measure is a convex function of $\mathbf{w}$ like lower semi/standard deviation of above, and most portfolios on the efficient frontier generated by mean – lower semi-absolute deviation model is consistent with the principle of *maximization of expected utility*.

Below –Target Risk

Let τ be the target rate of return. Then the below target risk of order α is defined in the form

$$T_B^{\alpha}(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - \tau\right|_{-}^{\alpha}\right]^{1/\alpha}$$

One can show that this measure is again consistent with the principle of *maximization of expected utility*. The measure is also a convex function of **w** for all α.

Value at Risk and Conditional Value at Risk

VaR (value at risk) is a relatively new lower partial risk measure and it is widely used for the measurement of market risk. Let $L(\mathbf{w})$ (= $-R(\mathbf{w})$) be the loss associated with portfolio **w**. Then $VaR_\beta(\mathbf{w})$, $0 < \beta < 1$, is defined as the smallest number ω_β such that

$$\mathrm{P}\left\{L(\mathbf{w}) \geq \omega_\beta\right\} = 1 - \beta$$

When $R(\mathbf{w})$ follows normal distribution we seldom experience a loss exceeding $VaR_\beta(\mathbf{w})$, where β is over 0.95. Unfortunately it is not valid when $R(\mathbf{w})$ exhibits a longer tail distribution. Also $VaR_\beta(\mathbf{w})$ is not a convex function associated with portfolio **w**. Therefore $VaR_\beta(\mathbf{w})$ is not an adequate measure of risk for portfolio optimization.

Conditional value at risk (*CVaR*) is an alternative measure of risk which maintains advantages of *VaR*, yet free from computational disadvantages of *VaR*. The measure is defined as follows:

$$CVaR_\beta(\mathbf{w}) = \frac{1}{1-\beta}\mathrm{E}\left[L(\mathbf{w}) \middle| L(\mathbf{w}) \geq VaR_\beta(\mathbf{w})\right]$$

For example if we take an asset with following characteristics, maturity 10 years, 10 year loss: maximum 10% in 99% cases, on average 12% in 1% of other cases, then the risk is measured with the *VaR* = 10% and *CVaR* = 12% with a confidence level of 99%. From the above definition one can see that $CVaR_\beta(\mathbf{w})$ is always located to the left of $VaR_\beta(\mathbf{w})$. Rockafellar and Uryasev [4] proved that $CVaR_\beta(\mathbf{w})$ can be minimized over usual set of feasible portfolios using convex minimization algorithms.

3 Formulation of Mean – Lower Partial Risk Models

In this part two mean – lower partial risk models are formulated, namely model with lower semi – absolute deviation risk measure and model with conditional value at risk measure. But the same approach can be used for the other lower partial risk measures as well. Let us assume that $\mathbf{R} = (R_1, R_2, \ldots, R_n)$ is distributed over finite set of points $\mathbf{r}_t = (r_{1t}, r_{2t}, \ldots, r_{nt})$, $t = 1, 2, \ldots, T$. These are obtained for example by historical data. Let p_t be the probability that **R** attains $\mathbf{r}_t$, $t = 1, 2, \ldots, T$, and W is the set of feasible portfolios, where

$$W = \left\{ \mathbf{w} \middle| \sum_{j=1}^{n} E_j w_j \geq \tau, \sum_{j=1}^{n} w_j = 1, \quad l \leq w_j \leq u, \quad j = 1, 2, \ldots, n \right\}$$

and τ, l and u are some appropriate constants.

Mean -Lower Semi-Absolute Deviation Model

From the above definition we have

$$W_{-}(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - E(\mathbf{w})\right|_{-}\right] = \sum_{t=1}^{T} p_t \left|\sum_{j=1}^{n} (r_{jt} - E_j) w_j\right|_{-}$$

and the model can be written as

$$\min \quad \sum_{t=1}^{T} p_t \left|\sum_{j=1}^{n} (r_{jt} - E_j) w_j\right|_{-}, \mathbf{w} \in W$$

In Konno - Waki –Yuuki [5] it is shown that this problem can be transformed into equivalent linear programming problem in the form:

$$\min \quad \sum_{t=1}^{T} p_t z_t$$

subject to

$$z_t \geq -\sum_{j=1}^{n} (r_{jt} - E_j) w_j, t = 1, 2, \ldots T$$
$$z_t \geq 0, t = 1, 2, \ldots T, \mathbf{w} \in W$$

Mean – CvaR Model

Let us define the loss function in the form

$$L(\mathbf{w}) = \mathrm{E}\left[\left|R(\mathbf{w}) - \tau)\right|_{-}\right] = \sum_{t=1}^{T} p_t \left|\sum_{j=1}^{n} r_{jt} w_j - \tau\right|_{-}$$

where τ is some constant. Then $CVaR(\mathbf{w})$ can be minimized by solving the following linear programming problem[4]:

$$\min \quad \omega + \sum_{t=1}^{T} p_t z_t / (1 - \beta)$$

subject to

$$z_t \geq -\sum_{j=1}^{n} r_{jt} w_j + \tau - \omega, t = 1, 2, \ldots T$$
$$z_t \geq 0, t = 1, 2, \ldots T, \mathbf{w} \in W$$

These two problems can also be written in the form of the multiple criteria problems and used for efficient frontier analysis in the corresponding two dimensional space.

References

[1] Markowitz, H. (1959) Portfolio Selection: Efficient Diversification of Investments. John Wiley nad Sons.

[2] Zeleny, M. (1982) Multiple Criteria Decision Making. McGraw Hill, New York.

[3] Ogryczak, W and A. Ruszczynski (1999) From stochastic dominance to mean – risk models: semi-deviation as risk measures. *Euro. J. Operat. Res.* **116**, 35-50.

[4] Rockafellar, R. T. and S. Uryasev (2000). Optimization of conditional value at risk. J. Risk **2**, 21-41.

[5] Konno, H. - Waki , H.–A. Yuuki (2002): Portfolio Optimization under Lower Partial Risk Measures. *Asia/Pacific Financial Markets* **9**, 127-140

[6] Mlynarovič, V. (2001): Finančné investovanie. Teória a aplikácie. IURA Edition, Bratislava, ISBN 80-89047-16-5.

Author Address:
Vladimír Mlynarovič, doc., Ing. CSc.
Faculty of Social and Economic Sciences
Comenius University in Bratislava
Odbojárov 10/A
820 05 Bratislava
e-mail: vladimir.mlynarovic@fses.uniba.sk

Securing of Business Information

Jaroslav Mlýnek
Department of Mathematics and Didactics of Mathematics
Faculty of Pedagogy
Technical University of Liberec

Abstract

Information of every successful business company belongs to its most important assets. Companies have to protect their information due to their own interests and legal regulations. This article describes the methodology of information evaluation for the purpose of information securing from confidentiality, integrity and availability points of view. The information evaluation is an essential step to ensure the adequate securing of information. The information evaluation is based on the principle of the worst possible scenario estimation that can be expected in a company, in case of information disclosure, modification or unavailability.

Keywords

Availability, confidentiality, evaluation of information, impact area, information unit, integrity, securing of information

1 Introduction

Information (in electronic form, written document, phone message, spoken voice) is a part of the most important assets of the every successful business company. Therefore, it is necessary to protect the information and to establish secure principles of its use.

In a business company, the information is processed and used, which is of different importance and significance for a company. In order to offer a possibility to provide the adequate protection of information, it is necessary to accomplish the evaluation of company information. However, the positive implementation of an information evaluation process represents a very difficult and specific step for every company.

The implementation of an information evaluation is an essential step for providing the adequate securing of information. Otherwise, certain important company information may be insufficiently secured, or vice versa, financial resources wastage would occur, owing to an excessive safeguard of less important information.

2 Information Evaluation

Managers of many companies often require knowing the value of information expressed by its price. This approach is not usually the most suitable way how to secure information because its implementation often can lead to wrong results.

We shell briefly describe one of the possible approaches, verified by practical usage. This approach is used by some quantitative methods of risk analysis (for example in methodology CRAMM – CCTA Risk Analysis and Management Methodology).

2.1 Identification and Grouping of Information

First, we must carry out the identification of company information. We can use the organizational structure of a company and/or the given company services.

Usually, we group company information into larger units (for example client information, staff personnel information, strategic intentions and plans). Middle sized business companies (about 100 staff) use approximately from 10 to 20 units, large business companies (more then 1000 staff) use usually 100 and more information units.

2.2 Procedure of information unit evaluation

In this approach, the information unit evaluation is based on the principle of estimation the worst possible scenarios that could reasonably be expected, namely from following three points of view:

- Information disclosure and misusing (for example by a competitive company),
- Information modification (unplanned modification by a company – deliberate or non deliberate),
- Information unavailability (for some time) or information deletion.

Well-chosen respondent (usually middle level manager) will execute the estimates of the worst possible scenarios.

We can define the process of information unit evaluation by the following mathematical model.

2.2.1 Model of the information unit evaluation

We define a three variable function (*f*) by the formula

(2.1) $f(r,s,t) \rightarrow A,$

which domain of definition is

$$D(f) = \left\{ [r,s,t]\,; r,s,t \in N; 1 \le r \le 3, 1 \le s \le \overline{S}, 1 \le t \le T_s \right\},$$

N … the set of natural numbers

r … number of the point of view

(*r* = 1 for the point of view regarding disclosure and misusing of information unit,

r = 2 for the point of view regarding modification of information unit,

r = 3 for the point of view regarding unavailability of information unit),

s … number of the area of negative impacts for company,

$\overline{S}$ … the total number of all impact areas,

the most often used areas:

Loss of goodwill,
Management and operations of a company,
Legal and regulatory obligations,
Commercial and economic interests,
Financial loss,
Personal safety,
Disruption to activities,

t … the impact level in impact areas,

T_s … the total number of impact levels in the area with the *s* number,

$v \in N$ … the evaluation of the given impact level; the evaluation is defined for all impact levels of all areas

$A = \{v; 1 \le v \le V\},$

$V \in N$ … the maximum of negative impact evaluations.

The information unit evaluation is executed for all three points of view.

For *r*=1 (view of disclosure), 2 (view of modification), 3 (view of unavailability)

$$(2.2)\ F(r) = \max_{1 \le s \le S} f(r,s,t_s),$$

where t_s is the impact level chosen by the respondent in the *s* impact area.

The F(1), F(2), F(3) evaluations of possible negative impact are values from 1 to *V* and they are generally different. These three values, the importance of confidentiality, integrity and information unit availability, are important for a company.

The possible wrong impacts can be investigated in more detail for every of the above-mentioned three points of view (for example if information is disclosed to a company employee or outside person, the extent of modification, different time unavailability, etc.).

2.2.2 Examples

Below, you can see an example of the information evaluation used in methodology CRAMM.

The allocation of evaluation (v) for impact levels (t) of all impact areas (s) it is possible to define by a matrix (see example on Figure 1, $\overline{S}$=8, V=10).

Number (s) /Name of the impact area	Impact level of the area (t)									
	1	2	3	4	5	6	7	8	9	10
1/ Loss of goodwill	2	3	5	7						
2/ Management and operations of company	1	3	5	6	7					
3/ Legal and regulatory obligations	3	4	5	6	7					
4/ Commercial and economic interests	1	2	3	4	5	6	7	8	9	10
5/ Financial loss	1	2	3	4	5	6	7	8		
6/ Personal information	1	2	3	4	5	6				
7/ Law enforcement	3	4	7	8						
8/ Personal safety	2	3	4	6	7	8	9	10		

Figure 1 – Example of the evaluation matrix

Below, you can see an example of impact levels description regarding areas “Loss of Goodwill” and “Law Enforcement” on Figure 2.

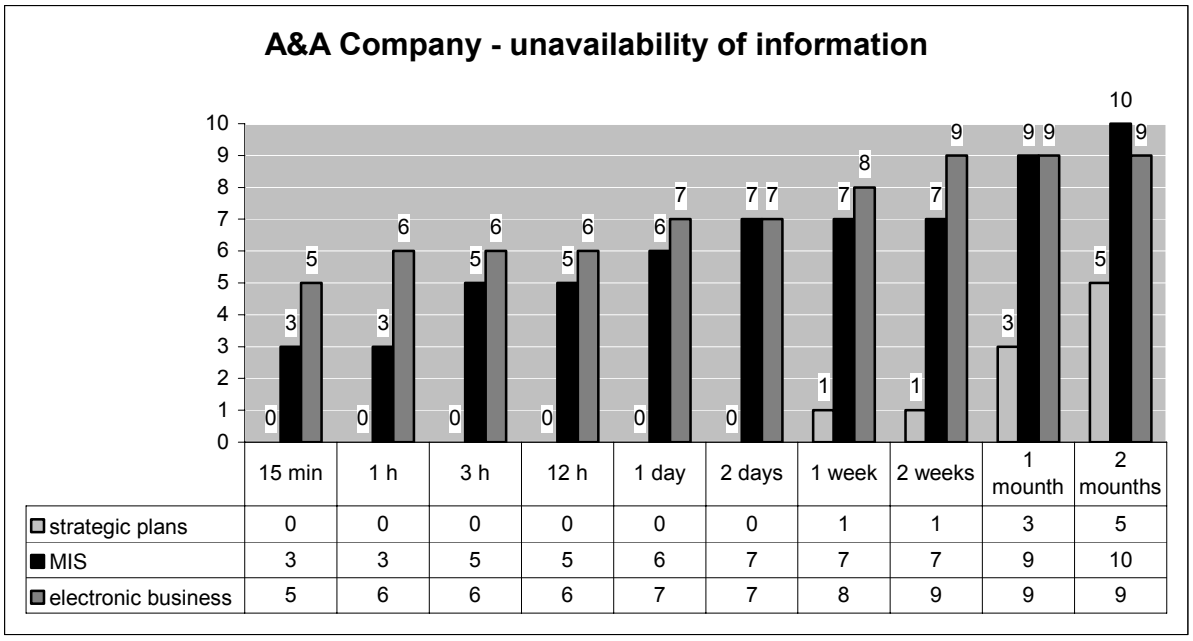
Impact area: **Loss of Goodwill**	
Number (t)/Description of the impact level	Evaluation (v)
1/ Partly problems inside a company, deterioration of worker relations inside various parts of a company.	2
2/ Deterioration of company relations with other organizations or the public, adverse publicity in regional extent.	3
3/ As point 2/, but with more widespread adverse publicity than regional.	5
4/ Significant deterioration of relations with other organizations or the public, with lasting effects.	7

Impact area: **Law Enforcement**	
Number (t)/Description of the impact level	Evaluation (v)
1/ Facilitating the commission of a crime, or prejudicing the investigation of a crime.	3
2/ Discontinuing the investigation or the trial of a crime cover up.	4
3/ Facilitating the commission of a serious crime, or impeding the investigation of a serious crime.	7
4/ Discontinuing the investigation or the trial of a serious crime cover up.	8

Figure 2 – Example of the impact levels of various areas

The possible wrong impact can be investigated in more detail for every from the above-mentioned three points of view (for example if information is disclosed to a company employee or to outside person, the extent of modification, different time unavailability).

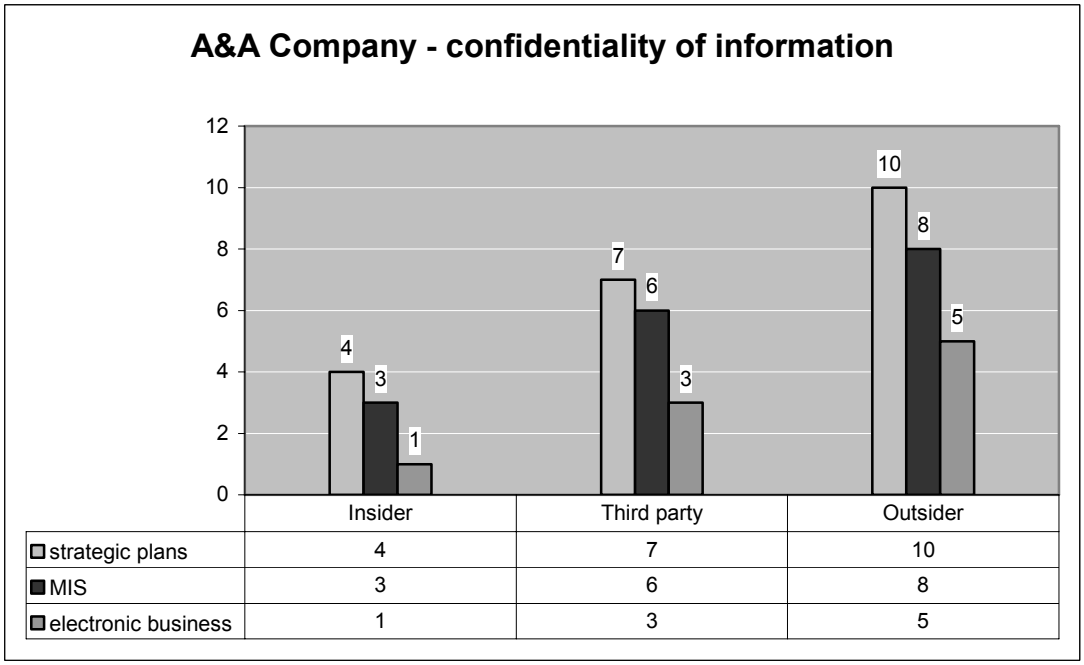
The example of the evaluation of three information units (strategic plans, management information system – MIS, electronic business) of „A&A Company" is shown on Figures 3, 4, 5. The maximum obtainable value is V=10.



	15 min	1 h	3 h	12 h	1 day	2 days	1 week	2 weeks	1 mounth	2 mounths
strategic plans	0	0	0	0	0	0	1	1	3	5
MIS	3	3	5	5	6	7	7	7	9	10
electronic business	5	6	6	6	7	7	8	9	9	9

Figure 3 – Unavailability point of view

We can see, on the Figure 3, that possible unavailability of strategic plans raises bigger problems after 2 months. On the other hand, the unavailability of electronic business brings problems to a company instantly.

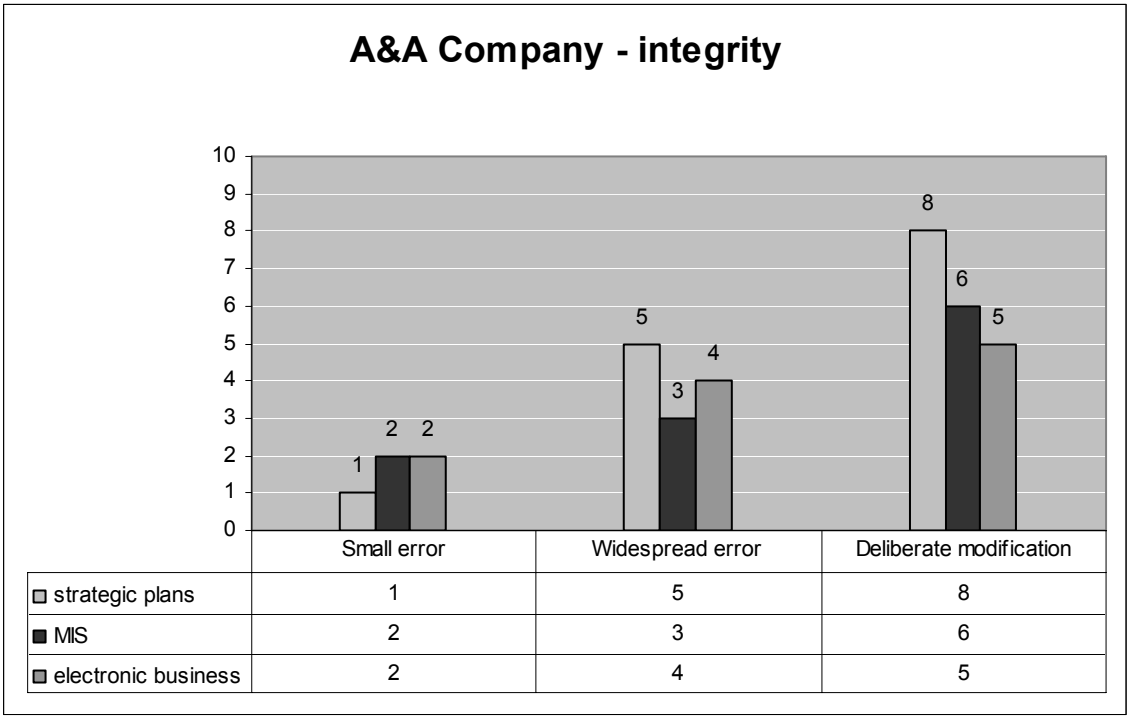


	Insider	Third party	Outsider
strategic plans	4	7	10
MIS	3	6	8
electronic business	1	3	5

Figure 4 – Confidentiality point of view

As opposed to unavailability, the ”Strategic plans” is very confidential information unit and its disclosure can cause even the dissolution of a company (see Figure 4).

Deliberate modification of all 3 units can cause problems for a company; the weightiest problems are produced in case of strategic plans (see figure 5).



	Small error	Widespread error	Deliberate modification
strategic plans	1	5	8
MIS	2	3	6
electronic business	2	4	5

Figure 5 – Integrity point of view

3 Securing of information

The determination of the required level of information unit securing against actual threats depends on the information unit evaluation, probability and weight of actual threats and on the vulnerability of processing and information unit saving. Set of security countermeasures is applied to information unit securing.

Mathematical cryptographic methods are important tool for ensuring the confidentiality and integrity of information in electronic form.

Now, business companies use symmetric cipher algorithms (DES, 3DES, AES) for confidentiality ensuring. The asymmetric cipher algorithms (RSA) are used within a digital signature method and for the management of secret keys for symmetric algorithms.

The digital signature enables to ensure information integrity and the principle of non-repudiation (unambiguous determination who is an author of information).

Many business companies use services of an internal or external certification authority for the certification of public keys for asymmetric cipher algorithms.

4 Conclusion

The important business information securing is a necessity to further successful existence of any business company.

If we want to achieve secured information, the important steps are the following: company information identification, grouping information into information units, information units’ evaluation

and the determination of important information units in a company. Otherwise, an adequate information securing cannot be ensured.

The necessary precondition of every successful information evaluation is choosing a convenient company manager.

The information evaluation required for securing the information is other view on information importance than usual business view.

The business companies protect their information due to their own interests and applicable legal regulations (especially Act n° 101/2000 Coll., on personal data protection, Act n° 227/2000 Coll., on electronic signature, Act n° 21/1992 Coll., on banks and subsequent provisions amending it).

References

[1] CRAMM User Guide, Issue 5.0, May 2003, London.

[2] ČSN/ISO 17799:2000 Code of Practice for Information Security Management.

[3] ISO/IEC TR 13335:2000 Guidelines for Information Security Management.

[4] ISO/TR 13569: 1997 Banking and related financial services – Information security guidelines.

[5] Sloupenský, V., Žehra, F.: Bankovní bezpečnost, Bankovní institut, Praha, 1997.

[6] Smejkal, V., Rais, K.: Řízení rizik, Grada Publishing, 2003.

Tests of Heteroskedasticity and Eventual Conflicts among Them

Dalibor Moravanský, Daniel Němec
Masaryk University Brno
Faculty of Economics and Administration
Department of Applied Mathematics and Informatics
dalibor@econ.muni.cz nemecd@econ.muni.cz

Abstract

The contribution presents the results of the authors' recent experience with the use of some tests of heteroskedasticity. First application refers to the investigations of the variability of university teachers' wages, while the second is the simulation study bringing knowledge of how often and how much the results of tests may differ when applied to samples of the different sizes. Some comments on properties of these tests relating to problems solved, which were met during applications are also amended.

Keywords

Heteroskedasticity, statistical testing, verification, econometric analysis

1. Heteroskedasticity as a common property of the model

The well-known *heteroskedasticity* problem relates to an unequal size of the diagonal elements of the covariance matrix Σ of the vector of random variables ε in the linear regression model. Although the matrix Σ is diagonal in this case, we cannot write $\Sigma = \sigma^2.I_T$, as in the opposite case of *homoskedasticity*.[1]

As is sufficiently known, when variances of disturbances are not of the same size, the *standard OLS method* is no longer efficient. At least *weighted least squares* should be used. As is obvious, *heteroskedasticity* can arise more often when we work with the *cross-section models* than with the *time series models*. Values of variables of the latter model (both explanatory and explained) are usually of the similar magnitudes, while in the former model, (exploiting, e.g., the firm data) they are characteristic of an unequal size (due to different intensities of indicators of economic activity, such as numbers of employees, revenues, etc).

Main occurrences causing heteroskedasticity pertain to:

- Wrong specification of the model, when some relevant explanatory variable is omitted.
- Accumulation of measure errors at the increasing value of the explanatory variable causing an increase in the disturbance variances as a result.
- Large absolute deviations in the values of the data within the same random sample and an appropriate dependence of the variance of the dependent variable on the variability of certain explanatory variable.
- The use of group averages (not original observations) computed from the sorted data.

2. Tests of heteroskedasticity and some their properties

Although the number of methods solving this problem is quite large, the answer to a question, when *heteroskedasticity* becomes of a serious importance was until recently uneasy to state. Some useful tests have been developed by econometricians until1980' but their practical use was usually restricted on a small number of them. Nor the recent sophisticated software packages do comprise them in a satisfactorily wide range. Therefore some actual experience how they behave in practice may be of some interest.

The following tests of heteroskedasticity are mentioned in the econometric literature*: Glejser test*, *Goldfeld-Quandt test*, *White general test*, *Breusch-Pagan test*, *Koenker-Basset* variant of the Breusch-Pagan test.

The possible contradiction among the inferences of these five tests has certain resemblance to the

[1] We maintain to write heteros*k*edasticity with "*k*" due to recommendation of McCulloch J.,H.: On Heteros*edasticity. *Econometrica* 53, No.2 /1985 p.483, although most of recent authors are not in any way influenced by this recommendation.

problem arising with a "test trio" based on the ML-principle, i.e. *Wald*, the *likelihood ratio* and the *Lagrange multiplier tests* used for different purposes in certain econometric areas. Those three tests have two common features: a) they all are based on the χ^2 – distribution and b) they are submitted to the well-known *Berndt-Savin inequality*, which enables to "rank" resulting test values.

2.1. GOLDFELD-QUANDT test [1965]

This test is probably the most often used *heteroskedasticity test* at all. Unlike other tests (with the exception of the *Glejser test* in its original representation), the researcher has to state the explanatory variable to which variations of the disturbance vector are related. This brings some arbitrariness into thinking, which variable would be the best choice, particularly when regressors are mutually correlated (what is a quite common case). Another, albeit the less serious problem of this method, is the proper choice of the number of dropped central observations, which are to be removed from the original sample, when the rest enters directly into the computations of the appropriate *SSE*'s. Therefore, the *Goldfeld-Quandt test* has its appeal when the number of explanatory variables is not too large and when we are able to suggest their relative importance.

As is well-known, the number of dropped observations T_2 influences the significance of the *Goldfeld-Quandt test*. At small T_2, the differences between the upper and the lower parts of the sample may not be distinctive enough, and the test does not lead to *heteroskedasticity* suggestions. If, on the other hand, T_2 is too large, the differences will usually be greater but (when T is too small) the number of degrees of freedom will be small and so will be the power of the test. Our experience is such that the ratio about 1/6 of dropped observations would be the moderate size. (The common recommendation lies between 1/8 and 1/4).

2.2. GLEJSER test [1969]

The idea of this test is again – similarly to the *Goldfeld-Quandt test* - that fluctuations of the random component are connected with the variability of certain j^*-th explanatory variable. Therefore we first formulate conventional regression equation between the dependent variable and the set of the explanatory variables

$$y_t = \beta_1 + \beta_1 . x_{t1} + \beta_2 . x_{t2} + \ldots \beta_{j^*} . x_{tj^*} + \ldots + \beta_k . x_{tk} + \varepsilon_t \tag{1}$$

Then we compute ${}_{OLS}\hat{\beta}$ and, successively, the residual vector $e = y - X\hat{\beta}$. The fit of the dependency between the vector of absolute residuals $|e_t|$ and supposed variable $x_{.j^*}$ is verified with the help of the conventional t-statistics. The most frequently used are, particularly, these dependencies:

$$(2A) \quad |e_t| = \alpha_1 + \gamma_1 x_{tj*} \qquad (2B) \quad |e_t| = \alpha_2 + \gamma_2 x_{tj*}^{-1}$$

$$(2C) \quad |e_t| = \alpha_3 + \gamma_3 . \ln x_{tj*} \qquad (2D) \quad |e_t| = \alpha_1 + \gamma_1 \sqrt{x_{tj*}}$$

Rather then estimates $\hat{\alpha}_i, \hat{\gamma}_i$ of the regression coefficients α_i, γ_i themselves, the appropriate t-*statistics* $t_{\alpha_i}, t_{\gamma_i}$ are important. If some t_{γ_i} value is statistically significant, then it is an indication of the existing relationship between $|e_t|$ and $x_{.j^*}$ If this is true for more regressions (2A) – (2D), then we usually take such dependency, in which parameters are mostly significant. Moreover, this evaluation also gives a hint for the shape of the transformation of model variables. *Glejser test* has usually the greater power than the *Goldfeld-Quandt test*.

To avoid the uncertainty, which explanatory variable $x_{.j^*}$ is to be chosen, the following change of the test was designed: Instead of only one explanatory variable, they all are taken, such that dependencies as $h(|e_t|) = \alpha + \sum_{j=1}^{k} \gamma_j x_{tj}$. with several choices of the function h are considered.

2.3. WHITE general test [1980]

To formulate plausible (and exercisable) tests, it is necessary to specify the nature of *heteroskedasticity*. We would have been in the best situation if we could test the general hypothesis of the form $H_0 : \sigma_t^2 = \sigma^2$ for all t against the alternative $H_1 : \sigma_t^2 \neq \sigma^2$ at least for one t. Such task, when we work with the model with T generally different parameters (namely $\sigma_1^2, \sigma_2^2, ..., \sigma_T^2$), makes such aim not generally attainable. *H. White* suggested a certain solution of the shape of the general test, which is based on the fact that the conventional *OLS estimator* of the covariance matrix Σ_b *in case of heteroskedasticity* is *inconsistent* as has been shown e.g. in [9].

The *exact covariance matrix of the OLS-estimator* in the model loaded *only* by *heteroskedasticity* has the form

$$Cov(_{OLS}\hat{\beta}) = \sigma^2.(X'X)^{-1}(X'VX)(X'X)^{-1}\ , \tag{3}$$

where V is the diagonal matrix with $tr(V) = T$. Its consistent estimate may be obtained with the help of the expression known as the *White's estimator*

$$\widehat{Cov}(_{White}\hat{\beta}) = (X'X)^{-1}(\sum_{t=1}^{T} e_t^2 \boldsymbol{x}_t \boldsymbol{x}_t')(X'X)^{-1} \tag{4}$$

while conventional *OLS-estimátor* of $Cov(\hat{\beta})$ *in the same model* has the well-known form

$$\widehat{Cov}(_{OLS}\hat{\beta}) = s^2(X'X)^{-1} \tag{5}$$

White suggested that the efficient test should have been based on the difference between both these estimates of the covariance matrix Σ in such a way that it should be tested whether such difference is statistically significant. The operational version of his test is given by the use of the statistics $T.\widetilde{R}^2$, where $\widetilde{R}^2$ is the coefficient of determination in the regression of the vector e_t^2 on *all different variables* involved in the $\breve{\boldsymbol{x}}_t \otimes \breve{\boldsymbol{x}}_t'$ [2] matrix (each element of this matrix is a T- vector). It means that all redundant elements, which are: (a) elements on symmetrical positions $\xi_{ij} = \xi_{ji}, i, j = 1,2,..,k$, (b) all "cross" explanatory variables identical to some other element[3] of the matrix, have to be dropped. This matrix looks like

$$\breve{\boldsymbol{x}}_t' \otimes \breve{\boldsymbol{x}}_t' = \begin{pmatrix} x_{t1}^2 & \times & \times & ... & \times \\ x_{t1}x_{t2} & x_{t2}^2 & \times & ... & \times \\ x_{t1}x_{t3} & x_{t2}x_{t3} & x_{t3}^2 & ... & \times \\ ... & ... & ... & ... & ..... \\ x_{t1}x_{tk} & x_{t2}x_{tk} & x_{t3}x_{tk} & ... & x_{tk}^2 \end{pmatrix} \tag{6}$$

Further, we include the one's vector (if that is not originally present in the matrix of regressors) into the regression. The appropriate regression equation will thus have the form

$$e_t^2 = \alpha_0 + \sum_{j=1}^{k}\sum_{l=j}^{k} \alpha_s x_{tj} x_{tk} + \eta_t \qquad \text{with disturbances } \eta_t \tag{7}$$

The coefficient of determination $\widetilde{R}^2$ computed from regression (7) has the usual form

$$\widetilde{R}^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum \eta_t^2}{a'Z'Za} \tag{8}$$

Under the null hypothesis, (which, however, represents more general situation than is

---

[2] The symbol $\otimes$ denotes the Kronecker's matrix product.

[3] When the regressor matrix contains, e.g., vectors $x_{t1} = 1, x_{t2} = z_t, x_{t3} = z_t^2$, then products $x_{t1}.x_{t3}$ and $x_{t2}.x_{t2}$, are both equal to z_t^2, and one of them must be omitted due to avoid of multicollinearity.

heteroskedasticity alone), the $T.\widetilde{R}^2$ statistics has asymptotically χ^2-distribution with so many degrees of freedom what is the number of explanatory variables in the regression (8) (neglecting the constant). If there are no redundancies in $\breve{\boldsymbol{x}}_t' \otimes \breve{\boldsymbol{x}}_t'$, and $\breve{\boldsymbol{x}}_t$ contains a constant, then the number of degrees of freedom d is equal $d = k.(k+1)/2 - 1$.

The *White test* is extremely general. To perform it, we do not need any special assumptions regarding the shape of *heteroskedasticity*. On the other hand, however, this advantage may cause a trouble. The test may only reveal the specification error, if such is present (the most often omitting some explanatory variable in the regression). It can also indicate other wrong specifications, such as wrongly chosen function $E[y_t] = x_t'.\beta$ or correlations between X a e when present in the model with stochastic regressors.

The power of the test cannot be cogently evaluated: against some alternatives it may be quite poor. The flaw is the unconstructiveness of the test: if we reject homoskedasticity, the result does not give any hint how to do next. Only if the researcher is self-confident of absence of such problems, its efficiency for detection of heteroskedasticity is sufficient[4].

2.4 The BREUSCH - PAGAN (also GODFREY's) test [1979]

The *Goldfeld–Quandt test* can be considered as a sufficiently powerful, if we are able to state the variable, according to division of the data sample can be carried out. In many situations, however, the variability of disturbances is connected not only to one explanatory variable.

T.S.Breusch and *A.R.Pagan* designed the test based on the Lagrange multipliers principle, which evaluates the hypothesis of the form $\sigma_t^2 = \sigma^2 f(\alpha_0 + \alpha' z_t)$, where z_t is a vector of the $k-1$ independent variables, $z_t = (x_{t2}, x_{t3}, ..., x_{tk})$, α_0 is a scalar, α_1 are vector constants and f is suitable transformation function on the explanatory variables.

The model is obviously *homoskedastic*, if the (vector) condition $\alpha = 0$ holds. The test can be carried out by the simple regression with the Lagrange multipliers statistics

LM = 1/2. explained sum of squares in the regression of $\dfrac{e_t^2}{e'e/T}$ *on the* z_t *vector.*

For the computation purpose, we take $\mathbf{Z}$ as the $[Txk]$ observations matrix of variables $(1, z_t)$, while g is a (column) vector with values $g_t = \dfrac{e_t^2}{e'e/T}$[5]. The test statistics then has the form $BP = 0{,}5.\boldsymbol{g}'.[Z(Z'Z)^{-1}Z']\boldsymbol{g}$ or, written in a detail,

$$BP = \frac{1}{2}.\begin{pmatrix} \dfrac{Te_1^2}{e'e} \\ \dfrac{Te_2^2}{e'e} \\ \cdots \\ \dfrac{Te_T^2}{e'e} \end{pmatrix}.\begin{pmatrix} \zeta_{11} & \zeta_{12} & .... & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & .... & \zeta_{2T} \\ .... & .... & .... & .... \\ \zeta_{1T} & \zeta_{2T} & .... & \zeta_{TT} \end{pmatrix}.\begin{pmatrix} \dfrac{Te_1^2}{e'e} & \dfrac{Te_2^2}{e'e} & .... & \dfrac{Te_T^2}{e'e} \end{pmatrix}^{6} \tag{9}$$

It may be shown that when the null hypothesis holds, the *BP -test* variable is asymptotically χ^2-distributed with k degrees of freedom. In spite of its usefulness, there has been shown that *Breusch-Pagan test* is too sensitive to the normality (of disturbances) assumption.

---

[4] Hsieh [1983] modified this test for the heteroskedasticity purpose and "*heterokurtosis*" investigated its power at the small sample by the means of the Monte Carlo experiments..

[5] In principle, that is the squared individual residuum from the square of the "average residual".

[6] ς_{ij} denotes the ij-[th] element of the matrix $Z(Z'Z)^{-1}Z'$

2.5. The KOENKER - BASSET test [1981,1982]

R. Koenker, and later along with *G.Basset Jr.* suggested that computation of the *LM-statistics* should be based on the more robust estimator $\hat{\sigma}^2$ of the variance of disturbances σ^2 than SSE/T, namely on the expression [7]

$$\Omega = \frac{1}{T}.\sum_{t=1}^{T}\left[e_t{}^2 - \frac{e'e}{T}\right]^2 \tag{10}$$

But if the vector of disturbances ε is not normally distributed, then dispersions of $\varepsilon_t{}^2$ will not be equal to $2\sigma^4$. We take the vector $\vartheta = (e_1{}^2, e_2{}^2, ..., e_T{}^2)$ and let $\iota = (1,1,....,1)$ is the [T,1] vector of ones. Further, we denote $\overline{\vartheta} = \frac{e'e}{T}$. After such change, the computation of *LM -variable* will be based on the statistics $KB = \Omega^{-1}(\vartheta - \overline{\vartheta}\,\iota)'.\left[Z(Z'Z)^{-1}Z'\right](\vartheta - \overline{\vartheta}\,\iota)$ or, written "structurally":

$$KB = \frac{1}{\Omega}.\begin{pmatrix} e_1{}^2 - \frac{e'e}{T} \\ e_2{}^2 - \frac{e'e}{T} \\ \cdots \\ e_T{}^2 - \frac{e'e}{T} \end{pmatrix}\begin{pmatrix} \zeta_{11} & \zeta_{12} & .... & \zeta_{1T} \\ \zeta_{12} & \zeta_{22} & .... & \zeta_{2T} \\ .... & .... & .... & .... \\ \zeta_{1T} & \zeta_{2T} & .... & \zeta_{TT} \end{pmatrix}.\begin{pmatrix} e_1{}^2 - \frac{e'e}{T} & e_2{}^2 - \frac{e'e}{T} & .... & e_T{}^2 - \frac{e'e}{T} \end{pmatrix} \tag{11}$$

Under the normality assumption, this amended statistics will have the same asymptotic distribution as the *Breusch-Pagan statistic*. At the absence of normality, there suggestions have been stated how to attain a greater power of the tests. *Waldman* [1983] showed that if we fill all columns of $\boldsymbol{Z}$ with the same regressors as in the case of *White test*, then all tests will coincide as to equality of the results. As is obvious, neither *White* nor *Breusch-Pagan* nor *Koenker-Basset* tests do require specifying the variable, on which the variability of disturbances depends. In this respect, they are more general then the *Goldfeld-Quandt* and *Glejser* tests (in their original versions).

3. Testing of homoskedasticity of the university teachers' wages

The data for 8 faculties of the Masaryk University Brno were obtained to inquire whether independence of the variability of individual teacher's wages on measurable personal factors exists or not. As a dependent variable, the reciprocal of the average wage [per 1 working hour] was chosen. As the explanatory variables, the following four were stated:

a) Sex of the teacher [man = 0, woman = 1]
b) Age of the teacher [in years]
c) Achieved professional status of the teacher at 4-stage level [professor, assistant professor, senior lecturer and lecturer]
d) Variable constructed as a product of square roots of variables (b) and (c) similarly as combined variables involved in the *generalized Leontief function*.

If we accept that the wage would consist of two main components: the tariff wage and the personal benefit (the sum of the personal surcharge and the bonus), we may test whether the second component depends on the first one. The null hypothesis stands that the wage level variability does not depend on component, the alternative then is that the higher is the first component the higher will be the second.

The analysis was performed on the sample of 2115 personal wages originating from the 2003 data. The main results are comprised in the following review[8]:

[7] The (sample) mean quadratic deviation of the variables $e_t{}^2$
[8] The thickly marked value denotes the rejecting of the hypothesis, while slim not rejecting

Faculty of Informatics	$Wg_t^* = -0{,}0177 - 0{,}0009.Sex_t + 0{,}0005.Age_t + 0{,}0038.Pos_t - 0{,}0021.\sqrt{Age_t.Pos_t}$ [−4,6007 [−1,4992] [4,5165] [6,6670] [−4,5123] $R^2 = 0{,}4855$, $DW = 1{,}6057$, $T = 111$ *Glejser:* 1,7101 with *F-crit*. = 3,0804 ; *Goldfeld-Quandt* 1,3578 with $F^*_{41,41}(0{,}95) = 1{,}6816$ *White:* **37,3405**, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* 1,4041, *Koenker-Basset*: 0,6293
Faculty of Social Sciences	$Wg_t^* = -0{,}0148 - 0{,}0002.Sex_t + 0{,}0004.Age_t + 0{,}0036.Pos_t + 0{,}0018.\sqrt{Age_t.Pos_t}$ [−4,239] [−0,535] [2,909] [6,239] [3,250] $R^2 = 0{,}753$, $DW = 1{,}697$, $T = 84$ *Glejser:* 3**,**164 with *F-crit*. = 3,1093 ; *Goldfeld-Quandt* **2,4826** with $F^*_{30,30}(0{,}95) = 1{,}8409$ *White:* 13,1921, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* 5,3635 *Koenker-Basset*: 1,3459
Faculty of Natural Sciences	$Wg_t^* = -0{,}0112 + 0{,}00015.Sex_t + 0{,}0041.Age_t + 0{,}0029.Pos_t - 0{,}0015.\sqrt{Age_t.Pos_t}$ [−8,8252] [0,7615] [5,7335] [9,9783] [−5,7534] $R^2 = 0{,}551$, $DW = 1{,}8018$, $T = 406$ *Glejser:* 0,5423 with *F-crit*. = 3,0181 ; *Goldfeld-Quandt* 1,0590 with $F^*_{164,164}(0{,}95) = 1{,}2939$ *White:* 13,9830, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* 1,9244 *Koenker-Basset*:1,0157
Faculty of Medicine	$Wg_t^* = -0{,}0209 + 0{,}0002Sex_t + 0{,}0006.Age_t + 0{,}0043.Pos_t - 0{,}0022.\sqrt{Age_t.Pos_t}$ [−17,211] [1,4324] [11,5875] [21,1075] [−11,8748] $R^2 = 0{,}6807$, $DW = 1{,}7152$, $T = 841$ *Glejser:* **13,4605** with *F-crit*. = 3,0065 ; *Goldfeld-Quandt* **1,8801** with $F^*_{346,346}(0{,}95) = 1{,}1937$ *White:* **45,4129,** while $\chi^2_{13}{}^*(0{,}95). = 22{,}362$, *Breusch-Pagan:* **40,9844** *Koenker-Basset*: **19,8395***
Faculty of Law	$Wg_t^* = -0{,}0154 + 0{,}0002.Sex_t + 0{,}0009.Age_t + 0{,}0051.Pos_t - 0{,}0036.\sqrt{Age_t.Pos_t}$ [−4,2723] [0,6079] [5,6171] [8,4161] [−6,0658] $R^2 = 0{,}7568$, $DW = 2{,}1286$, $T = 86$ *Glejser:* 1,7897 with *F-crit*. = 3,1065 ; *Goldfeld-Quandt* 1,3676 with $F^*_{31,31}(0{,}95) = 1{,}8221$ *White:* 8,2442, while $\chi^2_{13}{}^*(0{,}95). = 22{,}362$, *Breusch-Pagan:* 1,8415 *Koenker-Basset*:1,0416
Faculty of Philosophy	$Wg_t^* = -0{,}0106 + 0{,}0004.Sex_t + 0{,}0005.Age_t + 0{,}0037.Pos_t - 0{,}0022.\sqrt{Age_t.Pos_t}$ [−5,8539] [1,9163] [9,2146] [13,8146] [−9,9764] $R^2 = 0{,}7431$, $DW = 1{,}8580$, $T = 235$ *Glejser:* **7,2687** with *F-crit*. = 3,0348 ; *Goldfeld-Quandt* 1,1564 with $F^*_{93,93}(0{,}95) = 1{,}4090$ *White:* **32,6163**, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* **22,079** *Koenker-Basset*: **8,4344**
Pedagogical faculty	$Wg_t^* = -0{,}0236 + 0{,}00005.Sex_t + 0{,}00064.Age_t + 0{,}0047.Pos_t - 0{,}0024.\sqrt{Age_t.Pos_t}$ [−9,3592] [0,2007] [7,2353] [12,481] [−¨7,6160] $R^2 = 0{,}689$, $DW = 1{,}697$, $T = 238$ *Glejser:* 1,2183 with *F-statistics*. = 3,0342; *Goldfeld-Quandt* 1,3295 with $F^*_{35,35}(0{,}95) = 1{,}4064$ *White:* **28,9821**, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* 1,6627 *Koenker-Basset*: 0,9379
Faculty of Economics and Administration	$Wg_t^* = -0{,}191 + 0{,}0006.Sex_t + 0{,}0008.Age_t + 0{,}0049.Pos_t + 0{,}0031.\sqrt{Age_t.Pos_t}$ [−4,204] [1,5444] [6,2518] [8,3164] [−¨6,8314] $R^2 = 0{,}686$, $DW = 1{,}984$, $T = 114$ *Glejser:* 2,2714 with *F-statistics* = 3,0781; *Goldfeld-Quandt* 1,6319 with $F^*_{43,43}(0{,}95) = 1{,}6607$ *White:* 13,5284, while $\chi^2_{13}{}^*(0{,}95). = 22{,}3620$, *Breusch-Pagan:* 3,6158 *Koenker-Basset*: 2,7833

In cases of *Koenker-Basset* and *Breusch-Pagan* tests, the test statistics does not depend on the sample length, so we use $\chi^2_{0,95}(4) = 9{,}4877$ in both cases. As for the number of degrees of freedom in *White test*, we take value $d = k.(k+1)/2 - 1 - 1 = 13$, as we have to take into account that the product of variables Age and Pos is equal to square of the $\sqrt{Age.Pos}$ variable. Here $\chi^2_{0,95}(13) = 22{,}362$.

In the preceding, we have used the following modifications of the original *Glejser test*[9]: Instead of taking usual slightly non-linear specifications of the right-hand variables, we have taken these specifications of nature of disturbances' dispersions: (a): $\mathrm{var}[\varepsilon_t] = \sigma^2[\alpha'.z_t]$, (b): $\mathrm{var}[\varepsilon_t] = \sigma^2[\alpha'.z_t]^2$, (c): $\mathrm{var}[\varepsilon_t] = \sigma^2.\exp[\alpha'.z_t]$. For these three cases, testing the hypothesis that all the coefficients except the constant are zero constitutes a test of the homoskedasticity assumption in the context of the specific formulation. These tests are carried out by linearly regressing e_t^2, $|e_t|$ and $\log|e_t|$, respectively, on a constant and z_t vector and then using the *Wald test* to test to hypothesis. As a variable used for the sorting of the data, the variable $\sqrt{Age.Pos}$ was used in the *Goldfeld–Quandt test*.

4. Some simulation results

Further, the performances of these tests have been investigated on the artificial data sample, which was used in the linear regression model with two normally distributed regressors. Random variables have had dispersions $x_{.1}^2$ in the first, and 1 in the second case.

Model: $y_t = 1 + x_{t1} + x_{t2} + \varepsilon_t$; $x_{.1} \approx N(6;100)$, $x_{.2} \approx N(3;16)$; Number of iterations: 1000; Significance level: 95%; H_0: Hypothesis with presence of *homoskedasticity*.

Table 1: Simulation results - model with heteroskedasticity $\varepsilon_t \approx N\left(0;(x_{t1})^2\right)$

Sample length	50		100		500		1000	
H_0	accepted	rejected	accepted	rejected	accepted	rejected	accepted	rejected
White test	239	761	25	975	0	1000	0	1000
Goldfeld-Quandt test[10]	89	911	12	988	0	1000	0	1000
Breusch-Pagan test	166	834	50	950	0	1000	0	1000
Koenker-Bassett test	329	671	153	847	3	997	0	1000
Glejser test[11]	315	685	105	895	0	1000	0	1000
e^2	328	672	151	849	3	997	0	1000
|e|	298	702	93	907	0	1000	0	1000
log|e|	298	702	93	907	0	1000	0	1000

Table 2: Simulation results - model with homoskedasticity $\varepsilon_t \approx N(0;1)$

Sample length	50		100		500		1000	
H_0	accepted	rejected	accepted	rejected	accepted	rejected	accepted	rejected
White test	951	49	944	56	947	53	945	55
Goldfeld-Quandt test	894	106	894	106	892	108	898	102
Breusch-Pagan test	950	50	947	53	945	55	945	55
Koenker-Bassett test	947	53	945	55	954	46	949	51
Glejser test	955	45	956	44	955	45	965	35
e^2	947	53	945	55	954	46	949	51
|e|	949	51	950	50	948	52	956	44
log|e|	949	51	950	50	948	52	956	44

[9] These are, e.g., mentioned in [6]

[10] We assume that we know the source of heteroskedasticity, i.e. the variable x_1.

[11] The hypothesis H_0 is accepted when it is accepted at least in two specifications in the iteration.

5. Comments and remarks

- Results showed that at 2 faculties of the university (the *Faculty of Medicine* and the *Faculty of Philosophy*), most of tests reject the null hypothesis of *homoskedasticity* while at others the null hypothesis cannot be rejected.

- The regression coefficients always maintain the same (positive) signs standing at *Pos* and *Age* variables, but this is not true for the remaining variables *Sex* (with prevailing positive signs) and "*product variable"* (with prevailing negative signs). The t-statistics were almost everywhere sufficiently high in regressions.

- *Breusch-Pagan test* gave for all faculties larger values than *Koenker-Basset test* but nowhere (although differences were remarkable) opposite issues from the testing were signalized.

- The above mentioned problem with *White test* when applied to the given specification of the regression equation, was that the square of the "*product variable"* coincided exactly with the product of the *Pos* and *Age* variables. Therefore, appropriate corrections in the computer algorithm have to be made to avoid the exact multicollinearity.

- As is seen from Table 2, *Goldfeld-Quandt test* differed in simulations (no matter how large the sample length was) remarkably from other tests in numbers of accepted/rejected cases. This is probably due to the very nature of this test, which "forces" the algorithm to find contrasts in the both parts of the data sample, even if the chosen simulation scheme is entirely free of heteroskedasticity.

Generally reviewed, the results of testing of university teachers' wages were quite persuasive (bearing in mind that none corrections have been done with the raw data): In cases of 4 faculties, the perfect agreements of the testing were achieved: (*Faculties of Law*, those *of Economics* and *of Natural Sciences* where the null hypotheses were not rejected, while at *Faculty of Medicine* the alternative has been recognized). Further, in remaining 4 cases, always one test yielded the opposite inference: twice the White test leaded to the rejecting of the null, while other tests have spoken in favor of it, and twice the Goldfeld-Quandt test did the same, where in case of *Philosophical faculty* this test rejected the null but at *Faculty of Social Sciences* this test pointed to the opposite result.

References

[1] Breusch,T., Pagan A.: A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47/1979, p.1287-1294.

[2] Breusch,T., Pagan A.: The LM Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies* 47/1980, p.239-254.

[3] Engle, R.E.: The Wald, Likelihood Ratio and Lagrange Multipliers Tests in Econometrics. In: *Handbook of Econometrics.* Vol.II, Elsevier Science Publishers.

[4] Glejser, H.: A New Test for Heteroskedasticity. *Journal of the American Statistical Association* 64/1969, p.316-323.

[5] Goldfeld, S., Quandt, R,: Some Tests for Homoscedastity. Journal of the American Statistical Association, 60/1965, p. 539-47.

[6] Greene, W., H.: Econometric Analysis 4$^{ed.}$ Prentice-Hall. Inc. New Jersey 2000.

[7] Koenker,R.: A Note on Studentizing a Test of Heteroscedasticity. *Journal of Econometrics* 17/1981, p.107- 112.

[8] Koenker,R., Basset, G.: Robust Test for Heteroscedasticity Based on Regression Quantiles. E*conometrica* 50/1982, p.43-61.

[9] White,H.: A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica* 48/1980b, p. 817-838.

Figure 1: Explanatory variable $x_{.1}$ and residuals – model with heteroskedasticity (various lengths of samples)

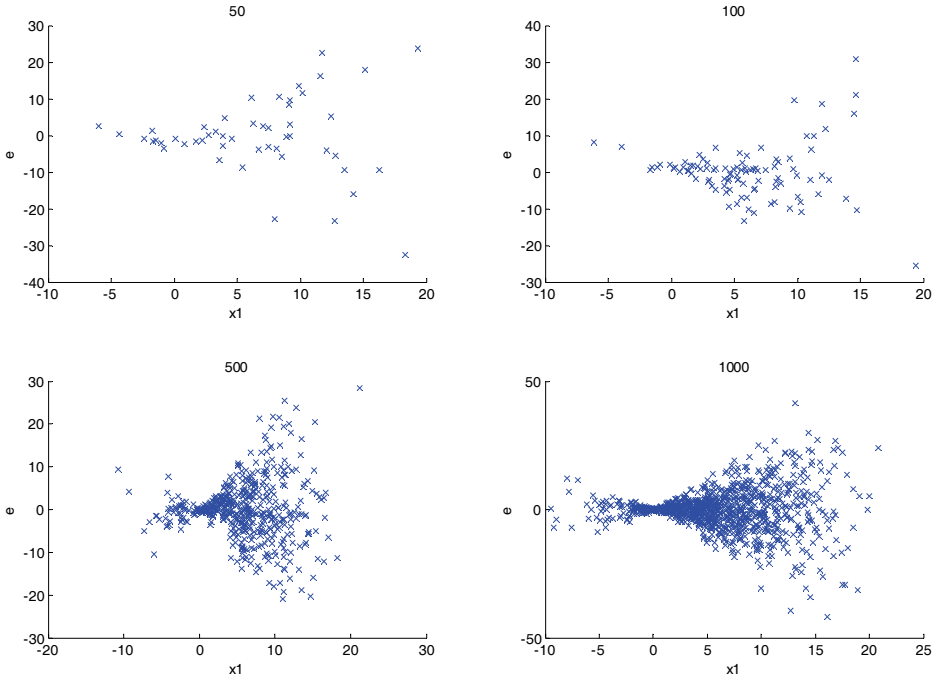
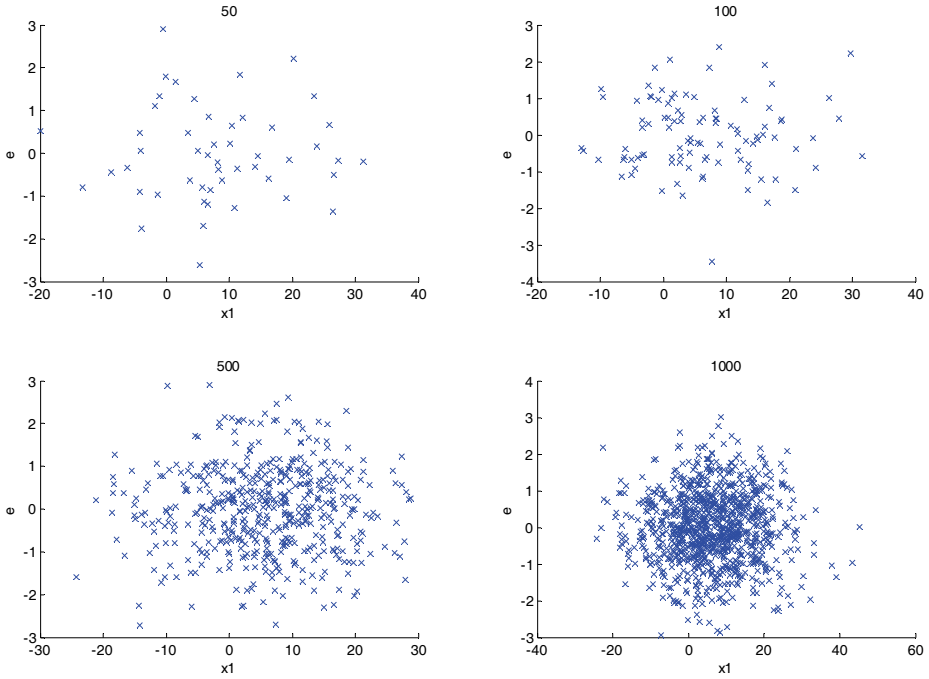


Figure 2: Explanatory variable $x_{.1}$ and residuals – model with homoskedasticity (various lengths of samples)



Czech Machinery in the Light of Tobin´s Q

Václava Pánková[1]
Vysoká škola ekonomická , Praha, pankova@vse.cz

Abstract

Investment expenditure relates to the optimization problem: to create an optimal capital stock as a function of expected profits. According to the *Q*-theory, investment depends on the ratio of the market value of business capital assets to their replacement value known as Tobin´s *Q* . *Q* is not observable but can be computed both on macroeconomic and microeconomic level. Macroeconomic *Q* is derived from the Keynes price level equation as a measure of the economy´s willing to invest. Expected value necessary for computing *Q* of a firm, even if not quoted on stock markets, uses a forecast from VAR model with panel data. Czech machinery is analyzed.

Keywords

Tobin´s *Q*, behavior of firms – investment, implementation of *Q*, panel data

1 Introduction

Investment is an important factor for increasing a production capacity. It is oriented towards the future and some aspects of it are irreversible. Investment is a source of maintaining or increasing of the capital stock in the economy. Supposing a process of partial adjustment, the actual level of capital is adjusted to a desired value. Alternative theories of investment differ in the way how is the desired value modeled and what factors are supposed to affect the speed of adjustment. Considering the investment being determined by relative prices and depreciation as a constant fraction of the level of capital stock, the neoclassical models are formulated based on well-defined optimizing techniques maximizing profit subject to a given production function. Another most usual one-equation investment models are the accelerator model, the cash flow model and the Tobin's *Q* model. None of these models has proved its superiority over the others; each describes and accents only a part of a complex reality. For details see e.g. Berndt (1991). According to the Tobin's theory, investment depends on the ratio *Q* of the market value of business capital assets to their replacement value. Under certain assumptions the neoclassical and the Tobin´s theories are equivalent what is a result of Hayashi (1982).

2 Investment as an optimization problem

To reach a level of investment I_t we need to describe a cost which is supposed to follow convex cost function $h(I_t)$, measured in production units. In case that the function $h(I_t)$, is strong convex, marginal cost is increasing. A difference between the real and the optimal capital stock is then better to be reduced by several smaller amounts than by one great investment. Let firm with capital K_t have a profit $\pi(K_t)$ from which it is necessary to subtract investment $P_t\ h(I_t)$, P_t being a unit production price. By an addition

[1] Financial support of GACR 402/04/0756 is gratefully acknowledged by the author

$$V_t = \sum_{\tau=0}^{\infty} \beta_{t+\tau} [\pi(K_{t+\tau}) - P_{t+\tau} h(I_{t+\tau})] ,$$

where $\beta_{t+\tau} = \prod_{\theta=1}^{\tau} (1 + r_{t+\theta})^{-\tau}$, r is an interest rate, the value of firm is given. The behavior of the firm is supposed to be formalized as an optimizing problem

$$V_t \rightarrow MAX$$

subject to

$$I_t - \delta K_t = K_t^* ,$$

the solution of which gives an optimal investment. The relation of relevant shadow price to the market value of the capital unit then gives a value of *Q*. In a more general fashion, the problem is described by Hayashi (1982) following a continuous dynamics.

A straightforward interpretation is that

Q:>1 indicates a marginal expected profit of a unit of capital to be higher than a unit of additional cost what is a motivation to invest

Q = 1 is an equilibrium state when there is no incentive for the firm to invest

Q<1 relates to an unprofitable environment and a recommendation is given (see e.g. Berndt, 1991) on behalf of the firm´s shareholders to sold off part of its capital plant and equipment.

3 How to compute Q

In practice, *Q* is not observable and the measurement of numerator as well as of denominator presents problems in empirical implementation. Looking for an appropriate approach, both macroeconomic and microeconomic level can be followed.

As for the macroeconomic approach let us briefly summarize a proposal of Mundschenk a.o. (2000). A way exists to derive *Q* from the Keynes equation for the price level. Tobin's *Q* is given here as a measure of the economy's willing to invest. Starting point is the equation

$$P = \frac{Y^{ef}}{Y} \qquad (1)$$

where *P* is the price level, *Y* is output in a quantity measure, Y^{ef} is effective demand equal to nominal income which can be decomposed into the sum of wages and profits as

$$Y^{ef} = wL + \bar{r}K .$$

Here, w is wage rate, L labor, $\bar{r}$ is return of capital K. Besides,

$$Y^{f} = wL + rK$$

with interest r, is a cost of production and

$$\Pi = Y^{ef} - Y^{f} = (\bar{r} - r)K$$

is a profit, which of course can be negative. In case of positive Π, an expansion is desirable. Having

$$\frac{\Pi}{rK} + 1 = \frac{\bar{r}}{r} = Q$$

or, in a forward looking theory,

$$Q^e = \frac{\bar{r}^e}{r}$$

with $\bar{r}^e$ being expected return on capital. Obviously, Q=1 when return on capital equals interest. Going back to the equation (1), we have

$$P = \frac{Y^{ef}}{Y} = \frac{Y^f + \Pi}{Y} = \frac{wL}{Y} + \frac{rK}{Y} + \frac{(Q-1)rK}{Y} = \frac{wL}{Y} + \frac{QrK}{Y}$$

and solving for Q

$$Q = \frac{P - \frac{wL}{Y}}{r\frac{K}{Y}}. \qquad (2)$$

Such a Q is expressed in terms of observable variables and its equivalence to average Tobin´s Q is given in Mundschenk a.o. (2000).

The calculation of Q proposed by Behr and Bellgardt (2002) relates to firms and reflects the fact that only a small part of economy's firms is quoted on stock markets. Starting by the formula

$$Q_{it} = \frac{V_{it} + D_{it} - N_{it}}{K_{it}}$$

the Q is computed for firm i at period t using the market value of equity market V_{it} , value of outstanding debt. D_{it}, remaining assets besides the capital stock N_{it} and replacement value of capital stock K_{it}. As for the V_{it}, the following approach enabling to process also non-quoted firms can be applied. Taking the values of pre-tax profits *PTP*, sales *S* and cash flow *CF*, a model can be formulated

$$\begin{aligned} PTP_{it} &= \alpha_{10} + \alpha_{11}PTP_{it-1} + \alpha_{12}S_{it-1} + \alpha_{13}CF_{it-1} + u_{1it} \\ S_{it} &= \alpha_{20} + \alpha_{21}PTP_{it-1} + \alpha_{22}S_{it-1} + \alpha_{23}CF_{it-1} + u_{2it} \\ CF_{it} &= \alpha_{30} + \alpha_{31}PTP_{it-1} + \alpha_{32}S_{it-1} + \alpha_{33}CF_{it-1} + u_{3it} \end{aligned} \qquad (3)$$

Obviously, the equations form a VAR model with one lag. Nevertheless, the idea is not to examine three sufficiently long time series but to work with panel data comprising short time series of a representative number of firms. As for the estimate of the VAR model proposed here, the dominant factor is the panel data structure. As there are the firms of different size comprised in the data set, the deviations from firms´ individual means can be used instead of levels. So, the equations do not contain firm specific effects. The VAR structure is important to help to produce forecasts of

$$\hat{V}_{it} = \sum_{\tau=1}^{\infty} \hat{PTP}_{it+\tau}\delta_{i\tau}^{\tau}$$

(4)

with $\delta_{i\tau}^{\tau} = \frac{1}{(1+r_{t\tau})^{\tau}}$, r being a market interest rate.

Looking for a firm's Q we than have

$$Q_{it} = \frac{\hat{V}_{it} + D_{it} - N_{it}}{K_{it}}. \tag{5}$$

In practice, the question of the number of summands in (4) arises. A common practice is to proceed according to an empirically verified rule that it is sufficient to consider four years as a relevant time horizon.

4 Applications

While the economy as a whole does relatively well in the Czech Republic, the performance of firms is very diversified. The machinery industry was chosen because of its long tradition and importance for the Czech economy. Unfortunately, there was no homogenous data source available. The choice had to be made between a system describing the period from 1995 to 2000 and another one concerning 2001 – 2003. The former one was used because it covers a longer period; during this time some firms appeared or disappeared, only 92 firms gave evidence about a permanent existence and were comprised into the data sample. So, a panel data sheet with 6 times 92 observations was processed. As it was shown in Kodera, Pánková (2002), the Czech machinery operates under constant returns to scale. In the small open economy of the Czech Republic, the firms are supposed to be rather the price – takers. To consider Q^a to be equal to Q^m, an information concerning an eventual homogeneity of cost functions should be given.

To characterize a macroeconomic background, the Q of Czech Republic was computed according to (2). With final consumption of households proxiing wL, Q=10.7 is an average value concerning the period in question.

As for the firms, the VAR model of type (3) was estimated as

$$\hat{PTP}_{it} = -4760.23 + 0.9582 PTP_{it-1} + 0.0236 S_{it-1} - 0.1751 CF_{it-1}$$
$$(1828) \quad (0.0682) \quad (0.0051) \quad (0.0752) \qquad R^2 = 0.98$$

$$\hat{S}_{it} = -4844.24 + 1.0124 PTP_{it-1} + 1.1341 S_{it-1} - 0.2692 CF_{it-1}$$
$$(17000) \quad (0.6379) \quad (0.0475) \quad (0.7032) \qquad R^2 = 0\text{-}96$$

$$\hat{CF}_{it} = -5907.10 - 0.1776 PTP_{it-1} + 0.0216 S_{it-1} + 1.0279 CF_{it-1}$$
$$(2167) \quad (0.0808) \quad (0.0060) \quad (0.0891) \qquad R^2 = 0.97$$

The Q´s were computed according to the (5) with an evident majority of negative values and a zero correlation to investment, when following the detailed panel structure. The non-favorable result of an investment model estimation was expected because of a high degree of zero investment values in the data sample.

To illustrate the situation more transparently, the results were averaged for each firm and 92 values of QP (= average of six Q^a of a firm) are plotted at the Figure 1.

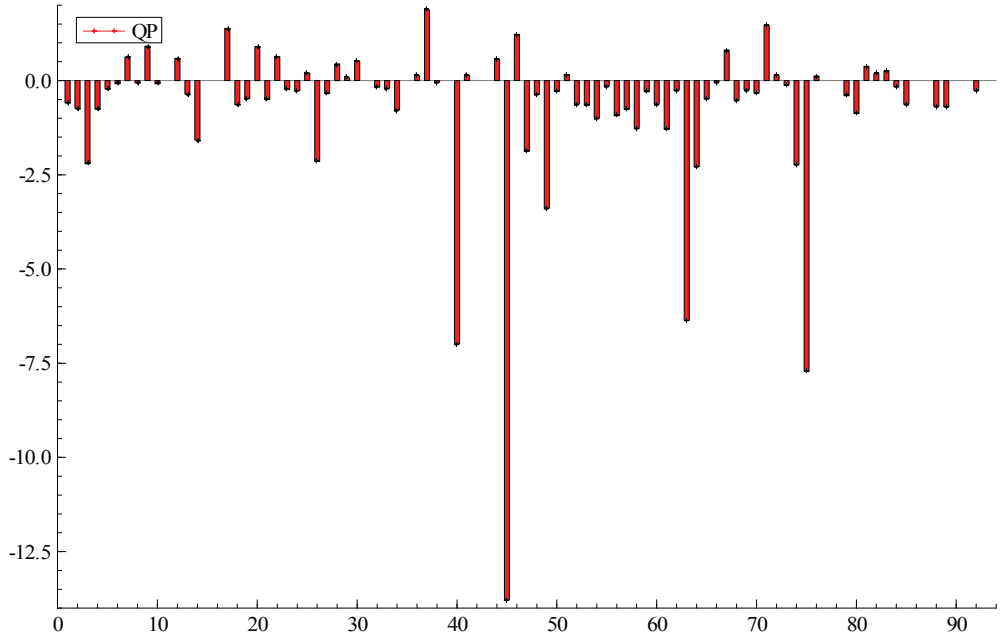


Figure 1.

5 Conclusions

The investment behavior, as well as other economic characteristics followed in the firms´ balance sheets, does not look very sane. A high degree of zero investment might be a signal for policy makers which seemed to be recognized by governmental authorities as a necessity of foreign – investment supporting policy. Actual conditions of the Czech industry are publicly summarized as good for the firms with foreign participation and poor for the others.

Using a new data sample with an eventual partitioning following home – foreign owner, and maybe also some other aspects as size of firms, indebted – no indebted firms, may give a more precise and actual information.

References

[1] Behr, A., Bellgardt, E.: Dynamic Q-investment functions for Germany using panel balance sheet data, Discussion paper 23/02, Economic Research Centre of the Deutsche Bundesbank, 2002

[2] Berndt, E.: The Practice of Econometrics, Addison-Wesley, 1991

[3] Hayashi, F.: Tobin´s marginal Q and Average Q: A Neoclassical Interpretation,, Econometrica, Vol. 50, No 1, 1982

[4] Kodera, J., Pánková, V.: Kapitálová výnosnost (možnosti využití produkční funkce pro ohodnocování firem v české ekonomice). *Politická ekonomie*, 2002, roč. L, č. 2, s. 209 – 216, ISSN 0032-3233

[5] Mundschenk, S., Schipper, B.C., Wannenwetsch, J., Wolff, G.B.: Estimating an Investment Function, internet, stefancollignon.de/Word/Tobins%20q.doc, 2000

Scheduling Serial - Parallel Processors – A Case Study

Jan Pelikán
Vysoká škola ekonomická v Praze

Abstract

Methods of Operations Research can play an important role in the process of increasing production productivity and decreasing production costs. We show in this case study how discrete optimization models can be used to find the optimal ordering production batches on lathes. The aim of the optimization is to suggest a processing order of production batches, which is connected with the minimum total processing time of the corresponding production order. For solving the problem, an original mathematical model including discrete variables is proposed. The model is solved by using of the system LINGO 8.0. Obtained results are very interesting since they lead to a significant decrease of the total processing time in comparison with the present practice.

Keywords

operations research, mathematical modeling, discrete models, scheduling models

1 Introduction

The subject of this case study is the production scheduling designed for the enterprise K-Bass Brandýs nad Orlicí. The scheduling consists in ordering batches, which are processed on lathes in the order satisfying given technological requirements. The lathes are considered as parallel processors. A production batch is a prescribed amount of a certain product. Each production order consists of a finite number of production batches. The aim of the scheduling is to find such processing schedule of batches, which minimizes the total processing time of the whole production order. It is assumed that before the processing of any batch on a lathe can begin, the corresponding lathe must be adjusted for the batch to be processed. Therefore the adjustment of a machine is the first phase of batch processing and the proper production is its second phase. We shall assume only one worker for adjusting the machines at our disposal. Therefore the adjustment process will be considered as a serial processing of batches on a unique processor. On the other hand, the proper production run, during which each batch is processed on different machines, can be considered as the parallel processing of batches on parallel processors. Therefore, delays can occur for the following reasons: (a) all machines are working so that no machine can be adjusted for some next batch (a delay in the adjustment process); (b) a machine must wait for adjustment while another machine is being adjusted, since the adjustment is carried out only by one worker (a delay in the production process). The minimization of the total processing time of all batches of a given production order is carried out in such a way that the resulting schedule leads at the same time to the maximum exploitation of the machines, minimum delays, and minimum production costs.

2 Description of technological conditions of the case study

The production sector consists of ten lathes, six from which are of the same type. The case study will find a schedule of batches for a one-week production. This production order could not be managed within the capacity of the two-shift working time, i.e. within 75 hours. The production order consists of 27 products. Types of products in production batches, numbers of pieces of the corresponding products as well as the corresponding processing and adjustment times are included in the Table 1.

Table 1. The production batches a their assignment to machines

Product number	Machine	Amount in pieces	Adjustment time (minutes)	Processing time (minutes)
1	1	1000	8,16	100,00
2	1	50	8,16	5,50
3	2	1000	10,20	260,00
4	2	50	10,20	16,50
5	2	50	10,20	20,50
6	3*)	1500	61,20	2370,00
7	4*)	2000	91,80	2440,00
8	5*)	2000	61,20	1380,00
9	6*)	667	61,20	1280,64
10	7*)	100	61,20	40,00
11	7*)	300	61,20	54,00
12	7*)	100	61,20	82,00
13	8*)	200	91,80	172,00
14	9	1000	91,80	2800,00
15	9	100	91,80	11,00
16	9	100	91,80	168,00
17	10	100	6,00	14,00
18	10	2000	6,00	460,00
19	10	1500	6,00	795,00
20	10	1000	6,00	140,00
21	10	100	6,00	14,00
22	10	100	6,00	45,00
23	10	300	6,00	18,00
24	10	2000	6,00	820,00
25	10	100	6,00	28,00
26	10	667	6,00	426,88
27	10	200	6,00	58,00

*) the batch can be assigned to any of machines 3 – 8.

3 A model with a priori assignment of batches to machines

First we shall assume that each production batch is assigned to a machine. This assignment mostly follows the fact that certain types of products can be produced on one certain machine (lathes), except the products 6 – 13, which can be produced on an arbitrary from the machines 3 - 8 (see Table 1). These products will be firstly assigned to lathes in such a way that all machines (lathes) are approximately equally loaded as far as the processing time is concerned. The result is given in Table 1.

We can assume that a machine will begin to produce the product immediately after it is adjusted to its production. Otherwise, the worker adjusting the machines will have to wait until the processed product is finished so that a delay in the adjustment process could occur. Further, we will assume that the worker will immediately start to adjust the machine for the next product as soon as the machine is free whenever he is not busy with adjusting some other machine. Since the products cannot be produced without the corresponding adjustment of the machines, which must precede the proper production process, the order, in which the adjustment is carried out, determines the order, in which the products are processed. It means that the products are processed in the order, corresponding to the order of adjusting the machines for them. The adjustment order is limited by the fact that we have only one worker adjusting the machines and it is assumed that he cannot adjust two or more different machines simultaneously so that he prepares the machines for the production of different products an a sequence similarly as in the one-processor case. In the other words, he can begin to adjust a machine only after the adjustment of a preceding machine had been finished. Another limitation for the

adjustment order follows the assumption that the adjustment of any machine can be commenced only after the machine have finished the production and it is idle.

We will introduce binary variables x_{ij} , which will provide the information about the adjustment order in the following sense: $x_{ij} = 1$ if the adjustment for product i precedes the adjustment for product j, $x_{ij} = 0$ otherwise.

Model parameters:
n the amount of products (batches) in pieces;
m – the number of machines;
t_i –the processing time of the i-th batch in minutes;
d_i – the adjustment time for the i-th batch in minutes;
$M >> 0$ – a large positive number;
S_k – the set of indices of those batches, which are assigned to machine k.

Model variables:
T – the time, at which all batches are finished;
x_{ij} – binary variables determining the adjustment order;
c_i – the time, at which the adjustment for product i is finished in minutes and starts the producing.

Model:

$T \rightarrow \min$		(1)
$x_{ij} + x_{ji} = 1$	$i, j = 1, ..., n, i < j$	(2)
$c_j - d_j \geq c_i + t_i - M(1 - x_{ij})$	$i,j = 1,\ldots,n, i \neq j , i \in S_k, j \in S_k, k=1,\ldots,m$	(3)
$c_j - d_j \geq c_i - M(1 - x_{ij})$	$i,j = 1,\ldots,n, i \neq j;$	(4)
$d_i \leq c_i$	$i = 1, \ldots, n.$	(5)
$c_i + t_i \leq T$	$i = 1, \ldots, n.$	(6)
$x_{ij} \in \{0,1\}$	$i,j = 1,\ldots,n, i \neq j;$	(7)
$c_i \geq 0$	$i = 1,\ldots,n.$	(8)

Equation (2) ensures that for each pair i, j either the adjustment for batch i precedes the adjustment for batch j or the other way round , i.e. the adjustment for batch j precedes the adjustment for batch i.

Inequality (4) ensures that the beginning time of the adjustment for batch j, i.e. $c_j - d_j$, must follow after the moment c_i, at which a preceding adjustment for batch i had been finished under the assumption that the adjustment for batch i precedes the adjustment for batch j (regardless on which machine the proper production process is carried out). This constraint follows the assumption that the worker cannot adjust more than one machine at the same time. Constraint (5) shifts the time moment, at which the adjustment is finished according to the corresponding adjustment. Inequality (6) determines the total processing time of all batches, which is (according to (1)) minimized. The model was solved using the program LINGO 8.0 for a one-week production program in Table 1, a number of production batches was $n = 27$ and a number of machines $m = 10$. The sets S_k are given in Table 1, from which it follows that $S_1 = \{1, 2\}$, $S_2 = \{ 3,4,5\}$, $S_3 = \{6\}$, $S_4 = \{7\}$, $S_5 = \{8\}$, $S_6 = \{9\}$, $S_7 = \{10,11,12\}$, $S_8 = \{13\}$, $S_9 = \{14,15,16 \}$, $S_{10} = \{17,18,19,20,21,22,23,24,25,26,27\}$. The mathematical model included 757 variables (729 of them were binary variables) and 1238 constraints. The computations took 5.4 minutes (PC 1.3 GHz). The resulting order, in which the machines are adjusted, is given in Table 2.

4 A model with a partial a priori assignment of batches to machines

Since a priori assignment of products to machines, which was used in the previous section, is not necessary to be optimal in general, we shall try in the sequel to optimize this assignment together with the total processing time of the schedule. As we have already mentioned, the assignment can be

chosen only at products 6 – 13, which can be processed on any of machines 3 – 8. The assignment of the other products to machines is fixed.

Table 2. Optimal ordering of production batches

order	batch	mashine	Start adj.	End adj.	Finish time
	i	k	c_i-d_i	c_i	c_i+t_i
1	14	9	0	91,8	2891,8
2	20	10	91,8	97,8	237,8
3	4	2	166,4	176,6	193,1
4	8	5	176,6	237,8	1617,8
5	18	10	237,8	243,8	703,8
6	2	1	603,84	612	617,5
7	7	4	612	703,8	3143,8
8	27	10	703,8	709,8	767,8
9	19	10	767,8	773,8	1568,8
10	6	3	823,2	884,4	3254,4
11	26	10	1568,8	1574,8	2001,68
12	5	2	1574,8	1585	1605,5
13	10	7	1585	1646,2	1686,2
14	12	7	1686,2	1747,4	1829,4
15	9	6	1912,56	1973,76	3254,4
16	22	10	2277,4	2283,4	2328,4
17	24	10	2328,4	2334,4	3154,4
18	3	2	2728,6	2738,8	2998,8
19	13	8	2738,8	2830,6	3002,6
20	11	7	2830,6	2891,8	2945,8
21	15	9	2891,8	2983,6	2994,6
22	16	9	2994,6	3086,4	3254,4
23	1	1	3146,24	3154,4	3254,4
24	21	10	3154,4	3160,4	3174,4
25	17	10	3174,4	3180,4	3194,4
26	23	10	3194,4	3200,4	3218,4
27	25	10	3218,4	3224,4	3252,4

Model 1.

To optimize at the same time also the assignment, we shall introduce binary variables y_{ik} , which will determine the assignment in the sense that $y_{ik} = 1$, if batch i is assigned to machine k and $y_{ik} = 0$ otherwise. Further we set $D = \{6,7,\ldots, 13\}$ and $S = \{3, \ldots, 8\}$, i.e. D contains the indices of those products and S the indices of those machines, where the assignment is not fixed. The remaining products are assigned a priori to corresponding machines according to fixed technological requirements. The mathematical model is based on the model (1) – (8) with the following additional constraints :

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ik}) - M(1 - y_{jk}), \quad i, j = 1, \ldots, n, i \neq j, i \in D, j \in D, k \in S \qquad (9)$$

$$\sum_{k \in S} y_{ik} = 1, \quad i \in D. \qquad (10)$$

Constraint (9) replaces constraint (3) for machines $k \in S$ and the corresponding sets of products S_k. It ensures that the adjustment process for a batch can be started only after finishing the production of the

preceding batch. Constraint (9) (unlike to (3)) includes variables y_{ik} and y_{jk}, which are both equal to 1, if both products are produced on the same machine k. Equation (10) ensures that each product from D is assigned to exactly one machine from S. This extension of the model (1) – (8) included 805 variables (777 of them were binary variables) and 1576 constraints. The computation time increased to 1 hour and 35 minutes. The resulting total processing time T was equal to 3254.4 minutes, i.e. the total processing time was the same as in the previous model, which means that a priori assignment in the previous model was optimal. The substantial increase of the computational time is remarkable in the case of extended model, which included the optimization of assignment batches to machines.

Model 2.

The assignment of products to machines have an alternative solution, since the machines, for which the assignment is not fixed are completely identical. Therefore, it is not necessary to assign products to these machines. It is sufficient to find out, which products can be assigned to an identical machine. In this way, we obtain a simpler model with a lower number of variables.

In Model 2, binary variable $y_{ij} = 1$, if product i is assigned to the same machine as product j and $y_{ij} = 0$ otherwise. For the indices of these variables always $i, j \in D$ and $i < j$ holds. Further, we shall replace constraint (9) from Model 1 with the following constraints (11) and (12), which ensure (similarly as constraint (9)) that the adjustment process for any batch can be commenced only after the production of a preceding batch had been finished.

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ij}),\ i < j, i \in D,\ j \in D \qquad (11)$$

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ji}),\ i > j, i \in D,\ j \in D \qquad (12)$$

Since we deal with an assignment of 8 products to 6 machines from the set S, it is necessary to assign at least 2 products to the same machine, i.e. at least 2 variables y_{ij} must be equal to 1, which is ensured by the following inequality (13):

$$\sum_{i \in D} \sum_{j \in D, i<j} y_{ij} \geq 2, \quad i \in D. \qquad (13)$$

Therefore inequality (13) replaces relation (10) from Model 1. Model 2 includes, in comparison with Model 1, 24 variables and 280 variables less.

5 Conclusion

The obtained schedule, in comparison with the schedule used in the practice, leads to the total processing time, which is 30% shorter. This result can substantially increase the productivity, exploiting the machines and lead to lower production costs. Nevertheless, it requires on the other hand the usage of sophisticated modeling techniques and a professional optimization software. It seems that an interconnection of these models with the information system would be possible too, because the program system LINGO 8.0 supports this interconnection. On the other side, disadvantage of this approach consists in the fact that we have to solve NP-hard problems and the increase of the size of problems we have to solve may cause a substantial increase of the necessary computational time. This happens also when we passed from Model 1 with a priori assignment to Model 2, which includes also the optimization of the assignment of products to machines. We obtained in this case a larger problem and the necessary computational time was higher too.

References

[1] PELIKÁN, Jan, An Optimization of the Production System – Case Study. New Delhi 08.12.2003-11.12.2003. In: APORS 2003. New Delhi: ORSI,2003,p. 48.

[2] PELIKÁN, Jan, Diskrétní modely v operačním výzkumu. 1. vydání, Praha: PROFESSIONAL PUBLISHING, 2001. 164 p., ISBN 80-86419-17-7.

[3] PELIKÁN, Jan, A minimax parallel processors scheduling problem. Edinburgh 08.07.2002-12.07.2002. In: IFORS 2002. OR in a globalized network world economy. Edinburgh: IFORS, 2002,155 p.

Split Delivery Problem

Jan Pelikán, Jan Fábry, Václav Kořenář
University of Economics, W. Churchill Sq. 4, 130 67 Prague 3
pelikan@vse.cz, fabry@vse.cz, korenar@vse.cz

Abstract

Delivery problem is a modification of the traveling salesman problem. It consists in the optimization of routes in a communication network, where all routes must start in one depot and each location must be visited. Demand for commodity is defined for each customer and total demand of all customers is greater than the capacity of the vehicle used for distribution. Each of the routes must be designed not to exceed the capacity of the vehicle. The objective is to minimize the total lengths of all routes. If demand of each customer is less than the capacity of the vehicle, delivery can be realized at once and it is possible to apply mathematical formulation analogous to the model of traveling salesman problem with two-indices variables. Otherwise it seems to be useful to split a delivery into several routes. In such situation the model with three-indices variables must be used what significantly increases a number of variables and decreases the real opportunity to find the optimum in reasonable time. The paper describes both approaches and shows the difference between them on the numerical example.

Keywords

Vehicle routing problem, integer programming, split delivery problem

1 Introduction

The delivery problem is a standard problem of operations research appearing in many practical applications. Solution of the problem brings significant savings, especially when the optimum is repeatedly applied in practice. Transportation cost represents the enormous part of the total cost and the product price. The indispensable effects of the transport optimization are savings on fuel consumption, decrease in the number of used vehicles and reduction of the impact on the environment.

The troubles connected with the solution of routing problems consist in their diversity, because of specific delivery conditions valid for concrete applications. Another barrier of the optimization process is the NP-hardness of the problem.

In the standard **Vehicle Routing Problem** (VRP) vehicles of given capacity are defined. The infinity number of vehicles of identical capacity is mostly considered. In modifications of VRP it is possible to limit the number of vehicles or to allow their different capacities. Further, it is defined the communication network consisting of vertices and edges and representing e.g. public road system. The depot, warehouse, production point is denoted the vertex 1 while other vertices are the locations of customers demanding for given amount. Each edge of the communication network is evaluated with distance between corresponding vertices, or transportation cost associated with the edge. Demand of the destination i is denoted q_i and the vehicle capacity is $V > 0$. Each destination is supposed to be served at once, i.e. $q_i \leq V$ for all i. Respecting the assumption $\sum_i q_i > V$, the vehicle has to serve all destinations in several routes beginning from and ending at the depot. The objective is to minimize the total distance traveled by the vehicle to serve all the destinations.

Mathematical model of VRP with a homogeneous and unlimited fleet of vehicles, with deliveries that cannot be split and with the demand which does not exceed the capacity of the vehicle is based on the Miler-Tucker-Zemlin formulation of Traveling Salesman Problem (see [5]). The binary variable x_{ij} corresponds to the decision to serve the destination i immediately before the destination j being served on any route. This variable is defined for $i \neq j$.

(MTZ):

minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{1}$$

subject to

$$\sum_{i=1}^{n} x_{ij} = 1,\ \ j = 2,3,...,n, \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} = 1,\ \ i = 2,3,...,n, \tag{3}$$

$$u_i + q_j - V(1 - x_{ij}) \le u_j,\ \ i = 1,2,...,n,\ \ j = 2,3,...,n,\ \ i \ne j, \tag{4}$$

$$u_1 = 0,\ \ u_i \le V,\ \ \ i = 1,2,...,n. \tag{5}$$

A different approach to the VRP is based on Dantzig-Fulkerson-Johnson formulation of Traveling Salesman Problem.

(DFJ):

minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{6}$$

subject to

$$\sum_{i=1}^{n} x_{ij} = 1,\ \ j = 2,3,...,n, \tag{7}$$

$$\sum_{j=1}^{n} x_{ij} = 1,\ \ i = 2,3,...,n, \tag{8}$$

$$\sum_{i,\,j\in S} x_{ij} \le |S| - \lambda_S,\ \ S \subset \{2,3,...,n\},\ \ 2 \le |S| \le n-1. \tag{9}$$

Parameter λ_S in the model corresponds to the minimal number of vehicles which have to be used for deliveries to destinations from *S*. Although the accurate value of this parameter can be calculated, it is mostly replaced by the estimation

$$\left\lceil \sum_{s\in S} q_s \Big/ V \right\rceil \tag{10}$$

because of troubles with the integer model. Since this number refers to the routes which have to be connected with the depot, directly or through the vertices outside *S*, the number of edges going from *S* is at least $2\lambda_S$ what is guaranteed by the **capacity cut**:

$$\sum_{i\in S}\sum_{j\notin S} x_{ij} \ge 2\lambda_S\ . \tag{11}$$

Both formulations contain the two-indices variable x_{ij} without any information about the corresponding route; nevertheless, all the routes can be recognized after obtaining the solution.

If the vehicles with different capacities are considered in the model, where the capacity of the *k*-th vehicle is denoted V_k, three-indices variables have to be introduced. They provide the original information about visited vertices *i* and *j*, but also new information about the vehicle and the route. This variable is defined for $i \ne j$. The example of such model follows.

minimize

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}^{k} \tag{12}$$

subject to

$$\sum_{i=1}^{n} x_{ij}^{k} = \sum_{i=1}^{n} x_{ji}^{k}, \quad j = 2,3,...,n, \quad k = 1,2,...K, \tag{13}$$

$$\sum_{j=2}^{n} x_{1j}^{k} \le 1, \; k = 1,2,...K, \tag{14}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{n} x_{ij}^{k} = 1, \; i = 2,3,...n, \tag{15}$$

$$u_i^k + q_j - V_k(1 - x_{ij}^k) \le u_j^k, \; i = 1,2,...,n, \; j = 2,3,...,n, \; i \ne j, \; k = 1,2,...,K, \tag{16}$$

$$u_1^k = 0, \; u_i^k \le V_k, \; i = 2,3,...,n, \; k = 1,2,...,K. \tag{17}$$

Binary variable x_{ij}^k=1, if the edge (i, j) belongs to the route k, x_{ij}^k=0, otherwise. The objective function (12) expresses total length of all routes in the schedule. As the alternative objective, the number of used vehicles can be applied, i.e. $\sum_{k=1}^{K} \sum_{j=1}^{n} x_{1j}^{k}$. If an edge enters a vertex on the route, another edge must leave it on the same route what is assumed for all vertices using the equation (13). Respecting the inequality (14), multiple edges leaving the depot during each route are forbidden. If the number of edges equals 1, it corresponds to the real route, otherwise the route is not realized at all. The equation (15) assures that each vertex lies exactly on one route. Relations (16) and (17) deal with the balancing of the load on the vehicle/route and assure it does not exceed the vehicle capacity.

2 Split Delivery Problem

If we omit the assumption considering the value q_i that must not exceed the vehicle capacity V, we have to admit the delivery can be split and therefore the vertex will lie on several routes. The possibility of splitting delivery can be considered also if $q_i < V$ for each destination i. The Figure 1 shows the graph for the vehicle routing problem with vehicle capacity V=10 and demands q=(0, 8, 8, 4), where there is more advantageous to apply splitting of delivery. The optimal routes are 1-2-3-1 and 1-3-4-1 with the total length 45, while in case the delivery cannot be split three routes 1-2-1, 1-3-1 and 1-4-1 are obtained with the total length 60.

In [1] it is proved that if the distance matrix fulfills the triangular inequality, then in the optimal solution two routes go maximally through one vertex.

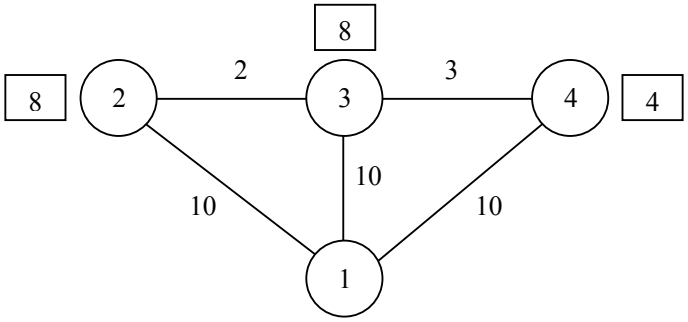


Figure 1: Example of delivery problem

Since it is necessary to identify all routes, three-indices variables must be introduced similarly to the delivery problem with different vehicle capacities. In contrast to the model without the possibility to split delivery to each vertex, we introduce variable q_i^k corresponding to the delivery to the i-th vertex

on the *k*-th route. Each vertex can belong to several routes and therefore the equation (15) in the previous model is not valid.

minimize

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}^{k} \tag{18}$$

subject to

$$\sum_{i=1}^{n} x_{ij}^{k} = \sum_{i=1}^{n} x_{ji}^{k}, \quad j = 2,3,...,n, \ k = 1,2,...K, \tag{19}$$

$$\sum_{j=2}^{n} x_{1j}^{k} \le 1, \quad k = 1,2,...K, \tag{20}$$

$$u_i^k + q_j^k - V_k(1 - x_{ij}^k) \le u_j^k, \ i = 1,2,...,n, \ j = 2,3,...,n, \ i \ne j, \ k = 1,2,...,K, \tag{21}$$

$$\sum_{k=1}^{K} q_i^k = q_i, \ i = 2,3,...,n, \tag{22}$$

$$q_i^k \le q_i \sum_{j=1}^{n} x_{ij}^k, \ i = 2,3,...,n, \ k = 1,2,...K, \tag{23}$$

$$u_1^k = 0, \quad u_i^k \le V_k, \quad i = 2,3,...,n, \ k = 1,2,...,K. \tag{24}$$

The objective function (18) and the relations (19) and (20) have the same meaning as (12) – (14) in the previous model. Relations (21) and (24) assure the balancing of the load on the vehicle. In contrast to the previous model, q_j^k is used instead of q_j. Equation (22) guarantees that the demand of the destination *i* is fully satisfied through partial deliveries q_i^k on all routes. Inequality (23) does not allow to deliver anything to the vertex which does not lie on the corresponding route.

Since the three-indices model represents enormous increase in the number of variables in comparison with the two-indices model, it will be impossible to obtain the optimal solution in practical problems because of their NP-hardness. Then it is possible to accept the sub-optimal solution provided by branch-and-bound method after the premature interruption of the optimization process, or it is practicable to suggest and use many heuristic approaches.

It is evident that each feasible solution of the problem without splitting delivery is also the feasible solution of the split delivery problem. In [3] it is suggested the algorithm which changes the optimal solution of VRP (or the solution obtained using any heuristics for VRP) on the base of the method of savings, where transportation cost is being decreased by splitting delivery to specific vertex into two different routes.

3 Reducibility of SDVRP

Let us assume the identical vehicle capacities, i.e. $V_k = V$ for all $k = 1,2,\ldots,K$. If at least one q_i is given as $q_i > V$, then it is possible to define the property of demand reducibility. For each vertex we express $q_i = n_i V + q_i'$, where n_i is an integer and $q_i' \in < 0, V)$. The SDVRP is called reducible if an optimal solution exists such that n_i deliveries to each vertex are realized directly (the routes 1-*i*-1) with fully filled vehicle capacity and remaining delivery q_i' is realized after solving SDVRP with reduced demand q_i'. If SDVRP is reducible it is possible to find such the optimal solution where first of all direct routes are realized (n_i for each vertex) and then reduced SDVRP is solved. Reducibility of SDVRP was proved only for stronger conditions (see [1]). Nevertheless, the reduction of demand is computed to simplify the process of finding solution of VRP and SDVRP by use of heuristic algorithms even in problems for which the reducibility is not proved.

4 Computational Experiments

In [3] it is possible to find the computational results comparing the solutions of VRP and SDVRP. The same vehicle fleet and network with the same demand of destinations are analyzed from the point of view of VRP and SDVRP. Since, the experiments are executed for the problems with 75, 115 and 150 vertices, the heuristic methods are used. Obtained results show that in case the demand is relatively low (up to a half of the vehicle capacity), the results of SDVRP in contrast to results of VRP are better only in few percents as to the total length and the number of routes. However, if demands are greater than 70% of vehicle capacity, savings based on splitting delivery are remarkable and reach 15% of the total length and 25% of the number of routes.

This publication was supported by the Grant agency CR under project 402/03/1283.

References

[1] Archetti, C., Mansini, R., and M. G. Speranza: *The Split Delivery Vehicle Routing Problem with Small Capacity*. November 7, 2001. Downloadable from http://www.eco.unibs.it/ ~www_mequ/ricerca/quaderni/201.pdf

[2] Bard, J. F., Kontoravdis. G., and G. Yu: A Branch-and-Cut Procedure for the Vehicle Routing Problem with Time Windows. *Transportation Science*. Vol. 36, No. 2 (2002).

[3] Dror, M., and P. Trudeau: Savings by Split Delivery Routing. *Transportation Science* 23, pp. 141-145 (1989).

[4] Nemhauser, G. L., and L. A. Wolsey: *Integer and Combinatorial Optimization.* John Wiley & sons Inc., New York 1998, ISBN 0-471-82819-X.

[5] Pelikán, J., Kořenář, V., and S. Mělníček: Using Cuts for the Traveling Salesman Problem and the Vehicle Routing Problem. In Bauer, L. (ed.). Mathematical Methods in Economics 2004. Brno, pp.248-255. ISBN 80-210-3496-3.

Differential Equations in Health Insurance Disability Risk Models

František Peller, Lea Škrovánková
Department of Mathematics, Faculty of Economic Informatics, University of Economics, Dolnozemská cesta 1, 852 35 Bratislava, e-mail: leaskrov@post.sk

Abstract

A basic probabilistic multistage structure is defined, which provides the possibility of a systematic modelling for Health insurances (disability annuities and lump sum). A time-continuous general model is proposed, which allows us to take into consideration a wide range of different policy conditions.

Keywords

Health insurance modeling, Markov processes, actuarial modeling of disability.

1 Introduction

The purpose of this paper is to illustrate how the mathematics of Markov and Semi-Markov stochastic processes can be used to develop a general approach for the actuarial modelling of disability and related benefits.

The use of Markov chains in life contingencies and their extensions has been proposed by several authors, in both the time-continuous case and the time-discrete case (Haberman 1984, Gatenby 1991, Pitacco 1993). Though an explicit and systematic use of the mathematics of multistate (Markov and Semi-Markov) models dates back to the end of the 1960s, it must be stressed that the basic mathematics of what we now call a Markov chain model were developed during the eighteenth century (seminal contributions by Bernoulli and Laplace demonstrate this for the continuous-time case).

In this paper we present the general aspects of a disability annuity cover, concerning policy conditions as well as single premium calculation. Then, the basic probabilistic structure is defined. After that, a time-continuous general model is proposed, which allows us to take into consideration a wide range of different policy conditions. This multistate model allows a well organized presentation of the mathematics of the insurances of the person and, in particular, of health insurances.

2 General policy conditions

Now we describe the development of an individual insurance policy which pays out a continuous annuity of dt units in the interval $(t, t+dt)$, if the insured is disabled. Let x be the age of the insured at policy issue. Let $S(x+t)$ represent the random state occupied by the policyholder at age $x+t$, for any t, $t > 0$.

The possible realizations of $S(x+t)$ are:

- (a) active member,
- (i) invalid or ill or disabled,
- (d) dead.

For example, the statement $S(x+t) = a$ means: the policyholder is in the state (a) at age $x+t$.

We assume $S(x) = a$. The graph of figure 1 illustrates the set of states and the set of possible transitions between states.

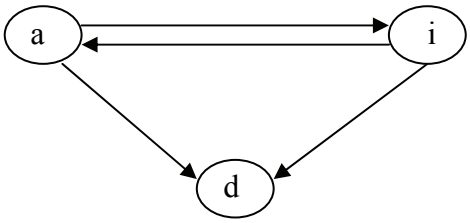


Figure 1. The three-state model

The random present value Yx of the benefits paid by the insurer is given by

$$Y_x = \int_0^\infty \left[S(x+t) = i \,\middle|\, S(x) = a\right] v^t dt \tag{1}$$

Where $\left[S(x+t) = i \,\middle|\, S(x) = a\right]$ denotes the indicator of the event S(x+t) = i conditioned to S(x) = a.
Let

$$\Phi(x,t) = P\left[S(x+t) = i \,\middle|\, S(x) = a\right] \tag{2}$$

Then the net single premium $\bar{a}_x^{ai}$ is given by

$$\bar{a}_x^{ai} = E(Y_x) = \int_0^\infty \Phi(x,t) v^t dt \tag{3}$$

where v is the annual discount factor. (An continuous annuity of 1 p. a. payable during the lifetime of a person now aged x is denoted $\bar{a}_x$ (Tutt 1991)).

In a more realistic situation, the benefits are paid out if and only if certain conditions concerning the duration of the disability are fulfilled. Let us define:

$\Phi^\Gamma(x,t)$ = probability that the insured is disabled and annuity is payable at age x+t, according to policy conditions. Of course, we have

$$\Phi^\Gamma(x,t) \le \Phi(x,t).$$

The set of policy conditions can be formally represented by a set of five parameters

$$\Gamma = [n_1, n_2, f, m, r],$$

Where:
(n_1, n_2) denotes the insured period (an annuity is payable if disability inception time belongs to this interval); for example n_1 = c = waiting period (from policy issue for new members),
n_2 = n = policy term;

f is the deferred period (from disability inception) – usually the first week;

m is the maximum number of years of annuity payment (from disability inception);

r is the stopping time (from policy issue) of annuity payment; for example, if ξ = retirement age, then $r = \xi - x$.

For example, we have:

$$\Phi^{[0,\infty,0,\infty,\infty]}(x,t) = \Phi(x,t)$$

$$\Phi^{[0,n,0,\infty,n]}(x,t) = \Phi(x,t), \quad \text{if } t < n$$

Thus, the net single premium (denoted according to the usual actuarial notation (Sakálová 1997)) for the second case is given by

$$\bar{a}^{ai}_{x:\overline{n}|} = \int_0^\infty \Phi^{[0,n,0,\infty,n]}(x,t)\, v^t dt = \int_0^n \Phi(x,t)\, v^t dt . \tag{4}$$

Remark. An continuous annuity on a life aged x, payable for n years certain and thereafter throughout life is denoted $\bar{a}_{x:\overline{n}|}$.

3 The Probabilistic model

Firstly, we assume that the stochastic process {S(x+t); t ≥ 0} is a time-continuous, time-inhomogeneous, three-state Markov chain. This means that we assume

$$P[S(y_n) = h_n \mid S(y_{n-1}) = h_{n-1} \wedge S(y_{n-2}) = h_{n-2} \wedge \ \dots \ \wedge S(y_1) = h_1] = \tag{5}$$

$$= P[S(y_n) = h_n \mid S(y_{n-1}) = h_{n-1}]$$

for any n, h_1, ..., h_n, y_1, ..., y_n.

Let ${}_t p^{\overline{gh}}_y$ denote the transition probabilities

$${}_t p^{\overline{gh}}_y = P[S(y+t) = h \mid S(y) = g], \quad h = a, i, d, \quad g = a, i. \tag{6}$$

The transition intensities are defined as

$$\mu^{gh}(y) = \lim_{t \to 0^+} \frac{{}_t p^{\overline{gh}}_y}{t}, \qquad h = a, i, d, \ g = a, i, \ h \neq g \tag{7}$$

These limits are assumed to exist for all relevant y and the intensities are assumed to be integrable functions of y.

Moreover, we are interested in the following probabilities:

$${}_t p^{\overline{hh}}_y = P[S(y+u) = h \text{ for all } u \in (0,t) \mid S(y) = h], \quad h = a, i. \tag{8}$$

The transition probabilities satisfy the Chapman-Kolmogorov relation:

$${}_t p^{gh}_y = \sum_{k=a,i} {}_\tau p^{gk}_y \ {}_{t-\tau} p^{kh}_{y+\tau}, \qquad h = a, i, d, \quad g = a, i, \tag{9}$$

This, for the probabilities ${}_t p_y^{hh}$, simply becomes

$$ {}_t p_y^{hh} = {}_\tau p_y^{hh} \; {}_{t-\tau} p_{y+\tau}^{hh}, \qquad h = a, i. \tag{10}$$

Moreover, the transition probabilities satisfy the Kolmogorov forward differential equation:

$$ \frac{d}{dt} {}_t p_y^{aa} = - \; {}_t p_y^{aa} \left[\mu^{ai}(y+t) + \mu^{ad}(y+t) \right] + \; {}_t p_y^{ai} \; \mu^{ia}(y+t) \tag{11}$$

4 The practical applications

Let us consider a disability cover in which just one state illness (i) is assumed. A time-continuous annuity of 1 monetary unit per year is paid when the insured is in state (i). At policy issue, the random present value of this annuity is:

$$ Y_x = \int_0^\infty \left[S(x+t) = i \mid S(x) = a \right] v^t dt \, . $$

Denoting by M the random number disability claims, the random present value Y_x can be also expressed as follows:

$$ Y_x = \sum_{h=1}^{M} v^{T_h} \; {}_{T_h|}\bar{a}_{D_h} \quad \text{if } M \geq 1, $$

Where:

$x + T_1$	random age at first claim,
D_1	random duration of first disability claim,
$x + T_2$	random age at second claim,
D_2	random duration of second disability claim,

Remark. A continuous annuity deferred m years on the life of a person now aged x is denoted ${}_{m|}\bar{a}_x$ (Tutt 1991).

Of course, this model allows for recovery; a simpler model can be built up by considering the following random variable:

$$ N(x, T_1) = \left(\sum_{h=1}^{M} D_h \right) / T_1 \quad \text{if } M \geq 1 $$

For example $N(x, T_1)$ is the random time spent in illness by a person aged x and healthy at policy issue, whose first disability claim occurs at time T_1.

Now, let us define the following random present value:

$$ Z_x = v^{T_1} \; \bar{a}_{\overline{N(x,T_1)|}} \, , $$

we have:

$$ Z_x \geq Y_x $$

and then

$$ E(Z_x) \geq E(Y_x) $$

because Zx reflects the annuity payments in a regular stream, hence at earlier times (on average) than the interrupted stream of payments in Y_x (see figure 2).

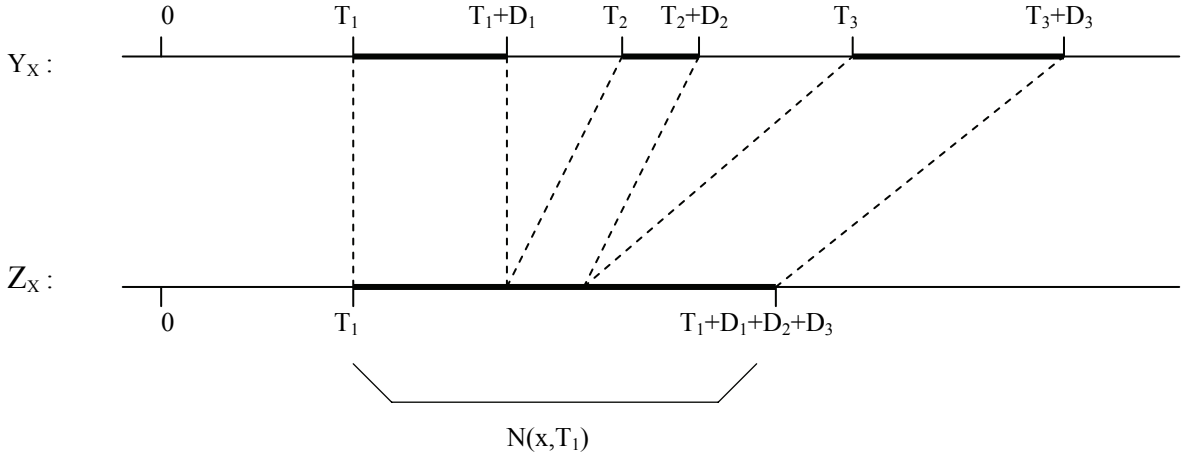
The net single premium is given by

$$\bar{a}_x^{ai} = E(Y_x)$$

So that it can be approximated by

$$\bar{a}_x^{ai} \cong E(Z_x) = \int_0^{\infty} {}_u p_x^{aa}\, \mu^{ai}(x+u) v^u E\left(\bar{a}_{\overline{N(x,u)}|}\right) du$$

Note that this approximation involves an implicit safety loading. Of course, the health story of an individual can in general be important in forecasting future changes in health. Thus, a Semi-Markov process may be more appropriate.



F*igure 2.* Random variables Yx and Zx

References

[1] Haberman, S.: Decrement tables and the measurement of morbidity, City University London 1984.

[2] Modul H, Actuarial science, The Institute of Actuaries, Oxford 1996.

[3] Neill, A.: Life contingencies, The Institute of Actuaries, Oxford 1977.

[4] Pitacco, E.: Disability risk models; towards a unifying approach, Universita di Trieste, Trieste 1993.

[5] Sakálová, K.: Recursive formulas in life insurance, Conference on Methods in Economics, Ostrava 1997.

[6] Tutt, L. W. G. and Tutt, S. I. M.: Financial aspects of Long Term Business, The Chartered Insurance Institute, Cambridge 1991.

On an Optimal Advertising Model with Lagged effect of Advertising

Pavel Pražák
Faculty of Informatics and Management
University of Hradec Králové
Nám. Svobody 301, 500 02 Hradec Králové, Czech Republic
pavel.prazak@uhk.cz

Abstract

This paper deals with a modification of the Vidale-Wolfe advertising model. The model is enriched by an assumption that the current intensity of sales depends mainly on the past intensity of advertising. Mathematical formulation results in a system of two ordinary differential equations. First we study special case when the intensity of advertising is constant. Then we concentrate on finding an optimal strategy of advertising with regard to a given cost function. For this purpose we use the Pontryagin maximum principle.

Keywords

ordinary differential equations, optimal control, Pontryagin maximum principle, singular solution.

1 Introduction

If firms want to introduce their new product into a market or they want to keep their position at the market they carry out advertising. Firms usually try to choose such advertising budgets that increase their profit best of all.

One of the first mathematical models of advertising comes from 1957. Its authors are M.L. Vidale and H.B.Wolfe, see [8]. Later in 1973 this model was modified by S. Sethi so that the theory of optimal control can be applied, see [4],[2]. The idea that the current intensity of sales depends mainly on the past intensity of advertising, which can be found in [7], is studied here. The qualitative theory of ordinary differential equations given in [1] is used in this paper too. This article follows up with [3].

2 Model Development

Let us consider a market with one type of goods and several producers/firms and choose one of these firms. Suppose that the total profit from all the sales at the market is known and it will be denoted P. At time t the chosen firm can attract only a fraction $y(t) \in [0,\, 1]$ of all the sales at the market and its profit is therefore $P\, y(t)$.

Further let us denote $u = u(t)$ the advertising effort at time t measured in a money unit per time unit. Since the firm has limited resources of incomes it is assumed that

$$u \in [0,\, \overline{u}], \tag{1}$$

where $\overline{u} > 0$ is a maximal budget for advertising expenditures. It is assumed that the effect $x = x(t)$ of advertising expenditures $u(t)$ is not immediate but it lags behind the actual advertising expenditures $u(t)$. The easiest way how to describe this feature is to state the difference equation

$$x(t+\tau) - x(t) = (u(t) - x(t))\tau,\ x(0) = x_0, \tag{2}$$

where x_0 represents the initial effect of advertising that can be interpreted as the initial acquiantance of the product. Dividing the latter equation by $\tau > 0$ and putting $\tau \to 0$ we gain the initial value problem (further referred to as IVP)

$$\dot{x} = u - x,\ x(0) = x_0, \tag{3}$$

where we disregard the time argument t in $x = x(t), u = u(t)$.

When a customer buys the product of the firm he tries it and decides to buy it once more or to buy a similar product from another firm. This process is independent of advertising and the firm cannot affect the rate of repeat sales. It means that if the firm doesn't promote its product, its repeated sales are decreasing at a constant proportionate rate $b > 0$, which the firm cannot affect:

$$y(t + \tau) - y(t) \approx -by(t)\tau. \tag{4}$$

The firm can affect its market share when it attracts new sales by advertising. It is considered to be proportionate to the advertising effect $x(t)$ and the fraction $1 - y(t)$ of customers that do not buy the product at the current time at a constant proportionate rate $a > 0$ that reflects the efficacy of advertising in attracting sales:

$$y(t + \tau) - y(t) \approx ax(t)(1 - y(t))\tau. \tag{5}$$

The above mentioned features act together and it is possible to write the difference equation

$$y(t + \tau) - y(t) = (ax(t)(1 - y(t)) - by(t))\tau,\ y(0) = y_0, \tag{6}$$

where $y_0 \in [0, 1]$ represents the initial market share of the firm.

Dividing the latter equation by $\tau > 0$ and putting $\tau \to 0$ we gain

$$\dot{y} = ax(1 - y) - by,\ y(0) = y_0. \tag{7}$$

The equations (3), (7) make up a system that represents the basic model for advertising with a lagged effect of the advertising expenditures.

3 Optimal advertising expenditure

The firm wishes to maximize its discounted profit stream which can be considered as the difference between its gross profit of $P\,y(t)$ and advertising expenditure $u(t)$, with the constant discount factor $r \in [0, 1]$. The introduced problem can be considered as an optimal control problem with the objective functional

$$\mathcal{J}(u(\cdot)) = \int_0^\infty e^{-rt}(Py(t) - u(t))dt, \tag{8}$$

that is to be maximized subject to the system (3), (7) and the condition (1).

The admissible control functions $u = u(t)$ are to be piecewise continuous. To solve the given control problem we use the Pontryagin maximum principle. We don't formulate it generally but only for the problem (8), (3), (7) and (1). First let us introduce its Hamiltonian function

$$H(t, y, u, \mu_0, \mu_1, \mu_2) = \mu_0 e^{-rt}(Py - u) + \mu_1(u - x) + \mu_2(ax(1 - y) - by), \tag{9}$$

where we disregard the time argument t in $y = y(t), u = u(t), \mu_1 = \mu_1(t)$ and $\mu_2 = \mu_2(t)$.

Theorem. *Let $(\widehat{u}, \widehat{x}, \widehat{y})$ be the optimal process to the problem (8), (3),(7) and (1), then there exist such a constant $\mu_0 \in \mathbf{R}$ and such continuous functions $\mu_1(t), \mu_2(t) : \mathbf{R} \to \mathbf{R}$ that the following conditions are valid*

(i) $(\mu_0, \mu_1(t), \mu_2(t)) \neq 0$ and $\mu_0 \geq 0$,

(ii) adjoint equation:

$$\dot{\mu_1}(t) = -\frac{\partial H}{\partial x}(t, x, y, u, \mu_0, \mu_1, \mu_2) = \mu_1(t) - \mu_2(t)a(1 - y(t)), \tag{10}$$

$$\dot{\mu_2}(t) = -\frac{\partial H}{\partial y}(t, x, y, u, \mu_0, \mu_1, \mu_2) = -\mu_0 e^{-rt}P + \mu_2(t)(ax(t) + b), \tag{11}$$

(iii) maximum principle: for every $t \in [0,\infty)$, where the function $\widehat{u}(t)$ is continuous

$$H(t,\widehat{x},\widehat{y},\widehat{u},\mu_0,\mu_1,\mu_2) = \max\{H(t,\widehat{x},\widehat{y},u,\mu_0,\mu_1,\mu_2)|\ u \in [0,\ \overline{u}]\}$$

(iv) stationary condition : $H(t,\widehat{x}(t),\widehat{y}(t),\widehat{u}(t),\mu_0,\mu_1(t),\mu_2(t)) = q(t)$, where

$$q(t) = r\mu_0 \int_t^\infty e^{-rs}(Py(s) - u(s))ds.$$

For the general formulation and the proof see [6]. In our discussion we concentrate on the possibility $\mu_0 \neq 0$ and put $\mu_0 = 1$. Since the Hamiltonian function (9) is linear in the variable u it reaches its maximum at the edges of domain of u. For all $u \in [0,\overline{u}]$ the maximum principle gives

$$(\widehat{u} - u)(e^{-rt} - \mu_1) \leq 0. \tag{12}$$

There exist three cases

$$\begin{array}{l} \text{If } \mu_1(t) > e^{-rt},\ t \in I_1 \text{ then } \widehat{u}(t) = \overline{u} \text{ on } I_1. \\ \text{If } \mu_1(t) = e^{-rt},\ t \in I_2 \text{ then it is not possible to determine} \\ \quad \widehat{u} \text{ on } I_2 \text{ by the maximum principle.} \\ \text{If } \mu_1(t) < e^{-rt},\ t \in I_3 \text{ then } \widehat{u}(t) = 0 \text{ on } I_3. \end{array} \tag{13}$$

4 Singular solution

In this item we suppose the existence of a time interval $I_2 \subset [0,\infty)$ where the condition $\mu_1(t) = e^{-rt}$ is valid. The problem to determine an optimal control with the mentioned condition is reffered to as the problem of singular control. [1]

Taking the derivation of the latter condition we gain

$$\dot{\mu}_1 = -re^{-rt}. \tag{14}$$

Putting the right side of the adjoint equation (10) into (14) we gain

$$e^{-rt} - \mu_2 a(1-y) = -re^{-rt}. \tag{15}$$

Taking the derivation of the latter condition and after some rearranging we gain

$$-r(r+1)e^{-rt} = \dot{\mu_2}a(1-y) + \mu_2 a(-\dot{y}) \tag{16}$$

Putting the right side of the equation (7) and the right side of the adjoint equation (11) into (16), using (15) for elimination of μ_2 and after some rearranging we finally gain the quadratic equation with a variable $(1-y)$

$$Pa(1-y)^2 - r(r+1)(1-y) - b(r+1) = 0. \tag{17}$$

Since $y \in [0,1]$ and so $1-y \in [0,1]$ we choose the positive sign before the radical in the roots of (17). Now we gain the singular value of sales

$$y_s = y_s(t) = 1 - \frac{r(r+1) + \sqrt{r^2(r+1)^2 + 4Pab(r+1)}}{2Pa},\ t \in I_2. \tag{18}$$

[1] Singular control problems are generally difficult to solve, see e.g. [2]. According to [5] it is desirable to check additional necessary conditions which are called generalized *Legendre-Clebsch conditions*. These conditions can be written in the form

$$(-1)^k \frac{\partial}{\partial u}\left[\frac{d^{2k}}{dt^{2k}}\frac{\partial}{\partial u}H\right] \leq 0,\ k = 0,1,2,\ldots$$

Since $y_s(t)$ is constant in I_2 then $\dot{y}_s(t) = 0$ in I_2. From (7) we gain the effect of advertising

$$x_s = x_s(t) = \frac{by_s}{a(1-y_s)},\ t \in I_2. \tag{19}$$

Since $x_s(t)$ is constant in I_2 then $\dot{x}_s(t) = 0$ in I_2. From (5) we gain the singular control

$$u_s = u_s(t) = x_s,\ t \in I_2. \tag{20}$$

To be $y_s > 0$ we use (18) to get $Pa > (r+b)(r+1)$. To keep $u_s < \overline{u}$ we use (20), (19) and (18) to get $Pa < (r+1)(1+\frac{a}{b}\overline{u})(r+b+a\overline{u})$.

There remains the question whether and how to reach the stationary solution $(x_s,\, y_s)$ if $y(0) \neq y_s$. For the next reasoning it is useful to notice that $u_s < \overline{u}$ implies [2] $y_s < \overline{y}$, where $\overline{y} = \frac{a\overline{u}}{a\overline{u}+b}$ is the stationary solution to (7) for $u(t) = \overline{u}$, $t \in I_2$.

If $y(0) < y_s$, then the value of sales y increases as rapidly as possible with the control $u(t) = \overline{u}$. On the other side if $y(0) > y_s$, then the value of sales y decreases as rapidly as possible with the conrol $u(t) = 0$. This suggests that for the adjustment to the singular stationary solution $(x_s,\, y_s)$ of IVP

$$\begin{array}{ll} \dot{x} = -x + u_s, & x(0) = x_0 \in [0,\, \overline{u}] \\ \dot{y} = ax(1-y) - by, & y(0) = y_0 \in [0,\, 1] \end{array} \tag{21}$$

we could use the next process:

$$u(t) = \begin{cases} \overline{u}, & \text{if} \quad y(t) < y_s \\ u_s, & \text{if} \quad y(t) = y_s \\ 0, & \text{if} \quad y(t) > y_s. \end{cases} \tag{22}$$

For the graphical illustration see figure 1.

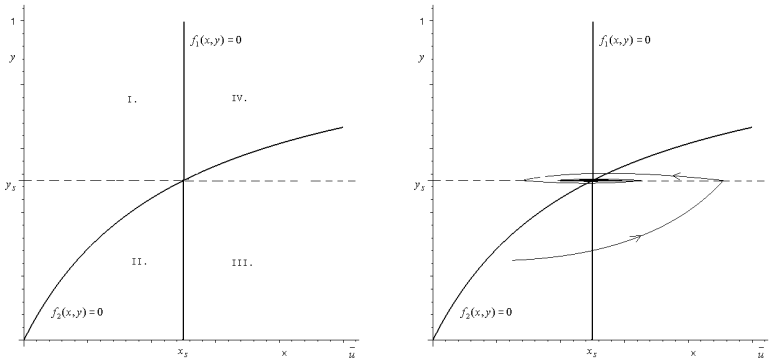


Figure 1: On left: An example of the phase diagram of IVP (21) with the singular point (x_s, y_s), where $x_s = u_s$ and nullclines $f_1(x,y) = -x + x_s = 0$, $f_2(x,y) = ax(1-y) - by = 0$. On right: An example of the path of the adjustment to the singular point (x_s, y_s).

To show that the described process really attains the singular steady state of the problem (21) we deal with both IVP

$$\begin{array}{ll} \dot{x} = -x + \overline{u}, & x(0) = x_0 \in [0,\, \overline{u}] \\ \dot{y} = ax(1-y) - by, & y(0) = y_0 \in [0,\, 1] \end{array} \tag{23}$$

and IVP

$$\begin{array}{ll} \dot{x} = -x, & x(0) = x_0 \in [0,\, \overline{u}] \\ \dot{y} = ax(1-y) - by, & y(0) = y_0 \in [0,\, 1] \end{array} \tag{24}$$

[2] Let $u(t) = C$ be constant in $[0, \infty)$ and let C be an arbitrary but fixed point in $[0, \overline{u}]$. Then the stationary solution to the equation (7) is $y^\circ = \frac{aC}{aC+b}$, which is an increasig function of C on $[0, \overline{u}]$. Since $u_s < \overline{u}$ we gain the observation.

Our next step is to divide the region $[0, \overline{u}] \times [0, 1]$ of the phase diagram of IVP (21) into four regions. It is accomplished by the x–nullcline $n_1 : x - u_s = 0$ and y–nullcline $n_2 : ax(1-y) - by = 0$, see the figure 1. Now observe that $ax(1-y) - by > 0$ in regions II and III for both the problems (23) and (24). Consequently, $\dot{y}(t) > 0$ in regions II and III for both the problems (23) and (24). Similarly, $\dot{y}(t) < 0$ in regions I and IV for both the problems (23) and (24). Observe further that $-x + \overline{u} > 0$ in all the regions I–IV for the problem (23). Consequently, $\dot{x}(t) > 0$ in all the regions I–IV for the problem (23). Finally observe that $-x < 0$ in all the regions I–IV. Consequently, $\dot{x}(t) < 0$ in all the regions I–IV for the problem (24).

Observation 1 *Let $x^1(t), y^1(t)$ be the solution to the problem (23) and $x_0^1 < x_s$, $y_0^1 = y_s$. Then there exists one and only one time $t_1 > 0$ such that $x^1(t_1) > x_s$ and $y^1(t_1) = y_s$.*

The initial point (x_0^1, y_0^1) is in the region I where for the problem (23) the conditions $\dot{x}^1(t) > 0$ and $\dot{y}^1(t) < 0$ are valid. Since the y–nullcline $ax(1-y) - by = 0$ is an increasing function in the variable x, the orbit $(x^1(t), y^1(t))$ cuts it at some point. At this intersection point the conditions $\dot{x}^1(t) > 0$ and $\dot{y}^1(t) = 0$ are valid. It means that the orbit $(x^1(t), y^1(t))$ moves to region II. The conditions $\dot{x}^1(t) > 0$ and $\dot{y}^1(t) > 0$ in regions II and III for the problem (23) are valid. Notice also that the orbit $(x^1(t), y^1(t))$ cannot intersect the line $n_3 : -x + \overline{u} = 0$ since this is another orbit of the problem (23) and the local uniqueness theorem for autonomous IVP is valid. It has been already shown that $y_s < \overline{y} = \frac{a\overline{u}}{a\overline{u}+b}$ thus there exists time $t_1 > 0$ such that $y^1(t_1) = y_s$ and $x^1(t_1) > x_s$.

Observation 2*Let $x^2(t), y^2(t)$ be the solution to the problem (24) and $x_0^2 > x_s$, $y_0^2 = y_s$. Then there exists one and only one time t_2 such that $x^2(t_2) < x_s$ and $y^2(t_2) = y_s$.*

It is possible to write a similar argumentation as in the observation 1.

Observation 3*Let $x^1(t), y^1(t)$ be the solution to the problem (23) with the initial condition $x_0^1 < x_s$, $y_0^1 = y_s$. Let $x^2(t), y^2(t)$ be the solution to the problem (24) with the initial condition $x_0^2 = x^1(t_1)$, $y_0^2 = y_s$. Then $x_0^1 < x^2(t_2)$.*

First let us consider the solution $x^*(t)$, $y^*(t)$ of the IVP (24) with the initial condition $x_0 = x_s, y_0 = 1$. On the orbit $(x^*(t), y^*(t))$ there is possible to find a point (x^*, y^*) such that $x^* < x_s$ and $y^* = y_s$. Let us also recall that according to the theorem of local uniqueness of IVP for autonomous problem it is not possible that any two different orbits of this IVP can intersect each other.

Case1 Let $x_0^1 < x^*$. We apply observation 1 and observation 2. Since the orbit $(x^2(t), y^2(t))$ of the IVP (23) can not intersect the orbit $(x^*(t), y^*(t))$ of the same IVP, the condition $x^* < x^2(t_2)$ is valid. But it means that $x_0^1 < x^2(t_2)$.

Case2 Let $x_0^1 > x^*$ and $x_0^1 > x^2(t_2)$ then it is possible to consider the solution $x^3(t)$, $y^3(t)$ to IVP (23) with the initial condition $x_0^3 = x^2(t_2)$, $y_0^3 = y_s$. According to observation 1 there exits time $t_3 > 0$ such that $x^3(t_3) > x_s$ and $y^3(t_3) = y_s$. The orbits $(x^1(t), y^1(t))$ and $(x^3(t), y^3(t))$ of the IVP (23) cannot intersect. With regards to the initial condition this implies that $x^3(t_3) > x^1(t_1)$. Now we can apply observation 2 and consider solution $x^4(t), y^4(t)$ of IVP (24) with the initial condition $x_0^4 = x^3(t_3), y_0^4 = y_s$. Then there exists time $t_4 > 0$ such that $x_0^4(t_4) < x_s$, $y^4(t_4) = y_s$. With the same argumentation as above is $x^4(t_4) < x_2(t_2)$. Similarly it is possible to continue with the construction of subsequent solutions and orbits. For all these orbits $x^{2n+2}(t_{2n+2}) < x^{2n}(t_{2n})$, $n \in \mathbf{N}$ are valid. Since x^* is not a stationary point of the IVP (23), there exists $n \in \mathbf{N}$ such that the orbit $(x^{2n}(t), y^{2n}(t))$ of the IVP (23) intersects the orbit $(x^*(t), y^*(t))$. But this is impossible and we can can conclude that the supposal $x_0^1 > x^2(t_2)$ can not be accepted and instead of it the condition $x_0^1 < x^2(t_2)$ is valid with regards to the uniqueness theorem of IVP.

A similar observation as observation 3 is possible to be found for an arbitrary initial point $(x_0, y_0) \in [0, \overline{u}] \times [0, 1]$. This suggests that the process (22) is successful and really converges to the singular stationary solution (x_s, y_s). For a geometrical illustration see figure 2.

5 Conclusion

From the above mentioned results it is evident that if the effectivity of advertising is high enough – the condition $Pa > (r+b)(r+1)$ is valid and if the potential advertising expenditure $\overline{u}$ is large enough – the condition $Pa < (r+1)(1+\frac{a}{b}\overline{u})(r+b+a\overline{u})$ is valid, it is possible to find a value u_s of advertising expenditure and a value of market share y_s which gives the maximal profit of the firm. As has been shown the level of market sales y_s can be reached by the process (22) that is in [2] called the most rapid approach path.

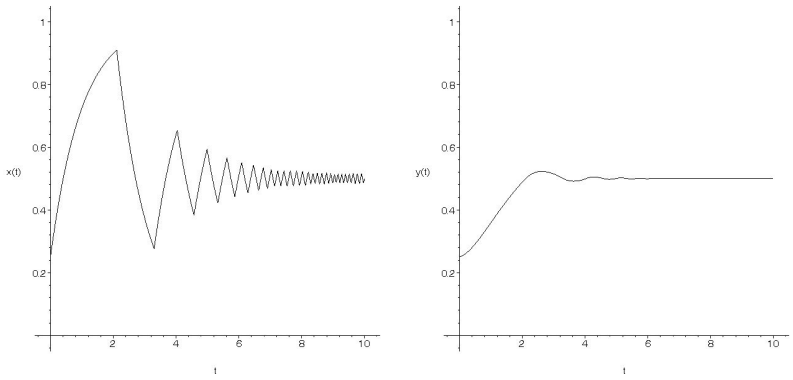


Figure 2: An example of the x–component graph and the y–component graph of the adjustment process (22). The x–component graph illustrates the lagged effect of advertising expenditure u and the y–component graph illustrates the adjustment process to the singular level y_s of sales.

Some empirical advertisement studies can be found in the book [7]. One of the given advertising strategies which is called pulsing advertising corresponds to the process (22). If the chosen firm uses this process it can avoid wasting of financial resources at the time when the effect of the previous advertising is still good. It is important to mention that the firm should modify its advertising strategy in order to avoid decreasing effectivity of advertising. But this process is out of the scope of our mathematical model.

The discussion of the given control problem (8), (3),(7) and (1) is not complete in this paper. The main reason for this incomleteness is that conditions for the zero advertising expenditure or the maximal possible advertising expenditure have not been found yet.

References

[1] Amann, H.: Ordinary Differential Equation, de Gruyter, Berlin, 1990

[2] Kamien, M.I., Schwartz, N.I.: Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management, North-Holland, 1991

[3] Pražák, P.: On the Vidale-Wolfe Advertising Model, *in MME'04,* , Brno 2004, 269–274

[4] Sethi, S.P.: Optimal control of the Vidale-Wolfe advertising model, *Operations Research,* **21,** (1973), 998-1013

[5] Sethi, S.P., Thomson, G.L.: Optimal Control Theory, Applications to Management Science and Economics, Kluwer, 2003

[6] Štorová, J.: Úlohy optimálneho riadenia na nekonečnom časovom horizonte, diploma thesis, FMFI University of Comenius, Bratislava 2001

[7] Tellis, G.J.: Reklama a podpora prodeje, Grada Publishing, Prague 2000

[8] Vidale, M.L., Wolfe, H.B.: An operations research study of sales response to advertising, *Operations Research,* **5** (1957), 370–381

Application of the Bootstrap filter method on a small economy model*

Hana Pytelová, Osvald Vašíček

Department of Applied Mathematics and Computer Science
Faculty of Economics and Administration
Masaryk University
Lipová 41a, 602 00 Brno
Czech Republic
e-mail: pytelova@mail.muni.cz, osvald@econ.muni.cz

Abstract

The paper shows the monetary policy problem in a simple framework and it illustrates the behaviour of the model on the Czech economy data. Model parameters are estimated by the weighted Bootstrap algorithm which represents an important alternative approach to model estimation. Its power lies in its generality because it is usable for non-local systems. It is especially important in the case that classical methods like extended Kalman filter diverge or are not applicable; or when only the lack of data is available (which is the case of the Czech Republic). Conditional probability density functions of the parameters and states are analyzed.

Keywords

weighted Bootstrap algorithm, monetary policy model, conditional probability density functions, impulse responses, rational expectations

1 Bootstrap Filter Method

The weighted Bootstrap method represents an important alternative estimation approach to model estimation, especially for non-Gaussian (even non local) systems. Its power lies in its generality because it is usable for non-local systems. It is especially important in the case that classical methods like extended Kalman filter diverge or are not applicable; or when only the lack of data is available (which is the case of the Czech Republic).

It is a Monte Carlo method based on the fact that information about the random variable distribution is carried by a random sample from this distribution. The larger

*This paper has been worked as a part of research activities at the grant project of GA CR No. 402/05/2172

size of the random sample, the more accurate information. Empirical probability denstity function gained from the random sample is an approximation of the real probability denstity function.

The computation process consists of three steps. In the prediction step, conditional probability density functions (CPDFs) and samples are recomputed (according to their importance), weights remain the same. In the filtration step, CPDFs and weights are recomputed, samples are the same. In the smoothing step, CPDFs and weights are recomputed, samples are the same. At the beginning, the weights are uniform. Then, they are computed as normalized likelihood.

2 Small Economy Model

Monetary policy significantly influences the short-term course of the real economy due to temporary nominal price rigidities. The baseline framework is a forward-looking dynamic general equilibrium model of economy with money and temporary nominal price rigidities based on [1]. The aggregate behaviorial equations evolve from optimization by households and firms. It is very important that current economic behaviour depends on expectations of the future course of monetary policy, as well as on current and past policy.

Let x_t be the output gap, which is the deviation of actual output from its potential level (in natural logarithms). Let π_t^n be the period t deviation of the net inflation rate from its target, let i_t be the deviation of the nominal interest rate from its equilibrial level and let E_t be the expected mean based on information available at time t. ϕ, δ, λ, β, γ_π, γ_x, μ and ρ are parameters. Parameters are specified in details in section 3. It is possible to form the following model:

$$x_t = -\phi(i_t - E_t\pi_{t+1}^n) + \delta E_t x_{t+1} + (1-\delta)x_{t-1} + g_t \quad (1)$$

$$\pi_t^n = \lambda x_t + \beta E_t \pi_{t+1}^n + v_t \quad (2)$$

$$i_t = \gamma_\pi E_t \pi_{t+1}^n + \gamma_x x_t \quad (3)$$

$$g_t = \mu g_{t-1} + \hat{g}_t \quad (4)$$

$$v_t = \rho v_{t-1} + \hat{v}_t \quad (5)$$

Equation (1) is a forward-looking IS curve. It describes demand side of economy. Equation (2) is a forward-looking Phillips curve. It represents the supply side of economy, it is a relation between nominal and real variables. Forward-looking Taylor rule (3) represents a reaction of the monetary authority to the state of the economy. It demonstrates real behaviour of many central banks. It is simple and much more transparent than rules based on optimization. The last two equations describe the random shocks that appear in the first two equations – it is a demand shock (4) and a supply shock (5), respectively. g_t and v_t are i.i.d. (independent, identically distributed) random variables with zero mean and variances σ_g^2 and σ_v^2, respectively, and where $\mu, \rho \in \langle 0, 1)$.

3 Results of the Estimation by the Bootstrap Filter

The model is identified on the Czech economy behaviour. Data of the output gap, the net inflation rate, the net inflation target and the short-term nominal interest rate are quarterly values of the data from the Czech National Bank (years 1994 — 2004) . The system is nonlinear and interdependent. It consists of equations (1), (2), (3), (4) and (5). The simultaneous estimation of parameters and states was realized by the Extended Bootstrap Filter Smoother (EBSFS) using [4]. 500 samples were used for each model time.

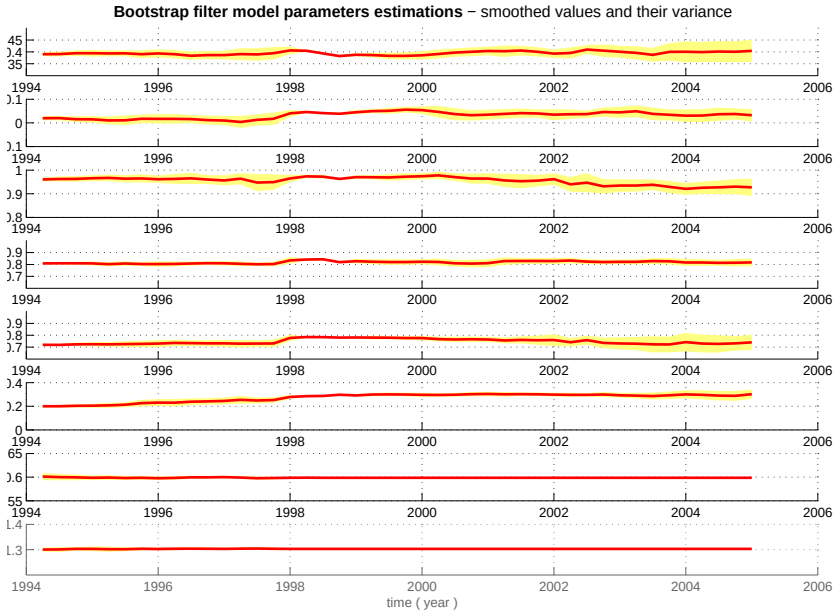


Figure 1: Samples, and filtered and smoothed values of parameters and states

Figure 1 represents development of smoothed values of parameters (ϕ, λ, β, μ, ρ, δ, γ_x, γ_π) and their variance. The interest elasticity (ϕ) should be greater than zero and it is between $0.38 - 0.41$. It represents a negative relation between the output and the real interest rate in the Czech economy. The Phillips curve coefficient (λ) should be greater than zero, and it is $0.01 - 0.05$. It represents a very slight relation between the output and the rate of inflation. The discount factor (β) should be between zero and unity, near to unity, and it is between $0.92 - 0.97$, which means, that agents do not discount future — "today is more than tomorrow". The random shock parameters (μ and ρ) should be between zero and unity, and they are between $0.80 - 0.84$ and $0.72 - 0.78$ respectively. It implies a great persistence in demand and supply shocks, i. e. that demand and supply shocks do not disappear immediately, but relatively slowly. Parameter of looking forward in production (δ)

is between $0.20 - 0.30$. It means that agents prefer predicitions of future based on the past development. Monetary policy coefficients γ_x and γ_π are around 0.6 and 1.30 respectively. It shows, that the central bank is more concerned on the inflation rate deviations from its target than on the output gap deviations.

4 CPDF Analysis

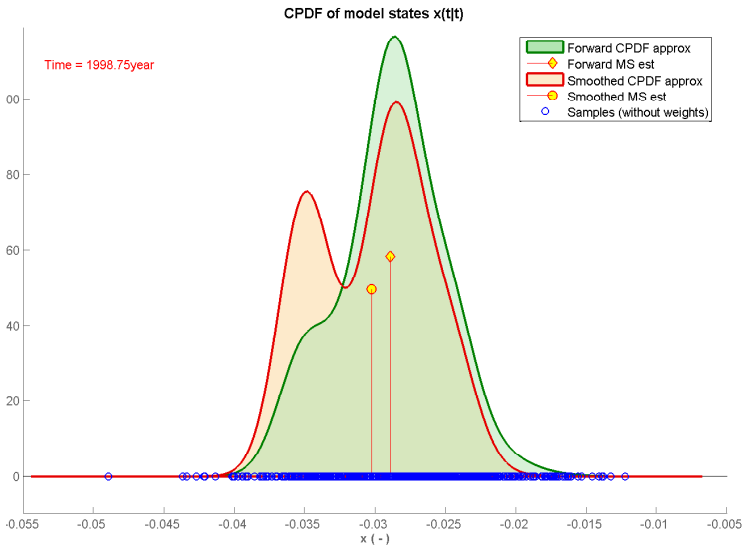


Figure 2: CPDF of the output gap at time t=18

The shape of the conditional probability density functions (CPDF) of the parameters is unimodal in each model time. On the other hand, Figure 2 shows that especially the CPDF of the state x_t (the output gap) is not unimodal. This fact magnifies the importance of using EBSFS instead of some classical methods based on strong presumptions like iterative extended Kalman filter.

5 Impulse Responses Analysis

Rational expectations were implemented to the model according to [3]. The Figure 3 shows the accommodation process of the economy to a unity demand shock (left part) and supply shock (right part). The upper two frames shows the the case of unexpected shocks. The lower two frames show the case of expected shocks in the fifth period. Periods of time are measured on the horizontal axis. Reaction of the variables is measured on the vertical axis. In the case of expected shocks, we can see accommodation reactions of the economy before the shock hit it.

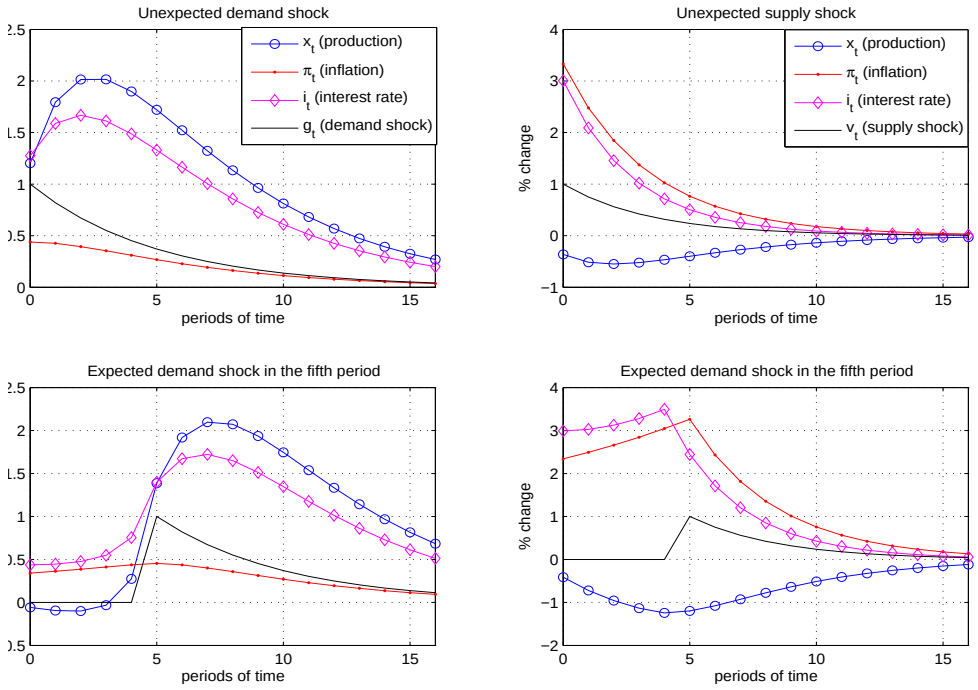


Figure 3: Impulse response under rational expectations - demand and supply shock

In the case of an unexpected demand shock at time 0, rising production rises prices and the central bank react by rising the interest rate. In the case of an unexpected supply shock at time 0 (for example rising price of oil etc.), prices rise and production decline, the central bank react by rising the interest rate (to fight with inflation). If the shock is expected, the accommodation processes in the economy begin at the moment of getting the information about the future shock.

6 Conclusion

The paper presents parameters and states estimation of the small economy model. The model is nonlinear and only limited data set is available. The weighted bootstrap filter method was used for estimation and bootstrap smoothing was also realized. EBSFS estimation procedure was used for the Czech economy data. The estimated model can be acceptably considered as a rough approximation of the behaviour of the real economic system. According to the estimated parameters, all of the causal relations are identified correctly. The estimation gained by the Bootstrap filtration is not constant. However, it can be considered acceptable from the economic point of interpretation. The course of the deviations of the samples of the parameters and unobservable states from the filtered development was in this case

unimodal with considerable evidence of a mean value. On the other hand, some of the observable states, especially the output gap, does not seem unimodal. This fact magnifies the significance of the Bootstrap filtration. The further reason for its utilization is unsufficient length of the available time series and dramatical changes in the economic development in transitive countries. The model incorporates rational expectations and the properties of the model were analyzed by impulse response analysis in the case of expected and unexpected demand and supply shocks, and it seems to be in accordance with reality. An extension of the model with the exchange rate and some other foreign characteristics like foreign output, price of oil etc., is an object of current research.

References

[1] Clarida, R.; Gali, J.; Gertler, M.: *The Science of Monetary Policy: A New Keynesian Perspective*. NBER Working Paper 7147, National Bureau of Economic Research, May 1999.

[2] Pytelová, H.; Vašíček, O.: *Estimation of Alternative Monetary Policy Rules and their Comparison*. In Bulletin of the Czech Econometric Society, July 2004, Volume 11, Issue 20, pp. 15-28. ISSN 1212 - 074x.

[3] Söderlind, P.: *Lecture Notes for Monetary Policy (PhD course at UNISG)*. University of St. Gallen, Switzerland, 2003.

[4] Trnka, P.: *Matlab toolbox for Sequential Monte Carlo methods (SMCtool)*. Czech Technical University in Prague, Faculty of Electrical Engineering, Depratment of Control Engineering, Prague, 2004.

[5] Walsh, C. E.: *Monetary Theory and Policy*. The M.I.T. Press, 1998. ISBN 0-262-23199-9.

Duality in Fuzzy Multiple Objective Linear Programming with Possibility and Necessity Relations

Jaroslav Ramík
Department of Mathematical Methods in Economics
Faculty of Business Administration, Silesian University,
Karviná, Czech Republic
e-mail: ramik@opf.slu.cz

Abstract
A class of fuzzy multiple objective linear programming (FMOLP) problems with fuzzy coefficients based on fuzzy relations is introduced, the concepts of feasible and (α, β)-maximal and minimal solutions are defined. The class of crisp (classical) MOLP problems can be embedded into the class of FMOLP ones. Moreover, for FMOLP problems a new concept of duality is introduced and the weak and strong duality theorems are derived. The introduced concepts and results are illustrated and discussed on a simple numerical example.

Keywords
Fuzzy relations, Fuzzy multiple objective linear programming, Duality

1 Introduction

The problem of duality has been investigated since the early stage of fuzzy linear programming (FLP), see [1], [4], [10]. In this paper we first introduce a broad class of fuzzy multiple objective linear programming problems (FMOLP problems) and define the concepts of $\beta-$feasible and $(\alpha, \beta)-$maximal and minimal solutions of FMOLP problems. The class of classical MOLP problems can be embedded into the class of FMOLP ones, moreover, for FMOLP problems we define the concept of duality and prove the weak and strong duality theorems - generalizations of the classical ones. The results are compared to the existing literature, see [7], [8] and [9]. To illustrate the introduced concepts and results we present and discuss a simple numerical example.

2 Preliminaries

Let X be an n-dimensional Euclidean space $\mathbf{R}^n$. By $\mathcal{F}(X)$ we denote the set of all fuzzy subsets A of X, where every fuzzy subset A of X is uniquely determined by the membership function $\mu_A : X \to [0,1]$, and $[0,1] \subset \mathbf{R}$ is a unit interval, $\mathbf{R}$ is the Euclidean space of real numbers. We say that the fuzzy subset A is *crisp* if μ_A is a characteristic function of A, i.e. $\mu_A : X \to \{0,1\}$. It is clear that the set of all subsets of X, $\mathcal{P}(X)$, can be isomorphically embedded into $\mathcal{F}(X)$.

Let

$$\begin{aligned} [A]_\alpha &= \{x \in X | \mu_A(x) \geq \alpha\} \text{ for } \alpha \in (0,1], \\ [A]_0 &= cl\{x \in X | \mu_A(x) > 0\} \ . \end{aligned}$$

where clB means a usual closure of B, $B \subset X$. For $\alpha \in [0,1]$, $[A]_\alpha$ are called *α-cuts.* $[A]_0$ is usually called a *support of A.*

A is called the *fuzzy quantity* if there exist $a, b, c, d \in \mathbf{R}$, $-\infty < a \leq b \leq c \leq d < +\infty$, such that

$$\begin{aligned} \mu_A(t) &= 0 && \text{if } t < a \text{ or } t > d, \\ \mu_A(t) &&& \text{is strictly increasing if } a < t < b, \\ \mu_A(t) &= 1 && \text{if } b \leq t \leq c, \\ \mu_A(t) &&& \text{is strictly decreasing if } c < t < d. \end{aligned}$$

The set of all fuzzy quantities is denoted by $\mathcal{F}_0(\mathbf{R})$.This set contains well known classes of fuzzy numbers: crisp (real) numbers, crisp intervals, triangular fuzzy numbers, trapezoidal and bell-shaped fuzzy numbers etc.

Let A, B be fuzzy sets with the membership functions $\mu_A : \mathbf{R} \to [0,1]$, $\mu_B : \mathbf{R} \to [0,1]$, respectively. We shall consider

$$\mu_{Pos}(A,B) = \sup\{\min(\mu_A(x), \mu_B(y)) | x \leq y, x, y \in \mathbf{R}\}, \tag{1}$$

$$\mu_{Nec}(A,B) = \inf\{\max(1-\mu_A(x), 1-\mu_B(y)) | x > y, x, y \in \mathbf{R}\}. \tag{2}$$

Here (1) is called the *possibility relation*, (2) is called the *necessity relation.*

The possibility and necessity relations have been originally introduced as *possibility and necessity indices* in [2], where also mathematical analysis and interpretation has been discussed. We write alternatively

$$\mu_{Pos}(A,B) = (A \preceq^{Pos} B),\ \mu_{Nec}(A,B) = (A \prec^{Nec} B), \tag{3}$$

where μ_{Pos} and μ_{Nec} are the membership functions of the fuzzy relations on $\mathbf{R}$. By $A \succeq^{Pos} B$ or $A \succ^{Nec} B$ we mean $B \preceq^{Pos} A$ or $B \prec^{Nec} A$, respectively.

It can be easily verified that all possibility and necessity relations are fuzzy extensions of the classical binary relation $\leq$, see [8].

3 Multiple Objective Linear Programming Problem with Fuzzy Coefficients

In this section we introduce a fuzzy multiple objective linear programming problem (FMOLP problem) where coefficients are fuzzy quantities.

Let $\mathcal{K} = \{1,2,...,k\}$,$\mathcal{M} = \{1,2,...,m\}$, $\mathcal{N} = \{1,2,...,n\}$, k, m, n be positive integers. The *multiple objective linear programming problem* (MOLP problem) is a problem

$$\begin{array}{ll} \text{maximize} & z_q = c_{q1}x_1 + \cdots + c_{qn}x_n\ ,\ q \in \mathcal{K}, \\ \text{subject to} & \\ & a_{i1}x_1 + \cdots + a_{in}x_n\ \leq\ b_i,\ i \in \mathcal{M}, \\ & \qquad\qquad x_j \geq 0,\ j \in \mathcal{N}. \end{array} \tag{4}$$

Here, we assume that the weights of the criteria are known, i.e. we have positive numbers $v_q > 0$, for all $q \in \mathcal{K}$, such that $\sum_{q\in\mathcal{K}} v_q = 1$. In practical situations the weights are interpreted as the relative importances of the objectives (or, criteria) and can be identified e.g. by using the well known Saaty's pairwise comparison method.

Setting

$$c_j = \sum_{q\in\mathcal{K}} v_q c_{qj}, j \in \mathcal{N}, \tag{5}$$

we obtain an associated LP problem

$$\begin{array}{ll} \text{maximize} & z = c_1x_1 + \cdots + c_nx_n\ , \\ \text{subject to} & \\ & a_{i1}x_1 + \cdots + a_{in}x_n\ \leq\ b_i,\ i \in \mathcal{M}, \\ & \qquad\qquad x_j \geq 0,\ j \in \mathcal{N}. \end{array} \tag{6}$$

The following result is well known, see e.g. [3].

Theorem 1 *If $x = (x_1, x_2, ..., x_n)$ is an optimal solution of LP problem (6), then it is a Pareto-optimal solution of MOLP problem (4).*
Moreover, if $x = (x_1, x_2, ..., x_n)$ is a Pareto-optimal solution of MOLP problem (4), then there exist positive numbers $v_q \geq 0$, for all $q \in \mathcal{K}$, with $\sum_{q\in\mathcal{K}} v_q = 1$ such that x is an optimal solution of LP problem (6) with (5).

Knowing relative importances, i.e. the weights $v_q, q \in \mathcal{K}$, of the objectives in some MOLP problem one can solve this problem as an associated LP problem. Here, we utilize this result in building duality theory for fuzzy MOLP problems.

Applying the Extension principle we can easily obtain the following property: If $\tilde{c}_{qj},\ \tilde{a}_{ij} \in \mathcal{F}_0(\mathbf{R}), x_j \geq 0$, then the fuzzy sets $\tilde{c}_{q1}x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_{qn}x_n$, $\tilde{a}_{i1}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in}x_n$ are again fuzzy quantities.

Let $\tilde{P}$ be a fuzzy relation - fuzzy extension of the usual binary relation $\leq$ on $\mathbf{R}$. The *fuzzy multiple objective linear programming problem* (FMOLP problem) is denoted as

$$\begin{array}{l} \text{"maximize" } \tilde{z}_q = \tilde{c}_{q1}x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_{qn}x_n \ , q \in \mathcal{K}, \\ \text{"subject to"} \\ \qquad (\tilde{a}_{i1}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in}x_n) \ \tilde{P} \ \tilde{b}_i, \ i \in \mathcal{M}, \\ \qquad\qquad x_j \geq 0, \ j \in \mathcal{N}. \end{array} \tag{7}$$

In (7) the value $\tilde{a}_{i1}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in}x_n \in \mathcal{F}_0(\mathbf{R})$ is "compared" with a fuzzy quantity $\tilde{b}_i \in \mathcal{F}_0(\mathbf{R})$ by some fuzzy relation $\tilde{P}$. The "maximization" of the objective functions denoted by "maximize" $\tilde{z}_q = \tilde{c}_{q1}x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_{qn}x_n$ (in quotation marks) will be investigated later on.

Setting

$$\tilde{c}_j = \sum_{q \in \mathcal{K}} v_q \tilde{c}_{qj}, j \in \mathcal{N}, \tag{8}$$

we obtain an assocoated FLP problem

$$\begin{array}{l} \text{"maximize"} \quad \tilde{z} = \tilde{c}_1x_1 + \cdots + \tilde{c}_nx_n \ , \\ \text{"subject to"} \\ \qquad (\tilde{a}_{i1}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in}x_n) \ \tilde{P} \ \tilde{b}_i, \ i \in \mathcal{M}, \\ \qquad\qquad x_j \geq 0, \ j \in \mathcal{N}. \end{array} \tag{9}$$

Now, we shall deal with the constraints of FMOLP problem (7), or, (9), see also [6], [5], or [9].

4 Feasible Region, β-Feasible Solution

A fuzzy set $\tilde{X}$, whose membership function $\mu_{\tilde{X}}$ is defined for all $x \in \mathbf{R}^n$ by

$$\mu_{\tilde{X}}(x) = \begin{cases} \min\{\mu_{\tilde{P}}(\tilde{a}_{11}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{1n}x_n, \tilde{b}_1), \cdots, \mu_{\tilde{P}}(\tilde{a}_{m1}x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{mn}x_n, \tilde{b}_m)\} \\ \qquad \text{if } x_j \geq 0 \text{ for all } j \in \mathcal{N}, \\ 0 \quad \text{otherwise,} \end{cases} \tag{10}$$

is called the *fuzzy set of feasible region* of the FMOLP problem (7).
For $\beta \in (0, 1]$, a vector $x \in [\tilde{X}]_\beta$ is called the β-*feasible solution* of the FMOLP problem (7), or, FLP (9).

Notice that the feasible region $\tilde{X}$ of a FMOLP problem is a fuzzy set. On the other hand, β-feasible solution is a vector belonging to the β-cut of the feasible region $\tilde{X}$. It is not difficult to show, that if all coefficients $\tilde{a}_{ij}$ and $\tilde{b}_i$ are crisp fuzzy quantities, i.e. they are isomorfic to the corresponding real numbers, then the fuzzy feasible region is isomorfic to the set of all feasible solutions of the corresponding classical LP problem, see [6], or [7].

5 Maximizing the Objective Function

Now we look for the "best" fuzzy quantities $\tilde{z}_q$ with respect to the given fuzzy constraints, or, in other words, with respect to the fuzzy set of feasible region of (7). Knowing the weights $v_q, q \in \mathcal{K}$, of the objectives we shall deal with the associated problem (9), particularly, with the single objective function $\tilde{z} = \tilde{c}_1x_1 + \cdots + \tilde{c}_nx_n$, where $\tilde{c}_j = \sum_{q \in \mathcal{K}} v_q \tilde{c}_{qj}, j \in \mathcal{N}$. We define special relations.

Let $\tilde{P}$ be a fuzzy relation on $\mathbf{R}$, let $\alpha \in (0, 1]$. Let $\tilde{a}, \tilde{b}$ be fuzzy quantities, we write

$$\tilde{a} \ \tilde{P}_\alpha \ \tilde{b}, \text{ if } \mu_{\tilde{P}}(\tilde{a}, \tilde{b}) \geq \alpha. \tag{11}$$

and call $\tilde{P}_\alpha$ the α-*relation on* $\mathbf{R}$ *associated to* $\tilde{P}$. We also write

$$\tilde{a} \ \tilde{P}^*_\alpha \ \tilde{b}, \text{ if } \tilde{a}\tilde{P}_\alpha\tilde{b} \text{ and } \mu_{\tilde{P}}(\tilde{b}, \tilde{a}) < \alpha, \tag{12}$$

and call $\tilde{P}^*_\alpha$ the *strict* α-*relation on* $\mathbf{R}$ *associated to* $\tilde{P}$.

Notice that $\tilde{P}_\alpha$ and $\tilde{P}^*_\alpha$ are binary relations on the set of fuzzy quantities $\mathcal{F}_0(\mathbf{R})$ being constructed from a fuzzy relation $\tilde{P}$ on the level $\alpha \in (0, 1]$. $\tilde{P}^*_\alpha$ is a strict relation to the relation $\tilde{P}_\alpha$.

If $\tilde{a}$ and $\tilde{b}$ are crisp fuzzy numbers corresponding to real numbers a and b, respectively, and $\tilde{P}$ is a fuzzy extension of relation $\leq$, then $a\tilde{P}_\alpha b$ if and only if $a \leq b$. Then for $\alpha \in (0, 1)$, $a\tilde{P}^*_\alpha \ b$ if and only if $a < b$.

Now, modifying the well known concept of efficient solution in multi-criteria optimization we define "maximization" (or "minimization") of the objective function of FLP problem (9). We allow for independent, i.e. different satisfaction levels: $\alpha \neq \beta$,where α is considered for the objective functions and β for the constraints, $\alpha, \beta \in (0,1]$.

Let $\tilde{c}_j$, $\tilde{a}_{ij}$ and $\tilde{b}_i$, $i \in \mathcal{M}$, $j \in \mathcal{N}$, be fuzzy quantities on $\mathbf{R}$. Let $\tilde{P}$ be a fuzzy relation on $\mathbf{R}$, being also a fuzzy extension of the usual binary relation $\leq$ on $\mathbf{R}$, let $\alpha, \beta \in (0,1]$. A β-feasible solution of (9) $x \in [\tilde{X}]_\beta$ is called the (α,β)-*maximal solution of* (9) if there is no $x' \in [\tilde{X}]_\beta$, $x \neq x'$, such that $\tilde{c}^T x \tilde{P}^*_\alpha \tilde{c}^T x'$. Here, $\tilde{P}^*_\alpha$ is the strict α-relation on $\mathbf{R}$ associated to $\tilde{P}$.

Clearly, if all coefficients of FLP problem (9) are crisp fuzzy quantities, then (α,β)-maximal solution of this problem is isomorphic to the classical Pareto-optimal solution of the corresponding LP problem (4).

6 Dual Problem and Duality Theorems

In this section we shall investigate the well known concept of duality in LP for FMOLP problems based on possibility and necessity fuzzy relations $\preceq^{Pos}$, and $\prec^{Nec}$. Here, we present some innovation of the weak and strong duality theorems which extend the known results for LP problems.

Consider the following FLP problem

$$\begin{array}{lll} & \text{"maximize"} & \tilde{z} = \tilde{c}_1 x_1 \tilde{+} \cdots \tilde{+} \tilde{c}_n x_n, \\ \text{(P)} & \text{"subject to"} & (\tilde{a}_{i1} x_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{in} x_n) \ \tilde{P} \ \tilde{b}_i, i \in \mathcal{M}, \\ & & x_j \geq 0, \quad j \in \mathcal{N}, \end{array} \tag{13}$$

which is an associated problem to FMOLP problem (7), where $\tilde{c}_j$ is defined by (8), $\tilde{a}_{ij}$ and $\tilde{b}_i$ are fuzzy quantities with membership functions $\mu_{\tilde{c}_j} : \mathbf{R} \to [0,1]$, $\mu_{\tilde{a}_{ij}} : \mathbf{R} \to [0,1]$ and $\mu_{\tilde{b}_i} : \mathbf{R} \to [0,1]$, $i \in \mathcal{M}$, $j \in \mathcal{N}$.

FLP problem (13) will be called the *primal FMOLP problem (P)*. The feasible region of (P) and (α,β)-maximal solution has been defined before. The *dual FMOLP problem (D)* can be formulated as follows

$$\begin{array}{lll} & \text{"minimize"} & \tilde{w} = \tilde{b}_1 y_1 \tilde{+} \cdots \tilde{+} \tilde{b}_m y_m \\ \text{(D)} & \text{"subject to"} & \tilde{c}_j \ \tilde{Q} (\tilde{a}_{1j} y_1 \tilde{+} \cdots \tilde{+} \tilde{a}_{mj} y_m), j \in \mathcal{N}, \\ & & y_i \geq 0, \quad i \in \mathcal{M}. \end{array} \tag{14}$$

Here, either $\tilde{P} = \preceq^{Pos}$, $\tilde{Q} = \prec^{Nec}$, or $\tilde{P} = \prec^{Nec}$, $\tilde{Q} = \preceq^{Pos}$. In problem (P), "maximization" is considered with respect to fuzzy relation $\tilde{P}$, in problem (D), "minimization" is considered with respect to fuzzy relation $\tilde{Q}$, which can be formulated analogically.

In the following duality theorems we present always two versions: (i) for fuzzy relation $\preceq^{Pos}$ in problem (P) and (ii) for fuzzy relation $\prec^{Nec}$in problem (P). In order to prove duality results we assume that the level of satisfaction α of the objective function is equal to the level of satisfaction β of the constraints, i.e. $\alpha = \beta$. Otherwise, the duality theorems in our formulation do not hold, see [9]. Moreover, we assume that each objective function is associated with a weight $w_q > 0, q \in \mathcal{K}$, such that $\sum_{q \in \mathcal{K}} w_q = 1$, where w_q may be interpreted as a relative importance of the q-th objective function. The corresponding proofs can be found in [9].

Theorem 2 Weak Duality Theorem. *Let $\tilde{c}_{qj}$, $\tilde{a}_{ij}$ and $\tilde{b}_i$ be fuzzy quantities, $q \in \mathcal{K}, i \in \mathcal{M}$ and $j \in \mathcal{N}$, $\alpha \in (0,1)$.*
(i) Let $\tilde{X}$ be a feasible region of FMOLP problem (13) with $\tilde{P} = \preceq^{Pos}$, $\tilde{Y}$ be a feasible region of FMOLP problem (14) with $\tilde{Q} = \prec^{Nec}$.
If a vector $x = (x_1, \ldots, x_n) \geq 0$ belongs to $[\tilde{X}]_\alpha$ and $y = (y_1, \ldots, y_m) \geq 0$ belongs to $[\tilde{Y}]_{1-\alpha}$, then

$$\sum_{j \in \mathcal{N}} \tilde{c}^{\mathrm{R}}_{qj}(\alpha) x_j \leq \sum_{i \in \mathcal{M}} \tilde{b}^{\mathrm{R}}_i(\alpha) y_i. \tag{15}$$

(ii) Let $\tilde{X}$ be a feasible region of FMOLP problem (13) with $\tilde{P} = \prec^{Nec}$, $\tilde{Y}$ be a feasible region of FMOLP problem (14) with $\tilde{Q} = \preceq^{Pos}$.
If a vector $x = (x_1, \ldots, x_n) \geq 0$ belongs to $[\tilde{X}]_{1-\alpha}$ and $y = (y_1, \ldots, y_m) \geq 0$ belongs to $[\tilde{Y}]_\alpha$, then

$$\sum_{j \in \mathcal{N}} \tilde{c}^{\mathrm{L}}_{qj}(\alpha) x_j \leq \sum_{i \in \mathcal{M}} \tilde{b}^{\mathrm{L}}_i(\alpha) y_i. \tag{16}$$

Notice that in case of single objective problem, (P1) and (D1) are classical dual LP problems and the same holds for (P2) and (D2).

Theorem 3 Strong Duality Theorem. *Let $\tilde{c}_{qj}$, $\tilde{a}_{ij}$ and $\tilde{b}_i$ be fuzzy quantities for all $q \in \mathcal{K}, i \in \mathcal{M}$ and $j \in \mathcal{N}$, let $w_q > 0$, for all $q \in \mathcal{K}$, such that $\sum_{q\in\mathcal{K}} w_q = 1$.*

(i) Let $\tilde{X}$ be a feasible region of FMOLP problem (13) with $\tilde{P} =\preceq^{Pos}$, $\tilde{Y}$ be a feasible region of FMOLP problem (14) with $\tilde{Q} =\prec^{Nec}$. If for some $\alpha \in (0,1)$, $[\tilde{X}]_\alpha$ and $[\tilde{Y}]_{1-\alpha}$ are nonempty, then there exists x^*- an (α,α) -maximal solution of FMOLP problem (P), and there exists y^* - an $(1-\alpha,1-\alpha)$-minimal solution of FMOLP problem (D) such that

$$\sum_{q\in\mathcal{K}}\sum_{j\in\mathcal{N}} w_q \tilde{c}^{\mathrm{R}}_{qj}(\alpha)x^*_j = \sum_{i\in\mathcal{M}} \tilde{b}^{\mathrm{R}}_i(\alpha)y^*_i. \tag{17}$$

(ii) Let $\tilde{X}$ be a feasible region of FMOLP problem (13) with $\tilde{P} =\prec^{Nec}$, $\tilde{Y}$ be a feasible region of FMOLP problem (14) with $\tilde{Q} =\preceq^{Pos}$. If for some $\alpha \in (0,1)$, $[\tilde{X}]_{1-\alpha}$ and $[\tilde{Y}]_\alpha$ are nonempty, then there exists x^*- an $(1-\alpha,1-\alpha)$-maximal solution of FLP problem (P), and y^* - an $(\alpha\ ,\alpha)$-minimal solution of FLP problem (D) such that

$$\sum_{q\in\mathcal{K}}\sum_{j\in\mathcal{N}} w_q \tilde{c}^{\mathrm{L}}_{qj}(\alpha)x^*_j = \sum_{i\in\mathcal{M}} \tilde{b}^{\mathrm{L}}_i(\alpha)y^*_i. \tag{18}$$

Remarks.

1. In the crisp and single-objective case, the above stated theorems are standard LP (Weak, Strong) Duality Theorems.

2. Usually, $\alpha \geq 0,5$. Then $[\tilde{X}]_\alpha \subset [\tilde{X}]_{1-\alpha}$, $[\tilde{Y}]_\alpha \subset [\tilde{Y}]_{1-\alpha}$, hence we can assume $x \in [\tilde{X}]_\alpha$ and $y \in [\tilde{Y}]_\alpha$. Evidently, the statement of the theorem remains unchanged.

3. Theorem 3 provides only the existence of the (α,α)-maximal solution (or $(1-\alpha,1-\alpha)$-maximal solution) of FMOLP problem (P), and $(1-\alpha,1-\alpha)$-minimal solution ((α,α)-minimal solution) of FMOLP problem (D) such that (17) or (18) holds. However, the proof of the theorem gives also the method for finding the solutions by solving (MO)LP problems (P1) and (D1).

7 Illustrative Example

In this section we discuss a simple illustrative example to clarify the introduced concepts and results, to provide some interpretation and features of possible applications. Last but not least, to solve the multi-objective FLP problem (P) by the single-objective FLP problem (D).

Let two new products A and B be manufactured. The manufacturing process is composed of two subprocesses, Processes 1 and 2. The estimated processing resources (e.g. processing time, materials) for manufacturing a batch of Product A for each process are the following: $\tilde{a}_{11}$ units for Process 1 and $\tilde{a}_{21}$ units for Process 2. On the other hand, the processing resources for manufacturing a batch of Product B for each process are as follows: $\tilde{a}_{12}$ units for Process 1, $\tilde{a}_{22}$ units at Process 2. The working resource for Process 1 is restricted by $\tilde{b}_1$ units, for Process 2 by $\tilde{b}_2$ units. The "profit" rates (1000 CZK/batch) of Products A and B are estimated as $\tilde{c}_{11}$ and $\tilde{c}_{12}$, respectively. The "utility" rates (1000 CZK/batch) of Products A and B are estimated as $\tilde{c}_{21}$ and $\tilde{c}_{22}$, respectively. The weights of the criteria are $w_1 = 0,6$ and $w_2 = 0,4$. All mentioned parameters $\tilde{a}_{ij}$, $\tilde{b}_i$ and $\tilde{c}_{qj}$ are subjected to uncertainty and they are expressed by fuzzy quantities. We shall investigate what quantity of Products A and B should be manufactured in order to "maximize" the total "profit" and total "utility". For this purpose we formulate the following FMOLP problem (primal problem)

$$\begin{array}{lll} & \text{"maximize"} & \tilde{z}_1 = \tilde{c}_{11}x_1 \tilde{+} \tilde{c}_{12}x_2, \\ & & \tilde{z}_2 = \tilde{c}_{21}x_1 \tilde{+} \tilde{c}_{22}x_2 \\ \text{(PE)} & \text{"subject to"} & (\tilde{a}_{11}x_1 \tilde{+} \tilde{a}_{12}x_2)\ \tilde{P}\ \tilde{b}_1, \\ & & (\tilde{a}_{21}x_1 \tilde{+} \tilde{a}_{22}x_2)\ \tilde{P}\ \tilde{b}_2, \\ & & x_1, x_2 \geq 0, \end{array} \tag{19}$$

where $\tilde{c}_{qj} = (c^L_{qj}, c_{qj}, c^R_{qj})$, $\tilde{a}_{ij} = (a^L_{ij}, a_{ij}, a^R_{ij})$ and $\tilde{b}_i = (b^L_i, b_i, b^R_i)$ are triangular fuzzy quantities (with triangular piecewise linear membership functions) given by the triples, as usual. Here, we are given the following triangular fuzzy quantities

$$\begin{array}{ll} \tilde{c}_{11} = (3,4,5), & \tilde{c}_{12} = (2,4,6), \\ \tilde{c}_{21} = (2,3,4), & \tilde{c}_{21} = (3,4,5), \\ \tilde{a}_{11} = (1,3,5), & \tilde{a}_{12} = (1,1,1), \\ \tilde{a}_{21} = (1,3,5), & \tilde{a}_{22} = (3,3,3), \\ \tilde{b}_1 = (8,11,14), & \tilde{b}_2 = (11,12,15). \end{array} \tag{20}$$

Notice that $\tilde{a}_{12}$ and $\tilde{a}_{22}$ are crisp fuzzy numbers.

Here, $\tilde{P}$ and $\tilde{Q}$ is a pair of dual fuzzy relations, particularly $\tilde{P} = \preceq^{Pos}$ and $\tilde{Q} = \prec^{Nec}$, see (1), (2), (3).

Given $\alpha, \beta \in (0,1)$, $\alpha = \beta$, we obtain the following couple of dual problems, see [9]:

$$\begin{array}{ll} \text{maximize} & z_1 = (5-\alpha)x_1 + (6-2\alpha)x_2, \\ & z_2 = (4-\alpha)x_1 + (5-\alpha)x_2, \\ \text{subject to} & (1+2\alpha)x_1 + x_2 \leq 14 - 3\alpha, \\ & (1+\alpha)x_1 + 3x_2 \leq 15 - 3\alpha, \\ & x_1, x_2 \geq 0, \end{array} \tag{21}$$

$$\begin{array}{ll} \text{minimize} & w = (14-3\alpha)y_1 + (15-3\alpha)y_2, \\ \text{subject to} & (1+2\alpha)y_1 + (1+\alpha)y_2 \geq 5 - \alpha, \\ & y_1 + \quad 3y_2 \geq 6 - 2\alpha, \\ & (1+2\alpha)y_1 + (1+\alpha)y_2 \geq 4 - \alpha, \\ & y_1 + \quad 3y_2 \geq 5 - \alpha, \\ & y_1, y_2 \geq 0. \end{array} \tag{22}$$

Suppose $\alpha = \beta = 0,7$ is an appropriate level of satisfaction (degree of satisfaction or, necessity degree) for the objective function and for the constraints. By Simplex method we obtain the following numerical results. The optimal solutions of problems (21), (22), i.e. "optimistic - pessimistic" dual couple are displayed in Table 1.

Table 1.

$\tilde{P} =$	$\preceq^{Pos}$		$\tilde{Q} =$	$\prec^{Nec}$
$x_1^* =$	4,15		$y_1^* =$	0,74
$x_2^* =$	1,95		$y_2^* =$	1,25
$z^* =$	24,91		$w^* =$	24,91

The optimal solutions of problems with $\tilde{P} = \prec^{Nec}$ and $\tilde{Q} = \preceq^{Pos}$, i.e. the "pessimistic - optimistic" dual couple are displayed in Table 2.

Table 2.

$\tilde{P} =$	$\prec^{Nec}$		$\tilde{Q} =$	$\preceq^{Pos}$
$x_1^{**} =$	2,19		$y_1^{**} =$	0,39
$x_2^{**} =$	2,22		$y_2^{**} =$	1,00
$z^{**} =$	15,65		$w^{**} =$	15,65

As is evident from Table 1, the value $z^* = 24{,}91$ of the optimal solution of the "optimistic" primal problem is greater than the value $z^{**} = 15{,}65$ of the optimal solution of the "pessimistic" primal one. This result is in a correspondence with our expectation. By Strong Duality Theorem $x^* = (4,15; 1,95)$ is a (0,7;0,7)-maximal solution of FLP problem (PE), and $y^* =$(0,74;1,25) is a (0,3;0,3)-minimal solution of FLP problem (DE) such that (17) holds, i.e. $z^* = w^*$. Moreover, $y^* =$(0,74;1,25) is a vector of dual (shadow) prices of the resources $\tilde{b}_i$ at disposition. The vector y^* is a $(1-\alpha, 1-\alpha)$-minimal solution of the "pessimistic" dual problem with the meaning that the smallest value of $\tilde{a}_{1j}y_1 \tilde{+} \tilde{a}_{2j}y_2$ with the degree of satisfaction at least $1-\alpha$, is less or equal to the largest value of $\tilde{c}_j$ with the degree of satisfaction at least $1 - \alpha = 0{,}3$.

Analogical explanation could be formulated for the other dual couple (PE) and (DE) with $\tilde{P} = \prec^{Nec}$ and $\tilde{Q} = \preceq^{Pos}$, i.e. for "pessimistic - optimistic" dual couple.

8 Conclusion

It is possible to investigate duality in FLP problems even in more general settings. There exist several ways of generalization. For instance, it is possible to extend the duality results to some other classes of fuzzy relations, or, to find some necessary conditions that fuzzy relations for comparing fuzzy numbers should satisfy in order to provide a duality result, or, eventually a duality gap. Moreover, in [7], the concept of dual couples of t-norms and t-conorms has been formulated and dual fuzzy relations have been defined. The role of dual relations in the couple of dual FLP problems should be also clarified and a more general duality theory could be derived. The other way of generalization is based on introducing interactive fuzzy coefficients, or oblique fuzzy vectors, see e.g. [7].

References

[1] G.B. Dantzig, *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963.

[2] D. Dubois and H. Prade, Ranking fuzzy numbers in the setting of possibility theory. Inform. Sci. 30, (1983) 183 - 224.

[3] M. Ehrgott, *Multicriteria optimization*. Springer, Lecture Notes in Economics and Math. Systems 491, Berlin-Heidelberg 2000.

[4] H. Hamacher, H. Leberling and H.-J. Zimmermann, Sensitivity analysis in fuzzy linear programming. *Fuzzy Sets and Systems* 1 (1978), 269-281.

[5] M. Inuiguchi, H. Ichihashi and Y. Kume, Some properties of extended fuzzy preference relations using modalities. Inform. Sci. 61, (1992) 187 - 209.

[6] M. Inuiguchi, J. Ramik, T. Tanino and M. Vlach, Satisficing solutions and duality in interval and fuzzy linear programming. *Fuzzy Sets and Systems* 135 (2003), 151-177.

[7] J. Ramik and M. Vlach, *Generalized Concavity in Fuzzy Optimization and Decision Analysis*. Kluwer Acad. Publ., Dordrecht - Boston - London, 2002.

[8] J. Ramik, Duality in Fuzzy Linear Programming: Some New Concepts and Results. *Fuzzy Optimization and Decision Making,* Vol.4, (2005), 25-39.

[9] J. Ramik, Duality in Fuzzy Linear Programming with Possibility and Necessity Relations. *Fuzzy Sets and Systems,* to appear.

[10] W. Rodder and H.-J. Zimmermann, Duality in fuzzy linear programming. *Extremal Methods and System Analysis*, A.V. Fiacco and K.O. Kortanek, Eds., Berlin - New York, 1980, 415-429.

Analysis of change point in the Cox regression model with application to unemployment data

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

Abstract

The paper deals with the Cox regression model for hazard function and considers the case of abrupt change of the effect of some covariates. The estimation procedure is based on the maximum likelihood approach. The consistent estimators of all parameters are provided with their asymptotic distributions. The method is adapted to the analysis of real monthly observed unemployment data from 1/1998 - 6/2003 modeled as a discrete-time series of Poisson counts.

Keywords

Change-point; Cox regression model; Hazard rate; Maximum likelihood estimator

1 Introduction

Suppose we are interested in the analysis of a quantity that can generally be expressed as the time to some event. Such a data can be modeled by intensity of the event. If there is an evidence of influenced covariates, we can use with advantage the Cox regression model. The important role is played by a hazard function $\lambda(t)$, which is defined conditionally on a vector of covariates as the instantaneous failure rate function, since it describes the immediate risk of failure of an individual at time t, given that it has not failure up to that time. We consider the possibility of abrupt change of the regression part of hazard function at an unknown time point.

Let T be an absolutely continuous nonnegative random variable which represents time to the event and Z be a vector of covariates. The hazard function defined in (1) has the form (2) in the case of classical Cox regression model, for more details see [2].

$$\lambda(t \mid Z) = \lim_{\Delta t \to 0+} \frac{P(T \le t + \Delta t \mid T \ge t, Z(s), 1 \le s \le t)}{\Delta t} \tag{1}$$

$$= \lambda_0(t) \exp\{\beta^T Z(t)\} \tag{2}$$

We will follow change point model defined in [4] with additional assumption of constant baseline hazard $\lambda_0(t) = 1$ for all t. We get hazard function on $< 0, 1 >$ with change point at time τ

$$\lambda(t \mid Z) = \exp\{\alpha^T Z_1(t) + (\beta + \gamma I(t > \tau))^T Z_2(t)\} = \exp\{r_\theta^T(t) Z(t)\}, \qquad t \in < 0, 1 >, \tag{3}$$

where $Z = (Z_1^T, Z_2^T)^T$, $\theta = (\tau, \alpha^T, \beta^T, \gamma^T)^T, r_\theta(t) = (\alpha^T, \beta^T + \gamma^T I(t > \tau))^T$ and $I(.)$ is the indicator function. Covariate Z_1 is a p-dimensional and Z_2 is a q-dimensional left-continuous process with right-hand limits, θ is a vector of unknown parameters. The hazard rate of failure depends on covariate Z_1 through parameter α, on covariate Z_2 through parameter β till the time τ and then through parameters β and γ.

In the classical Cox model, the asymptotic properties of the regression coefficients have been considered by [1]. In the Poisson processes with a change point in a hazard function of the form

[1] This work was supported by the Grant Agency of the Czech Republic (GACR 402/04/1294).

$\lambda(t) = \lambda_1 + \lambda_2 I(t > \tau)$, the asymptotic properties of estimators have been proved by [6]. In the Cox regression change point model, the asymptotic properties of all parameters have been well discussed in [4]. In our case it was necessary to derive a modified theory for special form of Cox regression model, based on the standard likelihoood function instead of partial likelihood function. The rest of the paper is organized as follows. At first we propose procedure for estimating the parameters of the model, then we derive a global consistency and asymptotic distribution of the estimators. Application to the real data is included in the end of the paper.

2 Estimation of model parameters

Let us have a sample of n individuals from population observed in the time interval $< 0, 1 >$ then $(T_i), 1 \leq i \leq n$, be the times to the event of each individual and $(Z_i), 1 \leq i \leq n$, be the related i.i.d. covariate sample. Let $\xi = (\alpha^T, \beta^T, \gamma^T)^T$ and let $\xi_0 = (\alpha_0^T, \beta_0^T, \gamma_0^T)^T$ and τ_0 be the actual parameter values. Let $\theta = (\tau, \xi^T)^T$ and $\theta_0 = (\tau_0, \xi_0^T)^T$ belong to $\Theta = (0, 1) \times \Xi$, where $\Xi = A \times B \times C$, $\alpha \in A \subset \mathbb{R}^p$, $\beta \in B \subset \mathbb{R}^q$, $\gamma \in C \subset \mathbb{R}^q$ and A, B, C be open and bounded convex sets. We suppose $\gamma_0 \neq 0$ and $\tau_0 \neq \{0, 1\}$, so that the change point exists. Let $P_{\theta,H}$ be the probability distribution of the variables (T_i, Z_i) for Z_i having a distribution H. Under the true parameter values, we denote $P \equiv P_{\theta_0,H_0}$ and $\mathbb{E}$ the expectation of variables.

Then similarly as [4] we obtain the following log-likelihood function:

$$l_n(\theta) = \sum_{i=1}^{n} r_\theta^T(T_i) Z_i(T_i) - \sum_{j=1}^{n} Y_j(T_i) \exp\{r_\theta^T(T_i) Z_j(T_i)\} = \sum_{i=1}^{n} r_\theta^T(T_i) Z_i(T_i) - S_n^{(0)}(T_i; \theta), \quad (4)$$

where $Y_i(t) = I(T_i \geq t)$ indicates whether individual i is still under the risk of event at time t and $S_n^{(0)}(t; \theta) = \sum_{i=1}^{n} Y_i(t) \exp\{r_\theta^T(t) Z_i(t)\}$. For fixed τ we have $\hat{\xi}_n(\tau) = \operatorname{argmax}_\Xi l_n(\tau, \xi)$, then we can define $l_n(\tau) = l_n(\tau, \hat{\xi}_n(\tau))$. As l_n is a discontinuous function of τ, the maximum of l_n may not be unique, we consider estimator

$$\hat{\tau}_n = \inf \left\{ \tau \in (0, 1) : \max\{l_n(\tau), l_n(\tau^+)\} = \sup_{\tau \in (0,1)} l_n(\tau) \right\},$$

where $l_n(\tau^+)$ is the right-hand limit of l_n at τ. Then the maximum likelihood estimator is defined $\hat{\xi}_n = \hat{\xi}_n(\hat{\tau}_n)$ and $\hat{\theta}_n = (\hat{\tau}_n, \hat{\xi}_n)$.

3 Global consistency of the estimators

At first, let us define some helpful processes and functions on $< 0, 1 > \times \mathbb{R}^{p+q}$ as [4]:

$$\begin{aligned}
S_n^{(1)}(t; \theta) &= \sum_{i=1}^{n} Y_i(t) Z_i(t) \exp\{r_\theta^T(t) Z_i(t)\}, \\
S_n^{(2)}(t; \theta) &= \sum_{i=1}^{n} Y_i(t) Z_i(t) Z_i(t)^T \exp\{r_\theta^T(t) Z_i(t)\}, \\
s^{(0)}(t; \theta) &= \mathbb{E}[Y_i(t) \exp\{r_\theta^T(t) Z_i(t)\}], \\
s^{(1)}(t; \theta) &= \mathbb{E}[Y_i(t) Z_i(t) \exp\{r_\theta^T(t) Z_i(t)\}], \\
s^{(2)}(t; \theta) &= \mathbb{E}[Y_i(t) Z_i(t) Z_i(t)^T \exp\{r_\theta^T(t) Z_i(t)\}].
\end{aligned}$$

Now we will study behavior of processes

$$\begin{aligned}
X_n(\theta) &= n^{-1}\{l_n(\theta) - l_n(\theta_0)\} \qquad \text{and} \\
X(\theta) &= \int_0^1 \left\{ (r_\theta(t) - r_{\theta_0}(t))^T \frac{s^{(1)}(t; \theta_0)}{s^{(0)}(t; \theta_0)} - (s^{(0)}(t; \theta) - s^{(0)}(t; \theta_0)) \right\} \mathrm{d}N(t),
\end{aligned}$$

where $N(t) = \sum_{i=1}^{n} N_i(t)$ is the number of events observed till the time t in the whole population and $N_i(t) = I(T_i \le t)$. The process X_n is written as a sum $X_n = X_{1n} + X_{2n} + X_{3n}$ according to the sign of $\tau - \tau_0$, where

$$\begin{aligned}
X_{1n}(\theta) &= n^{-1}\sum_{i=1}^{n} I(T_i \le \tau \wedge \tau_0)\{(\alpha-\alpha_0)^T Z_{1i}(T_i) + (\beta-\beta_0)^T Z_{2i}(T_i) - \\
&\quad -(S_n^{(0)}(T_i;\alpha,\beta) - S_n^{(0)}(T_i;\alpha_0,\beta_0))\}, \\
X_{2n}(\theta) &= n^{-1}\sum_{i=1}^{n} I(T_i \ge \tau \vee \tau_0)\{(\alpha-\alpha_0)^T Z_{1i}(T_i) + (\beta-\beta_0+\gamma-\gamma_0)^T Z_{2i}(T_i) - \\
&\quad -(S_n^{(0)}(T_i;\alpha,\beta+\gamma) - S_n^{(0)}(T_i;\alpha_0,\beta_0+\gamma_0))\}, \\
X_{3n}(\theta) &= n^{-1}\sum_{i=1}^{n} I(\tau < T_i \le \tau_0)\{(\alpha-\alpha_0)^T Z_{1i}(T_i) + (\beta-\beta_0+\gamma)^T Z_{2i}(T_i) - \\
&\quad -(S_n^{(0)}(T_i;\alpha,\beta+\gamma) - S_n^{(0)}(T_i;\alpha_0,\beta_0))\} + \\
&+ n^{-1}\sum_{i=1}^{n} I(\tau_0 < T_i \le \tau)\{(\alpha-\alpha_0)^T Z_{1i}(T_i) + (\beta-\beta_0-\gamma_0)^T Z_{2i}(T_i) - \\
&\quad -(S_n^{(0)}(T_i;\alpha,\beta) - S_n^{(0)}(T_i;\alpha_0,\beta_0+\gamma_0))\}.
\end{aligned}$$

Using the same arguments as [1] and using the similar conditions as [4], p.104, processes X_{kn} converge in probability and uniformly on paramatric space to the functions X_k defined by

$$\begin{aligned}
X_1(\theta) &= \int_0^{\tau\wedge\tau_0} \{((\alpha-\alpha_0)^T,(\beta-\beta_0)^T)\frac{s^{(1)}(t;\alpha_0,\beta_0)}{s^{(0)}(t;\alpha_0,\beta_0)} - (s^{(0)}(t;\alpha,\beta) - s^{(0)}(t;\alpha_0,\beta_0))\}\mathrm{d}N(t), \\
X_2(\theta) &= \int_{\tau\vee\tau_0}^{1} \{((\alpha-\alpha_0)^T,(\beta-\beta_0+\gamma-\gamma_0)^T)\frac{s^{(1)}(t;\alpha_0,\beta_0+\gamma_0)}{s^{(0)}(t;\alpha_0,\beta_0+\gamma_0)} - \\
&\quad -(s^{(0)}(t;\alpha,\beta+\gamma) - s^{(0)}(t;\alpha_0,\beta_0+\gamma_0))\}\mathrm{d}N(t), \\
X_3(\theta) &= I(\tau<\tau_0)\int_{\tau}^{\tau_0} \{((\alpha-\alpha_0)^T,(\beta-\beta_0+\gamma)^T)\frac{s^{(1)}(t;\alpha_0,\beta_0)}{s^{(0)}(t;\alpha_0,\beta_0)} - \\
&\quad -(s^{(0)}(t;\alpha,\beta+\gamma) - s^{(0)}(t;\alpha_0,\beta_0))\}\mathrm{d}N(t) + \\
&+ I(\tau_0<\tau)\int_{\tau_0}^{\tau_0} \{((\alpha-\alpha_0)^T,(\beta-\beta_0-\gamma_0)^T)\frac{s^{(1)}(t;\alpha_0,\beta_0+\gamma_0)}{s^{(0)}(t;\alpha_0,\beta_0+\gamma_0)} \\
&\quad -(s^{(0)}(t;\alpha,\beta) - s^{(0)}(t;\alpha_0,\beta_0+\gamma_0))\}\mathrm{d}N(t).
\end{aligned}$$

Similarly as [4], there is possible to show that the maximum value of the function $X = X_1 + X_2 + X_3$ is zero and it can not be reached if $\tau \neq \tau_0$. It follows that $\mid X(\hat\tau_n, \hat\xi_n) \mid$ converges in probability to zero. We get the weak consistency of $\hat\tau_n$ and $\hat\xi_n$ from the fact that X has a unique maximum at θ_0, with value zero.

Now we will briefly mention inference for the rates of convergence of the estimators. We follow the approach of [3] based on the local asymptotic behaviour of the likelihood ratio process for change point in nonhomogenous Poisson processes. By investigation of the asymptotic behaviour of the process $u \longmapsto \{l_n(\theta_{n,u}) - l_n(\theta_0)\}$, where $\theta_{n,u} = (\tau_0 + n^{-1}u_1, \xi_0 + n^{-1/2}u_2)$ and $u = (u_1, u_2)$, we get the process bounded in probability, which provides the convergence rates of $\hat\tau_n$ and $\hat\xi_n$. Let $\mathcal{U}_{n,\varepsilon} = \{u : \parallel \theta_{n,u} - \theta_0 \parallel \le \varepsilon\}$ then for $\varepsilon > 0$ sufficiently small

$$\begin{aligned}
\lim_{n\to\infty, A\to\infty} P(\sup_{u\in\mathcal{U}_{n,\varepsilon}, \|u\|>A} X_n(\theta_{n,u}) \ge 0) &= 0, \\
\limsup_{n\to\infty, A\to\infty} P(n \mid \hat\tau_n - \tau_0 \mid > A) &= 0, \\
\limsup_{n\to\infty, A\to\infty} P(n^{1/2} \parallel \hat\xi_n - \xi_0 \parallel > A) &= 0.
\end{aligned}$$

4 Asymptotic distribution of the estimators

In order to obtain asymptotic properties we study the limiting distribution of $\{l_n(\theta_{n,u}) - l_n(\theta_0)\}$ on sufficiently large compact sets. From previous section we derived

$$\hat{u}_n = (n(\hat{\tau}_n - \tau_0), n^{1/2}(\hat{\xi}_n - \xi_0)) = \mathrm{argmax}_u\{l_n(\theta_{n,u}) - l_n(\theta_0)\}.$$

As $\{l_n(\theta_{n,u}) - l_n(\theta_0)\}$ is not continuous with respect to the first component of u we concentrate all jumps to a process Q_n. Now as in the previous section we rewrite $\{l_n(\theta_{n,u}) - l_n(\theta_0)\}$ into the sum of three terms:

$$\begin{aligned}
l_{1n}(\theta_{n,u}) &= \sum_{i=1}^{n} I(T_i \le \tau_{n,u})[n^{-1/2}\{n^{1/2}(\alpha_{n,u} - \alpha_0)^T Z_{1i}(T_i) + n^{1/2}(\beta_{n,u} - \beta_0)^T Z_{2i}(T_i)\} - \\
&\quad -(S_n^{(0)}(T_i; \alpha_{n,u}, \beta_{n,u}) - S_n^{(0)}(T_i; \alpha_0, \beta_0))], \\
l_{2n}(\theta_{n,u}) &= \sum_{i=1}^{n} I(T_i > \tau_{n,u})[n^{-1/2}\{n^{1/2}(\alpha_{n,u} - \alpha_0)^T Z_{1i}(T_i) + n^{1/2}(\beta_{n,u} - \beta_0 + \gamma_{n,u} - \gamma_0)^T \\
&\quad Z_{2i}(T_i)\} - (S_n^{(0)}(T_i; \alpha_{n,u}, \beta_{n,u} + \gamma_{n,u}) - S_n^{(0)}(T_i; \alpha_0, \beta_0 + \gamma_0))], \\
l_{3n}(\theta_{n,u}) &= Q_n(u_1) = Q_n^-(u_1) - Q_n^+(u_1),
\end{aligned}$$

where Q_n^- and Q_n^+ are the processes defined on $\mathbb{R}$ by

$$\begin{aligned}
Q_n^-(u_1) &= \sum_{i=1}^{n} I(\tau_0 + n^{-1}u_1 < T_i \le \tau_0)\{\gamma_0^T Z_{2i}(T_i) - (S_n^{(0)}(T_i; \alpha_0, \beta_0 + \gamma_0) - S_n^{(0)}(T_i; \alpha_0, \beta_0))\} \\
&\quad \text{for} \quad u_1 < 0 \qquad \text{and} \quad Q_n^-(u_1) = 0 \quad \text{for} \quad u_1 \ge 0, \\
Q_n^+(u_1) &= \sum_{i=1}^{n} I(\tau_0 < T_i \le \tau_0 + n^{-1}u_1)\{\gamma_0^T Z_{2i}(T_i) + (S_n^{(0)}(T_i; \alpha_0, \beta_0) - S_n^{(0)}(T_i; \alpha_0, \beta_0 + \gamma_0))\} \\
&\quad \text{for} \quad u_1 > 0 \qquad \text{and} \quad Q_n^+(u_1) = 0 \quad \text{for} \quad u_1 \le 0.
\end{aligned}$$

Let us describe the asymptotic distribution of the process Q_n. At first, similarly as [4], we need to define some helpful functions and processes:

$$\begin{aligned}
d_0(t) &= s^{(0)}(t; \alpha_0, \beta_0) - s^{(0)}(t; \alpha_0, \beta_0 + \gamma_0), \\
p_0^+(t; Z_2) &= e^{(\beta_0+\gamma_0)^T Z_2(t)}\mathbb{E}\{Y(t)e^{\alpha_0^T Z_1(t)} \mid Z_2(t)\}, \\
p_0^-(t; Z_2) &= e^{\beta_0^T Z_2(t)}\mathbb{E}\{Y(t)e^{\alpha_0^T Z_1(t)} \mid Z_2(t)\}.
\end{aligned}$$

Let ν^+ and ν^- be independent real jump processes which satisfy $\nu^+(s) = 0$ for $s < 0$, $\nu^+(s)$ be a Poisson variable with parameter $p_0^+(\tau_0; Z_2)s$ for $s > 0$ and $\nu^-(s) = 0$ for $s > 0$, $\nu^-(s)$ be a Poisson variable with parameter $-p_0^-(\tau_0; Z_2)s$ for $s < 0$. Let $(V_k^+)_{k\ge1}$ and $(V_k^-)_{k\ge1}$ be independent sequences of i.i.d. variables with the conditional characteristic functions

$$\begin{aligned}
\varphi^+(t; Z_2) &\equiv \mathbb{E}\{e^{itV_k^+} \mid Z_2(\tau_0^+)\} = \exp\{it\gamma_0^T Z_2(\tau_0^+)\}, \\
\varphi^-(t; Z_2) &\equiv \mathbb{E}\{e^{itV_k^-} \mid Z_2(\tau_0)\} = \exp\{it\gamma_0^T Z_2(\tau_0)\},
\end{aligned}$$

and let $(V_k^+)_{k\ge1}$ be independent of ν^+ conditionally on $Z_2(\tau_0^+)$ and of ν^-, $(V_k^-)_{k\ge1}$ be independent of ν^- conditionally on $Z_2(\tau_0)$ and of ν^+. Finally, let $Q = Q^- - Q^+$ be the right-continuous jump process defined on $\mathbb{R}$ by

$$Q^+(s) = \nu^+(s)d_0(\tau_0) + \sum_{k=1}^{\nu^+(s)} V_k^+, \qquad Q^-(s) = \nu^-(s)d_0(\tau_0) + \sum_{k=1}^{\nu^-(s)} V_k^-$$

and let $\hat{s}_Q = \inf\{s; Q(s) = \max Q\}$. We will define the matrix I by

$$I = \begin{pmatrix} s^{(2)}(1;\theta_0) & s_q^{(2)}(1;\theta_0) - s_q^{(2)}(\tau_0;\theta_0) \\ (s_q^{(2)}(1;\theta_0))^T - (s_q^{(2)}(\tau_0;\theta_0))^T & s_{qq}^{(2)}(1;\theta_0) - s_{qq}^{(2)}(\tau_0;\theta_0) \end{pmatrix},$$

where $s_q^{(2)}(\tau_0;\theta_0)$ and $s_{qq}^{(2)}(\tau_0;\theta_0)$ are, respectively, the matrix made of the last q rows of $s^{(2)}(\tau_0;\theta_0)$ and its right lower $q \times q$ sub-matrix. Then there is possible to derive asymptotic properties of the estimators. Under similar conditions as [4], p.104, $n(\hat{\tau}_n - \tau_0)$ converges weakly to $\hat{s}_Q$, $n^{1/2}(\hat{\xi}_n - \xi_0)$ converges weakly to the Gaussian variable $\mathcal{N}_{p+q}(0, I^{-1})$ and they are asymptotically independent.

The asymptotic confidence intervals for the components of ξ_0 in model (3) are the same as in the model with a change point at a known time τ_0, so we can maximise the log-likelihood function (4) with respect to the parameter ξ for successive values τ_k on a grid in $(0, 1)$.

5 Analysis of unemployment data in the Czech Republic

For the numerical example we use grouped real unemployment data of population in the Czech Republic from the Czech Statistical Office. Data contain increases of unemployed people during 66 month, from 1/1998 - 6/2003. They are modeled as discrete-time series of Poisson counts (similarly as [5]), with piecewise log-linear trend of intensity. The whole analysis was computed in Matlab software version 6.0.

We suppose for $t = 1, \ldots, 66$ the data follow the model

$$N_t \sim Poiss\{N_0 \lambda(t \mid Z)\} \quad \text{and} \quad \lambda(t \mid Z) = \exp\{a + (b + c\mathrm{I}(t > \tau))t + \beta Z\},$$

where a, b, c, β, τ are unknown parameters, covariates Z are indicators of seasons (i.e. months January, ..., November) and N_0 is the average value of the monthly increases of unemployed people. The parameters of the model are estimated by the ML approach on grid for the parameter τ. The maximum of log-likelihood function is reached for $\tau = 9$. Our aim was to find the model where all parameters are statistically significant. That is why we eliminated parameter which corresponded to the month August. It means that this parameter is set to zero and have the same relative unemployment risk as the month December. New model has all parameters statistically significant. Following Table 1 shows the estimated parameters a, b, c, β with (asymptotic) 95 % confidence intervals.

parameters	estimates	confidence intervals	
a	-0.1980	-0.2047	-0.1913
b	0.0353	0.0346	0.0361
c	-0.0368	-0.0376	-0.0360
January	0.3597	0.3557	0.3637
February	-0.1193	-0.1239	-0.1147
March	-0.0955	-0.1001	-0.0909
April	-0.0913	-0.0959	-0.0868
May	-0.1662	-0.1709	-0.1616
June	0.0232	0.0188	0.0275
July	0.1961	0.1917	0.2004
September	0.2933	0.2891	0.2975
Octomber	0.0179	0.0133	0.0225
November	-0.0212	-0.0259	-0.0166

Figure 1 displays the ratio N_t/N_0 (as original data), the estimated hazard function $\hat{\lambda}(t \mid Z)$ (as a model) and linear trend with change point, $\exp\{\hat{a} + (\hat{b} + \hat{c}\mathrm{I}(t > \hat{\tau}))t\}$. There is possible to see quite good fit of the model to the data.

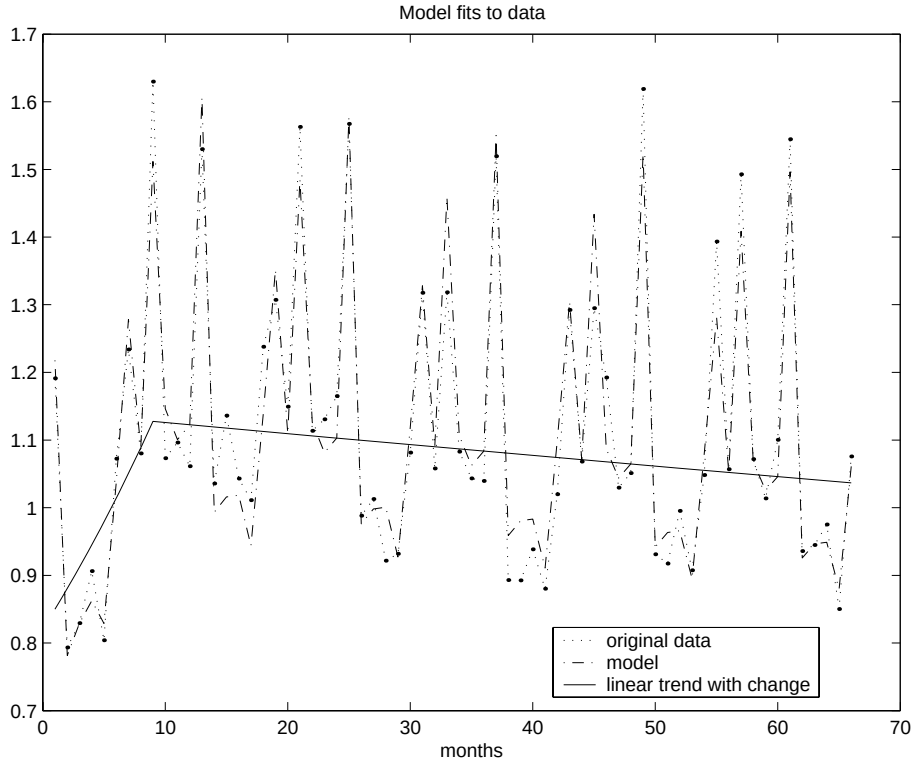


Figure 1

There are several interpretations of the change in the intensity of becoming unemployed in September 1998. Decreasing unemployment can be caused by change of the government, because it stabilized political and also economical situation in the Czech Republic. Another explanation can be connected with the economic background, there should be an upper boundary for possible amount of unemployed people in situation of relatively stable economy.

Relevant theory of change point suggests also possibility to derive the likelihood ratio test for a change point. It means to test the null hypotheses $H_0 : \gamma_0 = 0$ versus the alternative $H_1 : \gamma_0 \neq 0$. Building asymptotic confidence intervals for τ_0 is quite interesting theoretical and computational problem. It involves studying the behaviour of Q on compact sets and bootstrap confidence intervals with a resampling of the individuals could be considered. All these problems are subjects of author's future interest.

References

[1] Andersen,P.K. and Gill,R.D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, 10 : 1100-1120.

[2] Cox,D.R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B*, 34 : 187-220.

[3] Kutoyants,Y.A. (1998). Statistical Inference for Spatial Poisson Processes *Statistical Inference for Spatial Poisson Processes, Lecture Notes in Statistics*. Springer, New York.

[4] Pons, O. (2002). Estimation in a Cox regression model with a change-point at an unknown time. *Statistics*, 36(2) : 101-124.

[5] Volf, P. (2003). Cox's Regression Models for Dynamics of Grouped Unemployment Data. *Bulletin of the Czech Econometric Society*, Volume 10, Issue 19 : 151-162.

[6] Yao, Y. (1986) Maximum likelihood estimation in hazard rate models with a change-point. *Common. Statist. - Theoret. Methods*, 15 : 2455-2466. Issue 19 : 151-162.

Are Reduced Forms of Dornbusch Monetary Model Really Reduced?

Eva Rublíková
Katedra štatistiky FHI
Ekonomická univerzita
Bratislava

Abstract

Slovakia and Poland belong to the countries that joined the European Union in May 2004. The both countries are in the perspective of introducing Euro in 2009. In these circumstances it is highly welcome to carefully examine what are the determinants of exchange rates in Poland and in Slovakia.
Since we can observe systematic increase in trade between Poland and Slovakia it seems to be necessary to study the influence that these economies have on each other. In this article we present the results of estimates made for Poland and Slovakia. We would like to examine whether there is a correlation between monetary policy in these countries and the exchange rate of Slovak crown against Polish zloty.

Keywords

Exchange rates, monetary reduced form model, purchasing power parity, Dickey-Fuller test.

1 Dornbusch Reduced Forms of Monetary Model

In the study of floating exchange rates it is desire to find acceptable model that explains the movement of the nominal spot exchange rate in terms of other macroeconomic variables. The theoretical models are based on the monetary approach.
A two country model is considered with identical structural parameters in the domestic and foreign countries; domestic and foreign assets are assumed to be perfect substitutes, so that risk consideration is avoided. Domestic money and real outputs are assumed to be exogenous. Domestic and foreign goods are not considered as perfect substitutes and goods prices adjust to a new equilibrium with lag. The lagged response is due to costs adjustment, or due to lack of complete information. However, the maintenance of long-run equilibrium PPP implies that the long-run effect of a change in money supply is identical to that of monetary model.
Dornbusch monetary model is formulated by following equations [1]:

$$(m_t - m_t^*) - (p_t - p_t^*) = (k - k^*) + \phi(y_t - y_t^*) - \lambda(r_t - r_t^*) + \varepsilon_{1t} \quad (1)$$

$$(d_t - d_t^*) = \gamma(y_t - y_t^*) - \sigma(r_t - r_t^*) - \omega(s_t - p_t - p_t^*) \quad (2)$$

$$(p_t - p_t^*) - (p_{t-1} - p_{t-1}^*) = \delta\left[(d_t - d_t^*) - \lambda(y_t - y_t^*)\right] + \varepsilon_{2t} \quad (3)$$

$$(r_t - r_t^*) = E_t s_{t+1} - s_t + \varepsilon_{3t} \quad (4)$$

$$\bar{s}_t = \bar{p}_t - \bar{p}_t^* \quad (5)$$

$$E_t s_{t+1} - s_t = \alpha(\bar{s}_t - s_t) \quad (6)$$

Where
m is money supply, *p* is the price level, *y* is real income, *d* is domestic demand for goods and all are in natural logarithms. The variable *r* is nominal interest rates and asterisks denote foreign quantities. Variables ε_{1t}, ε_{2t}, ε_{3t} are mutually uncorrelated white noise disturbances. Variables $\bar{s}_t$, $\bar{p}_t$, $\vec{p}_t^*$ are the long-run equilibrium nominal exchange rate and prices.

The equation (1) represents the standard demand for real balances put in relative form, while the equation (2) specifies that the relative demand for goods depend on the level of relative real incomes, the interest differential and the terms of trade. Equation (3) implies price adjustment to be proportional to current excess demand and equation (4) is the uncovered interest parity condition with a white noise disturbances. The equation (5) implies PPP to hold absolutely in the long-run and equation (6) specifies next periods expectation of the exchange rate to be proportional to the current periods exchange rate deviation from long-run equilibrium.
The equations (1), (2), and (3) could be rewritten to the reduced form of monetary model:

$$(p_t - p_t^*) = \beta_0 + \beta_1 (p_{t-1} - p_{t-1}^*) + \beta_2 s_t + \tau_t \qquad (7)$$

where

$$\beta_1 = [1 + \delta(\omega + \sigma/\lambda)]^{-1}$$
$$\beta_2 = \beta_1 \delta \omega$$

$$\tau_t = \beta_3 (y_t - y_t^*) + \beta_4 (m_t - m_t^*) + \upsilon_t \qquad (8)$$

where

$$\beta_3 = \beta_1 \delta (\gamma - 1 - \sigma\phi/\lambda)$$
$$\beta_4 = \beta_1 \delta \sigma / \lambda$$
$$\beta_0 = -\beta_1 \sigma \delta (k - k^*)/\lambda$$

and $\upsilon_t = \beta_1 \varepsilon_{2t} - (\beta_1 \delta \sigma / \lambda)\varepsilon_{1t}$.

The coefficients are expected to satisfy the restrictions that $0 \le \beta_1 \le 1$, $0 \le \beta_2 \le 1$, $\beta_3 \le 0$ and $0 \le \beta_4 \le 1$.

Other reduced form solution for s_t the nominal exchange rate could be found as follows from [1]:

$$s_t = b_0 + b_1 (m_t - m_t^*) + b_2 (y_t - y_t^*) + b_3 (p_{t-1} - p_{t-1}^*) \qquad (9)$$

where

$$b_0 = \mu[-(k - k^*)(1 + 1/\alpha\lambda) - \beta_0/\alpha\lambda]$$
$$b_1 = \mu[1 + 1/\alpha\lambda - \beta_4/\alpha\lambda]$$
$$b_2 = -\mu[\phi(1 + 1/\alpha\lambda) + \beta_3/\alpha\lambda] \quad b_3 = -b_1$$
$$\mu = (1 + \beta_2/\alpha\lambda)^{-1}$$

and

$$\nu_t = -(\beta_1 \delta \sigma/\lambda^2)\varepsilon_{1t} - (\beta/\lambda)\varepsilon_{2t}.$$

Hence (7) and (9) are the reduced form equations for the endogenous variables $(p_t - p_t^*)$ and s_t and the long-run exchange rate equilibrium is given by

$$\bar{s}_t = -(k - k^*) + (m_t - m_t^*) - \phi(y_t - y_t^*). \qquad (10)$$

The hypothesis is, that the Dornbush model could be tested for Slovak and Polish Republics so we will estimate the reduced forms (7) and (9).

2 Estimation of Reduced Forms of Monetary Model

We have tried to estimate simple reduced forms of Dornbusch Monetary Model for quarterly data during the years 1998 till 2003, with 24 observations. As prices we will use consumer price indices CPI (%), previous year = 100, real income is given by GDP (bill. SKK or bill. PLN), at constant prices 1995 = 100, money supply is given by M2 (bill. SKK and bill. PLN). We have used the seasonally adjusted GDP which is specified as for example $y_t sadj$. For estimation purposes the reduced forms (7) and (9) of the Dornbusch monetary models were used.

The estimated results of the long-run equilibrium model (8) with standard errors in parenthesis are as follows:

$$s_t = \underset{(0,27)}{0,266}(m_t - m_t^*) - \underset{(0,11)}{1,22}(y_t - y_t^*)sadj \qquad (11)$$

$\rho = 0,865$ disturbance process is AR (1)

$R^2 = 0,62$; D-W = 1,72; MAPE = 1,47 %; LM = 1,05; JB=0,41; n = 22.

Estimation results of the reduced form of monetary model (7) with standard error in parenthesis are:

$$(p_t - p_t^*) = \underset{(0,163)}{0,158} + \underset{(0,133)}{0,726}(p_{t-1} - p_{t-1}^*) - \underset{(0,069)}{0,063}s_t$$

$R^2 = 0,66$; D-W = 2,04; LM=0,06; JB=0,05; n=23.

Since coefficient with s is not statistically significant the estimation of (7) should be:

$$(p_t - p_t^*) = \underset{(0,005)}{0,009} + \underset{(0,12)}{0,771}(p_{t-1} - p_{t-1}^*) \qquad (12)$$

$R^2 = 0,65$; D-W = 1,99 LM=0,002; JB=4,6; n=23.

Estimation of the reduced form model (9) with standard error in parenthesis presents equation (13):

$$s_t = \underset{(2,56)}{-2,243} + \underset{(0,34)}{0,511}(m_t - m_t^*) - \underset{(1,33)}{2,413}(y_t - y_t^*)sadj - \underset{(0,48)}{0,929}\,(p_{t-1} - p_{t-1}^*) \qquad (13)$$

$R^2 = 0,20$; D-W = 0,68; MAPE = 2,43 %; LM = 10,02; JB = 0,94; n = 22.

The coefficient of determination is only 20 % and there is positive autocorrelation of residuals. All estimated coefficients are not statistically significant at 5 % level. Moreover, the direction of the influence is not appropriate as the estimates of parameters are not of high precision.
Since the above equations are not satisfactory we suppose that the influence of the difference in money supply, the difference in real income and the difference in price indices is not immediately and there must be some lags before the results of monetary or real impulses will arise. Next, we therefore should check up what is the dynamic structure of the above mentioned model. We will try to find the best possible shape of equation (7) and (9).
The best estimation for equation (7) is given by equation (14):

$$(p_t - p_t^*) = \underset{(0,23)}{0,65} + \underset{(0,21)}{0,739}(p_{t-4} - p_{t-4}^*) - \underset{(0,097)}{0,254}s_{t-1} \qquad (14)$$

$\rho = 0,56$ disturbance process is AR (1)

$R^2 = 0,70$; D-W = 2,49; LM = 3,82; JB = 0,05; n = 19.

All coefficients are statistically significant at the 2 % level, the degree of determination is 70 % and there is no autocorrelation of residuals, residuals are normally distributed. The difference in consumer price negatively depends on the nominal exchange rate from previous quarter and the price adjustment should take about a year. We can easily agree with these results.
The best estimation of equation (9) is given by the equation (15):

$$s_t = \underset{(0.95)}{1,03} - \underset{(0.22)}{0,44}(m_{t-3} - m_{t-3}^*) - \underset{(0.49)}{0,92}(y_{t-4} - y_{t-4}^*)sadj - \underset{(0.30)}{0,93}\,(p_{t-1} - p_{t-1}^*) \qquad (15)$$

$\rho = 0{,}655$ disturbance process is AR (1)

$R^2 = 0{,}92$; D-W = 1,73; LM = 0,24; JB = 1,28; n = 19.

All coefficients are statistically significant at the 10 % level (most of them also at 5 %), $R^2 = 92$ %, residuals are normally distributed, there is no autocorrelation. Level of SKK/PLN nominal exchange rate depends on the difference in money supply with three quarters lag, difference in real GDP level with four quarter lag and on the difference in CPIs from previous quarter.

3 Conclusion

The results of the estimated monetary models for exchange rate determination show that econometricians have to manage a lot of work to design properly some structural model, but estimates of its parameters are not highly precise. The reason for this is that not only PPP is very strong condition but also there is a great problem of hidden multicolinearity as well.
It was also shown that well-known reduced forms of Dornbusch monetary model does not contain enough dynamics to explain relationship among variables. It was revealed by means of cross–correlation function between of exogenous variables and residuals. Reduced monetary models with more dynamics are better like that proposed by Dornbusch in [1].

References

[1] Baillie, R.T., McMahon, P.C.: The Foreign Exchange Market. Theory and evidence. Cambridge University Press, Cambridge, 1989.

[2] Diebold, F. X.: Empirical Modeling of Exchange Rate Dynamics. Editors Beckmann, M. and Krelle, W. Springer-Verlag Berlin, 1988.

[3] MacDonald, R., Taylor, M. P.: Exchange rates and Open Economy Macroeconomics, Basil Blackwell, Cambridge, 1989.

[4] Enders, W.: Applied Econometric Time Series. John Wiley & Sons, Inc. New York, 1995.

[5] Ivaničová, Z.: Discrete Optimization of Exchange Rate Model. Proceeding from 9th International Scientific Conference, Quantitative Methods in Economy – Compatibility of Methodology and Practice with the EU Conditions (on CD), Nov. 13.–14th 2003, Bratislava, p. 38.

[6] Rublíková, E.: Purchasing Power Parity and Cointegration. Economic and Informatics N.1 2003. FEI EU Bratislava, p. 147-154.

[7] National Bank of Poland Bulletins, NBP, Warszawa 1998-2004.

[8] Polish Statistical Bulletins, GUS, Warszawa 1998-2004.

Equity and Efficiency in a Measure Space with Nonadditive Preferences: The Problem of Cake Division*

Nobusumi Sagara[1] and Milan Vlach[2,3]

[1] Faculty of Economics, Hosei University
4342, Aihara, Machida, Tokyo, 194–0298 Japan
nsagara@mt.tama.hosei.ac.jp
[2] School of Mathematics and Physics, Charles University
Malostranske namesti 25, 118 00 Praha 1, Czech Republic
mvlach@ksi.ms.mfff.cuni.cz
[3] Kyoto College of Graduate Studies for Informatics
7, Monzen-cho, Sakyo-ku, Kyoto, 606-8225 Japan
m_vlach@kcg.ac.jp

Abstract. This paper considers a classical problem of cake division in a nonatomic finite measure space among finitely many individuals. We investigate a nonadditive continuous preference relation in a Borel σ-field and prove the existence of Pareto optimal envy-free partitions, Pareto optimal α-equitable partitions, and α-Rawls optimal partitions. We also show that Pareto optimal α-equitability is equivalent to α-Rawls optimality, but Pareto optimality does not imply Rawls optimality.

1 Metric on a Borel σ-Field

Let $(X, \mathscr{B}_X, \mu)$ be a measure space, where X is a topological space, $\mathscr{B}_X$ is the Borel σ-field of X, and μ is a measure on $\mathscr{B}_X$. A measure μ is a *Borel measure* if $\mu(\Omega) < \infty$ for each compact subset Ω of X. Let μ be a Borel measure and $\Omega \in \mathscr{B}_X$ be a compact subset of X. When Ω is endowed with the relative topology from X, the Borel σ-field $\mathscr{B}_\Omega$ of Ω is given by $\mathscr{B}_\Omega = \{E \cap \Omega | \, E \in \mathscr{B}_X\}$ and the restriction of μ, which we denote again μ, to the Borel measurable space $(\Omega, \mathscr{B}_\Omega)$ makes $(\Omega, \mathscr{B}_\Omega, \mu)$ a finite Borel measure space. Each element f in $L^1(\Omega, \mathscr{B}_\Omega, \mu)$ is identified with an element $\tilde{f}$ in $L^1(X, \mathscr{B}_X, \mu)$ by the embedding $f \mapsto \tilde{f}$ satisfying $\tilde{f} = f$ on Ω and $\tilde{f} = 0$ on $X \setminus \Omega$. This embedding yields an isometry on $L^1(\Omega, \mathscr{B}_\Omega, \mu)$ into $L^1(X, \mathscr{B}_X, \mu)$ and under this identification $L^1(\Omega, \mathscr{B}_\Omega, \mu)$ is a closed vector subspace of $L^1(X, \mathscr{B}_X, \mu)$. Note that the norm

* The first author gratefully acknowledges the financial support by Grant-in-Aid for Scientific Research (No. 14730021) from the Japan Society for the Promotion of Science, and by Special Assistance for Promoting the Advancement of the Education and Research of the Private Universities on "International Research Project on Aging (Japan, China and Korea)" from the Ministry of Education, Culture, Sports, Science and Technology.

topology on $L^1(\Omega, \mathscr{B}_\Omega, \mu)$ coincides with its relative topology induced by the norm topology on $L^1(X, \mathscr{B}_X, \mu)$.

Two measurable sets A and B in $\mathscr{B}_\Omega$ are *μ-equivalent* if $\mu(A \triangle B) = 0$, where $A \triangle B = (A \cup B) \setminus (A \cap B)$ is the symmetric difference of A and B. The μ-equivalence defines an equivalence relation on $\mathscr{B}_\Omega$. We denote the equivalence class of $A \in \mathscr{B}_\Omega$ by $[A]$ and the set of equivalence classes in $\mathscr{B}_\Omega$ by $\mathscr{B}_\Omega[\mu]$. If, for any two μ-equivalence classes $\mathbf{A}$ and $\mathbf{B}$, we define the metric d by $d(\mathbf{A}, \mathbf{B}) = \mu(A \triangle B)$ where A and B are arbitrarily selected elements of $\mathbf{A}$ and $\mathbf{B}$, then $\mathscr{B}_\Omega[\mu]$ becomes a complete metric space (see Dunford and Schwartz [3, Lemma III.7.1]). Since $\mu(A \triangle B) = \int |\chi_A - \chi_B| d\mu$ where χ_A and χ_B are characteristic functions of A and B respectively, we know that two measurable sets A and B are μ-equivalent if and only if their characteristic functions differ by a μ-null function. Therefore, the mapping $\mathbf{A} \mapsto \chi_A$ where A is an arbitrarily selected element of $\mathbf{A}$ is an isometry on $\mathscr{B}_\Omega[\mu]$ into $L^1(\Omega, \mathscr{B}_\Omega, \mu)$, and hence $\mathscr{B}_\Omega[\mu]$ can be identified with $\mathscr{X}_\Omega = \{\chi_A \in L^1(\Omega, \mathscr{B}_\Omega, \mu) | \, [A] \in \mathscr{B}_\Omega[\mu]\}$.

A topological space X is *locally compact* if each point in X has a neighborhood whose closure is compact. A Borel measure μ on $\mathscr{B}_X$ is *regular* if it is both outer regular and inner regular, that is, $\mu(E) = \inf\{\mu(U) | \, U \text{ is open and } E \subset U\} = \sup\{\mu(K) | \, K \text{ is compact and } K \subset E\}$ for each $E \in \mathscr{B}_X$. A *Haar measure* is a Borel measure μ on a locally compact topological group X such that $\mu(U) > 0$ for each nonempty open set U, and $\mu(xE) = \mu(E)$ for each $E \in \mathscr{B}_X$ and $x \in X$. Every locally compact topological group has a regular Haar measure (see Halmos [4, Theorem 58.B]).

Theorem 1. *Let $(X, \mathscr{B}_X, \mu)$ be a Borel measure space with X a locally compact topological group and μ a regular Haar measure. If Ω is a compact subset of X and $(\Omega, \mathscr{B}_\Omega, \mu)$ is the finite measure space induced by the restriction of $(X, \mathscr{B}_X, \mu)$, then $\mathscr{B}_\Omega[\mu] = \mathscr{X}_\Omega$ is a compact metric space.*

A *partition* of Ω is an n-tuple of disjoint elements in $\mathscr{B}_\Omega$ whose union is Ω. We denote the set of partitions of Ω by $\mathscr{P}$. The n-times Cartesian product of $\mathscr{B}_\Omega[\mu]$ is denoted by $\mathscr{B}^n_\Omega[\mu]$. Define the set of equivalence classes of partitions by

$$\mathscr{P}[\mu] = \left\{ (\mathbf{A}_1, \ldots, \mathbf{A}_n) \in \mathscr{B}^n_\Omega[\mu] \,\middle|\, \begin{array}{l} \exists (A_1, \ldots, A_n) \in \mathscr{P} : \\ A_i \in \mathbf{A}_i, \ i = 1, \ldots, n \end{array} \right\}.$$

Theorem 2. *Let $(X, \mathscr{B}_X, \mu)$ be a Borel measure space with X a locally compact topological group and μ a regular Haar measure. If Ω is a compact subset of X and $(\Omega, \mathscr{B}_\Omega, \mu)$ is the finite measure space induced by the restriction of $(X, \mathscr{B}_X, \mu)$, then $\mathscr{P}[\mu]$ is a compact metric space.*

2 Existence of a Continuous Utility Function

Let $(\Omega, \mathscr{B}_\Omega, \mu)$ be the finite Borel measure space introduced in the previous section. A *preference relation* $\succsim$ on $(\Omega, \mathscr{B}_\Omega, \mu)$ is a complete transitive binary relation on $\mathscr{B}_\Omega$. The strict preference $A \succ B$ means that $A \succsim B$ and $B \not\succsim A$. The

indifference $A \sim B$ means that $A \succsim B$ and $B \succsim A$. A real-valued set function u on $\mathscr{B}_\Omega$ *represents* $\succsim$ if $u(A) \geq u(B)$ holds if and only if $A \succsim B$ does, and such u is called a *utility function*.

Definition 1. *A preference relation $\succsim$ on $\mathscr{B}_\Omega$ is:*

(1) *μ-monotone if $A \supset B$ and $\mu(A) > \mu(B)$ implies $A \succsim B$.*
(2) *μ-strictly monotone if $A \supset B$ and $\mu(A) > \mu(B)$ implies $A \succ B$.*
(3) *μ-indifferent if $\mu(A \triangle B) = 0$ implies $A \sim B$.*

The μ-indifference of a preference relation $\succsim$ *induces* a preference relation $\succsim_\mu$ on $\mathscr{B}_\Omega[\mu]$ defined by $\mathbf{A} \succsim_\mu \mathbf{B}$ if and only if there exist $A \in \mathbf{A}$ and $B \in \mathbf{B}$ such that $A \succsim B$. This is equivalent to saying that $\mathbf{A} \succsim_\mu \mathbf{B}$ if and only if $A \succsim B$ for each $A \in \mathbf{A}$ and $B \in \mathbf{B}$. Thus, any utility function u representing $\succsim$ on $\mathscr{B}_\Omega$ induces a utility function u_μ representing $\succsim_\mu$ on $\mathscr{B}_\Omega[\mu]$ by $u_\mu(\mathbf{A}) = u(A)$ with $A \in \mathbf{A}$.

Definition 2. *A preference relation $\succsim$ on $\mathscr{B}_\Omega$ is μ-continuous if it is μ-indifferent and for each $\mathbf{A} \in \mathscr{B}_\Omega[\mu]$ both the upper contour set $\{\mathbf{B} \in \mathscr{B}_\Omega[\mu] \,|\, \mathbf{B} \succsim_\mu \mathbf{A}\}$ and the lower contour set $\{\mathbf{B} \in \mathscr{B}_\Omega[\mu] \,|\, \mathbf{A} \succsim_\mu \mathbf{B}\}$ are closed in $\mathscr{B}_\Omega[\mu]$.*

Definition 3. *A function f on $\mathscr{B}_\Omega$ is:*

(1) *μ-monotone if $A \supset B$ and $\mu(A) > \mu(B)$ implies $f(A) \geq f(B)$.*
(2) *μ-strictly monotone if $A \supset B$ and $\mu(A) > \mu(B)$ implies $f(A) > f(B)$.*
(3) *μ-indifferent if $\mu(A \triangle B) = 0$ implies $f(A) = f(B)$.*
(4) *μ-continuous if it is μ-indifferent and induces a continuous function f_μ on $\mathscr{B}_\Omega[\mu]$.*

The next theorem guarantees the existence of a μ-continuous utility function representing μ-continuous preferences. The proof is based on the celebrated theorem of Debreu [1].

Theorem 3. *Let $(X, \mathscr{B}_X, \mu)$ be a Borel measure space with X a locally compact topological group and μ a regular Haar measure, and let Ω be a compact subset of X and $(\Omega, \mathscr{B}_\Omega, \mu)$ be the finite measure space induced by the restriction of $(X, \mathscr{B}_X, \mu)$. Then, for any μ-continuous preference relation $\succsim$ on $\mathscr{B}_\Omega$, there exists a μ-continuous utility function representing $\succsim$.*

Example 1. A typical example of a μ-continuous and μ-monotone (resp. μ-strict monotone) preference relation on $\mathscr{B}_\Omega$ is the one defined by $A \succsim B$ if and only if $f(\mu(A)) \geq f(\mu(B))$, where f is a continuous and increasing (resp. strictly increasing) function on $[0, \mu(\Omega)]$ and μ is a finite measure of a measurable space $(\Omega, \mathscr{B}_\Omega)$.

3 Structure of Pareto Optimal Partitions

Let $(X, \mathscr{B}_X, \mu)$ be a Borel measure space with X a locally compact topological group and μ a nonatomic regular Haar measure. Let Ω be a compact subset of X and $(\Omega, \mathscr{B}_\Omega, \mu)$ be the nonatomic finite measure space induced from $(X, \mathscr{B}_X, \mu)$ as in Section 1. Denote the finite set of individuals by $I = \{1, \ldots, n\}$. A utility function of individual $i \in I$ on $\mathscr{B}_\Omega$ is denoted by u_i.

Definition 4. *A partition $(A_1, \ldots, A_n)$ is Pareto optimal if there exists no partition $(B_1, \ldots, B_n)$ such that $u_i(B_i) \geq u_i(A_i)$ for each $i \in I$ and $u_i(B_i) > u_i(A_i)$ for some $i \in I$.*

Let u_i be μ-continuous for each $i \in I$. Since the range of u_i is bounded for each $i \in I$ by Theorem 1, without loss of generality we may assume that $(u_1(A_1), \ldots, u_n(A_n)) \in [0,1]^n$ for any $(A_1, \ldots, A_n) \in \mathscr{P}$, where $[0,1]^n$ is the n-times Cartesian product of the unit interval $[0,1]$. The *utility possibility set* is defined as follows.

$$U = \{(x_1, \ldots, x_n) \in [0,1]^n | \, \exists (A_1, \ldots, A_n) \in \mathscr{P} : x_i \leq u_i(A_i) \ \forall i \in I\}.$$

Denote the set of Pareto optimal partitions by $\mathscr{P}^*$. Define the *Pareto frontier* of U by

$$U^* = \{(u_1(A_1), \ldots, u_n(A_n)) | \, (A_1, \ldots, A_n) \in \mathscr{P}^*\}$$

and the equivalence classes of the set of Pareto optimal partitions by

$$\mathscr{P}^*[\mu] = \left\{ (\mathbf{A}_1, \ldots, \mathbf{A}_n) \in \mathscr{B}_\Omega^n[\mu] \,\middle|\, \begin{array}{l} \exists (A_1, \ldots, A_n) \in \mathscr{P}^* : \\ A_i \in \mathbf{A}_i \quad \forall i \in I \end{array} \right\}.$$

Denote the $(n-1)$-dimensional *unit simplex* by

$$\Delta^{n-1} = \left\{ (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n \,\middle|\, \sum_{i=1}^n \alpha_i = 1 \text{ and } \alpha_i \geq 0, \, i = 1, \ldots, n \right\}.$$

Define the function $\rho : \Delta^{n-1} \to (0, \infty)$ by $\rho(s) = \sup\{\lambda > 0 | \, \lambda s \in U\}$.

Lemma 1. *If u_i is μ-continuous and μ-strictly monotone for each $i \in I$, then the map on Δ^{n-1} to U^* defined by $s \mapsto \rho(s)s$ is a homeomorphism between Δ^{n-1} and U^*.*

Theorem 4. *If u_i is μ-continuous and μ-strictly monotone for each $i \in I$, then $\mathscr{P}^*[\mu]$ is homeomorphic to Δ^{n-1}.*

Remark 1. It is well known that for a finite dimensional commodity space, the set of Pareto optimal allocations is homeomorphic to the unit simplex if preferences of each individual are continuous, monotone, and strictly convex (see Varian [9]). For an infinite dimensional commodity space, the existence of the homeomorphism between the set of Pareto optimal allocations and the unit simplex was demonstrated by Magill [6] and Mas-Colell [7]. In our setting the key argument is the construction of ρ and its continuity. The continuity of ρ follows from the similar argument to that of Mas-Colell [7], who worked in a commodity space with topological vector lattices.

4 Equity and Efficiency

In this section we examine the compatibility between equity and efficiency (Pareto optimality). Three notions of equity are introduced: *Envy-freeness*, *α-equity*, and *α-Rawls optimality*. All these notions have already appeared in the classical paper by Dubins and Spanier [2] before economists presented an analytical framework for the problem of fair division.

Envy-Free Partition

The first notion of equity under investigation is envy-freeness along the line of Varian [9].

Definition 5. *A partition $(A_1, \dots, A_n)$ is envy-free if $u_i(A_i) \geq u_i(A_j)$ for each $i, j \in I$.*

Theorem 5. *If u_i is μ-continuous and μ-strictly monotone for each $i \in I$, then there exists a Pareto optimal envy-free partition.*

The following theorem due to Scarf [8, Theorem 3] plays a crucial role to prove Theorem 5.

Intersection Theorem (Scarf [8]). *Let $\Delta_i = \{(\alpha_1, \dots, \alpha_n) \in \Delta^{n-1} | \, \alpha_i = 0\}$ for each $i \in I$. If the collection $\{F_1, \dots, F_n\}$ is a closed covering of Δ^{n-1} satisfying $\Delta_i \subset F_i$ for each $i \in I$, then $\bigcap_{i \in I} F_i \neq \emptyset$.*

α-Equitable Partitions

The next notion of equity under investigation is α-equitability. When referring to $\alpha = (\alpha_1, \dots, \alpha_n)$ in the following definitions, we mean that α belongs to $\operatorname{int} \Delta^{n-1}$.

Definition 6. *A partition $(A_1, \dots, A_n)$ is α-equitable if $\alpha_i^{-1} u_i(A_i) = \alpha_j^{-1} u_j(A_j)$ for each $i, j \in I$. An α-equitable partition for $\alpha = (1/n, \dots, 1/n)$ is simply said to be equitable.*

Theorem 6. *Let u_i be μ-continuous and μ-strictly monotone for each $i \in I$. Then for any $\alpha \in \operatorname{int} \Delta^{n-1}$ there exists a Pareto optimal α-equitable partition.*

α-Rawls Optimal Partitions

Another important criterion proposed by Dubins and Spanier [2] is the anticipation of Rawls' criterion of maximizing the welfare of the least well-off individual. To define this formally, consider the following maximization problem:

$$\sup\Big\{ \min_{i \in I} \alpha_i^{-1} u_i(A_i) \Big| \, (A_1, \dots, A_n) \in \mathscr{P} \Big\}. \tag{R_α}$$

Definition 7. *A partition $(A_1, \ldots, A_n)$ is α-Rawls optimal if it is a solution to (R_α). An α-Rawls optimal partition for $\alpha = (1/n, \ldots, 1/n)$ is simply said to be Rawls optimal.*

Theorem 7. *Let u_i be μ-continuous and μ-strictly monotone for each $i \in I$. Then for any $\alpha \in \operatorname{int} \Delta^{n-1}$ every α-Rawls optimal partition is Pareto optimal.*

As was pointed out by Weller [10], Rawls optimality is much stronger than Pareto optimality for the case of additive preferences with a nonatomic probability measure. This is also true for the case of nonadditive continuous preferences.

Theorem 8. *If u_i is μ-continuous and μ-strictly monotone for each $i \in I$, then there exists a Pareto optimal partition which is not Rawls optimal.*

However, Pareto optimality implies α-Rawls optimality for some $\alpha \in \operatorname{int} \Delta^{n-1}$. To show this, we introduce the following notion: A partition $(A_1, \ldots, A_n)$ is *μ-positive* if $\mu(A_i) > 0$ for each $i \in I$.

Theorem 9. *Let u_i be μ-continuous and μ-strictly monotone for each $i \in I$. We then have the following.*

(1) *A μ-positive partition is Pareto optimal if and only if it is α-Rawls optimal for some $\alpha \in \operatorname{int} \Delta^{n-1}$.*
(2) *For any $\alpha \in \operatorname{int} \Delta^{n-1}$ a partition is α-Rawls optimal if and only if it is Pareto optimal and α-equitable.*

For the case of additive preferences with a nonatomic probability measure on a σ-field, Legut and Wilczyński [5] used a minimax theorem to show the existence of an α-Rawls optimal partition. The same result is true for our setting.

Corollary 1. *Let u_i be μ-continuous and μ-strictly monotone for each $i \in I$. Then for any $\alpha \in \operatorname{int} \Delta^{n-1}$ there exists an α-Rawls optimal partition.*

References

1. Debreu, G.: Continuity properties of Paretian utility. Internat. Econom. Rev. **5** (1964) 285–293
2. Dubins, L. E., Spanier, E. H.: How to cut a cake fairly. Amer. Math. Monthly **68** (1961) 1–17
3. Dunford, N., Schwartz, J. T.: Linear Operators, Part I: General Theory. John Wiley & Sons, New York (1958)
4. Halmos, P. R.: Measure Theory. Springer-Verlag, Berlin Heidelberg New York (1950)
5. Legut, J., Wilczyński, M.: Optimal partitioning of a measurable space. Proc. Amer. Math. Soc. **104** (1988) 262–264
6. Magill, M. J. P.: An equilibrium existence theorem. J. Math. Anal. Appl. **84** (1981) 162–169
7. Mas-Colell, A.: The price equilibrium existence problem in topological vector lattices. Econometrica **54** (1986) 1039–1053
8. Scarf, H.: The approximation of fixed points of a continuous mapping. SIAM J. Appl. Math. **15** (1967) 1328–1343
9. Varian, H. R.: Equity, envy, and efficiency. J. Econom. Theory **9** (1974) 63–91
10. Weller, D.: Fair division of a measurable space. J. Math. Econom. **14** (1985) 5–17

Comparison the Quality of Classification Algorithms [1]

Hana Skalska
Faculty of Informatics and Management, University of Hradec Kralove
hana.skalska@uhk.cz

Abstract

This article summarises and describes different measures of the predictive quality (performance) of classification models. Quantitative measures of quality can be complemented with the visual representation. One of the best known is the ROC (Receiver Operating Characteristic) curve that represents the performance of binary classification model within the full range of conditions (costs and class distributions) of discrimination. The duality between ROC and expected costs of a classifier (cost curve) is described in more details here. Cost curve measures the difference in performance of two classifiers directly in expected costs (EC). The use of ROC and EC is highly important when asymmetric misclassification costs, imbalanced probabilities of classes or changing conditions occur. A hybrid classifier can be found on the convex hull (ROCCH) of the ROC curves of different classifiers. This is potentially the best classification model for any mixture of outside constraints. Lower envelope of cost curves of different classifiers corresponds to ROCCH in ROC space.

Key words: Classification, Predictive Accuracy Measures, ROC, Cost Curve

1 Classification

Classification solves the problem of classifying the individuals into one of several categories on the basis of a number of measurements made on each of individuals. An individual is considered to be a random observation given from any of finite number of categories or populations. We will concentrate to only binary classification and to the activity of classification that is known as discrimination or class assignment. This problem can be solved as statistical decision problem with linear or nonlinear discrimination rule, Bayesian methods, or logistic regression [1]. Also semi-statistical or non-statistical solutions, such as decision trees, decision rules, neural networks, logic programming approaches, multi-criteria classification, etc can be used.

Classification tasks have different applications, for example fraud detection, prediction the quality of loan, detection of abnormal observations, classifying consumers as risky or reliable, automatic recognition of letters or digits from handwritten source, speech recognition etc.

In the process of classification it is desirable to minimize on the average the bad effects of misclassification, usually the average costs of misclassification. Statistical methods of discrimination are usually designed to be optimal for data with categorical or quantitative predictor variables (measurements). Conceptual methods of classification are built in the form of decision trees and characterize classes by conjunction of logical terms and partially optimal rules are based on variables chosen in each step with the threshold that is locally optimal. Often these algorithms are developed also for classification of individuals with non-numerical measurements as are web pages, text, audio, visual, or spatial data [7] and therefore they have very broad possibilities of practical use.

Classification algorithms are optimalised with respect to costs of misclassification and probabilities of classes. When external conditions are changing than classification model can lose its optimality. An evaluation and comparison of the models within the full range of conditions is necessary.

The aim of this article is to summarize the characteristics of classification accuracy in binary classification when only two classes are admitted. These measures are important for comparison of predictive quality of different algorithms as well as for setting the optimal parameters of any algorithm and understanding the problem of cost-sensitive classification.

[1] *This paper was partly supported by the Grant Agency of Czech Republic – grant No 402/04/1308*

Firstly we will concentrate on description of the measures of predictive accuracy. Then the possibilities of ROC analysis and dual representation with expected costs will be described.

2 Measures of accuracy

We will assume two classes G_0 and G_1 of individuals. Alternative notation Neg (negatives) and Pos (positives) for the classes is commonly used in binary classification. Let FP (false positive) denotes number of individuals from G_0 incorrectly classified to G_1 and FN (false negative) denotes number of individuals from G_1 incorrectly classified to G_0. Correct classifications of individuals from G_0 (negatives) or G_1 (positives) will be TN (true negatives) and TP (true positives) respectively. Lower cases (tp, fp, tn, fn) will be used for rates. Classification of an individual is based on vector of measurements $x' = (x_1, x_2, \ldots, x_p)$ on that individual. Confusion matrix (Table1) describes errors of classification when model is evaluated over a training dataset of *n* individuals with known classes. PNeg and PPos are the total numbers of individuals that are predicted as negative or positive one.

	Predicted as		Total
	G0	G1	
Actually G0	TN	FP	Neg
Actually G1	FN	TP	Pos
Total	PNeg	PPos	n

Table 1. Confusion matrix

2.1 Description of measures of performance

Measures of accuracy can be defined with respect to group of negative individuals or with respect to group of positive ones (usually positive individuals are of our interest). Different measures [2, 3, 4, 5, 6, 8, 9] are described in Table 2.

Formula	Additional description
$^{P}\mathrm{Recall} = \frac{TP}{Pos} = tp = \frac{Pos - FN}{Pos} = 1 - fn$	Sensitivity
$^{P}\mathrm{Precision} = \frac{TP}{TP + FP} = \frac{TP}{PPos} = ppv$	Positive predictive value
$^{N}\mathrm{Recall} = \frac{TN}{Neg} = tn = \frac{Neg - FP}{Neg} = 1 - \frac{FP}{Neg} = 1 - fp$	Specificity
$^{N}\mathrm{Precision} = \frac{TN}{TN + FN} = \frac{TN}{PNeg} = npv$	Negative predictive value
$\mathrm{Accuracy} = \frac{TP + TN}{n} = tp \cdot \frac{Pos}{n} + tn \cdot \frac{Neg}{n} \sim$ $\sim tp \cdot P(G_1) + (1 - fp) \cdot P(G_0)$	Weighted mean of rates of true positives and negatives. Not appropriate in problems with imbalanced priors of classes or highly asymmetric costs
$^{P}F_{\lambda} = \frac{1}{\lambda \cdot {}^{P}\mathrm{Recall} + (1 - \lambda) \cdot \frac{1}{{}^{P}\mathrm{Precision}}}$	F-measure with respect to G_1 $0 \le \lambda \le 1$

$^{P}F_{0.5} = \frac{2 \cdot {}^{P}\mathrm{Recall} \cdot {}^{P}\mathrm{Precision}}{{}^{P}\mathrm{Recall} + {}^{P}\mathrm{Precision}}$	$\lambda = 0.5$ if Recall and Precision are of the same importance
$^{P}GMPR = \sqrt{{}^{P}\mathrm{Recall} \cdot {}^{P}\mathrm{Precision}}$	Geometric mean of Recall and Precision with respect to G_1
$SOR = {}^{P}\mathrm{Recall} + {}^{N}\mathrm{Recall} = tp + tn$	Sum of Recalls
$GMOR = \sqrt{{}^{P}\mathrm{Recall} \cdot {}^{N}\mathrm{Recall}} = \sqrt{tp \cdot tn}$	Geometric mean of Recalls
$I_\alpha = \frac{TP}{n}\log\left(\frac{Pos}{n}\right) + \frac{FN}{n}\log\left(1 - \frac{Pos}{n}\right) + \frac{FP}{n}\log\left(1 - \frac{Neg}{n}\right) + \frac{TN}{n}\log\left(\frac{Neg}{n}\right)$	Information score
ROC, PR (Precision-Recall) Curve, DET Curve, EC - Expected Costs Curve	Graphs
ROCCH is build under ROC space of different classifiers. Each classifier with ROC points on convex hull is optimal for specific condition (costs and class ratio).	ROC Convex Hull
AUC Area Under (ROC) Curve. Equivalent to the probability that randomly chosen positive individual will be ranked higher than a randomly chosen negative one.	Equivalent to Wilcoxon statistic. Scalar value representing expected performance.
$2 \cdot AUC - 1$	Gini coefficient

Table 2. Measures of performance of classifiers

2.2 ROC and Expected Costs

ROC is the plot of points with coordinates (x,y) = (fp, tp). For discrimination function or logistic regression, empirical ROC can be produced with varying threshold value on discrimination score. We will assume that the smaller values on discriminating variable are associated with individuals from group G_0 of negative individuals. For a non-statistical cost – sensitive classification methods (classification trees, neural network, genetic algorithms) sequence of ROC points is produced with varying of conditions of classification (changing the costs of misclassification or probabilities of classes). Each condition will produce one confusion matrix as well as one point on ROC curve.

Let c_{ij}, $i \neq j$ for $i,j = 0,1$ are the costs of misclassification an individual from the group G_j to the group G_i and $P(G_0) = q_0$, $P(G_1) = q_1$ are the probabilities of groups G_0, G_1, $q_0 + q_1 = 1$.

Expected loss from costs of misclassification or expected cost (EC) is given with equation:

$$EC = c_{10} \cdot P\left(G_1 \middle| G_0\right) \cdot P\left(G_0\right) + c_{01} \cdot P\left(G_0 \middle| G_1\right) \cdot P\left(G_1\right) = c_{10} \cdot P(FP) \cdot q_0 + c_{01} \cdot P(FN) \cdot q_1 \qquad (2.1)$$

Where c_{10} is the cost of false positive and c_{01} is the cost of false negative. It follows that

$$EC = c_{10} \cdot fp \cdot q_0 + c_{01} \cdot fn \cdot q_1 = c_{10} \cdot fp \cdot q_0 + c_{01} \cdot \left(1 - tp\right) \cdot q_1 \qquad (2.2)$$

EC is in association with the coordinates (fp, tp) of ROC curve. Other properties of ROC are:

- The slope c of the line connecting two different ROC points (fp_1,tp_1) and (fp_2, tp_2) with the same expected costs is given by

$$c = \frac{q_0 \cdot c_{10}}{q_1 \cdot c_{01}} \tag{2.3}$$

  Class probabilities (priors) and misclassification costs are interchangeable for the purpose of evaluation the performance of classification model. Change in prior rates has the same effect as the change (in the same proportion) of costs rate.
- On the ROC, point A_1 dominates to point A_2 (has a lower expected cost), if A_1 has higher true positive rate (tp) and lower false positive rate (fp). Then point A_1 has a lower expected cost than point A_2 also for all possible cost ratios and class distributions.
- Set of the points of any ROC curve A dominates to the set of points on ROC curve B (in ROC space) if A has lower expected costs than B for all operating points (A will have higher true positive rate for the same values of false positive rates).
- ROC visualises the performance of a classifier independently on cost or class ratios [3,5,10].

Secondary importance of ROC is in their use for selection of the set of parameters of individual classifier that gives the best performance for a particular costs ratio and class frequency. Only Webb and Ting [11] cautioned against ROC as the only measure of predictive accuracy under varying class distribution. Firstly, when the classes contain causally relevant subclasses whose frequencies may vary at different rates and secondly, if there are attributes upon which the classes are causally dependent.

2.3 Dual Representation of ROC Curve

Cost curve is based on a dual representation between a point (fp, tp) on the ROC curve and the iso-accuracy line that connects ROC points with the same expected cost. The slope of this line is given with ratio of priors and costs of misclassification (Eq. 2.3). Each point on the ROC can be converted into a line in expected costs space. Therefore cost curve was suggested [2] as an alternative to the ROC. Lower convex hull of these lines represents the operating range in which a classifier has minimal expected costs.

Cost curve is given with normalised expected costs (NEC) on vertical axis and probability cost function for the group of positive individuals (pcf(G_1)) on horizontal axis. NEC equals EC normalised with respect to the maximal expected costs when EC_{MAX} is obtained when tp = 0 and fp = 1.

$$NEC = \frac{EC}{EC_{MAX}} = \frac{c_{10} \cdot fp \cdot q_0 + c_{01} \cdot (1 - tp) \cdot q_1}{c_{10} \cdot q_0 + c_{01} \cdot q_1} = fp.pcf(G_0) + (1 - tp).pcf(G_1)$$

$$NEC = fp \cdot (1 - pcf(G_1)) + (1 - tp) \cdot pcf(G_1) = (1 - tp - fp) \cdot pcf(G_1) + fp \tag{2.4}$$

where

$$pcf(G_1) = \frac{c_{01} \cdot q_1}{c_{10} \cdot q_0 + c_{01} \cdot q_1} \qquad \text{and} \qquad pcf(G_0) = 1 - pcf(G_1) = \frac{c_{10} \cdot q_0}{c_{10} \cdot q_0 + c_{01} \cdot q_1}$$

For a point (tp,fp) on a ROC, (relative) normalised expected misclassification costs are given with equation 2.4. If pcf(G_1) = 0, then NEC = fp, if pcf(G_1) = 1, then NEC = 1-tp.

2.4 PR Curve and DET Curve

PR and DET curves are other possibilities how to visualise the performance of classifiers.

PR curve or ^{P}Precision (on Y axis) – ^{P}Recall (on X axis) curve visualises the ratio of correct predictions from positive predictions (positive predictive value) against ratio of individuals that were classified correctly from all positive individuals (sensitivity).

Detection Error Trade-off (DET) curve describes association between false positive and false negative ratios of incorrect classification when decision threshold changes. DET describes relationship between (x.y) = (fp, fn) rates. Usually both coordinates are transformed to normal deviation scale (z-scale).

When underlying distributions of discriminating variable are normal than DET curves are closed to linear. Normal deviation scale spreads out the plots and two or more classifiers can be compared. This way DET can distinguish more precisely between different systems they are performing well.

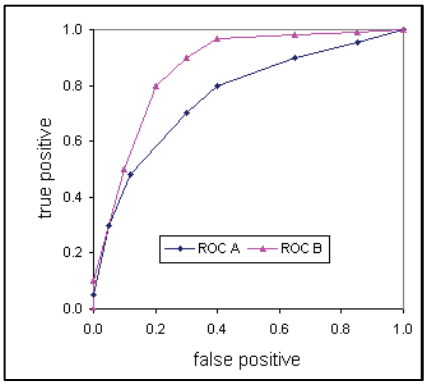
2.5 Example

Data for ROC curves in Graph 1 is given in Table 3.

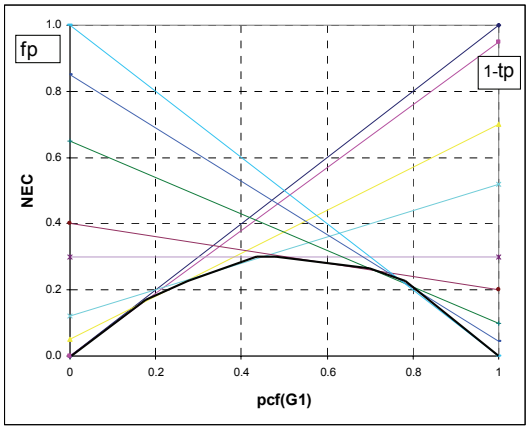
Expected costs with lower envelopes are presented in Graphs 2a, 2b. Relative expected cost of misclassification is on the vertical axis. A perfect classification is represented with the line connecting the points (0,0) and (0,1) and the worst one with the line connecting (0,1) and (1,0).

ROC A		ROC B	
fp	tp	fp	tp
0.000	0.000	0.000	0.000
0.000	0.050	0.000	0.100
0.050	0.300	0.100	0.500
0.120	0.480	0.200	0.800
0.300	0.700	0.300	0.900
0.400	0.800	0.400	0.970
0.650	0.900	0.650	0.980
0.850	0.954	0.850	0.990
1.000	1.000	1.000	1.000

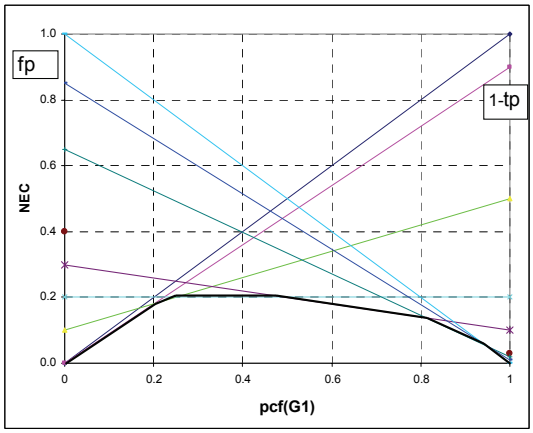
Table 3. Data



Graph 1. ROC curves



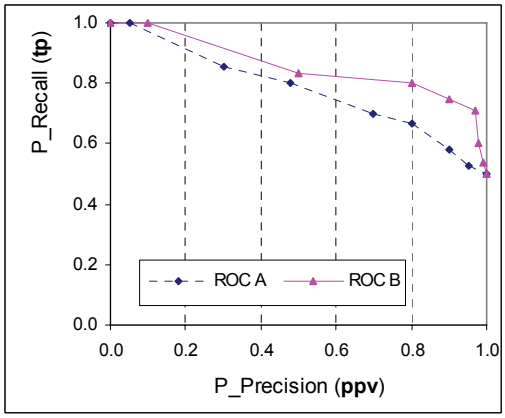
Graph 2a. EC_ROC A



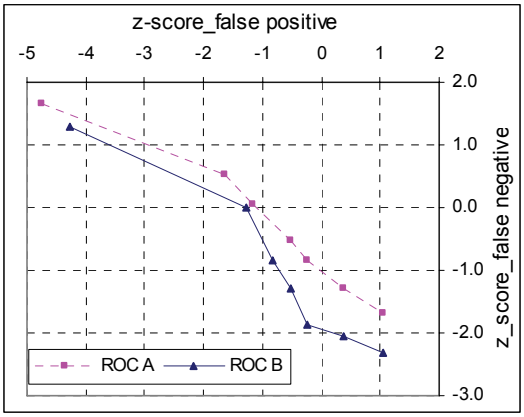
Graph 2b. EC_ROC B

Graphs 2a and 2b describe NEC lines for individual points of ROC. Each line describes normalised expected costs for a ROC point (fp, tp) when costs or classes prior settings change. NEC quantifies immediately the lower achievable expected costs. Also NEC can quantify the difference between relative expected costs of different classifiers for a particular pcf(G_1) point (this value depends on misclassification costs and priors – external conditions of classification). Shapes of lower envelopes (bold lines) of NEC for both models are in accordance with the data. ROC of classifier B dominates to ROC of classifier A and mostly B can reach lower expected costs. If probabilities of groups are the same and misclassification costs are not different (pcf(G_1)=0.50), then expected costs of model A are 0.3 and 0.2 for model B. The difference in expected costs can be given immediately.

Graph 3. represents the Precision - Recall Curve of positive individuals. Graph 4. describes DET curve with coordinates transformed to normal scale deviation (z-scale). Transformation the scale on DET curve spreads out mainly the points close to (0,0) and (1,1) on the ROC and therefore classifiers that are performing well can be distinguished better. But there are some difficulties with z- transformation that is too sensitive when fp or fn are close to 0 or 1.



Graph 3. PR Curve



Graph 4. DET Curve

3 Summary

Measures of classification accuracy in binary classification are important for comparison of predictive quality of different algorithms. This article summarises measures that allow comparison of classification methods of different forms, statistical or non statistical. This is important as the possibility to make decisions about setting the optimal parameters of any algorithm and understanding the problem of cost-sensitive classification. ROC curves and their dual representation with expected cost curve are described in more details. The difference in expected costs between different models can be immediately given from the EC graphs. ROC and EC methods are useful mainly for comparison of classifiers when asymmetric misclassification costs, imbalanced probabilities of classes or changing conditions occur. Lower convex envelope of cost curves can be used for detection the classifiers with lower expected cost.

References

[1] Anderson, T.W.: *An Introduction to Multivariate Statistical Analysis*. J.Wiley & Sons, New Jersey, 2003.

[2] Drummond, C., Holte, R.C.: *Explicitly Represented Expected Costs: An Alternative to ROC representation*. Proceedings of the Sixth ACM SIGKDD International Conference on KDDM, pp. 198-207, 2000.

[3] Fawcett, T: *ROC Graphs: Notes and Practical Considerations for Researchers*. Technical Report HP Laboratories, MS 1143, 2004, 38p.

[4] Hand, D.J., Till, R.J.: *A Simple Generalisation of the Area Under the ROC Curve to the Multiple Class Classification Problems*. Machine Learning 45, pp.171-186, 2001.

[5] Hanley, J.A., McNeil B.J.: *The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve*. Radiology 143, pp.29-36, 1982.

[6] Joshi, M.V.: *On Evaluating Performance of Classifiers for Rare Classes*. Proceedings of the ICDM'02, pp.641-644, 2002.

[7] Klosgen, W., Zytkov, J.M. (Eds.): *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, pp. 254-444, 2002.

[8] Martin A., Doddington G., Kamm T., Ordowski M., Przybocki M.: *The DET Curve in Assessment of Detection Task Performance*. Proc.Eurospeech'97, Rhodes, pp.1895-1898, 1997.

[9] Skalska, H.: *ROC Curve and AUC Estimates*. Aplimat 2003, Bratislava, 2003 (in Czech).

[10] Swets, J.A., Dawes, R.M., Monahan, J.: *Better decisions through science*. Scientific American 283, 82-87, 2000.

[11] Webb, G.I., Ting, K.M.: *On The Application of ROC Analysis to Predict Classification Performance Under Varying Class Distributions*. Machine Learning 58, pp. 15-29, 2005

Mean Variance Optimality in Markov Decision Chains

Karel Sladký, Milan Sitař

Department of Econometrics
Institute of Information Theory and Automation, AS ČR
Pod Vodárenskou věží 4, 182 08 Praha 8
sladky@utia.cas.cz, milan.sitar@tiscali.cz

Abstract

In this note, we consider discrete-time Markov decision processes with finite state space. Recalling explicit formulas for the growth rate of expected value and variance of the cumulative (random) reward, algorithmic procedures for finding optimal policies with respect to various mean variance optimality criteria are discussed.

Keywords

Markov decision chains with finite state space, expectation and variance of cumulative rewards, asymptotic behavior, mean variance optimality

1 Introduction

Motivation, objective and results. The usual optimization criteria examined in the literature of Markov decision processes, such as a total discounted or mean reward structure, can be quite insufficient to fully capture the various aspects for a decision maker. It may be preferable if not necessary to select or to include more sophisticated criteria that also reflect variability-risk features of the problem. Most notably, the variance of the cumulative rewards can be indicative and the so-called mean-variance optimality criteria may be of interest.

In this note we aim to discuss some algorithmic procedures for finding optimal policies with respect to various mean variance optimality criteria. To this end, after preliminaries in Sections 2 and 3 we study reward variance of (uncontrolled) Markov reward chains and present some formulas for the growth rate and asymptotic behaviour of the variance of total cumulative reward. Finally, in Section 4 algorithmic procedures for finding optimal policies with respect to various mean variance optimality criteria are discussed. The technical details and computational experience with large scale numerical examples are summarized in the Appendix.

Formulation and Notation. Consider a Markov decision chain $X = \{X_n,\, n = 0, 1, \ldots\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$, finite set $\mathcal{A}_i = \{1, 2, \ldots, K_i\}$ of possible decisions (actions) in state $i \in \mathcal{I}$ and the following transition and reward structure:

p_{ij}^k : transition probability from $i \to j$ $(i, j \in \mathcal{I})$ if action $k \in \mathcal{A}_i$ is selected in state i,

r_{ij}^k : one-stage reward for a transition from $i \to j$ if action k is selected in state i,

r_i^k : expected value of the one-stage rewards incurred in state i if action k is selected,

$r_i^{(2),k}$: second moment of the one-stage rewards incurred in state i if action k is selected.

Obviously, $r_i^k = \sum_{j\in\mathcal{I}} p_{ij}^k \cdot r_{ij}^k$, $r_i^{(2),k} = \sum_{j\in\mathcal{I}} p_{ij}^k \cdot [r_{ij}^k]^2$ and hence the corresponding one-stage reward variance $\sigma_i^k = r_i^{(2),k} - [r_i^k]^2$.

A policy controlling the chain is a rule how to select actions in each state. In this note, we restrict on stationary policies, i.e. the rules selecting actions only with respect to the current state of the Markov chain X. Then a policy, say π, is fully identified by some decision vector f whose ith element $f_i \in \mathcal{A}_i$ identifies the action taken if the chain is in state X. Stationary policy $\pi \sim (f)$ then completely identifies the transition probability matrix $\boldsymbol{P}(f)$. Observe that the ith row of $\boldsymbol{P}(f)$ has elements $p_{i1}^{f_i}, \ldots, p_{iN}^{f_i}$.

Let the vectors $\boldsymbol{R}^\pi(n)$, $\boldsymbol{S}^\pi(n)$, and $\boldsymbol{V}^\pi(n)$ denote the first moment, the second moment and the variance of the (random) total reward ξ_n respectively received in the n next transitions of the considered Markov chain X if policy $\pi \sim (f)$ is followed, given the initial state $X_0 = i$. More precisely, for the elements of $\boldsymbol{R}^\pi(n)$, $\boldsymbol{S}^\pi(n)$, and $\boldsymbol{V}^\pi(n)$ we have

$$R_i^\pi(n) = \mathsf{E}_i^\pi[\xi_n], \quad S_i^\pi(n) = \mathsf{E}_i^\pi[\xi_n^2], \quad V_i^\pi(n) = \boldsymbol{\sigma}_i^{2,\pi}[\xi_n]$$

where $\xi_n = \sum_{k=0}^{n-1} r_{X_k,X_{k+1}}$ and E_i^π, $\boldsymbol{\sigma}_i^{2,\pi}$ are standard symbols for expectation and variance if policy π is selected and $X_0 = i$.

In what follows $\boldsymbol{I}$ denotes identity matrix, and $\boldsymbol{e}$ is reserved for a unit column vector. By $\varepsilon(n)$ we denote a function in n such that $\varepsilon(n) \to 0$ geometrically fast as $n \to \infty$, i.e. for some $\alpha > 0$ and $\delta \in (0,1)$ $|\varepsilon(n)| \leq \alpha \cdot \delta^n$. By $\boldsymbol{\varepsilon}(n)$ we denote a vector function such that each component $\boldsymbol{\varepsilon}_i(n) \to 0$ as $n \to \infty$ geometrically.

Assumptions. We make the following assumptions:

AS 1. For any stationary policy $\pi \sim (f)$, the transition probability matrix $\boldsymbol{P}(f)$ has a single class of recurrent states, state N is recurrent.

To avoid some technical complications we make the following additional assumption:

AS 2. The transition probability matrix $\boldsymbol{P}(f)$ is aperiodic.

Preliminaries. (For the sake of brevity we omit the argument f at $\boldsymbol{P}(f)$.) It is well-known that under AS 1 the rows of the limiting matrix $\boldsymbol{P}^* = \frac{1}{n}\sum_{k=0}^{n-1} \boldsymbol{P}^k$ of the Markov chain X are identical and equal to the (row) vector of steady state probabilities $\boldsymbol{p}^* = [p_1^*, \ldots, p_N^*]$ as determined by $\boldsymbol{p}^* = \boldsymbol{p}^* \cdot \boldsymbol{P}$ along with $\boldsymbol{p}^* \cdot \boldsymbol{e} = 1$. Moreover, if $\boldsymbol{P}$ is aperiodic (i.e. under AS 2) then $\boldsymbol{P}^* = \lim_{k\to\infty} \boldsymbol{P}^k$ and the convergence is geometrical.

Under Assumption AS 1, using policy $\pi \sim (f)$ the average reward g^π of the Markov chain X is well defined, independently of the initial state i at time 0, by $g^\pi = \sum_{j\in\mathcal{I}} p_j^*(\pi) \cdot r_j^{f_j}$ and obviously, for the $\boldsymbol{g}^\pi$, the constant vector with elements g^π, $\boldsymbol{g}^\pi = \boldsymbol{P}^*(\pi) \cdot \boldsymbol{r}(f)$.

From the theory of dynamic programming (e.g. [5], [6]) it is well-known that the vectors of expected reward and times in the n next transitions fulfil the recursive formulas (see also (2.1) and (2.3) of Section 2)

$$\boldsymbol{R}^\pi(n+1) \;=\; \boldsymbol{r}(f) + \boldsymbol{P}(f) \cdot \boldsymbol{R}^\pi(n) \;\text{ with }\; \boldsymbol{R}^\pi(0) = 0 \tag{1.1}$$

and that under assumptions AS 1, AS 2 there exists vectors $\boldsymbol{w}^\pi$ such that

$$\boldsymbol{R}^\pi(n) = \boldsymbol{g}^\pi \cdot n + \boldsymbol{w}^\pi + \boldsymbol{\varepsilon}(n) \Rightarrow \lim_{n\to\infty} \frac{1}{n}\boldsymbol{R}^\pi(n) = \boldsymbol{g}^\pi = \boldsymbol{P}^*(\pi) \cdot \boldsymbol{r}(f), \tag{1.2}$$

hence $\boldsymbol{R}^\pi(n)$ possesses a linear growth rate g^π in n up to a geometric convergence to the null vector and vectors $\boldsymbol{w}^\pi$. The constant vector $\boldsymbol{g}^\pi$ along with vectors $\boldsymbol{w}^\pi$ are uniquely determined by

$$\boldsymbol{w}^\pi + \boldsymbol{g}^\pi \;=\; \boldsymbol{r}(f) + \boldsymbol{P}(f) \cdot \boldsymbol{w}^\pi, \quad \boldsymbol{P}^*(\pi) \cdot \boldsymbol{w}^\pi = 0. \tag{1.3}$$

2 Reward Variance for Finite Time and Transient Case

Obviously, for any integers $m < n$ we get (recall that ξ_n denotes the total random reward obtained in the n next transitions of the Markov chain X, $\xi^{(m,n)}$ is reserved for the reward obtained from the m-th up to the n-th transition)

$$\begin{aligned} \mathsf{E}_i^\pi[\xi_n] &= \mathsf{E}_i^\pi[\xi_m] + \mathsf{E}_i^\pi\{\sum_{j\in\mathcal{I}} \mathsf{P}(X_m = j)\cdot \mathsf{E}_j^\pi[\xi^{(m,n)}]\} \\ &= \mathsf{E}_i^\pi[\xi_m] + \mathsf{E}_i^\pi\{\sum_{j\in\mathcal{I}} \mathsf{P}(X_m = j)\cdot \mathsf{E}_j^\pi[\xi_{n-m}]\}. \end{aligned} \tag{2.1}$$

Similarly from $[\xi_n]^2 = [\xi_m]^2 + [\xi^{(m,n)}]^2 + 2\cdot\xi_m\cdot\xi^{(m,n)}$ we get for every $i\in\mathcal{I}$

$$\begin{aligned} \mathsf{E}_i^\pi[\xi_n]^2 &= \mathsf{E}_i^\pi[\xi_m]^2 + \mathsf{E}_i^\pi[\xi^{(m,n)}]^2 + 2\cdot\mathsf{E}_i^\pi[\xi_m\cdot\xi^{(m,n)}] \\ &= \mathsf{E}_i^\pi[\xi_m]^2 + \mathsf{E}_i^\pi\{\sum_{j\in\mathcal{I}} \mathsf{P}(X_m = j)\cdot \mathsf{E}_j^\pi[\xi_{n-m}]^2\} \\ &\quad +2\cdot\mathsf{E}_i^\pi\{\xi_m\sum_{j\in\mathcal{I}} \mathsf{P}(X_m = j)\cdot\mathsf{E}_j^\pi[\xi_{n-m}]\}. \end{aligned} \tag{2.2}$$

In particular, for $m = 1$, $n := n+1$ if policy $\pi\sim(f)$ is followed (2.1), (2.2) take on the forms:

$$R_i^\pi(n+1) = r_i^{f_i} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot R_j^\pi(n), \tag{2.3}$$

$$S_i^\pi(n+1) = r_i^{(2),f_i} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot R_j^\pi(n) + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot S_j^\pi(n). \tag{2.4}$$

Since $V_i(\cdot) = S_i(\cdot) - [R_i(\cdot)]^2$ by (2.3), (2.4) we arrive after some algebra at

$$\begin{aligned} V_i^\pi(n+1) &= r_i^{(2),f_i} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot R_j^\pi(n) \\ &\quad - \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[R_i^\pi(n+1)]^2 - [R_j^\pi(n)]^2\} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot V_j^\pi(n). \end{aligned} \tag{2.5}$$

For the so-called transient case, i.e. when $\lim\limits_{n\to\infty} R_i^\pi(n) = R_i^\pi$, $\lim\limits_{n\to\infty} S_i^\pi(n) = S_i^\pi$, $\lim\limits_{n\to\infty} V_i^\pi(n) = V_i^\pi$, (2.3), (2.4) and (2.5) take on the form

$$R_i^\pi = r_i^{f_i} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot R_j^\pi, \tag{2.6}$$

$$S_i^\pi = r_i^{(2),f_i} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot R_j^\pi + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot S_j^\pi, \tag{2.7}$$

$$V_i^\pi = r_i^{(2),f_i} - [R_i^\pi]^2 + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{2\cdot r_{ij}^{f_i}\cdot R_j^\pi + [R_j^\pi]^2\} + \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot V_j^\pi. \tag{2.8}$$

3 Infinite Horizon: General Case

In general, the values $R_i^\pi(n)$, $S_i^\pi(n)$, $V_i^\pi(n)$ are not bounded. In this section we present formulas for the asymptotic behaviour of the reward variance $V_i^\pi(n)$ similar to (1.2) and show that the growth rate of the variance is asymptotically linear in time. The material is taken from [7] and [8]. To this end, we present here only the resulting formulas and sketches of the proofs, for details see [7], [8].

In particular, by (1.2), (1.3), (2.3) for the third term of (2.5), we have after a lot of algebra

$$\begin{aligned}&\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[R_i^\pi(n+1)]^2-[R_j^\pi(n)]^2\}\\ =\;& 2\cdot n\cdot g^\pi\cdot r_i^{f_i}+\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[g^\pi+w_i^\pi]^2-[w_j^\pi]^2\}+\varepsilon(n).\end{aligned}\tag{3.1}$$

Similarly by (1.2) for the second term of (2.5) we have

$$\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot R_j^\pi(n)=n\cdot g^\pi\cdot r_i^{f_i}+\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot w_j^\pi+\varepsilon(n).\tag{3.2}$$

Substitution from (3.1), (3.2) in (2.5) yields

$$V_i^\pi(n+1)=\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot V_j^\pi(n)+s_i(\pi)+\varepsilon(n),\tag{3.3}$$

where

$$\begin{aligned}s_i(\pi) &= \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[r_{ij}^{f_i}+w_j^\pi]^2\}-[g^\pi+w_i^\pi]^2 &(3.4)\\ &= \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[r_{ij}^{f_i}-g^\pi+w_j^\pi]^2\}-[w_i^\pi]^2. &(3.5)\end{aligned}$$

Hence, in matrix form we have:

$$\boldsymbol{V}^\pi(n+1)=\boldsymbol{s}(\pi)+\boldsymbol{P}(f)\cdot\boldsymbol{V}^\pi(n)+\boldsymbol{\varepsilon}^{(1)}(n)\tag{3.6}$$

where for elements of the vector $\boldsymbol{s}(\pi)$ are given by (3.4), (3.5) and elements of the vector $\boldsymbol{\varepsilon}^{(1)}(n)$ converge to zero geometrically (i.e., for some numbers $c>0$, $\delta\in(0,1)$ $\|\boldsymbol{\varepsilon}^{(1)}(n)\|\le c\delta^n$.)

In order to investigate the behavior of $\boldsymbol{V}^\pi(n)$ for large n, observe that for a fixed stationary policy (3.6) is similar to (1.1) up to the term $\boldsymbol{\varepsilon}^{(1)}(n)$ converging geometrically to the null vector. Due to the geometrical convergence, it can be shown (see [7], [8]) that the growth rate of the reward variance defined recursively

$$\boldsymbol{W}^\pi(n+1)=\boldsymbol{s}(\pi)+\boldsymbol{P}(f)\cdot\boldsymbol{W}^\pi(n)\tag{3.7}$$

is the same. Hence in virtue of (1.2) we can conclude that under assumptions AS 1, AS 2 there exists vectors $\boldsymbol{w}^{(2),\pi}$ such that

$$\boldsymbol{W}^\pi(n)=\boldsymbol{g}^{(2),\pi}\cdot n+\boldsymbol{w}^{(2),\pi}+\boldsymbol{\varepsilon}(n)\tag{3.8}$$

hence $\boldsymbol{W}^\pi(n)$ possesses a linear growth rate $g^{(2),\pi}$ in n up to a geometric convergence to the null vector and vectors $\boldsymbol{w}^{(2),\pi}$. The constant vector $\boldsymbol{g}^{(2),\pi}$ along with vector $\boldsymbol{w}^{(2),\pi}$ are uniquely determined by

$$\boldsymbol{w}^{(2),\pi}+\boldsymbol{g}^{(2),\pi}=\boldsymbol{s}(\pi)+\boldsymbol{P}(f)\cdot\boldsymbol{w}^{(2),\pi},\quad \boldsymbol{P}^*(\pi)\cdot\boldsymbol{w}^{(2),\pi}=0\tag{3.9}$$

implying that

$$\boldsymbol{g}^{(2),\pi}=\lim_{n\to\infty}\frac{\boldsymbol{W}^\pi(n)}{n}=\lim_{n\to\infty}\frac{\boldsymbol{V}^\pi(n)}{n}=\boldsymbol{P}^*(\pi)\cdot\boldsymbol{s}(\pi).\tag{3.10}$$

Moreover, it can be shown that (3.10) still holds if elements of the vector $\tilde{\boldsymbol{s}}(\pi)$ take on the following simplified forms

$$\begin{aligned}\tilde{s}_i(\pi) &= \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[r_{ij}^{f_i}]^2+2\cdot r_{ij}^{f_i}\cdot w_j^\pi\}-[g^\pi]^2\\ &= \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot\{[r_{ij}^{f_i}-g^\pi]^2+2\cdot r_{ij}\cdot w_j^\pi\} &(3.11)\\ &= r_i^{(2),f_i}-[g^\pi]^2+2\cdot\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot w_j^\pi. &(3.12)\end{aligned}$$

Hence for the vector $\tilde{\boldsymbol{s}}(\pi)$ with elements $\tilde{s}_i(\pi)$ we can write ($[\cdot]_{\mathrm{sq}}$ denotes that elements of the vector are squared)

$$\boldsymbol{g}^{(2),\pi} = \boldsymbol{P}^*(\pi)\cdot\tilde{\boldsymbol{s}}(\pi) = \boldsymbol{P}^*(\pi)\cdot\{\boldsymbol{r}^{(2)}(f) - [\boldsymbol{g}^\pi]_{\mathrm{sq}} + 2\cdot\tilde{\boldsymbol{r}}(f,\pi)\} \tag{3.13}$$

$$= \bar{\boldsymbol{g}}^{(2),\pi} + 2\cdot\boldsymbol{P}^*(\pi)\cdot\tilde{\boldsymbol{r}}(f,\pi) \tag{3.14}$$

where

$\boldsymbol{r}^{(2)}(f)$ is a column vector with elements $r_i^{(2),f_i} = \sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot[r_{ij}^{f_i}]^2$,
$\tilde{\boldsymbol{r}}(f,\pi)$ is a column vector with elements $\sum_{j\in\mathcal{I}} p_{ij}^{f_i}\cdot r_{ij}^{f_i}\cdot w_j^\pi$,
$[\boldsymbol{g}^\pi]_{\mathrm{sq}}$ is a constant vector with elements $[g^\pi]^2$,
$\tilde{\boldsymbol{g}}^{(2),\pi} = \boldsymbol{P}^*(\pi)\cdot\boldsymbol{r}^{(2)}(f)$ is a constant vector with elements $\sum_{i\in\mathcal{I}} p_i^*(\pi)\cdot r_i^{(2),f_i}$, and
$\bar{\boldsymbol{g}}^{(2),\pi} = \tilde{\boldsymbol{g}}^{(2),\pi} - [\boldsymbol{g}^\pi]_{\mathrm{sq}}$ is a constant vector with elements $\bar{g}^{(2),\pi} = \tilde{g}^{(2),\pi} - [g^\pi]^2$.

Obviously, $\tilde{g}^{(2),\pi}$ averages expected values of the second moments of one-stage rewards, $\bar{g}^{(2),\pi}$ denotes the average "one-stage reward variance" considered with respect to the mean reward g^π instead of the one-stage expected reward $r_i^{f_i}$ in state $i\in\mathcal{I}$, and the last term in (3.14) expresses the Markov dependence that occurs if the second moment and variance of cumulative rewards are considered (cf. also (2.2)).

4 Mean Variance Selection Rules

First observe that by (1.1), (3.10) in the class of stationary policies limiting behaviour of the ratio of total reward variance to total reward is well defined, in particular, $\lim_{n\to\infty} V_i^\pi(n)/R_i^\pi(n) = g^{(2),\pi}/g^\pi$ for every $i\in\mathcal{I}$. This justifies the introduction of the following mean variance selection rules (considered for simplification only in the class of stationary policies):

Stationary policy $\hat{\pi}\sim(\hat{f})$ is called *δ-mean variance optimal* (with $\delta\in\langle 0,1\rangle$) if for every stationary policy $\pi\sim(f)$

$$h^\delta(\hat{\pi}) = \delta\cdot\frac{g^{(2),\hat{\pi}}}{g^{\hat{\pi}}} - (1-\delta)\cdot g^{\hat{\pi}} \le h^\delta(\pi) = \delta\cdot\frac{g^{(2),\pi}}{g^\pi} - (1-\delta)\cdot g^\pi. \tag{4.1}$$

In particular, 1-mean variance optimal policy is the mean variance optimal policy, i.e. it minimizes the ratio $g^{(2),\pi}/g^\pi$, 0-mean optimal policy maximizes the average reward.

Similarly, the *square mean variance optimal policy* $\pi^\star$ minimizes the ratio of the mean variance to the squared mean reward in the class of stationary policies, i.e. stationary policy $\pi^\star\sim(f^\star)$ is called *square mean variance optimal* if for every policy $\pi\sim(f)$

$$h^{(2)}(\pi) = \frac{g^{(2),\pi}}{[g^\pi]^2} \ge \frac{g^{(2),\pi^\star}}{[g^{\pi^\star}]^2} = h^{(2)}(\pi^\star). \tag{4.2}$$

Using formulas (1.2) and (3.10), (3.13) for a given stationary policy $\pi\sim(f)$ we can easily calculate the values of $h^\delta(\pi)$ and $h^{(2)}(\pi)$ to evaluate the policy according to δ-mean variance and square mean variance optimality criteria. However, this approach is not possible when we are seeking δ-mean variance optimal and square mean variance optimal policies; some algorithmic procedures must be employed not to evaluate all stationary policies. Unfortunately, there is the following substantial difference in the both formulas: in (1.2) the ith element of the vector $\boldsymbol{r}(f)$ depends only on the action selected in state i, however in (3.4), (3.5) and in the simplified form given by (3.11), (3.12) the ith elements of the vectors $\boldsymbol{s}(\pi)$, $\tilde{\boldsymbol{s}}(\pi)$ depend on the decisions taken in all states through the last term occurring in the right-hand side of $\tilde{s}_i(\pi)$. For this reason for the mean variance tradeoff in the literature only "one-stage reward variance" (called also the myopic variance) instead of the variance of total (cumulative) rewards is considered,

i.e., we ignore the last term in (3.11), (3.12) and employ the fact that $\boldsymbol{g}^\pi$ is a constant vector. This "one-stage reward variance" simplification enables to employ some techniques of stochastic dynamic programming, however, considering "one-stage reward variance" simplification may select policies not optimal with respect to the above criteria.

In [7] a policy iteration algorithm was suggested for finding the convex hull of all stationary policies (see Figure 1) and it was shown that if the "one-stage reward variance" is considered the δ-mean variance optimal and square mean optimal policies can be selected in the vertices of the south-east boundary of this convex hull. It it important that the stationary policies corresponding to adjacent vertices of the convex hull select different actions only in one state.

In the Appendix we modify this procedure to incorporate also the term $2 \cdot \boldsymbol{P}^*(\pi) \cdot \tilde{\boldsymbol{r}}(f, \pi)$ (i.e. the last term in (3.14)) to work with the variance of cumulative rewards.

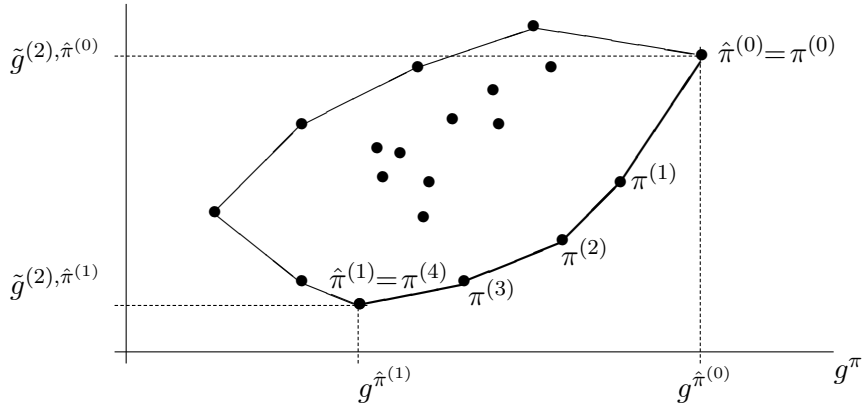


Figure 1: Convex hull of the set of stationary policies.

References

[1] F. Benito: Calculating the variance in Markov processes with random reward. Trabajos de Estadistica y de Investigacion Operativa *33* (1982), 3, 73–85.

[2] R. A. Howard: Dynamic Programming and Markov Processes. MIT Press, Cambridge, Mass. 1960.

[3] H. Kawai: A variance minimization problem for a Markov decision process. European J. Oper. Research *31* (1987), 140–145.

[4] P. Mandl: On the variance in controlled Markov chains. Kybernetika *7* (1971), 1–12.

[5] M. L. Puterman: Markov Decision Processes – Discrete Stochastic Dynamic Programming. Wiley, New York 1994.

[6] S. M. Ross: Applied Probability Models with Optimization Applications. Holden–Day, San Francisco, Calif. 1970.

[7] K. Sladký and M. Sitař: Optimal solutions for undiscounted variance penalized Markov decision chains. In: Dynamic Stochastic Optimization (LNEMS 532, K. Marti, Y. Ermoliev, and G. Pflug, eds.) Springer, Berlin 2004, pp. 43–66.

[8] K. Sladký: On mean reward variance in semi-Markov processes. Mathematical Methods of Operations Research, to appear.

[9] M. J. Sobel: The variance of discounted Markov decision processes. J. Appl. Probab. *19* (1982), 794–802.

[10] M. J. Sobel: Maximal mean/standard deviation ratio in an undiscounted MDP. Oper. Research Lett. *4* (1985), 157–159.

[11] D. J. White: Mean, variance and probability criteria in finite Markov decision processes: A review. J. Optim. Theory Appl. *56* (1988), 1–29.

Appendix

In what follows we modify algorithmic procedure for finding convex hull of stationary policies by adding to the values $\tilde{g}^{(2),\pi}$ also the term $2 \cdot \boldsymbol{p}\ (\pi) \cdot \tilde{\boldsymbol{r}}(f,\pi)$. This algorithmic procedure will generate a finite sequence of stationary policies

$$\hat{\pi}^{(1)} \equiv \pi^{(0)} \sim (f^{(0)}),\ \ \pi^{(1)} \sim (f^{(1)}),\ \ldots,\ \pi^{(k)} \sim (f^{(k)}),\ \ldots,\ \pi^{(K)} \sim (f^{(K)})$$

with decreasing mean rewards that are suboptimal with respect to the growth rate of the variance of cumulative rewards.

To this end, let for stationary policies $\tilde{\pi} \sim (\tilde{f})$, $\pi \sim (f)$

$$\boldsymbol{\gamma}^{(1)}(f,\tilde{\pi}) \ = \ \boldsymbol{r}(f) + [\boldsymbol{P}(f) - \boldsymbol{I}] \cdot \boldsymbol{w}^{\tilde{\pi}} - \boldsymbol{g}^{\tilde{\pi}}, \tag{4.3}$$

$$\boldsymbol{\gamma}^{(2)}(f,\tilde{\pi}) \ = \ \boldsymbol{r}^{(2)}(f) + 2 \cdot \tilde{\boldsymbol{r}}(f,\tilde{\pi}) + [\boldsymbol{P}(f) - \boldsymbol{I}] \cdot \hat{\boldsymbol{w}}^{(2),\tilde{\pi}} - \hat{\boldsymbol{g}}^{(2),\tilde{\pi}} \tag{4.4}$$

where $\boldsymbol{w}^{\tilde{\pi}}$, $\boldsymbol{g}^{\tilde{\pi}}$ are calculated from (1.3) and $\hat{\boldsymbol{w}}^{(2),\tilde{\pi}}$, $\hat{\boldsymbol{g}}^{(2),\tilde{\pi}}$ are obtained by

$$\hat{\boldsymbol{w}}^{(2),\pi} + \hat{\boldsymbol{g}}^{(2),\pi} = \boldsymbol{r}^{(2)}(f) + 2 \cdot \tilde{\boldsymbol{r}}(f,\pi) + \boldsymbol{P}(f) \cdot \hat{\boldsymbol{w}}^{(2),\pi}, \quad \boldsymbol{P}^{*}(\pi) \cdot \hat{\boldsymbol{w}}^{(2),\pi} = 0. \tag{4.5}$$

For $\ell = 1, 2$ $\gamma_i^{(\ell)}(\cdot,\tilde{\pi})$ denotes the ith element of the vector $\boldsymbol{\gamma}^{(\ell)}(\cdot,\tilde{\pi})$, in particular,

$$\gamma_i^{(1)}(f_i,\tilde{\pi}) \ = \ r_i^{f_i} + \sum_{j \in \mathcal{I}} [p_{ij}^{f_i} - \delta_{ij}] \cdot\ w_j^{\tilde{\pi}} - g^{\tilde{\pi}}, \tag{4.6}$$

$$\gamma_i^{(2)}(f_i,\tilde{\pi}) \ = \ r_i^{(2),f_i} + 2 \cdot \sum_{j \in \mathcal{I}} p_{ij}^{f_i} \cdot r_{ij}^{f_i} \cdot w_j^{\tilde{\pi}} + \sum_{j \in \mathcal{I}} [p_{ij}^{f_i} - \delta_{ij}] \cdot\ \hat{w}_j^{(2),\tilde{\pi}} - \hat{g}^{(2),\tilde{\pi}} \tag{4.7}$$

(δ_{ij} is reserved for the Kronecker symbol).

Moreover, let

$$\bar{\gamma}_i^{(2)}(f_i,\hat{\pi},\tilde{\pi}) = r_i^{(2),f_i} + 2 \cdot \sum_{j \in \mathcal{I}} p_{ij}^{f_i} \cdot r_{ij}^{f_i} \cdot w_j^{\hat{\pi}} + \sum_{j \in \mathcal{I}} [p_{ij}^{f_i} - \delta_{ij}] \cdot\ \hat{w}_j^{(2),\tilde{\pi}} - \hat{g}^{(2),\tilde{\pi}}, \tag{4.8}$$

i.e., in contrast to (4.7) we select in (4.8) different policy for the values $w_j^{\hat{\pi}}$.

Now we are ready to suggest a heuristic algorithmic procedure for generating a sequence of "efficient" stationary policies $\{\pi^{(k)} \sim (f^{(k)}), k = 0, 1, \ldots, K\}$ (similar to that in Figure 1) using the complete formula for the growth rate of the variance of cumulative rewards.

Algorithm. Let (stationary) policy $\hat{\pi}^{(0)} \sim (\hat{f}^{(0)})$ maximizes the mean reward.

Construct a (finite) sequence of stationary policies (for $k = 0, 1, \ldots$)

$$\hat{\pi}^{(1)} \equiv \pi^{(0)} \sim (f^{(0)}),\ \ \pi^{(1)} \sim (f^{(1)}),\ \ldots,\ \pi^{(k)} \sim (f^{(k)}),\ \ldots \tag{4.9}$$

such that

(i)

$$f_i^{(k)} \neq f_i^{(k+1)} \text{ for one } i_k = i, \quad f_i^{(k)} = f_i^{(k+1)} \text{ for every } i \neq i_k, \tag{4.10}$$

(ii)

$$a^{(k+1)} := \frac{\gamma_{i_k}^{(2)}(f_{i_k}^{(k+1)},\pi^{(k)})}{\gamma_{i_k}^{(1)}(f_{i_k}^{(k+1)},\pi^{(k)})} = \max_{i \in \mathcal{I}} \max_{\substack{k \in \mathcal{A}_i \\ \gamma_i^{(1)}(k,\pi^{(k)}) < 0}} \frac{\gamma_i^{(2)}(k,\pi^{(k)})}{\gamma_i^{(1)}(k,\pi^{(k)})} \tag{4.11}$$

(iii) For the selected decision vector $f^{(k+1)}$ calculate the values $\boldsymbol{w}^{\tilde{\pi}}$, $\hat{\boldsymbol{g}}^{(2),\tilde{\pi}}$ by (1.3) and the values $\hat{\boldsymbol{w}}^{(2),\tilde{\pi}}$, $\boldsymbol{g}^{\tilde{\pi}}$ by (4.5).

(iv) Check whether also

$$\frac{\bar{\gamma}_{i_k}^{(2)}(f_{i_k}^{(k+1)},\pi^{(k+1)},\pi^{(k)})}{\gamma_{i_k}^{(1)}(\pi_{i_k}^{(k+1)},\pi^{(k)})}=\max_{i\in\mathcal{I}}\max_{\substack{k\in\mathcal{A}_i\\ \gamma_i^{(1)}(k,\pi^{(k)})<0}}\frac{\bar{\gamma}_i^{(2)}(k,\pi^{(k+1)},\pi^{(k)})}{\gamma_i^{(1)}(k,\pi^{(k)})} \tag{4.12}$$

If (4.12) is fulfilled, fix the improved policy $\pi^{(k+1)}\sim(f^{(k+1)})$ and go to **(i)**, else improve policy $\pi^{(k+1)}\sim(f^{(k+1)})$ in one state using

$$a^{(k+1)}:=\frac{\gamma_{i_k}^{(2)}(f_{i_k}^{(k+2)},\pi^{(k+1)},\pi^{(k)})}{\gamma_{i_k}^{(1)}(f_{i_k}^{(k+2)},\pi^{(k)})}=\max_{i\in\mathcal{I}}\max_{\substack{k\in\mathcal{A}_i\\ \gamma_i^{(1)}(k,\pi^{(k)})<0}}\frac{\gamma_i^{(2)}(k,\pi^{(k+1)},\pi^{(k)})}{\gamma_i^{(1)}(k,\pi^{(k)})}. \tag{4.13}$$

and go to **(i)**.

Remark. The sequence of stationary policies fulfilling conditions (4.9), (4.10), (4.11) is generated by an algorithmic procedure of a policy iteration type. Since the decision vectors $f^{(k)}$, $f^{(k+1)}$ $(k=0,1,\ldots)$ differ only in one state, to calculate the values $\boldsymbol{w}^{\pi^{(k)}}$, $\boldsymbol{w}^{(2),\pi^{(k)}}$, $\boldsymbol{g}^{\pi^{(k)}}$, $\boldsymbol{g}^{(2),\pi^{(k)}}$ for an improved policy $\pi^{(k+1)}\sim(f^{(k+1)})$ it is not necessary to solve repeatedly the set of equation (1.3), (3.9), but only to update the current values of $\boldsymbol{w}^{\pi^{(k)}}$, $\hat{\boldsymbol{w}}^{(2),\pi^{(k)}}$, $\boldsymbol{g}^{\pi^{(k)}}$, $\hat{\boldsymbol{g}}^{(2),\pi^{(k)}}$.

Observe that steps (i), (ii), (iii) of the Algorithm well corresponds to the policy iteration algorithms for finding the convex full of all stationary policies (depicted in Figure 1) on condition that in (4.4), (4.5), (4.7) the term $2\cdot\tilde{\boldsymbol{r}}(f,\pi)$ is ignored.

Unfortunately, if it is necessary to employ also condition (4.13) is some steps, it may happen that the Algorithm need not hold the nice properties as for the "myopic" ("one-stage reward variance") case, e.g., the subsequent members of the sequence of the policies $\pi^{(k)}$ need not differ only in one state.

Computational experience. The above algorithms were tested on many large scale numerical examples (with dimensions of the state space $\mathcal{I}$ and each action set $\mathcal{A}_i$ up to 100 elements, hence there exists up to $100^{100}=10^{200}$ stationary policies). Using a standard Pentium computer generating the set of all efficient stationary policies (i.e. finding all edges of the convex hull of the set of stationary policies located on the south-east boundary of the convex hull) along with evaluating respective mean and variance of each efficient policy took less than 8 minutes. The number of efficient policies (located on the south-east boundary of the convex hull of all stationary policies, cf. Figure 1) is less than 300 elements. Moreover, numerical examples proved that the number of efficient policies grows linearly in the number of states and logarithmically in the number of admissible actions.

Forecasting in Continuous Double Auction

Martin Šmíd
mailto:martin@klec.cz, http://www.klec.cz/martin

Institute of Information Theory and Automation,
Academy of Sciences of the Czech Republic
182 08 Prague, Pod Vodárenskou věží 4, Czech Republic

Abstract. Recently, the continuous double auction, i.e. the trading mechanism used in the majority of the financial markets, is the subject of an extensive study. In the present paper, a model of the continuous double auction with the completely random flow of the limit orders is studied. The main result of the paper is an approximate formula for the distribution of the market price and the traded volume at the time τ given the information available at $t < \tau$.[1]

Keywords: limit order markets, continuous double auction, price and volume, forecasting, market microstructure

AMS classification: 91B26,
JEL classification: C51,G10

1 Introduction

In the present paper, the trading mechanism with the following rules is studied:

1. At any time instant, each agent may place a *buy (limit) order* or a *sell (limit) order*, each limit order containing a *limit price* and an *order size* (i.e. the required/offered amount of the commodity). For simplicity, we assume the order size to be unit.[2]
2. If a newly arrived limit order matches with the best waiting limit order of the opposite type (i.e. the one with the most favorable limit price, let us call it *best counterpart*) then a trade is made for the limit price of the best counterpart (if there is more then one counterpart with the best limit price then the oldest one, i.e. the one with the earliest placement date, is executed).
3. If a newly arrived limit order finds no counterpart then it remains waiting until it is executed or canceled by its submitter.

The trading mechanism, described here, is usually called *continuous double auction* (CDA), the list of all the currently waiting buy orders is called *buy limit order book*, the highest limit price of the orders contained in the buy limit order book is called *(best) bid*, the list of all the currently waiting sell orders is called *sell limit order book* and the lowest limit price of the orders contained in the sell limit order book is called *(best) ask*.

In reality, many markets possess the structure described above: many financial markets, first of all, various marketplaces, real estate markets, trading made by means of the advertising in newspapers etc.

In the present paper, the complete randomness of the agents' actions is assumed. In particular, the times of the arrivals of the limit orders are assumed to follow a Poisson process, their limit prices are regarded as i.i.d. random variables independent of the arrival times and the lifetimes of the limit orders are supposed to be exponentially distributed independent both of the arrival times and of the limit prices.

The model introduced by the present paper is a generalization of the model of Smith et al. [2003]; the generalizations consist in possibly non-uniform density of the limit prices, possibly continuous price space (the lack of ticks) and in continuous (not discretized) time domain. The main result of the present paper is an approximate formula for the future distribution of the market price and of the total traded volume.

[1] This work was supported by grant no. 402/04/1294 and by grant no. 402/03/H057 of the Czech Science Foundation and by grant no. 454/2004/AEK/FSV of the Grant Agency of the Charles University.

[2] Cf. Smith et al. [2003] for a partial justification of the assumption of the unit order size.

The paper is organized as follows: in Section 2, the model of the CDA with complete randomness is defined, in Section 3, the forecast of the price and the volume is constructed. Section 4 concludes the paper.

2 Continuous Double Auction with Complete Randomness

2.1 Definitions

Denote $\tau^1_{b\star} < \tau^2_{b\star} < \dots$ the times of the arrivals of the buy orders. For each $i \in \mathbb{N}$, denote x^i the limit price of the i-th buy order[3] and denote $\tau^i_{b\dagger}$ the time at which the i-th buy order is canceled provided that it is not executed until $\tau^i_{b\dagger}$ ($\tau^i_{b\dagger}$ will be called *cancelation time*). Analogously, denote $\tau^1_{s\star} < \tau^2_{s\star} < \dots$ the arrival times of the sell orders, y^i the limit price of the i-th sell order and $\tau^i_{s\dagger}$ its cancelation time for each $i \in \mathbb{N}$.

According to the informal description of the CDA, given in Introduction, a buy order may find itself in four possible states: *prenatal* (not yet arrived), *waiting*, *executed* and *canceled*.

Denote $X^i_t, Y^i_t \in \{prenatal, waiting, executed, canceled\}$, $t \in \mathbb{R}^+$, the state of the i-th buy order, sell order respectively, at the time t for each $i \in \mathbb{N}$, $t \in \mathbb{R}^+$. Further, denote $\mathcal{N}_{\mathbb{R}}$ the space of all the counting measures[4] on $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$. The *buy limit order book* may be mathematically described as

$$B_t \in \mathcal{N}_{\mathbb{R}}, \qquad B_t(A) \stackrel{\triangle}{=} |\{i : x^i \in A, X^i_t = waiting\}|, \qquad A \in \mathbb{B}(\mathbb{R})$$

(the symbol $|\bullet|$ denotes the number of elements of the set, $\mathbb{B}(\Xi)$ stands for the Borel σ-algebra of a metric space Ξ). Symmetrically, we describe the *sell limit order book* as

$$S_t \in \mathcal{N}_{\mathbb{R}}, \qquad S_t(A) \stackrel{\triangle}{=} |\{i : y^i \in A, Y^i_t = waiting\}|, \qquad A \in \mathbb{B}(\mathbb{R}).$$

Finally, define the *i-th (best) bid* as $b^i_t = \max\{p \in \mathbb{R} : B_t[p, \infty) \geq i\}$ and the *i-th (best) ask* as $a^i_t = \min\{p \in \mathbb{R} : S_t(-\infty, p] \geq i\}$ for each $i \in \mathbb{N}$ (it is understood that $\max \emptyset = -\infty$ and $\min \emptyset = \infty$).

2.2 The Dynamics of the System

Assume, throughout the present subsection, that no pair of the events (i.e. the arrivals of the orders and their cancelations) happens at the same time. According to the informal definition, the process $(X^1_t, Y^1_t, X^2_t, Y^2_t, \dots)$ evolves according to the following rules:

- The i-th buy order is in the state *prenatal* at the time 0 for each $i \in \mathbb{N}$.
- For each $i \in \mathbb{N}$, X^i_t may jump only at the times $\tau^i_{b\star}, \tau^i_{b\dagger}, \tau^1_{s\star}, \tau^2_{s\star}, \dots$ as follows:
  - *When $t = \tau^i_{b\star}$:* If $x^i < a^1_{t-}$ then $X^i_t = waiting$, otherwise $X^i_t = executed$.
  - *When $t = \tau^i_{b\dagger}$:* If $X^i_{t-} = waiting$ then $X^i_t = canceled$, otherwise $X^i_t = X^i_{t-}$ ($= executed$).
  - *When $t = \tau^j_{s\star}$ for some $j \in \mathbb{N}$:* If the i-th buy order is currently the best buy order (i.e. the oldest of all the waiting buy orders with the limit price b^1_{t-}) and $x^j \leq b^1_{t-}$ then $X^i_t = executed$, otherwise $X^i_t = X^i_{t-}$.
- The symmetric rules hold for the processes Y^i, $i \in \mathbb{N}$.

2.3 The Market Price and the Traded Volume

We naturally define the *market price* p_t and the *traded volume* q_t as follows:

- $p_0 = undefined$,[5] $q_0 = 0$.
- The process (p_t, q_t) is piecewise constant right continuous and it they may jump only at the times $\tau^1_{b\star}$, $\tau^1_{s\star}$, $\tau^2_{b\star}$, $\tau^2_{s\star}$, ... as follows:
  - *When $t = \tau^i_{b\star}$ for some $i \in \mathbb{N}$:* If the newly buy arrived order was executed at t then $p_t = a^1_t$ and $q_t = q_{t-} + 1$, otherwise $(p_t, q_t) = (p_{t-}, q_{t-})$.
  - *When $t = \tau^i_{s\star}$ for some $i \in \mathbb{N}$:* If the newly arrived sell order was executed at t then $p_t = b^1_t$ and $q_t = q_{t-} + 1$, otherwise $(p_t, q_t) = (p_{t-}, q_{t-})$.

[3] i.e. the one with the arrival time $\tau^i_{b\star}$

[4] The measure is *counting* if its values on the measurable sets are nonnegative integers.

[5] We take p_t as a process taking values in the space $\mathbb{R}^{undefined} \stackrel{\triangle}{=} \mathbb{R} \cup \{undefined\}$.

2.4 The Stochastic Properties of the Order Flow

Assume that there exist a probability measure π and positive constants $u, c \in \mathbb{R}^+$ such that $x^i \sim \pi$, $\Delta\tau^i_{b\star} \triangleq (\tau^i_{b\star} - \tau^{i-1}_{b\star}) \sim \mathrm{Exp}(u)$ and $\Delta\tau^i_{b\dagger} \triangleq (\tau^i_{b\dagger} - \tau^i_{b\star}) \sim \mathrm{Exp}(c)$ for each $i \in \mathbb{N}$ (the symbol $\mathrm{Exp}(m)$ denotes the exponential distribution with mean $1/m$).

Analogously, suppose that there exist a probability measure ρ and positive constants $v, d \in \mathbb{R}^+$ such that $y^i \sim \rho$, $\Delta\tau^i_{s\star} \triangleq (\tau^i_{s\star} - \tau^{i-1}_{s\star}) \sim \mathrm{Exp}(v)$ and $\Delta\tau^i_{s\dagger} \triangleq (\tau^i_{s\dagger} - \tau^i_{s\star}) \sim \mathrm{Exp}(d)$ for each $i \in \mathbb{N}$.

Finally, assume that $\Delta\tau^1_{b\star}, \Delta\tau^1_{b\dagger}, x^1, \Delta\tau^1_{s\star}, \Delta\tau^1_{s\dagger}, y^1, \Delta\tau^2_{b\star}, \Delta\tau^2_{b\dagger}, x^2, \Delta\tau^2_{s\star}, \Delta\tau^2_{s\dagger}, y^2, \ldots$ are mutually independent.

Since both the arrivals of the buy orders and the arrivals and the sell orders follow the time-spatial Poisson processes [Šmíd, 2005a] and since, in both the cases, the cancelation times are independent on the arrivals, we call our setting *complete random arrival of the orders*.

Denote $(\tau^i)_{i=1}^{\infty}$ the increasing sequence of the elements of the set $\{\tau^i_{b\star} : i \in \mathbb{N}\} \cup \{\tau^i_{b\dagger} : i \in \mathbb{N}\} \cup \{\tau^i_{s\star} : i \in \mathbb{N}\} \cup \{\tau^i_{s\dagger} : i \in \mathbb{N}\}$ and put $\tau^0 = 0$. From the absolute continuity of the exponential distribution it follows that the times $\tau^0, \tau^1, \ldots$ mutually differ almost sure hence the dynamics of our system is well defined by subsection 2.2.

3 Forecasts of the Market Price and the Traded Volume

As it was already mentioned, our aim is a forecast of (p_τ, q_τ) given the state of the system up to the time $t < \tau$. Since $\Xi_t \triangleq (B_t, S_t, p_t, q_t)$ is a continuous time Markov chain [Šmíd, 2005b, Theorem 1.], the forecast can be based solely on the state of the system at the time t. Moreover, when we modify our definition so that Ξ_0 may take other values then $(0, 0, undefined, 0)$, we may assume that $t = 0$.

Fix $\tau \geq 0$ and $\Xi_0 = (B_0, S_0, p_0, q_0) \in \mathcal{N}_{\mathbb{R}} \times \mathcal{N}_{\mathbb{R}} \times \mathbb{R}^{undefined} \times \mathbb{Z}^+$, $\mathbb{R}^{undefined} \triangleq \mathbb{R} \cup \{undefined\}$. To construct the forecast for the time τ, we shall use the usual technique, i.e. the expansion according to the number of the events:

$$\mathbb{P}((p_\tau, q_\tau) \in A) = \sum_{k=0}^{\infty} \mathbb{P}((p_\tau, q_\tau) \in A | n_\tau = k) \mathbb{P}(n_\tau = k) \tag{1}$$

where n_τ is the number of the jumps of Ξ until τ. However, since the inter-event times of the process Ξ are dependent on the state of the process, the evaluation of (1) could be quite complicated. Hence, we have to modify the process Ξ so that its inter-event times are i.i.d. first.

3.1 Uniformization

Let $N \in \mathbb{N}$ be a sufficiently large constant and let $(\theta^i_v)_{i=1}^{\infty}$ be i.i.d. exponential variables with mean one independent of Ξ. Let v be a counting process[6] starting from zero. Denote $(\bar\tau^i)_{i=1}^{\infty}$ the times of the jumps of the process $\bar\Xi \triangleq (\Xi, v) = (B, S, p, q, v)$ defined by the following rules:

- $\bar\tau^0 = 0$,
- if there is no event of Ξ between $\bar\tau^{i-1}$ and $\bar\tau^{i-1} + \Delta\tau^i_v$, where $\Delta\tau^i_v = [(N - |B_{\bar\tau^{i-1}}|)c + (N - |S_{\bar\tau^{i-1}}|)d]^{-1}\theta^i_v$, then let the process v jump at $\bar\tau^{i-1} + \Delta\tau^i_v$, otherwise leave v unchanged at $\bar\tau^{i-1} + \Delta\tau^i_v$.

It could be shown [Šmíd, 2005b, Theorem 2] that $\bar\Xi_t$ is a Markov chain with

$$\Delta\bar\tau^i \triangleq \bar\tau^i - \bar\tau^{i-1} \sim \mathrm{Exp}(\bar r), \quad \bar r = u + v + N(c + d), \quad \text{for each } i \leq \tilde N \tag{2}$$

where $\tilde N = \tilde N(N) \triangleq N - \max(|B_0|, |S_0|) + 1$ such that

$$(\Delta\bar\tau^i)_{i=1}^{\tilde N} \text{ are i.i.d and independent of } \bar\Xi_{(k)} = (B_{(k)}, S_{(k)}, p_{(k)}, q_{(k)}, v_{(k)}) \triangleq \bar\Xi_{\bar\tau^k}.^{7} \tag{3}$$

[6] i.e. piecewise constant right continuous with unit increments

[7] The process $(\bar\Xi_{(k)})_{k=1}^{\infty}$ is usually called *embedded chain*).

3.2 The Expansion and a Truncation

Denote $\bar{n}_\tau$ the number of the jumps of $\bar{\Xi}$ up to the time τ and let $A \in \mathbb{R}^{undefined} \times \mathbb{Z}^+$. Clearly, we may write

$$\mathbb{P}((p_\tau, q_\tau) \in A) = \sum_{k=0}^{\infty} \mathbb{P}((p_{(k)}, q_{(k)}) \in A | \bar{n}_\tau = k) \mathbb{P}(\bar{n}_\tau = k) = U_{A,\tilde{N}} + \eta_1$$

where

$$U_{A,\tilde{N}} = \sum_{k=0}^{\tilde{N}} \mathbb{P}((p_{(k)}, q_{(k)}) \in A | \bar{n}_\tau = k) \mathbb{P}(\bar{n}_\tau = k) \overset{(2),(3)}{=} \sum_{k=0}^{\tilde{N}} \mathbb{P}((p_{(k)}, q_{(k)}) \in A) \mathbb{P}(\text{Poisson}(\bar{r}) = k)$$

and where

$$\eta_1 = \eta_1(\tilde{N}) \leq \sum_{k=\tilde{N}+1}^{\infty} \mathbb{P}(\bar{n}_\tau = k) = 1 - \sum_{k=1}^{\tilde{N}} \mathbb{P}(\bar{n}_\tau = k) = \mathbb{P}(\text{Poisson}(\bar{r}) \geq \tilde{N} + 1).$$

It is straightforward that, to compute $\mathbb{P}((p_\tau, q_\tau) \in A)$ with a required accuracy, it suffices to choose N sufficiently large and to evaluate $U_{A,\tilde{N}}$. We deal with the latter task in the next subsection.

3.3 Forecasting of the Embedded Chain

Fix $k \leq \tilde{N}$ all through the present subsection. It follows from the basic probability theory that, to specify the distribution of $(p_{(k)}, q_{(k)})$, it suffices to evaluate $\mathbb{P}((p_{(k)}, q_{(k)}) \in A)$ for all the sets $A = I \times \{0, 1, \ldots, \hat{q}\}$ where $\hat{q} \in \mathbb{N}$ and where $I = (-\infty, \hat{p})$ for some $\hat{p} \in \mathbb{R}$.

Hence, fix $A = I \times \{0, 1, \ldots, \hat{q}\}$. Further, choose a disjoint partition $\mathcal{M} = \{M^1, M^2, \ldots, M^m\}$ of $\mathbb{R}$ containing all the points of $B_{(0)}$ and all the points of $S_{(0)}$[8] such that $I = M^1 \cup M^2 \cup \cdots \cup M^{m'}$ for some $m' \leq m$. Clearly,

$$\mathbb{P}((p_{(k)}, q_{(k)}) \in A) = \mathbb{P}((\phi_{(k)}, q_{(k)}) \in B)$$

where $B = B(A) \triangleq \{1, 2, \ldots, m'\} \times \{0, 1, \ldots, \hat{q}\}$ and

$$\phi_{(k)} = \begin{cases} 0 & \text{if } p_{(k)} = undefined \\ 1 & \text{if } p_{(k)} \in M^1 \\ \ldots . \end{cases}$$

Further, define the random elements $e_{(i)} \in E \triangleq \{v, b^1_\star, s^1_\star, b^1_\dagger, s^1_\dagger, b^2_\star, s^2_\star, b^2_\dagger, s^2_\dagger, \ldots, b^m_\star, s^m_\star, b^m_\dagger, s^m_\dagger\}$, $i = 1, 2, \ldots, k$, such that $e_{(i)} = v$ if a void event happened, i.e. v was incremented, at the "time" i and, for each $1 \leq j \leq m$,

$$e_{(i)} = \begin{cases} b^j_\star & \text{if a new buy order whose limit price lies in } M^j \text{ arrived at } i \\ s^j_\star & \text{if a new sell order whose limit price lies in } M^j \text{ arrived at } i \\ b^j_\dagger & \text{if one of the waiting buy orders whose limit price lies in } M^j \text{ was canceled at } i \\ s^j_\dagger & \text{if one of the waiting sell orders whose limit price lies in } M^j \text{ was canceled at } i. \end{cases}$$

After denoting $e_{(1,k)} \triangleq (e_{(1)}, e_{(2)}, \ldots, e_{(k)})$ and putting $\mathcal{E}^k \triangleq \otimes_{\nu=1}^k E^\nu$, we may write

$$\mathbb{P}((\phi_{(k)}, q_{(k)}) \in B) = \sum_{\epsilon \in \mathcal{E}^k} \mathbb{P}((\phi_{(k)}, q_{(k)}) \in B | e_{(1,k)} = \epsilon) \mathbb{P}(e_{(1,k)} = \epsilon).$$

Unfortunately, some of the summands of the latter formula are not easy to compute: Suppose, for instance, that $k = 4$, $p_{(0)} = b^1_{(0)}$, $\mathcal{M} = \{(-\infty, b^1_{(0)}), \{b^1_{(0)}\}, (b^1_{(0)}, a^1_{(0)}), \{a^1_{(0)}\}, (a^1_{(0)}, \infty)\}$ and

[8] i.e. $\mathcal{M} \supseteq \{\{y\} : B_{(0)}\{y\} > 0 \vee S_{(0)}\{y\} > 0\}$

$\epsilon = (b^3_\star, s^3_\star, b^3_\star, s^3_\star)$. Here, $\mathbb{P}(\phi_{(k)} = 3|e_{(1,k)} = \epsilon) = \mathbb{P}(\text{at least one pair of the new orders matches})$ which is not trivial but computable. However, after we add $s^3_\dagger$ or $b^3_\dagger$ into $e_{(1,k)}$, the evaluation starts to lead to untractable combinatorial problems.

Fortunately, by the refinement of the partition $\mathcal{M}$, the total probability of the "scenarios" $e_{(1,k)}$ for which the conditional distribution is problematic to compute may be arbitrarily decreased: Denote $\beta^j_\epsilon = |\{i \in \mathbb{N} : i \le k : \epsilon^i = b^j_\star\}|$ and $\sigma^j_\epsilon = |\{i \in \mathbb{N} : i \le k : \epsilon^i = s^j_\star\}|$ for each $1 \le j \le m$ (ϵ^i denotes the i-th component of ϵ) and put

$$\tilde{\mathcal{E}}^k \stackrel{\triangle}{=} \{\epsilon \in \mathcal{E}^k : \beta^j_\epsilon = 0 \vee \sigma^j_\epsilon = 0 \text{ for each } j \in C\}, \qquad C \stackrel{\triangle}{=} \{1 \le j \le m : |M^j| > 1\}.$$

Clearly,

$$\mathbb{P}((\phi_{(k)}, q_{(k)}) \in B) = \sum_{\epsilon \in \tilde{\mathcal{E}}^k} \mathbb{P}((\phi_{(k)}, q_{(k)}) \in B | e_{(1,k)} = \epsilon)\mathbb{P}(e_{(1,k)} = \epsilon) + \eta^k_2 \tag{4}$$

where $\eta^k_2 = \eta^k_2(\mathcal{M}) \le \mathbb{P}(e_{(1,k)} \in \mathcal{E}^k - \tilde{\mathcal{E}}^k)$. It is relatively easy to compute both the conditional and the unconditional probabilities in (4): For any real measure μ, denote $\mu^{\mathcal{M}}$ the restriction of μ to $\sigma(\mathcal{M})$. Since $e_{(i)}$ is conditionally independent of $e_{(1,i-1)}$ given $(B^{\mathcal{M}}_{(i-1)}, S^{\mathcal{M}}_{(i-1)})$ for each $1 \le i \le k$,[9] since

$$\mathbb{P}(e_{(i)} = \varepsilon | B^{\mathcal{M}}_{(i-1)} = B, S^{\mathcal{M}}_{(i-1)} = S) = \gamma(\varepsilon, B, S), \qquad \gamma(\varepsilon, B, S) = \begin{cases} \frac{u\pi(M^j)}{\bar{r}} \text{ if } \varepsilon = b^j_\star \text{ for some } j \\ \frac{v\rho(M^j)}{\bar{r}} \text{ if } \varepsilon = s^j_\star \text{ for some } j \\ \frac{cB(M_j)}{\bar{r}} \text{ if } \varepsilon = b^j_\dagger \text{ for some } j \\ \frac{dS(M_j)}{\bar{r}} \text{ if } \varepsilon = s^j_\dagger \text{ for some } j \\ \frac{(N-|B|)c+(N-|S|)d}{\bar{r}} \text{ if } \varepsilon = v \end{cases}$$

for each pair of counting measures B, S defined on $(\mathbb{R}, \sigma(\mathcal{M}))$, each $\varepsilon \in E$ and each $i \le k$ and since (by Šmíd [2005b, Theorem 3]) there exist (easily computable) mappings $G_1, G_2, \ldots, G_{k-1}$ such that, on the set $[e_{(1,k)} \in \tilde{\mathcal{E}}^k]$, $(B^{\mathcal{M}}_{(\nu)}, S^{\mathcal{M}}_{(\nu)}) = G_\nu(e_{(1,\nu)})$ for each $1 \le \nu < k$, we have

$$\mathbb{P}(e_{(i)} = \varepsilon | e_{(1,i-1)}) = \mathbb{E}\left[\mathbb{P}(\, e_{(i)} = \varepsilon | B^{\mathcal{M}}_{(i-1)}, S^{\mathcal{M}}_{(i-1)}, e_{(1,i-1)})\middle| \, e_{(1,i-1)}\right]$$
$$\mathbb{E}\left[\mathbb{P}(\, e_{(i)} = \varepsilon | B^{\mathcal{M}}_{(i-1)}, S^{\mathcal{M}}_{(i-1)})\middle| \, e_{(1,i-1)}\right] = \mathbb{P}(e_{(i)} = \varepsilon | B^{\mathcal{M}}_{(i-1)}, S^{\mathcal{M}}_{(i-1)})$$

for each $\varepsilon \in E$ (the last "=" follows from the fact that $\sigma(B^{\mathcal{M}}_{(i-1)}, S^{\mathcal{M}}_{(i-1)}) \subseteq \sigma(e_{(1,i-1)})$) which yields

$$\mathbb{P}(e_{(1,k)} = \epsilon) = \prod_{i=1}^{k} \mathbb{P}(e_{(k)} = \epsilon^i | e_{(1)} = \epsilon^1, \ldots, e_{(i-1)} = \epsilon^{i-1}) = \prod_{i=1}^{k} \gamma(\epsilon^i, G_{i-1}(\epsilon^1, \ldots, \epsilon^{i-1}))$$

(we take $G_0 = (B^{\mathcal{M}}_{(0)}, S^{\mathcal{M}}_{(0)})$).

The conditional probabilities in (4) are also easy to compute provided that $\epsilon \in \tilde{\mathcal{E}}^k$: Since $(\phi_{(k)}, q_{(k)}) = F(e_{(1,k)})$ on the set $[e_{(1,k)} \in \tilde{\mathcal{E}}^k]$ for some (easily computable) mapping F [Šmíd, 2005b, Theorem 3], we have

$$\mathbb{P}((\phi_{(k)}, q_{(k)}) \in B | e_{(1,k)} = \epsilon) = \mathbf{1}_B(F(\epsilon)).$$

It remains to show that η^k_2 may be made arbitrarily small by the refinement of $\mathcal{M}$: Indeed,

$$\mathbb{P}(e_{(1,k)} \in \mathcal{E}^k - \tilde{\mathcal{E}}^k) = \mathbb{P}(\beta^j_{e_{(1,k)}} > 0 \wedge \sigma^j_{e_{(1,k)}} > 0 \text{ for some } j \in C) \le \sum_{j \in C} \zeta^j \tag{5}$$

where

$$\begin{aligned} \zeta^j =& \mathbb{P}(\beta^j_{e_{(1,k)}} > 0 \wedge \sigma^j_{e_{(1,k)}} > 0) = 1 - \mathbb{P}(\beta^j_{e_{(1,k)}} = 0) - \mathbb{P}(\sigma^j_{e_{(1,k)}} = 0) + \mathbb{P}(\beta^j_{e_{(1,k)}} = 0, \sigma^j_{e_{(1,k)}} = 0) \\ =& 1 - (1-\mu_j)^k - (1-\nu_j)^k + (1-(\mu_j+\nu_j))^k, \qquad \mu_j = u\pi(M^j)/\bar{r}, \qquad \nu_j = v\rho(M^j)/\bar{r}. \end{aligned}$$

[9] The conditional independence could be shown similarly to the proof of the Markov property of the embedded chain.

Since, by an easy calculation,

$$\zeta^j = \sum_{i=2}^{k} \binom{k}{i} (-1)^i [(\mu_j + \nu_j)^i - (\mu_j^i + \nu_j^i)] \leq \sum_{i=2}^{k} \binom{k}{i} [2 \max_{j \in C}(\mu_j \vee \nu_j)]^i = o(\max_{j \in C}(\mu_j \vee \nu_j)) \quad (6)$$

as $\max_{j \in C}(\mu_j \vee \nu_j) \to 0$ and since the partition $\mathcal{M}$ may be constructed so that

$$\max_{j \in C}(\mu_j \vee \nu_j) \leq \frac{24 \max(u, v)}{\bar{r}(m - 2(|B_{(0)}| + |S_{(0)}|) - 1)} = O(m^{-1})$$

as $m \to \infty$ for each $j \in C$ [Šmíd, 2005b, Theorem 4], we get from (5) and (6) that

$$\mathbb{P}(e_{(1,k)} \in \mathcal{E}^k - \tilde{\mathcal{E}}^k) \leq m \cdot o(O(m^{-1})) = \frac{O(m^{-1})}{m^{-1}} \frac{o(O(m^{-1}))}{O(m^{-1})} \overset{m \to \infty}{\longrightarrow} 0$$

given a suitable choice of the partitions.

3.4 The Forecast

By summarizing the previous paragraphs, we are getting

$$\mathbb{P}((p_\tau, q_\tau) \in A) = \sum_{k=0}^{\tilde{N}} \sum_{\epsilon \in \tilde{\mathcal{E}}^k} \mathbf{1}_B(F(\epsilon)) \prod_{i=1}^{k} \Gamma_i(\epsilon^1, \ldots, \epsilon^i) \frac{\bar{r}^k}{k!} + \eta_1 + \eta_2$$

for some easily computable mappings $F, \Gamma_1, \ldots, \Gamma_k$. Moreover, η_1 may be made arbitrarily small by increasing $\tilde{N}$ while $\eta_2 \overset{\triangle}{=} \sum_{k=1}^{\tilde{N}} \eta_2^k$ may be arbitrarily decreased by a suitable choice of partition $\mathcal{M}$.

3.5 Possible Further Refinements

Let us mention two ways of reducing the (possibly very large) computational complexity of the algorithm designed in the previous paragraphs.

First, not all the conditional distributions of $(p_{(k)}, q_{(k)})$ given $e_{(1,k)} = \epsilon$, $\epsilon \in \mathcal{E}^k - \tilde{\mathcal{E}}^k$, are non-computable. As it was already mentioned, some of the distributions corresponding to the scenarios from $\mathcal{E}^k - \tilde{\mathcal{E}}^k$ may be computed after a more detailed analysis, some of them are even Dirac: when we modify our example (subsection 3.3) so that $\epsilon = (v, v, b_\star^5, s_\star^5)$, then $p_{(k)} = 4$ independently of the limit prices of the new orders. Hence, we may increase the number of the scenarios we take into account to decrease the errors η_2^k.

Second, it follows from the definition of the CDA that the quantities $p_{(i)}$, $q_{(i)}$, $b_{(i)}^\kappa$, …, $b_{(i)}^1$, $a_{(i)}^1$, …, $a_{(i)}^\kappa$, depend only on $p_{(i-1)}, q_{(i-1)}, b_{(i-1)}^{\kappa+1}, \ldots, b_{(i-1)}^1, a_{(i-1)}^1, \ldots, a_{(i-1)}^{\kappa+1}, \ldots$ and on independent random variables for each $i, \kappa \in \mathbb{N}$ which implies that only the quantities

$$p_{(k-1)}, q_{(k-1)}, b_{(k-1)}^1, a_{(k-1)}^1, p_{(k-2)}, q_{(k-2)}, b_{(k-2)}^2, b_{(k-2)}^1, a_{(k-2)}^1, a_{(k-2)}^2, \ldots, b_{(0)}^k, \ldots, a_{(0)}^k$$

are relevant for the distribution of $p_{(k)}, q_{(k)}$. Hence, to reduce the number of the branches of our computation, we may accumulate the scenarios having equal impact on the relevant quantities.

4 Conclusion

In the present paper, a way of computing the forecasts of the market price and of the traded volume in the model of continuous double auction with complete randomness was suggested. With a "sufficiently efficient" computer, the "future" distribution of the forecasted quantities may be evaluated with an arbitrarily accuracy.

Bibliography

M. Šmíd. Conditional distribution of the limit order book given the history of the best quote process, 2005a. `http://econwpa.wustl.edu:80/eps/em/papers/0503/0503015.pdf`.

M. Šmíd. Forecasting in continuous double auction, 2005b. Research report no. 2128, Institute of Information Theory and Automation, May 2005.

E. Smith, J. D. Farmer, L. Gillemot, and S. Krishnamurthy. Statistical theory of the continuous double auction. *Quantitative finance*, 3(6):481–514, 2003.

Live OR: IFORS tutORial Project

Moshe Sniedovich
Department of Mathematics and Statistics
The University of Melbourne
Parkville , Victoria 30052
Australia
m.sniedovich@ms.unimelb.edu.au

Abstract

In this paper we provide information about the tutORial project organised by the international Federation of OR Societies (IFORS). This includes details on the more than thirty on-line modules currently available on the project's web site (www.ifors.org/tutorial/) covering main stream OR/MS topics ranging from elementary linear algebra, to linear programming integer programming and dynamic programming. We also discuss issues related to the incorporation of the these modules in OR/MS courseware, including technical and pedagogical issues related to the use of the modules by lecturers, students and the public in general.

Keywords

IFORS, tutORial, education, online, modules

1 Introduction

The International Federation of OR Societies (IFORS) is an umbrella organisation comprising the national OR/MS societies of more than 45 countries with total membership exceeding 25,000. Its mission is to promote OR as an academic discipline and as a career path, internationally. Information on IFORS activities can be found on its website at www.ifors.org.

In recognition of the fact that education plays an important role in OR and that the World Wide Web offers effective tools for the distribution and delivery of courseware, in 1998 IFORS initiated a project called *tutORial* whose goal is to coordinate the development of on line tutorial modules for main stream OR topics.

At present the project's web site (www.ifors.org/tutorial/) offer more than thirty modules — all are freely accessible by the public — supporting teaching/learning in the following areas:

– **Linear Algebra** (eg row operation, linear equations, matrix inverse)

– **Linear Programming** (eg simplex method, dual simplex method)

– **Integer Programming** (eg Gomory's Cuts, branch and bound)

– **Dynamic Programming** (eg Shortest path, CPM, Knapsack)

– **Graphs and Networks** (eg Prim's Algorithm, Kruskal's Algorithm)

– **Simulation** (eg Random Numbers Tester, Queueing Models)

The complete list of modules is provided in the appendix.

In the next section we briefly describe how these students can be used by students, lecturers, tutors/markers and the general public.

2 How to Use the Modules

The web site of the project is open to the public so anyone can use the modules. All that is needed for this purpose is a web browser such as *MS Internet Explorer*, *Netscape Navigator*, and many others. All the modules are written in *JavaScript* — not *Java* — so they are very light. make sure that *JavaScript* is enabled on your browser.

The modules can be used in a variety of ways. But before we explain how they can be used by students and lecturers, it is important to stress what they were **not** designed to do:

– The modules are *numerically* oriented and therefore do not have any symbolic manipulation capabilities.

– The main objective of the modules in not to provide users " press-button " facilities for generating solutions to numerical problems. Rather, they were designed to enable users to *experiment* with algorithms and procedures where the user is doing the thinking and the modules merely execute simple operations. Thus, in this framework solving a problem is typically much more than just pressing a **SOLVE** button.

– The modules were not designed to solve " large " problems. So do not expect the Simplex Method module to solve for you large scale LP problems. Rather, the modules were designed to handle problems whose size is similar to the size of problems that typically used in introductory OR/MS textbooks to illustrate methods, procedures and algorithms.

– The modules were not designed to replace the use textbooks, lecture notes and so on. In other words, it is assumed that users of the modules are already familiar with the topics under consideration and are using the modules to experiment with these topics rather than to learn from scratch about these topics. For this reason the on-line help facilities provided by the modules are very limited.

– While every effort is being made to insure that the modules are *educationally rich*, you may no doubt soon find that they are not "fancy". We simply do not have the resources required to equip the modules with *"bells and whisles"*. So, ... do not expect bells and whistles on board the modules.

> It is also extremely important to realize that the modules vary greatly with regard to the level of interaction they provide, hence the intermediate results they generate. In due course most of the modules will be highly interactive and will be capable of generating, if/when necessary, substantial intermediate results. At present some modules have not yet reached this level of maturity and sophistication.

With this in mind let us now briefly discuss how they can be used by students, lecturers, tutors/markers and the public in general.

– **Students**
You can use the modules to experiment with algorithms described in class or in your textbooks. For example, you can use the Simplex Method module to solve linear programming problems. There is no "SOLVE" button. You have to tell the module how to solve the problem, the module will only save you the trouble of doing the arithmetic. So to use the module you have to know how simplex method works. The module will warn you if you do something wrong, so in effect using the module you can identify difficulties that you may have implementing the method. And of course, you can quickly check that the solutions to the assignments you submit are correct — including the intermediate results.

– **Lecturers**
You can use the modules to generate numerical questions (and solutions) for exams, assignments, and lectures. For example, you can

check how many iterations will be required, whether the intermediate results are student-friendly and so on. You can use the modules in the lecture theatre to illustrate how algorithms work as well as what can go wrong if they are not used correctly.

– **Tutors/markers**
You can use the modules to prepare examples for your tutorials and to check assignments submitted by students. For example, you can easily identify where exactly something went wrong in a solution submitted by a student for a LP problem.

– **Public in General**
You do not have to be an OR person to make use of the tutORial modules. The linear algebra modules, for instance, can be useful by students and lecturers from related disciplines. And many of the modules are relevant to computer science students (eg dynamic programming, spanning tree algorithms, topological sorting, Dijkstra's Algorithm, and so on). And some of the games we have on board may have a general appeal.

Some of the modules are linked to papers published in the electronic journal *INFORMS Transactions on Education* (ITE), for example Sniedovich [1,2,3,4,5].

3 Summary

The IFORS tutORial project provides students and lecturers in OR/MS and related fields a wide collection of educationally rich on line modules to support learning/teaching of main stream OR subjects. These modules can be used for a variety of purposes including demonstration of and experimentation with OR algorithms, generating and checking assignment problems, and ... recreation. If you have not done so already, visit the project's web site and let us know if there is anything we can do to make these modules more user-friendly!

www.ifors.org/tutorial/

References

[1] Sniedovich M., OR/MS Games: 1. A neglected educational resource, INFORMS Transactions on Education, Vol. 3(1), 2002.

[2] Sniedovich M., OR/MS Games: 2. The Towers of Hanoi, INFORMS Transactions on Education, Vol. 3(2), 2002.

[3] Sniedovich M., OR/MS Games: 3. The Counterfeit coin problem, INFORMS Transactions in Education, Vol. 4(1), 2003.

[4] Sniedovich, M., OR/MS Games: 4. The joy of egg-dropping in Braunschweig and Hong Kong, INFORMS Transactions on Education, 4(1), 48-64, 2003.

[5] Sniedovich, M., Towards an AoA-Free Courseware for the Critical Path Method, INFORMS Transactions on Education, in press, 2005.

Appendix

The following modules are currently (June 15, 2005) available on the *tutORial* web site at `www.ifors.org/tutorial/`. The list is taken "as it is " from the menu on the website.

Dynamic Programming

– Die Hard at the Pub
– Towers of Hanoi
– New Towers of Hanoi
– Shortest Path Problem
– Dijkstra's Algorithm
– Knapsack Problem
– Critical Path Method
– Chained Matrices
– Replacement Problem
– Bridge & Torch
– Dropping Eggs
– False Coin Problem
– Traveling Salesman Problem
– Transitive Closure

Graphs and Networks

- Dijkstra's Algorithm
- Prim's Algorithm
- Kruskal's Algorithm
- TopSort
- Transitive Closure

Integer Programming

- Branch and Bound (Knapsack)
- Gomory's Cuts
- Royal Optimization
- 8 Easy Pieces

Linear Algebra

- Row Operations
- The Equator
- The Invertor
- The Determinator

Linear Programming

- The Simplex Place
- Dual Simplex Method
- Hungarian Algorithm

Simulation

- M/M/1 Queues
- G/G/s Queues
- Queueing Networks
- Random Generator Tester

Other

- Geee Park

Portmanteau tests based on Kendall's autocorrelation coefficients

Miroslav Šiman
Charles University, Prague, Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Sokolovská 83, 186 75 Prague, Czech Republic
siman@karlin.mff.cuni.cz

Abstract

Portmanteau tests based on serial Kendall's autocorrelation coefficients will be proposed and their usefulness will be judged by a small but representative Monte Carlo study dealing with testing for conditional heteroskedasticity. Besides, acceptable approximations to variances of these coefficients at higher lags will be found.

Keywords

GARCH, portmanteau tests, rank tests, Kendall's tau, rank autocorrelation

1 Introduction

A testing of randomness usually stands at the beginning of any time series analysis and plays a key role in solving many problems occuring in practice. Portmanteau tests of independence are often preferred due to their simplicity and easy, fast implementation. When outliers are suspected to occur in examined data, portmanteau tests based on rank measures of dependence can often still be used correctly.

Let $R_1, R_2, \ldots, R_T$ be ranks of a given time series $Y_1, Y_2, \ldots, Y_T$ and let us follow the convention $R_k = R_{((k-1) \bmod T)+1},\ k = 1, 2, \ldots$

Sample rank autocorrelations $\widehat{r}_S(k)$ at lags $k = 1, 2, \ldots$ are defined as follows:

$$\widehat{r}_S(k) = \frac{\sum_{i=1}^{T-k}(R_i - \bar{R})(R_{i+k} - \bar{R})}{\sum_{i=1}^{T}(R_i - \bar{R})^2} \quad \left(\bar{R} = \frac{1}{T}\sum_{i=1}^{T} R_i\right)$$

It is well known (see e.g. Dufour and Roy [3, 4]) that (for $1 \le k < \frac{T}{2}$)

$$E\widehat{r}_S(k) = -\frac{(T-k)}{T(T-1)}, \quad \operatorname{var}(\widehat{r}_S(k)) = \frac{5T^4 - (5k+9)T^3 + 9(k-2)T^2 + 2k(5k+8)T + 16k^2}{5(T-1)^2T^2(T+1)}$$

Under the null hypothesis H_0 of independence and identical continuous distribution of random variables $Y_1, \ldots, Y_T$, the first few, say m, standardized sample rank autocorrelations

$$\tilde{r}_S(k) = \frac{(\widehat{r}_S(k) - E\widehat{r}_S(k))}{\sqrt{\operatorname{var}(\widehat{r}_S(k))}},\ k = 1, 2, \ldots, m$$

have jointly the multivariate standard normal asymptotic distribution. This property of theirs has been already used for constructing successful rank portmanteau tests (see e.g. Dufour and Roy [4] for the most established test of this type).

Ferguson, Genest and Hallin [5] introduced the serial Kendall's coefficient $\widehat{r}_K(k)$ and its

circular version $\widehat{r}^{\circ}_K(k)$, $k = 1, 2, \ldots$:

$$\widehat{r}_K(k) = 1 - \frac{4N_{k,T}}{(T-k)(T-k-1)}, \quad N_{k,T} = \sum_{i=1}^{T-k}\sum_{j=1}^{T-k} \mathrm{Ind}(R_i < R_j, R_{i+k} > R_{j+k})$$

$$\widehat{r}^{\circ}_K(k) = 1 - \frac{4N^{\circ}_{k,T}}{T(T-1)}, \quad N^{\circ}_{k,T} = \sum_{i=1}^{T}\sum_{j=1}^{T} \mathrm{Ind}(R_i < R_j, R_{i+k} > R_{j+k})$$

and proved the following useful formulae:

$$EN_{k,T} = \frac{(3T-3k-1)(T-k)}{12} - \frac{k}{6}, \quad EN^{\circ}_{k,T} = \frac{T(3T-1)}{12}, \quad \left(1 \le k < \tfrac{T}{2}\right)$$

$$\mathrm{var}(N_{1,T}) = \frac{10T^3 - 37T^2 + 27T + 74}{360}, \quad \mathrm{var}(N^{\circ}_{1,T}) = \frac{10T^3 - 7T^2 - 49T}{360}, \quad (T \ge 4)$$

$$\frac{\mathrm{var}(N_{k,T})}{T^3} \underset{T\to\infty}{\longrightarrow} \frac{1}{36}, \quad \frac{\mathrm{var}(N^{\circ}_{k,T})}{T^3} \underset{T\to\infty}{\longrightarrow} \frac{1}{36}, \quad k = 1, 2, \ldots$$

as well as the joint asymptotic normality of both $(\widehat{r}_K(1), \ldots, \widehat{r}_K(m))$ and $(\widehat{r}^{\circ}_K(1), \ldots, \widehat{r}^{\circ}_K(m))$ for any fixed positive integer m.

The accurately standardized coefficients

$$\tilde{r}_K(k) = \frac{(\widehat{r}_K(k) - E\widehat{r}_K(k))}{\sqrt{\mathrm{var}(\widehat{r}_K(k))}}$$

seem hot candidates to be used in portmanteau tests too. It will be done in this article; required acceptable approximations to $\mathrm{var}(\widehat{r}_K(k))$'s will be derived in the next section. In fact, empirical studies on various single sample autocorrelation coefficients (see e.g. Ferguson, Genest and Hallin [5] and references therein) indicate that some improvement could really be achieved this way although $\tilde{r}_K(k)$'s and $\tilde{r}_S(k)$'s are known to be asymptotically equivalent (see ibid.).

Concretely, the tests T_1, T_2, T_3, respectively based on the statistics S_1, S_2 and S_3:

$$S_1 = \sum_{k=1}^{m} \tilde{r}_S(k)^2 \sim \chi^2_m, \quad S_2 = \frac{3\sqrt{T}}{2}\sum_{k=1}^{m} (\widehat{r}_K(k) - E\widehat{r}_K(k))^2 \sim \chi^2_m, \quad S_3 = \sum_{k=1}^{m} \tilde{r}_K(k)^2 \sim \chi^2_m$$

will be considered. Approximative distributions used for testing H_0 are stated behind the $\sim$ sign. The threshold parameter m is often recommended not to be chosen pointlessly too high (see e.g. Battaglia [1]) and best results are often obtained for $m \approx 5$.

In a word or two, T_1 (introduced in Dufour and Roy [4]) is included as a rank portmanteau benchmark, T_2, T_3 are analogues to T_1 employing $\widehat{r}_K(k)$'s standardized by their exact means and by the asymptotic or newly derived variance formulae.

$T_1, \ldots, T_3$ will be investigated and compared in a small Monte Carlo study dealing with testing for conditional heteroskedasticity. Such testing is often performed by applying a test of independence to (possibly squared) absolute values of given observations because such values often exhibit serial autocorrelation if the null hypothesis H_0 does not hold but a conditionally heteroskedastic alternative does.

As most common alternatives of this type, only GARCH(1,1) models (say $\{\varepsilon_t\}$) will be considered here:

$$\varepsilon_t = z_t\sigma_t, \quad \sigma_t^2 = c + a_1\varepsilon_{t-1}^2 + b_1\sigma_{t-1}^2, \; 0 < c, a_1, b_1, \; a_1 + b_1 < 1,$$

where $\{z_t\}$ is a sequence of independent and identically distributed random variables with zero mean and unit variance, usually expected to have the standard normal distribution or, even more frequently, the standardized Student distribution with three or slightly more degrees of freedom (see Berkes, Horváth and Kokoszka [2] for more information on GARCH(1,1) processes).

2 Variance of Kendall's serial autocorrelations

Proposition 1. *Under the null hypothesis of independent continuous observations $Y_1, \ldots, Y_T$, the following formulae hold for any fixed $k \geq 1$ and $T \geq 8k+1$:*

$$\begin{aligned}
\mathrm{var}(\widehat{r}_K(k)) &= \frac{2[10T^3 + (-30k-7)T^2 + (30k^2+46k-49)T + (-10k^3-29k^2+113k)]}{45(T-k)^2(T-k-1)^2} \\
\mathrm{var}(\widehat{r}^{\circ}_K(k)) &= \frac{2[10T^3 - 7T^2 - 49T]}{45T^2(T-1)^2}
\end{aligned}$$

Proof. First let us focus on $\mathrm{var}(N_{k,T})$. The second moment

$$E(N_{k,T}^2) = \sum_i \sum_j \sum_p \sum_q P(R_i < R_j, R_{i+k} > R_{j+k}, R_p < R_q, R_{p+k} > R_{q+k})$$

can be decomposed as in Ferguson, Genest and Hallin [5]:

$$\begin{aligned}
E(N_{k,T}^2) &= A + 2B + 2C + D \\
A &= \sum_{i,j} P(R_i < R_j, R_{i+k} > R_{j+k}) \\
B &= \sum_{i,j,q\neq j} P(R_i < R_j, R_{i+k} > R_{j+k}, R_i < R_q, R_{i+k} > R_{q+k}) \\
C &= \sum_{i,j\neq p, p\neq i} P(R_i < R_j, R_{i+k} > R_{j+k}, R_p < R_i, R_{p+k} > R_{i+k}) \\
D &= \sum_{i,j,p,q \text{ distinct}} P(R_i < R_j, R_{i+k} > R_{j+k}, R_p < R_q, R_{p+k} > R_{q+k})
\end{aligned}$$

where, apparently,

$$A = E(N_{k,T}) = \frac{1}{12}[3T^2 + (-6k-1)T + (3k^2-k)]$$

As far as the terms B, C, D are concerned, several special cases must be treated separately, see Ferguson, Genest and Hallin [5] for more details. This proof generalizes that of theirs to general k and follows the same approach. Thus many details are omitted here.

In case of B and C, it can be easily deduced that

$$\begin{aligned}
B &= \frac{3}{24}a + \frac{5}{24}b + \frac{18}{120}c + \frac{11}{120}d + \frac{4}{36}e \\
C &= \frac{1}{24}a + \frac{0}{24}b + \frac{3}{120}c + \frac{6}{120}d + \frac{1}{36}e \\
a &= 4[T-3k] = 2b \\
b &= 2[T-3k] \\
c &= 4[(T-4k)(T-k-4) + 2k(T-k-3)] \\
&= 4[T^2 + (-3k-4)T + (2k^2+10k)] = 2d \\
d &= 2[(T-5k)(T-2k-4) + 2k(T-2k-3) + 2k(T-2k-2)] \\
&= 2[T^2 + (-3k-4)T + (2k^2+10k)] \\
e &= [(T-k)(T-k-1)(T-k-2) - a - b - c - d] \\
&= T^3 + (-3k-9)T^2 + (20+24k+3k^2)T + (-k^3-15k^2-44k)
\end{aligned}$$

Concretely, B and C can be expressed as follows:

$$\begin{aligned}
B &= \frac{1}{180}[20T^3 + (-60k-39)T^2 + (60k^2+57k+1)T + (-20k^3-18k^2+35k)] \\
C &= \frac{1}{180}[5T^3 + (-15k-9)T^2 + (15k^2+12k-14)T + (-5k^3-3k^2+50k)]
\end{aligned}$$

The computation of D is slightly more complicated:

$$\begin{aligned}
D &= \frac{59}{120}a' + \frac{4}{9}b' + \frac{35}{72}c' + \frac{5}{24}d' + \frac{1}{16}e' \\
a' &= 4[T-4k] \\
b' &= 4[(T-5k)(T-k-5) + 2k(T-k-4)] \\
&= 4[T^2 + (-4k-5)T + (3k^2+17k)] \\
c' &= 2[(T-6k)(T-2k-5) + 2k(T-2k-4) + 2k(T-2k-3)] \\
&= 2[T^2 + (-4k-5)T + (4k^2+16k)] \\
d' &= 4(\alpha_1 + \beta_1 + \alpha_2 + \beta_2 + \alpha_3 + \beta_3 + \alpha_4 + \beta_4) \\
&\quad \alpha_1 = (T-8k)(T-3k-6)(T-k-7) + (T-8k)2(T-k-6) \\
&\quad \beta_1 = 2(T-8k)k(T-k-6) \\
&\quad \alpha_2 = 2k(T-3k-5)(T-k-7) + 2k(T-k-6) \\
&\quad \beta_2 = 2k(k-1)(T-k-6) + 2k(T-k-5) + 2k^2(T-k-6) \\
&\quad \alpha_3 = 2k(T-3k-4)(T-k-7) + 2k(T-k-6) \\
&\quad \beta_3 = 2k(k-1)(T-k-6) + 2k^2(T-k-6) \\
&\quad \alpha_4 = 2k(T-3k-3)(T-k-6) + 2k(T-k-5) \\
&\quad \beta_4 = 2k(k-1)(T-k-5) + 2k^2(T-k-5) \\
(d') &= 4[T^3 + (-4k-11)T^2 + (5k^2+39k+30)T + (-2k^3-30k^2-88k)] \\
e' &= (T-k)(T-k-1)(T-k-2)(T-k-3) - 6a' - 6b' - 6c' - 3d' \\
&= T^4 + (-4k-18)T^3 + (6k^2+66k+107)T^2 \\
&+ (-4k^3-78k^2-346k-210)T + (k^4+30k^3+251k^2+558k)
\end{aligned}$$

To sum up, D can be written as

$$\begin{aligned}
D &= \frac{1}{720}[45T^4 + (-180k-210)T^3 + (270k^2+570k+195)T^2 \\
&\quad +(-180k^3-510k^2-90k+66)T + (45k^4+150k^3-65k^2+394k)]
\end{aligned}$$

It is now easy to verify that

$$\begin{aligned}
E(N_{k,T}^2) &= \frac{1}{720}[45T^4 + (-180k-10)T^3 + (270k^2-30k-9)T^2 \\
&\quad +(-180k^3+90k^2+102k-98)T + (45k^4-50k^3-53k^2+226k)] \\
\mathrm{var}(N_{k,T}) &= \frac{1}{360}[10T^3 + (-30k-7)T^2 + (30k^2+46k-49)T + (-10k^3-29k^2+113k)] \\
\mathrm{var}(\widehat{r}_K(k)) &= \frac{2[10T^3 + (-30k-7)T^2 + (30k^2+46k-49)T + (-10k^3-29k^2+113k)]}{45(T-k)^2(T-k-1)^2}
\end{aligned}$$

Besides, it also turned out that $\mathrm{var}(\widehat{r}_K^{\circ}(k))$ does not depend on k and that its form can be easily obtained as $\mathrm{var}(\widehat{r}_K(0))$. Thus

$$\mathrm{var}(\widehat{r}_K^{\circ}(k)) = \frac{2[10T^3 - 7T^2 - 49T]}{45T^2(T-1)^2},\ k = 1, 2, \ldots$$

Comments. These formulae correspond with the results obtained in Ferguson, Genest and Hallin [5] for $k = 1$. Besides, they were verified within the limits of error in a huge simulation experiment dealing with $k = 1, 2, \ldots, 5$. These formulae also appeared to hold exactly even in some cases when $T \leq 8k$. Besides, the following simulation study indicates that they can be successfully used even for sufficiently high T's not greater than $8k$.

In principle, exact formulae could be derived similarly even in the cases of relatively small T, but they would not be of such great interest in the portmanteau testing context where often $k \approx 5$ and $T \geq 50$.

3 Simulation study

All computations in this section were done in MATLAB. We always used $N = 10\,000$ independent data samples of length $T = 25, 50, 100, 200$ and 500. $S_1, \ldots, S_3$ were then applied to them with the threshold parameter m successively increasing from $m = 1$ to 12, and with a representative significance level $\alpha = 0.10$.

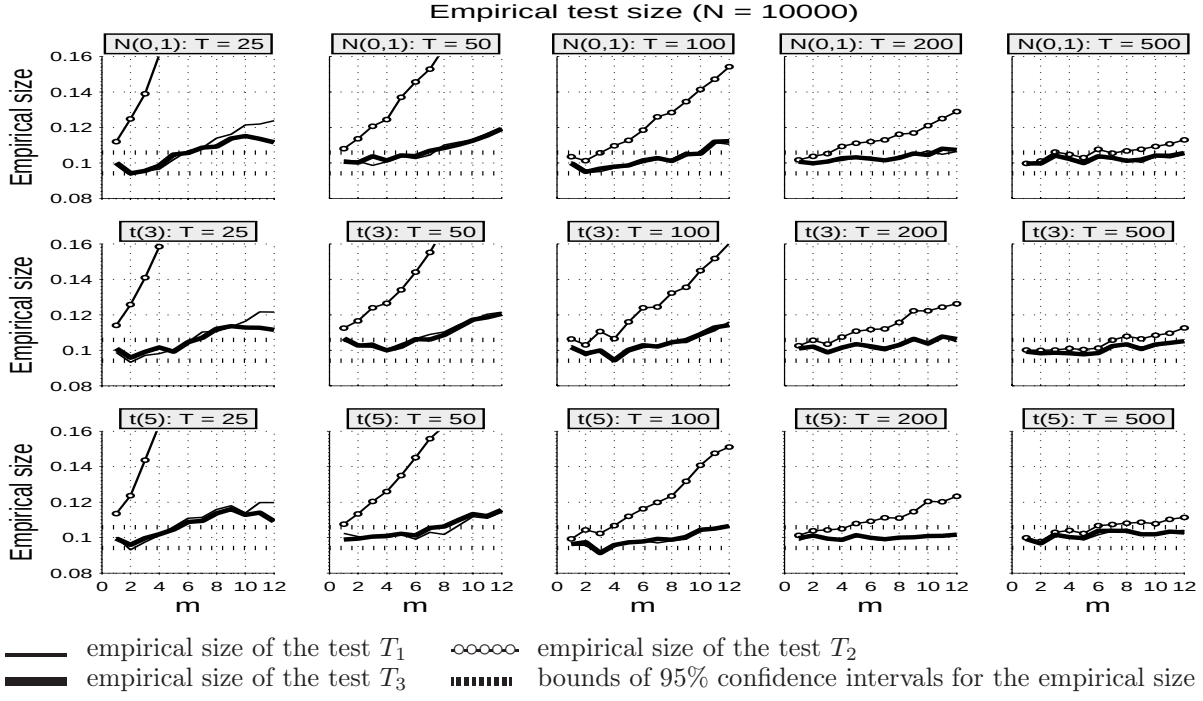
When testing size, we generated random samples from the standard normal distribution and the standardized Student distributions with three and five degrees of freedom. Selected results are presented in Figure 1 and numerically illustrated in Table 1.

Table 1: Empirical test size comparison taking into consideration both time series length and distribution of innovations z_t's ($N = 10\,000$, $\alpha = 0.10$).

		actually observed significance level (empirical test size)									
		$T = 25$		$T = 50$		$T = 100$		$T = 200$		$T = 500$	
$m =$		5	10	5	10	5	10	5	10	5	10
$\mathbf{T_1}$	N(0,1)	0.102	0.121	0.104	0.113	0.098	0.105	0.103	0.107	0.101	0.104
	t(3)	0.100	0.116	0.104	0.117	0.099	0.108	0.104	0.104	0.098	0.102
	t(5)	0.106	0.114	0.102	0.112	0.098	0.105	0.102	0.101	0.099	0.102
$\mathbf{T_2}$	N(0,1)	0.182	0.350	0.137	0.194	0.113	0.142	0.111	0.121	0.103	0.109
	t(3)	0.182	0.350	0.134	0.198	0.116	0.145	0.111	0.122	0.100	0.108
	t(5)	0.187	0.352	0.135	0.194	0.112	0.141	0.108	0.120	0.102	0.108
$\mathbf{T_3}$	N(0,1)	0.105	0.115	0.104	0.113	0.099	0.105	0.103	0.105	0.100	0.104
	t(3)	0.099	0.113	0.102	0.117	0.100	0.109	0.104	0.104	0.098	0.103
	t(5)	0.105	0.113	0.102	0.113	0.097	0.104	0.101	0.101	0.100	0.102

N ... number of replications α ... intended (asymptotic) significance level
T ... time series length m ... threshold parameter

Figure 1: Empirical test size comparison taking into consideration both time series length and distribution of innovations z_t's ($N = 10\,000$, $\alpha = 0.10$).



When testing power, we concentrated on various GARCH(1,1) models with innovations drawn from the same distributions that were considered in the test size study. This time we generated longer data samples and then trimmed them by the first 300 observations to the length T, all that in order to prevent initial conditions to have any effect.

Weakly stationary GARCH(1,1) alternatives with persistent volatility, i.e. with parameters satisfying additional conditions $a_1 \sim 0.1$, $b_1 \sim 0.9$, were slightly preferred because they are most common in practice. The remaining parameter c was always set to 1. Some results of the power test investigation are documented in Table 2.

Table 2: Empirical test power under some GARCH(1,1) alternatives ($N = 10\,000$, $\alpha = 0.10$).

		empirical test power									
		$T = 25$		$T = 50$		$T = 100$		$T = 200$		$T = 500$	
$m =$		5	10	5	10	5	10	5	10	5	10
$a_1 = 0.1, b_1 = 0.85$											
$\mathbf{T_1}$	N(0,1)	0.112	0.118	0.150	0.148	0.244	0.246	0.428	0.435	0.780	0.799
	t(3)	0.115	0.122	0.161	0.155	0.251	0.249	0.418	0.418	0.725	0.735
	t(5)	0.111	0.113	0.154	0.148	0.259	0.251	0.456	0.464	0.787	0.806
$\mathbf{T_3}$	N(0,1)	0.119	0.126	0.153	0.154	0.245	0.246	0.429	0.438	0.780	0.799
	t(3)	0.118	0.126	0.164	0.162	0.252	0.249	0.419	0.419	0.725	0.737
	t(5)	0.116	0.118	0.157	0.155	0.260	0.254	0.457	0.466	0.788	0.808

N ... number of replications α ... intended (asymptotic) significance level
T ... time series length m ... threshold parameter

It is obvious that T_3 clearly outperforms T_2 as for the test size. The newly derived variance formulae thus proves much better than the asymptotic ones, even in the cases not allowed for by the Proposition 1 (i.e. for k relatively high in comparison to T).

Moreover, T_1 and T_3 exhibit roughly the same behaviour. As the power of T_3 almost never exceeds that of T_1 by more than one percentage point (regardless the time series length considered), it is really debatable whether S_3 should be ever employed in practice. We cannot recommend its use, partly for the simple reason that its computation is usually more demanding than the evaluation of S_1.

Due to the limitation of the length of this article, it is not possible to include all results obtained. Nevertheless, all of them seem to support our findings.

4 Conclusions

We investigated the possibility of using portmanteau tests based on Kendall's autocorrelation coefficients and concluded that there is hardly any advantage in employing them. Besides, we derived explicit formulae for the variances of such coefficients at any fixed lag for all sufficiently high T's.

We hope that this contribution will be found useful and help the others to improve their portmanteau testing skills.

References

[1] F. Battaglia: Approximate power of portmanteau tests for time series. Statist. Probab. Lett. 9 (1990), pp. 337–341.

[2] I. Berkes, L. Horváth and P. Kokoszka: Probabilistic and statistical properties of GARCH processes. In: Asymptotic methods in stochastics: Festschrift for Miklós Csörgö (L. Horváth and B. Szyszkowicz, eds.), Fields Inst. Commun. 44 (2004), American Mathematical Society, pp. 409–429.

[3] J.-M. Dufour and R. Roy: Some robust exact results on sample autocorrelations and tests of randomness. J. Econometrics 29 (1985), pp. 257–273. Corrigendum in J. Econometrics 41 (1989), pp. 279–281.

[4] J.-M. Dufour and R. Roy: Generalized portmanteau statistics and tests of randomness. Comm. Statist. Theory Methods 15 (1986), pp. 2953–2972.

[5] T. S. Ferguson, C. Genest and M. Hallin: Kendall's tau for serial dependence. Canad. J. Statist. 28 (2000), pp. 587–604.

Fuzzy Probability Spaces in Decision-Making under Risk

Jana Talašová and Ondřej Pavlačka

Department of Mathematical Analysis and Applied Mathematics,
Faculty of Natural Sciences, Palacky University Olomouc,
tř. Svobody 26, 771 46 Olomouc, Czech Republic
talasova@inf.upol.cz, onpav@centrum.cz

Abstract

In this paper, two types of fuzzy probability spaces will be introduced and their possible applications in methods of decision-making under risk will be described. First, a fuzzy probability space that generalizes the classical probability space to the situation of fuzzy random events will be studied. It can be applied e.g. when given continuous probability distributions of risk factors are to be approximated by discrete ones. Second, a fuzzy probability space that enables an adequate mathematical modelling of expertly set uncertain probabilities of elementary events will be defined. Its application in fuzzy decision matrices will be shown.

Keywords

Fuzzy probability, fuzzy random event, fuzzy probability space, fuzzy risk analysis, fuzzy decision matrix.

1 Introduction

In decision-making under risk, decision matrices are often used. They describe how consequences of alternatives depend on the fact which of possible and mutually exhaustive states of the world occurs. The states of the world are supposed to be set exactly and their probabilities to be known. However, in practice, the states of the world are often specified only vaguely and their probabilities are based on experts' estimations. In some cases, the probability distribution of states of the world is obtained as a result of discretization of continuous probability distributions of risk factors. In this paper, it will be shown how the apparatus of fuzzy sets, especially two types of fuzzy probability spaces, can be used in such cases.

2 Basic Notion

A fuzzy set A on a universal set U, $U \neq \emptyset$, is defined by its membership function $A : U \to [0,1]$. A set $SuppA = \{x \in U \mid A(x) > 0\}$ is called a support of A, a set $A_\alpha = \{x \in U \mid A(x) \geq \alpha\}$, $\alpha \in (0,1]$, is an α-cut of A and a set $KerA = \{x \in U \mid A(x) = 1\}$ is a kernel of A. A fuzzy set A is called normal if $Ker\ A \neq \emptyset$. The family of all fuzzy sets on U is denoted by $\mathcal{F}(U)$.

A fuzzy number is a normal fuzzy set A on $\Re$ (the set of all real numbers), whose α-cuts A_α, $\alpha \in (0,1]$, are closed intervals, and whose support $SuppA$ is bounded. It is said, that a fuzzy number A is defined on $[a,b]$ if $SuppA \subseteq [a,b]$. The family of all fuzzy numbers on $[a,b]$ is denoted by $\mathcal{F}_{\mathcal{N}}([a,b])$.

A fuzzy scale on $[a,b]$ (see [3]) is a finite set of fuzzy numbers $A_1, A_2, \ldots, A_n$ that are defined on $[a,b]$ and form a fuzzy partition on $[a,b]$, i.e. $\sum_{i=1}^{n} A_i(x) = 1$ holds for any $x \in [a,b]$. If a fuzzy scale represents a mathematical meaning of a natural linguistic scale, then it is called a linguistic fuzzy scale.

Fuzzy numbers V_i, $i = 1, 2, \ldots, m$, defined on $[0,1]$ are called normalized fuzzy weights (see [2]) if for all $\alpha \in (0,1]$ and for all $i \in \{1, 2, \ldots, m\}$ the following holds: for any $v_i \in V_{i\alpha}$ there exist $v_j \in V_{j\alpha}$, $j = 1, 2, \ldots, m$, $j \neq i$, such that

$$v_i + \sum_{j=1, j\neq i}^{m} v_j = 1 \ . \tag{1}$$

A fuzzy weighted average (see [2]) of fuzzy numbers U_i, $i = 1, 2, \ldots, m$, defined on $[a,b]$ with normalized fuzzy weights V_i, $i = 1, 2, \ldots, m$, is a fuzzy number U on $[a,b]$ whose membership function is given by the following formula

$$\begin{aligned} U(u) = \max\{&\min\{V_1(v_1), \ldots, V_m(v_m), U_1(u_1), \ldots, U_m(u_m)\} | \\ &| \textstyle\sum_{i=1}^{m} v_i \cdot u_i = u, \ \sum_{i=1}^{m} v_i = 1, \ v_i \geq 0, \ u_i \in [a,b], \ i = 1, \ldots, m\} \ . \end{aligned} \tag{2}$$

The following denotation will be used for the fuzzy weighted average:

$$U = (\mathcal{F}) \sum_{i=1}^{m} V_i \cdot U_i \ . \tag{3}$$

3 Probabilities of Fuzzy Events and Fuzzy Discretization of Continuous Risk Factors

A decision matrix can be used as an instrument of risk analysis when risk factors affecting the consequences of alternatives are of discrete nature. Then possible combinations of risk factors values determine states of the world. If the considered risk factors are independent, then probabilities of states of the world are equal to products of probabilities of the risk factors values. If they are dependent, then it is necessary to deal with their conjugate probability distribution (probabilities of states of the world are calculated as products of conditional probabilities of the risk factors values).

In practice, the decision matrix method is applied also in cases when some of the risk factors are continuous. Then their continuous probability distributions have to be approximated by discrete ones. The approximation is more realistic when continuous domains of the risk factors (closed intervals) are replaced by fuzzy scales instead of real discrete scales.

For that purpose, it is necessary to generalize the classical probability space to the situation of fuzzy events. Let us suppose a probability space $(\Re^n, \mathcal{B}_n, P)$, where $\Re$ is the set of all real numbers, $\mathcal{B}_n$ is the σ-algebra of all Borel subsets of $\Re^n$, and P is a probability measure defined on $\mathcal{B}_n$, i.e. a mapping $P : \mathcal{B}_n \to \Re$ with the following properties: a) $P(\Re^n) = 1$, b) $P(A) \geq 0$ for all $A \in \mathcal{B}_n$, c) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ for any sequence $A_1, A_2, \ldots$, such that $A_i \in \mathcal{B}_n$ for $i = 1, 2, \ldots$, and $A_i \cap A_j = \emptyset$ if $i \neq j$. Then a field of fuzzy random events $\mathcal{F}_{\mathcal{B}}(\Re^n)$ can be defined as a family of all the fuzzy sets on $\Re^n$ whose membership functions are Borel measurable, and the probability of each fuzzy event $A \in \mathcal{F}_{\mathcal{B}}(\Re^n)$ can be given by the following formula

$$P(A) = \int_{\Re^n} A(x) dP(x). \tag{4}$$

It can be proved (see [1]) that $\mathcal{F}_{\mathcal{B}}(\Re^n)$ has the following properties: a) $\Re^n \in \mathcal{F}_{\mathcal{B}}(\Re^n)$, b) if $A \in \mathcal{F}_{\mathcal{B}}(\Re^n)$, then $\overline{A} \in \mathcal{F}_{\mathcal{B}}(\Re^n)$, c) if $A_i \in \mathcal{F}_{\mathcal{B}}(\Re^n)$, $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_{\mathcal{B}}(\Re^n)$, i.e. $\mathcal{F}_{\mathcal{B}}(\Re^n)$ is a σ-algebra of fuzzy sets on $\Re^n$. Moreover, the mapping $P : \mathcal{F}_{\mathcal{B}}(\Re^n) \to \Re$ given by (4) satisfies the classical axioms of probability (see [1]). Therefore, the triple $(\Re^n, \mathcal{F}_{\mathcal{B}}(\Re^n), P)$ can be called a fuzzy probability

space. Now, it will be shown how the fuzzy probability space can be used for a fuzzy discretization of continuous risk factors in decision making under risk.

First, let us suppose that consequences of alternatives are affected by only one continuous risk factor whose probability distribution is given by a density $f(x)$. Then we will define a fuzzy scale $A_1, A_2, \ldots, A_n$ on the domain of the risk factor. The elements of the fuzzy scale are fuzzy random events, that is why their probabilities $P(A_i)$, $i = 1, 2, \ldots, n$ are given by (4), i.e.

$$P(A_i) = \int_{Supp A_i} A_i(x) f(x) dx. \tag{5}$$

It can be shown (see [5]) that $\sum_{i=1}^{n} P(A_i) = 1$ and $P(A_i) \geq 0$, $i = 1, 2, \ldots n$, so, a discrete probability distribution on the elements of the fuzzy scale is defined by this procedure.

If the consequences of alternatives are affected by more than one independent continuous risk factor, then analogous procedure of fuzzy discretization is applied to each of them. All combinations of the risk factors fuzzy values determine states of the world, and probabilities of the states are given as products of probabilities of the risk factors fuzzy values.

Now, let us suppose that the consequences of alternatives are affected by more (e.g. two) mutually dependent continuous risk factors whose conjugate probability distribution is given by a density $f(x, y)$. Then, fuzzy scales $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_m$ will be defined on domains of both the risk factors. The Cartesian products $A_i \times B_j$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, where $(A_i \times B_j)(x, y) = A_i(x) B_j(y)$ for $x, y \in \Re$, form a fuzzy partition on the Cartesian product of domains of both the risk factors. A conjugate discrete probability distribution of the fuzzy discretized risk factors is given by the following probability function:

$$P(A_i, B_j) = P(A_i \times B_j) = \int_{Supp(A_i \times B_j)} A_i(x) B_j(y) f(x, y) dx dy, \tag{6}$$

for $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. Possible states of the world are determined by such combinations of risk factors fuzzy values for which $P(A_i, B_j) \neq 0$.

If a functional relation between risk factors and consequences of alternatives exists, then the extension principle is applied to calculation of the fuzzy consequences of alternatives under the fuzzy states of the world. If the consequences of alternatives are set expertly, e.g. by means of an evaluating linguistic fuzzy scale, then states of the world expressed by combinations of (linguistically described) fuzzy values of the risk factors mean an advantage again.

4 Expertly Defined Fuzzy Probabilities and Their Application in Decision Matrices

When the decision matrix method is applied to a problem of decision-making under risk, then probabilities of states of the world need not necessarily be results of an exact mathematical risk analysis. Especially if states of the world are specified only vaguely or if they are affected by a large amount of hardly describable risk factors, their probabilities are often set directly, on the basis of experts' knowledge and experience.

To enable a correct mathematical modelling of uncertain probabilities of a finite set of random experiment results (states of the world in decision matrices), it is necessary to extend the classical probability space $(\Omega, \mathcal{P}(\Omega), P)$, where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_r\}$, $P(\omega_i) = p_i$, $p_i > 0, i = 1, 2, \ldots, r$, $\sum_{i=1}^{r} p_i = 1$, $\mathcal{P}(\Omega)$ is the

family of all subsets of Ω, and $P(A) = \sum_{i:\omega_i \in A} p_i$ holds for all $A \in \mathcal{P}(\Omega)$, to the situation of fuzzy probabilities of elementary events.

If uncertain probabilities of elementary events $\omega_1, \omega_2, \ldots, \omega_r$ are modelled by normalized fuzzy weights $P_1, P_2, \ldots, P_r$, then fuzzy probabilities of random events $A \in \mathcal{P}(\Omega)$ can be defined by means of the fuzzy weighted average operation in the following way

$$P(A) = (\mathcal{F}) \sum_{i=1}^{r} P_i \cdot \chi_A(\omega_i), \tag{7}$$

where $\chi_A(.)$ is the characteristic function of A.

If the set of random events on Ω is extended to fuzzy random events on Ω, and the concept of fuzzy random event probability introduced in the previous section is applied, then a fuzzy probability space with a finite set Ω can be defined as a triple $(\Omega, \mathcal{F}(\Omega), P)$, where $\Omega = \{\omega_1, \omega_2, \ldots, \omega_r\}$ is a set of elementary events whose fuzzy probabilities are given by normalized fuzzy weights $P_1, P_2, \ldots, P_r$, $P_i \neq 0$, $i = 1, 2, \ldots, r$, $\mathcal{F}(\Omega)$ is the family of all fuzzy sets on Ω (fuzzy random events), and a mapping $P : \mathcal{F}(\Omega) \to \mathcal{F}_{\mathcal{N}}([0,1])$ assigns to each fuzzy random event $A \in \mathcal{F}(\Omega)$ its fuzzy probability $P(A)$ in the following way

$$P(A) = (\mathcal{F}) \sum_{i=1}^{r} P_i \cdot A(\omega_i). \tag{8}$$

It can be proved (see [5]) that the mapping $P(.)$ defined by (8) has the properties that represent a generalization of the classical axioms of probability: a) $P(\Omega) = 1$, b) $\forall A \in \mathcal{F}(\Omega)$: $P(A) \geq 0$, c) for any $A_1, \ldots, A_s \in \mathcal{F}(\Omega)$, $A_i \cap A_j = \emptyset$ for $i \neq j$: $P(\bigcup_{j=1}^{s} A_j) = (\mathcal{F})(P(A_1) \cdot 1 + P(A_2) \cdot 1 \ldots + P(A_s) \cdot 1 + P(\overline{A_1 \cup \ldots \cup A_s}) \cdot 0)$.

In a fuzzy probability space $(\Omega, \mathcal{F}(\Omega), P)$ with a finite set Ω any mapping $U : \Omega \to \mathcal{F}_{\mathcal{N}}(\Re)$, where $\mathcal{F}_{\mathcal{N}}(\Re)$ denotes the set of all fuzzy numbers, defines a discrete fuzzy random variable. E.g. fuzzy evaluations of an alternative under possible states of the world represent a discrete fuzzy random variable. The probability distribution of a discrete fuzzy random variable U is given by a mapping $P(U_i) = P_i$, $i = 1, 2, \ldots, r$, where $U_i = U(\omega_i)$ and $P_i = P(\omega_i)$ for $i = 1, 2, \ldots, r$.

An expected fuzzy value FEU of a discrete fuzzy random variable U is defined as a fuzzy weighted average of fuzzy values $U_1, U_2, \ldots, U_r$ with fuzzy probabilities $P_1, P_2, \ldots, P_r$, i.e.

$$FEU = (\mathcal{F}) \sum_{i=1}^{r} P_i \cdot U_i \ . \tag{9}$$

Now, let us consider the following problem of decision-making under risk: The best of alternatives $x_1, x_2, \ldots x_n$ is to be chosen with respect to degrees in which the alternatives fulfil an objective associated with a criterion K. The degrees of the objective fulfilment depend not only on the alternatives themselves but also on the fact which of the possible states of the world $S_1, S_2, \ldots S_r$ occurs. Fuzzy probabilities of the states of the world are set expertly, by normalized fuzzy weights $P_1, P_2, \ldots P_r$. A fuzzy evaluation of an alternative x_i in such a situation when a state S_k occurs is expressed by a fuzzy number $U_{i,k}$ that is defined on the interval $[0, 1]$. Fuzzy numbers $U_{i,k}$, $i = 1, 2, \ldots, n$, $k = 1, 2, \ldots, r$ give an uncertain information concerning the percentage of decision objective fulfilment. These fuzzy evaluations are set either directly by an expert, e.g. by means of an evaluating linguistic fuzzy scale (if the criterion K is qualitative), or they are calculated using measured values of the criterion K and a fuzzy objective defined for K (if the criterion is quantitative). For more details concerning such fuzzy evaluations of alternatives see [3].

This problem can be described by the following decision matrix:

Table 1. Decision Matrix

	S_1 P_1	S_2 P_2	... ...	S_r P_r	FEU
x_1	$U_{1,1}$	$U_{1,2}$	...	$U_{1,r}$	FEU_1
x_2	$U_{2,1}$	$U_{2,2}$	...	$U_{2,r}$	FEU_2
...	...	...	...	...	...
x_n	$U_{n,1}$	$U_{n,2}$	...	$U_{n,r}$	FEU_n

Fuzzy numbers FEU_i express expected fuzzy evaluations of alternatives x_i for $i = 1, 2, \ldots, n$; they are calculated according to the following formula

$$FEU_i = (\mathcal{F}) \sum_{k=1}^{r} P_k \cdot U_{i,k} \ . \tag{10}$$

The best of the alternatives will be chosen by means of the rule of maximal expected fuzzy evaluation. For that purpose some of the preference relations on $\mathcal{F}_{\mathcal{N}}(\Re)$ that are described in [3] can be used, e.g. ordering of fuzzy numbers according to their centers of gravity. It is also possible to approximate the expected fuzzy evaluations of alternatives linguistically, by means of linearly ordered elements of an evaluating linguistic fuzzy scale defined on $[0, 1]$.

A similar approach can be applied also to multi-criteria decision-making under risk. In [4], utilization of expertly defined fuzzy probabilities in three-dimensional decision matrices is presented, where multi-criteria evaluating procedures are based either on fuzzy weighted averages of partial fuzzy evaluations or on fuzzy expert systems (see also [3],[2]).

5 Conclusion

Models of decision-making under risk are more realistic when the mathematical apparatus of the fuzzy sets theory is applied. The fuzzy discretization of continuous risk factors is more realistic than the classical discretization using scales of real numbers. Similarly by setting probabilities of states of the world, the exploitation of experts' knowledge is higher when fuzzy probabilities are used instead of the classical real ones.

References

1. Negoita, C.V., Ralescu, D.A.: Applications of Fuzzy Sets to Systems Analysis. Stuttgart, Birkhuser Verlag - Edituria Technica 1975.
2. Pavlačka, O.: Normované fuzzy váhy (Normalized fuzzy weights). Diploma thesis. Faculty of Natural Sciences, Palacky University, Olomouc 2004.
3. Talašová, J.: Fuzzy metody vícekriteriálního hodnocení a rozhodování (Fuzzy methods of multi-criteria evaluation and decision making). Publishing house of Palacky University, Olomouc 2003.
4. Talašová, J.: Fuzzy množiny v rozhodování za rizika (Fuzzy Sets in Decision - Making under Risk). Proceedings of the 4th Conference Aplimat, Bratislava, February 1-4, 2005, 532-544.
5. Talašová, J., Pavlačka, O.: Fuzzy metody rozhodování za rizika (Fuzzy methods of decision making under risk). Unpublished technical report. Faculty of Natural Sciences, Palacky University, Olomouc 2005.

Basic Ways of Monte Carlo Simulation to Efficient Pricing of European Options

Tomáš Tichý[1]
Department of Finance, Faculty of Economics, VŠB-TU Ostrava, Sokolská 33, 701 21 Ostrava, Czech Republic. E-mail: tomas.tichy@vsb.cz

Abstract
We can distinguish many ways to price nonlinear financial derivatives. Monte Carlo simulation is the method which is very useful mainly in pricing of non-linear types of financial derivatives – options with complicated payoff functions or when complex underlying processes are considered. In this paper we suppose Variance Gamma process, which can be regarded as a subordinated Brownian motion. We run basic variance reduction techniques to increase the efficiency of the plain Monte Carlo method in pricing of options. Plain vanilla call option is supposed to allow the comparison of all methods with the "true" price. All results are also compared to the standard case of geometric Brownian motion.

Keywords
Monte Carlo simulation, Variance reduction methods, pricing of options, Variance gamma model

1 Introduction

The essential part of functional financial markets is the knowledge of methods to determine prices of secondary financial assets (financial derivatives). It is regularly supposed that markets are efficient, prices of primary assets are given by the interaction of demand and supply, and there are no models to determine the price of primary assets more or at least such precisely as the market do.

However, the definition of financial derivatives says that their prices are derived from the price of underlying assets. Therefore, there should exist some way to get the price of any financial derivative and this price must correspond with the market view. Otherwise the arbitrage opportunity will arise.

A very interesting type of financial derivatives is an option, since the payoff function is nonlinear. It causes the risk resulting from short positions is (almost) unlimited. By contrast, the profit from long positions can be also unlimited. It implies the requirement on efficient risk management of options. Of course, it also requires to know the ways to pricing and hedging.

By an option we generally mean nonlinear financial derivative that gives its owner (*long position*) the right to buy (*call options*) or the right to sell (*put options*) the underlying asset under predefined conditions. Simultaneously, the seller of the option has an obligation to meet the right of the owner. The predefined conditions concerns, for example, the underlying amount of assets, the maturity time, the exercise price, etc. Sometimes also other conditions can be defined (average price, barrier level, etc.) and such options are referred to as exotic options.

Usually, there is a plenty of methods to price, and subsequently hedge, any option, to name some of them: solving of PDE (partial differential equation), PIDE (partial integro-differential equations) or applying the notion of martingales, the expectation operator with useful probability density functions to get analytical formula. It is naturale, that each method must lead to the same result, respecting the same inputs.

[1]Substantial part of this research was done under the support provided by GAČR (Czech Science Foundation – Grantová Agentura České Republiky) within the project No. 402/05/P085. Some algorithms written in Mathematica® were developed under the project No. 402/04/1357.

However, in some cases we can only run numerical procedures, such as binomial or multinomial lattice models, apply FDM (finite difference method), solve PDE and PIDE numerically or apply Monte Carlo simulation. For example, it is the case of options with complicated payoff functions, multifactor models or complex processes.

Some complex processes allow us to model higher moments of the distribution of asset returns. Although the non-normality of asset returns is documented starting by Fama [10], see also [18] models incorporating skewness and kurtosis of asset returns were provided relatively recently, see e.g. Variance Gamma model (VG model) (Madan and Seneta [20] for the symmetric case and Madan and Milne [19] and Madan *et al.* [21] for the asymmetric case), Normal Inverse Gaussian (NIG) model (Barndorff-Nielsen [1]) and its generalization Hyperbolic Model by Eberlein and Keller [9], Meixner model introduced by Schoutens [27] or CGMY model (according to Carr, Geman, Madan and Yor [6]) which further generalized VG model. In this paper we will suppose the VG model which can be regarded as a Brownian motion subordinated by a gamma time.

The task of this paper is to study Monte Carlo simulation and methods aiming on increasing the efficiency of plain Monte Carlo method (PMC) in pricing of options under VG model. Since the complexity of the VG model has allowed to get analytical formula only for European plain vanilla option, we will apply the method of MC simulation on this simple type of financial derivative.

In the next Chapter we define the method of Monte Carlo simulation and its application in estimating the price of options under BS model (Black and Scholes [2]) and VG model [21]. Subsequently we briefly describe basic variance reduction methods aiming primary on the quality of generated random numbers. Finally we compare the efficiency of selected methods for the case of simple BS model and complex VG model.

2 Monte Carlo Simulation in option pricing

The application of Monte Carlo simulation in pricing of options was firstly analyzed by Boyle [3] and complexly treated in Boyle *et al.* [4]. Here, we will usually suppose a European plain vanilla call. The word "European" means that the option can be exercised just at the maturity time. "Plain vanilla call" indicates that the payoff function is following:

$$\Psi_{\mathcal{T}} = (\mathcal{S}_{\mathcal{T}} - \mathcal{K})^{+} \equiv \max(\mathcal{S}_{\mathcal{T}} - \mathcal{K}; 0). \tag{1}$$

Here, $\mathcal{T}$ defines the maturity time, $\mathcal{S}_{\mathcal{T}}$ is the price of the underlying asset at maturity, and $\mathcal{K}$ is the exercise price.

Depict first the value of an option f, whose payoff function $\Psi_{\mathcal{T}}$ is defined by equation (1), at time t by $\mathcal{V}_t$. Now, define the set of risk-neutral (martingale) probabilities $\mathcal{Q}$ – such probabilities of future states under which the (risky) stochastic process $\mathcal{S}(t)$ behave as it would be martingale. Hence, if

$$\dot{\mathcal{S}}(t) = \mathcal{S}(t) \cdot e^{-r(t)} \tag{2}$$

we get

$$\mathbb{E}_0^{\mathcal{Q}}[\dot{\mathcal{S}}(t)] = \dot{\mathcal{S}}_0 \text{ for all } t \geqq 0. \tag{3}$$

In this case, we can define the value of an option f as its payoff expected at maturity and discounted at the riskless rate r up to the beginning:

$$\mathcal{V}_t = e^{-r \cdot \tau} \cdot \mathbb{E}_t^{\mathcal{Q}}[\Psi_{\mathcal{T}}], \tag{4}$$

where $\tau = \mathcal{T} - t$.

Suppose also that ω depicts future states of the world, Ω is the set of all such states, $\omega \in \Omega$, and the option payoff at maturity is uniquely determined by ω, $\Psi_{\mathcal{T}}(\omega)$. Thus, we can rewrite (4) in more details as follows

$$\mathcal{V}_t = e^{-r\cdot\tau} \cdot \mathbb{E}_t^{\mathcal{Q}}[\Psi_{\mathcal{T}}(\omega)] = e^{-r\cdot\tau} \cdot \int_{\Omega} \Psi_{\mathcal{T}}(\omega) d\mathcal{Q}. \tag{5}$$

Hence, due to (5) it is clear that to get an estimate of the option value $\hat{\mathcal{V}}_t$ it is sufficient to generate (simulate) enough relevant future states ω. Note, that relevant states are all states which can affect the option payoff – the future evolution of stock prices, interest rates, volatility, dividend yields, foreign exchange rates, etc. However, here we will study the plain vanilla European call option – the only relevant state is the underlying asset price at the maturity time, $\omega \equiv \mathcal{S}_{\mathcal{T}}$.

Denote by N the number of generated future states ω – or random scenarios of the underlying asset price evolution. Then it holds that

$$\mathcal{V}_t \approx \hat{\mathcal{V}}_t = e^{-r\tau} \cdot \frac{1}{N} \sum_{n=1}^{N} \Psi_{\mathcal{T}}^{(\mathcal{Q})} \left(\mathcal{S}_{\mathcal{T}}{}^{(n)} \right) \text{ for } N \to \infty. \tag{6}$$

Obviously, in order to get an estimate of the price we first determine the option payoff for each relevant (risk neutral) state ω (price of the underlying asset at the maturity time $\mathcal{S}_{\mathcal{T}}$. Subsequently, we have to calculate an average payoff and discount its value to the beginning. Note, that to get reliable estimate we must realize sufficiently high number of different future paths.

Suppose first, that the underlying asset price follows geometric Brownian motion in time, which has been regularly supposed starting from Black and Scholes [2]. Thus, in continuous time

$$\mathcal{S}_{t+dt} = \mathcal{S} \cdot \exp\left[(\mu - \frac{\sigma^2}{2})dt + \sigma w_{dt} \right]. \tag{7}$$

Here, constant μ and σ denote true (statistic) mean value of the price return and its standard deviation (volatility) all per annum, dt is an infinitesimal time interval and w_{dt} is a random component determined by Wiener process. For example, if we define $\tilde{\varepsilon}$ as a random number from standard normal distribution, $\tilde{\varepsilon} \in \mathcal{N}[0;1]$, then $w_{dt} = \tilde{\varepsilon} \cdot \sqrt{dt}$.

Clearly, if we want to realize a random evolution of an asset $\mathcal{S}$ in order to price an option, we must change its statistical (risky) drift μ into the risk neutral one – the riskless rate r. Moreover, we are interested only in the price at time $\mathcal{T}$. Hence, an n-th risk neutral estimate of the future price is

$$\mathcal{S}_{\mathcal{T}}{}^{(n)} = \mathcal{S}_t \exp\left[\Delta \mathcal{S}_{\tau}^{(n)} \right] = \mathcal{S}_t \cdot \exp\left[\left(r - \frac{\sigma^2}{2} \right) \cdot \tau + \sigma \tilde{\varepsilon}^{(n)} \sqrt{\tau} \right]. \tag{8}$$

From (8) it is evident that the only source of uncertainty is $\tilde{\varepsilon}$. Therefore, the optimal N is such which ensure target probability distribution (unit variance and zero mean, skewness and excess kurtosis) of the random element with minimal time cost.

The second model on which we concern here is the Variance gamma process. The model is determined by three parameters $\{\theta, \vartheta, \nu\}$. It can be defined either as a subordinated Brownian motion or as a difference between two gamma processes. Here, we apply the first approach according which the geometric Brownian motion is driven by a gamma process $\tilde{g}(\tau)$, or in other words by a random gamma time:

$$\tilde{w}_{\tilde{g}(\tau)}, \text{ where } \tilde{g}(\tau) \in \mathcal{G}\left[\frac{\tau}{\nu}; \nu \right]. \tag{9}$$

And therefore

$$\tilde{w}_{\tilde{g}(\tau)} \in \mathcal{N}\left[\theta\mathbb{E}[\tilde{g}(\tau)]; \vartheta^2\mathbb{E}[\tilde{g}(\tau)]\right]. \tag{10}$$

On the basis of (9) and (10) we can define Variance gamma process as follows

$$\mathcal{VG}(\tau) = \theta\tilde{g}(\tau) + \vartheta\sqrt{\tilde{g}(\tau)} \cdot \tilde{\varepsilon}. \tag{11}$$

Note that θ is drift of the process and primary controls the skewness, ϑ primary controls the variance, ν primary control the kurtosis and correlation between $\tilde{g}(\tau)$, and $\tilde{\epsilon}$ is zero.

The evolution of the underlying asset price in VG model can be described by

$$\mathcal{S}_{\mathcal{T}}{}^{(n)} = \mathcal{S}_t \exp\left[\Delta\mathcal{S}_\tau^{(n)}\right] = \mathcal{S}_t \exp\left[(r-\omega)\tau + \mathcal{VG}(\tau)\right], \tag{12}$$

where

$$\omega = -\tfrac{1}{\nu} \cdot \ln\left(1 - \theta \cdot \nu - \tfrac{1}{2} \cdot \vartheta^2\nu\right)$$

ensures that setting the risk neutral drift r into equation (12), we obtain exactly the risk neutral expected value:

$$\mathbb{E}[\mathcal{S}_\mathcal{T}] = \mathcal{S}_t \cdot \exp(r \cdot \tau).$$

It is also worth noting, that unlike the simple Brownian motion, also all other parameters $\{\theta, \vartheta, \nu\}$ can (and should be) change into theirs risk-neutral counterparts.

3 Variance reduction techniques

In order to get reliable estimate of the price we should realize huge number of (independent) paths – usually we need $N \gg 100000$. Although it can be very time consuming to produce such huge number of paths, the result still need not be reliable. It is the reason why efficient improvements to plain Monte Carlo simulation (PMC) are still developed. These techniques are commonly referred to as *variance reduction techniques*, since applying them we aim on reduction of the variance (error term).

Brief review of most important methods include for example Charnes [7] or Hull [16]. More complete and rigorous treatment with many applications is provided by Boyle *et al.* [4] and Glasserman [13].

Next, we briefly describe the most important techniques focusing primary on efficient producing of random numbers. Since we study here the case of plain vanilla call option, the methods applying in case of more complicated payoffs or if the whole path of the underlying asset evolution must be monitored, such as *Importance sampling*, *Control variate technique*, *Bridge sampling*, including *Brownian bridge* or *Variance gamma bridge* are not relevant.

The simplest technique, both from the theoretical and application point of view, is commonly called the *Antithetic variate method* (AVM, AMC). The method was firstly applied in option pricing by Boyle [3]. The key idea is, that if

$$\tilde{\varepsilon} \in \mathcal{N}[0;1]$$

then the same must be true also for

$$-\tilde{\varepsilon} \in \mathcal{N}[0;1].$$

The perfectly negative correlation of these two samples substantially reduce the error in estimating the price.

The improvement is presented in two aspects. First, if we set the target number of independent paths to be N then we have to generate only $M = N/2$ paths. Hence, applying the method

we can significantly decrease the time cost. Second, since $(\tilde{\varepsilon}_m - \tilde{\varepsilon}_m)/2 = 0$, the method will also have positive effect on all symmetry measures (mean, skewness), whose values will be exactly as we need. The shortcoming of this approach is that it can be applied primarily for symmetric distributions.

More sophisticated method is the *Stratified sampling* (SS, SMC). In general, there exist two approaches to SS. The first way is direct. It consists of stratifying the interval of admissible values into equiprobable strata, that is with equal probabilities. Suppose random number from standard normal distribution: $\tilde{\varepsilon} \in \mathcal{N}[0;1]$. Therefore, the interval of admissible values is $\tilde{\varepsilon} \in (-\infty; +\infty)$. The next step is to divide this interval into M subintervals in such a way that

$$\Pr\{\tilde{\varepsilon} \in (\varepsilon_m; \varepsilon_{m+1})\} = p_m, \quad p_m = \frac{1}{M}, \quad m = 1, M \tag{13}$$

and for example

$$\varepsilon_1 = -\infty, \quad \varepsilon_{M+1} = \infty.$$

Subsequently, we need to generate M random numbers uniformly distributed between zero and one: $\tilde{\mathcal{U}} \in \mathcal{R}[0;1]$. Finally, we can generate random numbers from target distribution as follows:

$$\tilde{\varepsilon} = \varepsilon_m + \tilde{\mathcal{U}} \cdot (\varepsilon_{m+1} - \varepsilon_m). \tag{14}$$

Clearly, if $M = N$, then we take just one number from each equiprobable interval. Similarly, if $M = N/2$ we have to generate five random values from each interval $(\varepsilon_{m+1}; -\varepsilon_m)$.

However, sometimes it is difficult to handle with plus or minus infinity. In this case, we can prefer indirect stratification. Applying this procedure, we generate (stratify) 'probabilities', first. Hence, we stratify the unit interval into M subintervals of equal length. Next we produce uniformly distributed random numbers from these subintervals:

$$\tilde{u} = \frac{m-1}{M} + \frac{\tilde{\mathcal{U}}}{M}. \tag{15}$$

Finally, we transform each $\tilde{u}$ into its related value of the target distribution by inverse transform. For example, considering normal distribution:

$$\tilde{\varepsilon} = \mathcal{N}^{-1}(\tilde{u}).$$

Another very interesting approach, which extend the application of SMC also for more dimensions, is *Latin hypercube sampling* (LHS). This approach was firstly introduced by McKay *et al.* [22] and later analyzed by Stein [29].

Suppose, that we need to generate two-dimensional random processes. An obvious way for one dimension would be stratified sampling. However, here we need to get two independent coordinates for each random state. Hence, we cannot stratify the interval since results obtained in this way were not strictly independent.

Fortunately, the solution is easy. We can put randomly the subintervals of both dimensions together. Suppose that $M = N = 10$ and denote the coordinates by $\{x, y\}$. The first step is to stratify the unit interval into ten equiprobable subintervals. Next we randomly permutate these subintervals. Subsequently, we can generate both coordinates for each N applying the indirect SS method. That is, we first get values of the inverse distribution function for both coordinates, where x is based on the origin subintervals and y on theirs permutation, and finally we transform them into target two-dimensional random numbers. For example, for $n = 1$ we get $x \in \langle 0, 0.1)$ and $y \in \langle 0.3, 0.4)$ and therefore $\{x, y\} = \{0.043, 0.314\}$. Figure 1 illustrates location of all coordinates within stratified subintervals.

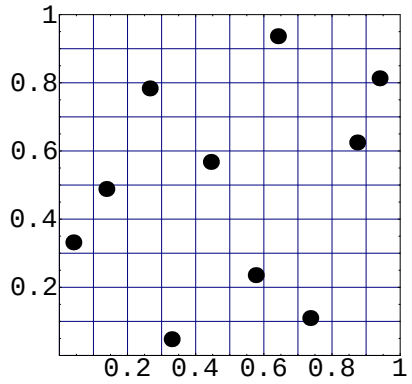


Figure 1: Illustrating LHS for two dimensions

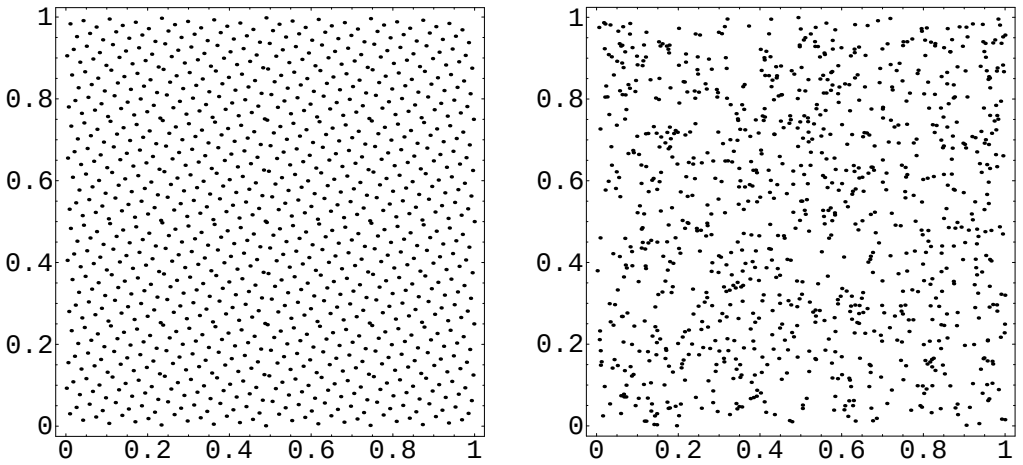


Figure 2: Comparing quasirandom (*left*) and pseudorandom (*right*) numbers

All methods described above can be regarded as special cases of Monte Carlo simulation. However, it is not the case of the Quasi Monte Carlo simulation. While applying Monte Carlo simulation we primary need to generate pseudorandom numbers uniformly distributed on the unit interval and fulfilling prespecified tests of randomness, applying Quasi Monte Carlo simulation we proceed according to the chosen algorithm and generate quasirandom numbers in deterministic rather than random sequence. The review of basic approaches has been provided e.g. by Niederreiter [23] or Glasserman [13]. Figure 2 shows the difference between pseudorandom and quasirandom numbers in two-dimensional cube for N=1000. In order to produce quasirandom numbers we have proceed according to Woźniakowski [30].

4 Pricing the European call option

In this section we will generate random evolution of the underlying asset in order to estimate the price of the European call option. First we will suppose BS model (and geometric Brownian motion), next we relax the assumption of Gaussian returns applying the Variance gamma model. All calculations are done in Mathematica® on Pentium 4 3.2 GHz HT PC with 512 MB RAM. We will subsequently suppose $N = 10^x$ random paths, where $x = 1, 2, ..., 6$.

For simplicity we have stated following parameters: $\mathcal{S} = 100, \mathcal{K} = 101, r = 10\%$ p.a., $\tau = 1$.

For the case of BS model we suppose volatility to be $\sigma = 21$. Similarly, for the case of VG model $\theta = -0.1436, \vartheta = 0.12136, \nu = 0.3$, which give us the volatility of 0.1446, the skewness of 0.8055 and the kurtosis of 4.14.

The true prices can be obtained by the BS model [2] and VG model [21]. In particular, $\mathcal{V}_{BS} = 13.0295$ and $\mathcal{V}_{VG} = 10.9815$. Results for the BS model and VG model are obvious from Table 1 and Table 2, respectively.

Method	PMC		AV MC		SMC		QMC	
N	price	time costs	price	time costs	price	time costs	price	time costs
10^2	17.7456	0.000	11.8243	0.000	12.9683	0.078	13.0268	0.016
10^3	12.8686	0.031	12.7975	0.016	13.0265	0.359	13.0766	0.078
10^4	13.0780	0.156	13.0897	0.125	13.0312	3.344	13.0382	1.156
10^5	12.9633	1.391	13.0267	1.015	13.0295	32.453	13.0305	13.782
10^6	13.0419	13.609	13.0303	10.219	13.0295	345.141	13.0296	164.296

Table 1: Estimating the price for the Black and Scholes model, $\mathcal{V}_{BS} = 13.0295$

Apparently, applying the plain Monte Carlo simulation we get pure results even for huge N. Results for AVM are slightly better with lower computational times. However, to get really reliable result we should apply SMC. Although results of simple QMC are obtained approximately in half time comparing with SMC, the results are worse. Also if we consider the estimated prices for comparable time cost, the best results are obtained by SS.

Method	PMC		AV MC		LHS MC		QMC (LHS)	
N	price	time costs	price	time costs	price	time costs	price	time costs
10^2	10.7796	0.047	10.7141	0.016	10.9701	0.265	10.7232	0.188
10^3	11.0139	0.062	11.0632	0.078	11.0486	1.735	10.9841	1.828
10^4	11.1802	0.516	11.0011	0.516	10.9736	16.781	10.9800	17.797
10^5	10.9996	4.828	10.9800	4.828	10.9727	167.422	10.9908	181.625
10^6	10.9665	49.031	10.9802	48.984	10.9840	1677.983	10.9801	1823.451

Table 2: Estimating the price for the VG model, $\mathcal{V}_{VG} = 10.9815$

Now we proceed to the case of Variance gamma process – the process which allows us to model also higher moments of the underlying distribution. Since the model consists of two independent processes, we will apply instead of SS and Simple QMC the method of Latin hypercube sampling, based either on pseudorandom numbers or quasirandom numbers.

Again, the results of PMC or with AVM are very plain. Both application of LHS method provide relatively good results. Albeit the results for BS model were satisfactory also for low number of scenarios, the complexity of the VG model caused that reliable results are obtainable only for high N with significant time costs.

4 Conclusions

In this paper, we have shown basic variance reduction methods to increase the efficiency of the Monte Carlo estimates to European option price. In particularly, we have presented and verified the efficiency of the plain Monte Carlo, antithetic variate Monte Carlo, stratified Monte Carlo and Quasi Monte Carlo. Since it does not make sense to price simple options in the Black and Scholes setting, we have concerned mainly on the Variance Gamma model, the model that allows

to model the higher moments of the underlying distributions, which is, however, related with higher computational times.

Since the simple Quasi Monte Carlo technique did not show important improvements comparing with the stratification of pseudo-random numbers in case of Black-Scholes model, we have decided to present only stratified QMC technique when studying the VG model. Clearly, it provide more accurate results, however, with higher time costs.

The efficiency of variance reduction techniques for the case of complex underlying models, such as VG model, rises when the exotic options are priced.

References

[1] BARNDORFF-NIELSEN, O. E. Normal inverse Gausian distributions and the modeling of stock returns, *Research report* No. 300, Department of Theoretical Statistics, Aarhus University, 1995.

[2] BLACK, F., SCHOLES, M. The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* **81** (May-June 1973), 637–659, 1973.

[3] BOYLE, P. Options: a Monte Carlo approach. *Journal of Financial Economics* **4**, 323–338, 1977.

[4] BOYLE, P., BROADIE, M., GLASSERMAN, P. Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21**, 1267–1321, 1997.

[5] BROADIE, M., DETEMPLE J. B. Option Pricing: Valuation Models and Applications. *Managament Science* **50** (9), 1145–1177, 2004.

[6] CARR, P., GEMAN, H., MADAN, D. B., YOR, M. Stochastic Volatility for Lévy Processes, Mathematical Finance **13**, 345–382, 2003.

[7] CHARNES, J. M. Using Simulation for Option Pricing in *Proceedings of the Winter Simulation Conference*. Orlando, FL, 2000.

[8] DUAN, J. C., SIMONATO, J. G. Empirical martingale simulation for asset prices. *Managament Science* **44**, 1218–1233, 1998.

[9] EBERLEIN, E., KELLER, U. Hyperbolic Distributions in Finance. *Bernoulli* **1**, 281–299, 1995.

[10] FAMA, E. F. The Behaviour of Stock Market Prices. *Journal of Business* **38**, 34–105, 1965.

[11] FAURE, H. Suggestions for Quasi-Monte Carlo Users, *Lecture Notes in Pure and Applied Mathematics*, Vol. 141, Marcel Dekker. 1993.

[12] GEMAN, H., MADAN, D. B., YOR, M. Time changes for Lévy processes, *Mathematical Finance* **11**, 79–96, 2001.

[13] GLASSERMAN, P. *Monte Carlo Methods in Financial Engineering*, Springer-Verlag, 2002.

[14] HALTON, J. H. On the Efficiency of Certain Quasi-Random Sequences of Points in Evaluating Multi-dimensional Integrals. *Numerical Mathematics* **2**, 84–90, 1960.

[15] HAMMERSLEY, J. M., HANDSCOMB, D. C. *Monte Carlo Methods*, Methuen, 1964.

[16] HULL, J. C. *Options, Futures, & other Derivatives.* Prentice Hall. 2002.

[17] KEMNA, A. G. Z., VORST, A. C. F. A pricing method for options based on average asset values. *Journal of Banking and Finance* 14, 113-129,1990.

[18] MACBETH, J. D., MERVILLE, L. J. An Empirical Examination of the Black-Scholes Call Option Pricing Model. *Journal of Finance* **34** (5), 1173-1186, 1979.

[19] MADAN, D. B., MILNE, F. Option pricing with VG martingale components, *Mathematical Finance* **1**, 39–56, 1991.

[20] MADAN, D. B., SENETA, E. The VG model for Share Market Returns, *Journal of Business* **63** (4), 511–524, 1990.

[21] MADAN, D. B., CARR, P., CHANG, E. C. The variance gamma process and option pricing, *European Finance Review* 2, 79–105, 1998.

[22] MCKAY, M. D., CONOVER, W. J., BECKMAN, R. J. A comparison of three methods for selecting input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245, 1979.

[23] NIEDERREITER, H. Random Number Generation and Quasi-Monte Carlo Methods. *CBMS-NSF Regional Conference Series In Applied Mathematics*, Vol. 63. Society for Industrial and Applied Mathematics. 1992.

[24] ÖTKEN, G. Quasi-Monte Carlo Methods in Option Pricing, *Mathematica in Education and Research* **8** (3-4), 52–57, 1999.

[25] RIBEIRO, C., WEBBER, N. Valuing path dependent options in the Variance Gamma model by Monte Carlo with a Gamma Bridge, *Journal of Computational Finance* **7** (2), winter 2003/2004.

[26] RUBINSTEIN, M. Implied binomial trees, *Journal of Finance* **69**, 771–818, 1994.

[27] SCHOUTENS, W. The Maixner process in Finance. *EURANDOM Report* 2001-2002, EURANDOM, Eindhoven, 2001.

[28] SCHOUTENS, W. *Lévy processes in Finance: Pricing Financial Derivatives*, Wiley, 2003.

[29] STEIN, M. Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29**, 143–151, 1987.

[30] WOŹNIAKOWSKI, H. Average-case complexity of multivariate integration. *Bull. Amer. Math. Soc. (new series)* **24** (1), 185–194, 1991.

[31] TICHÝ, T. Efficient pricing of barrier options in the Variance gamma model. In Dluhošová D. (ed.) *New Approaches and Financial Intruments in financial decision-making – Nové přístupy a finanční nástroje ve finančním rozhodování*, VŠB-TU Ostrava, 2004.

Sparse Parameter Estimation in Economic Time Series Models

Jaromír Tonner,
Faculty of Economics and Administration, Masaryk University of Brno,
Czech Republic

Abstract

The aim of this contribution is to study techniques and algorithms which are appropriate for modeling and analysis of data in economic models with a lot of parameters. So the aim is to reach a reduction of information underlying in data into the least possible number of parameters and to find their estimates with appropriately constructed and numerically stable algorithms. An attention will be devoted to predictions in economic time series and for estimation of parameters in models of small opened economics. An identification of redundant parameters and their displacement from the model will enable us an essential reduction of uncertainty of estimations of the rest of significant parameters. In this article we would like to explain and demonstrate the techniques based on ℓ_1 optimization for the estimation of parameters in models of univariate time series (ARIMA models). We will use simulated data as well as real data.

Keywords

overcomplete system, sparse system, ℓ_1 norm optimization, ARIMA models, stationary time series

1 Chapter - Introduction

Here a new approach to parameter estimation in time series models with large number of parameters is suggested. We used a modified version [1] of the Basis Pursuit Algorithm (BPA) by Chen et al [SIAM Review 43 (2001), No. 1][1] to verify its applicability to times series modeling. For simplicity we restrict to AR(I)MA models of stationary time series.

[1] [2] - Chen, Scott S., Donoho David L., Saunders Michael A.

We are also interested in testing this method for quantitative analysis of economic time series and in case of success, a continued research of quantitative analysis of economic systems is planned, especially models of small opened economics. These systems usually have a lot of zero parameters and that is why we suppose that our method will work very well.

2 Chapter - General algorithm

For the estimation of parameters the following algorithm has been applied .

1. We will choose a degree of overcompletness.For example for AR(p) it means to choose an AR process of order much higher than actual p. So we will estimate AR(m) where $m >> p$.

2. Now we will find the parameters of this process by PDSCO [1] algorithm solving Yule - Walker system in the least - squares sense

$$\begin{bmatrix} \gamma(1) \\ \vdots \\ \gamma(m) \end{bmatrix} = \begin{bmatrix} \gamma(0) & \dots & \gamma(m-1) \\ \vdots & & \vdots \\ \gamma(m-1) & \dots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_m \end{bmatrix},$$

where $\gamma(h)$ is autocovariance function. If we denote the matrix equation as

$$b = Ax$$

(for MA and ARMA processes we have different A and b but the procedure is the same) so we will have to find x^1 (exponent denotes index) which satisfies $||Ax^1 - b^1||_2 \to \min$ where we have put $b^1 := b$.

3. Having the solution x^1 from step 2, we put $b^2 := Ax^1$. Then we are going to minimize ℓ_1 norm of the parameter vector to approach to sparse solution. So we will look for x^2 satisfying $b^2 = Ax^2$ with $||x^2||_1$ minimized.

4.Further we will eliminate those components from vector x^2 which are close to zero. From matrix A and vector b^2 we have to eliminate the appropriate rows and columns.

[1] [2] - Chen, Scott S., Donoho David L., Saunders Michael A.

5. We conclude with step 2 where we use adjusted A and b^2. This algorithm can be applied due to MA or ARMA process analogically except that we must make provision for unknown Z_t. In this case A and b depend on estimated parameters and the above 5 - step procedure may be iterated several times possibly to arrive more accurate estimates.

Remark 1. Solutions in steps 2,3 and 5 was obtained via the PDSCO [1] algorithm setting its control parameters accordingly.

3 Chapter - Estimation on simulated data

A lot of computation on simulated processes was carried out. There are some of them, but only described by words. See [1] for more details regarding the examples mentioned below.

Example 1. *Let X_t is $ARMA(2,3)$ process with $\phi = [0.5,\ 0]$, $\theta = [0.8,\ 0.6,\ 0]$ and $\sigma = 1.5$, with length $n = 500$. We can consider this process overcomplete, because the last parameters are zero. We can show that ℓ_1 optimization from step 3 can find these zero parameters. Then we reduce the number of parameters and so improve our estimation in step 5.*

The estimation made by IDENT [2] is inaccurate in comparison with our procedure. It is because IDENT estimated more parameters. But our algorithm enabled us to remove redundant parameters.

The analysis of residuals of ARMA process confirms white noise.

Example 2. *Let X_t is $ARMA(4,2)$ process with $\phi = [0.5,\ 0,\ 0,\ -0.2]$, $\theta = [0.8,\ 0.6]$ and $\sigma = 1.5$, with length $n = 500$. This process has zero parameters and we can show that ℓ_1 optimization from step 3 can find these zero parameters and in further steps we put them equal zero.*

The estimation made by IDENT is inaccurate in comparison with PDSCO. It is because IDENT estimated all parameters even those equaling zero.

The residuum of ARMA process was tested by 4 procedures which did not refused the hypothesis that the residuum is white noise. We can also note that the estimation of standard deviation made by PDSCO is nearer the real standard deviation.

Example 3. *Let X_t is $ARMA(2,1)$ process with $\phi = [0.9,\ -0.8]$, $\theta = [0.6]$ and $\sigma = 1.5$, with length $n = 500$.*

[1] [2] - Chen, Scott S., Donoho David L., Saunders Michael A.

[2] MATLAB toolbox for parameter estimation

The matrix of this process is badly conditioned. Our algorithm does not converge so we have to take estimated parameters in the first step. These parameters are not worse than parameters estimated by IDENT.

The residuum of ARMA process was tested by 4 procedures which did not refused the hypothesis that the residuum is white noise. Again the estimation of standard deviation is nearer the real standard deviation.

Example 4. *Let X_t is $ARMA(2,1)$ process with $\phi = [0.9, -0.8]$, $\theta = [0.6]$ and $\sigma = 1.5$, with length $n = 500$. Here we would like to show the comparison with the previous example.*

Because of the bad condition number of estimated matrix our algorithm does not converge. So if we take any other estimation of parameters than in the first step, these parameters will be inaccurate.

The analysis of residuals refused the hypothesis that the residuum is white noise. The estimation of standard deviation does not converge and grows. The parameter estimates do not converge. Then it is not possible to simulate real process with bad parameters.

4 Chapter - Estimation on real data

Example 5. *Let x_t is the inventory investment. I used the real data of U.S. economy from [4].*

The figure 1 shows the original process and differenced process. The second figure shows the comparison of the real process and simulated process made by PDSCO and by IDENT. There $ARIMA(3,1,4)$ is simulated. We can see that simulation made by PDSCO is more accurate than that made by IDENT. It corresponds with simulated examples. We also carried out comparison with solution in [4]. Our procedure gave better simulation in sense of residuals behavior.

The second figure also shows that this method is very useful for simulation of processes with bad development and with a lot of shocks.

5 Chapter - Conclusion

After having accomplished and analyzed a lot of numerical simulations in [1] we can draw the following conclusions:

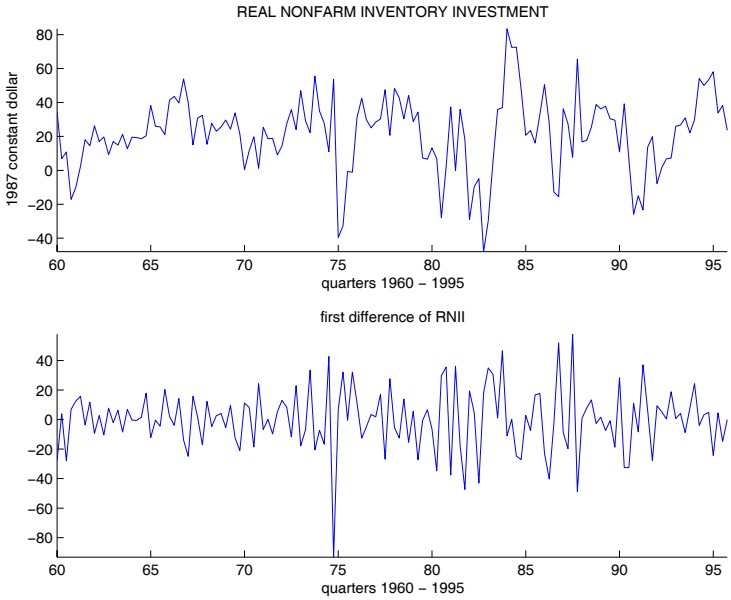


Figure 1: Real data and its first difference

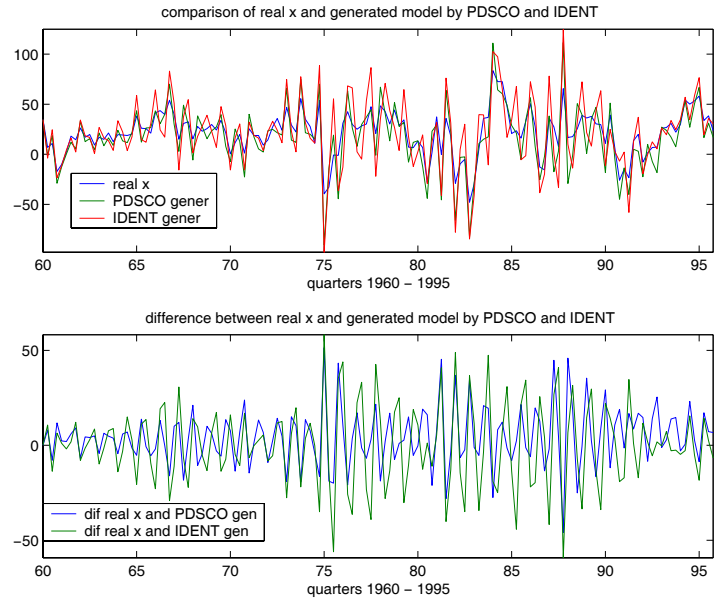


Figure 2: Comparison of processes

(1) We were able to reliably identify nearly zero parameters in the model allowing us to reduce the originally badly conditioned overparametrized model. Among others we need not take care about model orders the fixing of which is a common preliminary step used by standard techniques.
(2) As the model usually depends on the estimated parameters we tried to improve their accuracy by iterating BPA several times. In case of convergence we mostly arrived at much better estimates exhibiting narrower empirical confidence intervals compared with those obtained with MATLAB System Identification Toolbox (IDENT). If the convergence failed, the accuracy of the initial estimates could not be improved, yet still it was not worse than that from IDENT.
(3) For long memory time series having autocorrelation function with slow decay there was no benefit from the use of our BPA-based approach and the use of more precise ARFIMA models cannot be avoided.

These conclusions enable us to use our procedure to analyze processes behaving like ARIMA. Tests on real data show that this method is useful for quantitative analysis of univariate economic time series and so further research on economic systems is recommended.

References

[1] Tonner J., Veselý V.: Sparse Parameter Estimation in Overcomplete Time Series Models, Perspectives in Modern Statistical Inference III, Mikulov, July 18 - 22, 2005, Czech Republic: submitted as contributed paper.

[2] Chen, Scott S., Donoho David L., Saunders Michael A.: Atomic Decomposition by Basis Pursuit, SIAM Review Vol. 43, No. I, pp. 129 - 159.

[3] Brockwell, Peter J., Davis Richard A.: Time series: Theory and methods, 2 - nd edition., Springer - Verlag, New York, 1991.

[4] Pindyck Robert S., Rubinfeld Daniel L.: Econometric models and economic forecasts, 4 - th edition, Irwin/McGraw - Hill, 1998.

A Numerical Method for Solving Optimal Control Problems

A. Topchishvili[1], N. Jibladze[2]
1. Arbeit und Bildung e.V., Krummbogen 3, 35039 Marburg, Germany
2. Georian Technical University, M. Kostava Str. 77, 0175 Tbilisi

Abstract

A numerical method is developed for solving optimal control problems. There is realized an approach, so that an infinite-dimensional optimization problem is reduced to a finite-dimensional one, which is solved by a nonlinear programming method - the gravitation centers method (GCM). By applying GCM we find an optimal control, which can be of an arbitrary structure. We can solve two-point boundary value problems by applying algorithmic tools of random sampling. At the same time with GCM we obtain an optimal solution with an admissible accuracy and minimal consumption of computer time.
A numerical example illustrates the efficiency of the developed method for optimal control problems.

Keywords

Optimal control, space, reduction, gravitation centers method.

1 Problem Statement

Consider continuously working lumped-parameter optimal control nonstationary dynamic system. It is well-known that the state of such system varies according to the first order system of ordinary differential equations of the type

$$\dot{y} = \varphi\big(y(t), u(t), t\big) \tag{1}$$

where $y(t)$ is an n-dimensional state phase vector, $u(t)$ is an m-dimensional control vector, φ is an n-dimensional (in general case nonlinear) vector-function, t is time, $\dot{y} = dy/dt$ is a derivative of $y(t)$ with respect to t.

The vectors $y(t)$ and $u(t)$, which define the control process, are variable vectors and their variability is restricted to the given areas, such that

$$y(t) \in Y \subset R^n, \quad u(t) \in U \subset R^m. \tag{2}$$

The movement of the system (1) is restricted by the initial and the final states

$$y(t_0) = y_0, \quad y(t_f) = y_f. \tag{3}$$

where t_0 is the initial time, and t_f is the final time of the movement.

The control process over the system (1) is characterized by a criterion, which can be presented by one of the following functionals:

- the Lagrange integral functional

$$J\big[u(t)\big] = \int_{t_0}^{t_f} f\big(y(t), u(t), t\big)dt; \tag{4}$$

- the Mayer terminal functional

$$J\big[u(t)\big] = \Phi\big[y(t_f)\big]; \tag{5}$$

- the Bolz mixed functional

$$J\big[u(t)\big] = \int_{t_0}^{t_f} f\big(y(t), u(t), t\big)dt + \Phi\big[y(t_f)\big]. \tag{6}$$

Generally speaking f and Φ are nonlinear functions of the variable and control vectors.

The optimal control problem can be stated as follows. We have to find a control function $u^*(t),\ t \in [t_0, t_f]$, such that it transfers the given system from the initial state y_0 into the final state y_f, it provides the functional $J[u(t)]$ its extreme value and at the same time the dynamic equations (1), the boundary conditions (3) and the constraints (2) hold.

In the case of the Lagrange functional the given formulation can be presented in a compact way

$$\min\left\{ J[u(t)] = \int_{t_0}^{t_f} f\left(y(t), u(t), t\right) dt \;\middle|\; \begin{array}{l} \dot{y} = \varphi(y, u, t),\ y(t_0) = y_0; \\ y(t_f) = y_f;\ y(t) \in Y;\ u(t) \in U \end{array} \right\}. \qquad (7)$$

The designation $J[u(t)]$ (rather than the designation $J[y(t), u(t)]$) in the problem (7) emphasizes the fact that $u(t)$ is an independent variable and $y(t)$ is considered as a solution of the system of differential equations (1).

The function $u^*(t) = \left(u_1^*(t), u_2^*(t), \ldots, u_m^*(t)\right)^T$, which is the solution of the problem (7), is referred to as an optimal control, the function $y^*(t) = \left(y_1^*(t), y_2^*(t), \ldots, y_n^*(t)\right)^T$, which corresponds to $u^*(t)$, is called an optimal trajectory; the pair $\{u^*(t), y^*(t)\}$ defines an optimal control process.

The values t_0, t_f, y_0, y_f in the problem (7) are given, or in other words the left end $y(t_0)$ and the right end $y(t_f)$ of the trajectory $y = y(t)$ in the problem (7) are fixed. Therefore such control problem is called a problem with fixed ends. Here we can consider two groups of problems:

I. Problems with fixed control time, where the initial moment t_0 and the final moment t_f are given;
II. Problems with nonfixed control time, where either the initial moment t_0 or the final moment t_f is unknown.

Unlike the problem (7) in practice we frequently have problems, such that one or both ends of admissible trajectories are not fixed and can vary in given ranges of some sets. Such control problems are referred to as problems with free ends. Based on the form of an optimization functional, in practice we have several particular cases of optimal control problems.

As mentioned above, the control vector-function $u(t)$ can be chosen from the area U of admissible controls, which can be an arbitrary closed and bounded set from the Euclidean space R^m. Besides, we seek for the function $u(t)$ among a definite class of functions. Generally such class is limited to a class of piecewise continuous functions, such that they have only a finite number of first order break points over the interval $[t_0, t_f]$. As it is well known, such functions provide noninertial control of the object or instantaneous switching of a control function at definite moments. Therefore, selection of the piecewise continuous function class is most profitable by theoretical standpoint. At the same time, for practical application it is completely convenient too.

2 Problem Reduction

In fact, the optimal control problem (7) is a mathematical programming problem, which differs from an ordinary mathematical programming problem by the number of variables, as here we have an infinite number of variables. In other words, an optimal control problem of continuous dynamic systems is a mathematical programming infinite-dimensional problem in an infinite-dimensional space.

The later statement indicates to the formal homogeneousity of the mentioned problems. This fact exactly indicates that methods, which have been developed for solving one class of optimization problems, can be widely applied for solving optimization problems of the second class. In order that the formal homogeneousity of problems can become a reality and we can apply mathematical programming methods for solving optimal control problems it is necessary to have convenient tools, which allow reducing infinite-dimensional optimization problems to finite-dimensional ones. An approximation problem deals with this in the functional space. Thus, we can use mathematical programming

methods for solving optimal control problems in the case when we reduce an infinite-dimensional problem to a finite-dimensional one.
It is possible to reduce the problem (7) to the finite-dimensional one approximating the given infinite-dimensional functional space by a finite-dimensional subset. Based on such approximation, instead of the function $u(t)$ (with some definite accuracy) we can consider the set of linear functions

$$u_c(t) \approx u_c(\lambda) = \sum_{i=1}^{c} \lambda_i \omega_i, \ (c = 1,2,\ldots), \tag{8}$$

where $\lambda_i, i = 1,2,...,c,$ are unknown coefficients, which are to be found, and $\omega_i, i = 1,2,...,c,$ are some given analytic functions (for example, stepwise, polynomial, exponential, etc.).
Taking into account (8) the optimization functional $J[u(t)]$ with the fixed c is transformed into the ordinary function of the variables $\lambda_1, \lambda_2, \ldots, \lambda_c$: $J[u(t)] = J[\lambda_1, \lambda_2, \ldots, \lambda_c]$. The initial optimal control problem is reduced to the following finite-dimensional mathematical programming problem:

$$\min\left\{ J[\lambda] = \int_{t_0}^{t_f} f\left(y, u_c, t\right) dt \;\middle|\; \begin{array}{l} \dot{y} = \varphi(y, u_c, t),\ y(t_0) = y_0;\ y(t_f) = y_f; \\ g(y, u_c) \le 0;\ \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_c) \in Q_\lambda \subset R^c \end{array} \right\}. \tag{9}$$

Obviously, under the concrete structure of the control function $u_c(t)$, if we choose the values $\lambda_1, \lambda_2, \ldots, \lambda_c$ so that the equality $\min J[\lambda_1, \lambda_2, \ldots, \lambda_c] = J[\lambda_1^*, \lambda_2^*, \ldots, \lambda_c^*]$ takes place, the function $u_c^*(t) = \sum_{i=1}^{c} \lambda_i^* \omega_i$, which is defined by the optimal parameters $\lambda_1^*, \lambda_2^*, \ldots, \lambda_c^*$, can be considered as the solution of the initial problem in some sense.

3 Gravitation Centers Method

For solving the problem (9), and this is equivalent to finding extreme values of the parameters $\lambda_1, \lambda_2, \ldots, \lambda_c$, we are going to apply the gravitation centers method [1]. To describe this method briefly first consider the following finite-dimensional minimization problem:

$$\min\left\{ f(x) \;\middle|\; x = (x_1, x_2, \ldots, x_n) \in Q \subset R^n \right\}, \tag{10}$$

where $f(x)$ is a multimodal function, which is defined over the bounded and closed set $Q \subset R^n$. We suppose that $x^* = (x_1^*, x_2^*, \ldots, x_n^*) \subset Q$ is an absolute minimum point, where $f(x^*) < f(x) \quad \forall x \in Q$.
As it is known, the gravitation centers method, was developed based on the following statement: a sequence of gravitation centers $\{x_p\}$ of the bounded and closed sets Q_p, $p = 1,2,\ldots,\Re$, which are formed under the classic Lebesgue decomposition of the optimization function $f(x)$, converges to the absolute extreme point x^*.
Really, let's realize the standard Lebesgue decomposition of the optimization function $f(x)$ at some points ς_p, $p = 1,2,\ldots,\Re$; here ς_p, $p = 1,2,\ldots,\Re$, are scalar values from the value set of $f(x)$, and $\Re$ is the number of Lebesgue levels. As a result we get a sequence of nested one in another bounded and closed sets $\{Q_p\}$, where $Q_p = \{x \mid f(x) \le \varsigma_p\} \subset Q$, $p = 1,2,...,\Re$. When $\Re \to \infty$, the measure of the mentioned sets converges to zero. Then, by means of the well-known theorem from the functional analysis we get that $\bigcap_{p=1}^{\infty} Q_p = x^*$. As a result we get that an arbitrary sequence of points, such that each one belongs only to one set Q_p, $p = 1,...,\infty,$ converges to x^*. Therefore instead of this

arbitrary sequence of converging points we shall take the sequence $\{x(\varsigma_p)\}$, consisting of the gravitation centers of these sets, such that $x(\varsigma_p) \in Q_p$, $p = 1,2,...,\infty$. Thus

$$x^* = \lim_{p\to\infty} x(\varsigma_p), \tag{11}$$

where

$$x(\varsigma_p) = \frac{\int\!\!\!\!\!\!\int\limits_{Q_p} x[f(x)-\varsigma_p]^2\,dx}{\int\!\!\!\!\!\!\int\limits_{Q_p} [f(x)-\varsigma_p]^2\,dx}. \tag{12}$$

It is too difficult to calculate analytically the expression (12). In the most cases it is impossible to realize it practically, as we can not define in an explicit way the integration area Q_p. Therefore we introduce the indicator function

$$\theta(x,\varsigma_p) = \begin{cases} 1, & f(x) \le \varsigma_p, \\ 0, & f(x) > \varsigma_p. \end{cases} \tag{13}$$

Using (13) we get

$$x(\varsigma_p) = \frac{\int\!\!\!\!\!\!\int\limits_{Q} x[f(x)-\varsigma_p]^2\,\theta(x,\varsigma_p)dx}{\int\!\!\!\!\!\!\int\limits_{Q} [f(x)-\varsigma_p]^2\,\theta(x,\varsigma_p)dx}. \tag{14}$$

The algorithmic realization of the method takes into account calculation of coordinates of gravitation centers by the Monte-Carlo method. Based on generality and simplicity of realization the Monte-Carlo method is almost a unique tool when calculating high dimensional multiple integrals.

4 Numerical Example

Given a continuous dynamic system, such that its state varies according to the linear differential equation

$$\begin{pmatrix} \dot{y}_1 \\ \dot{y}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -0.16 & -0.032 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} 0 \\ 0.16 \end{pmatrix} u, \tag{15}$$

and the optimization functional is given by the integral of a quadratic form

$$J[u] = 0.5 \int_{t_0}^{t_f} \left(y_1^2 + y_2^2 \right) dt \to \min. \tag{16}$$

When $t = 0$, the system is at the initial state

$$y_0 = y(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{17}$$

and at the moment $t = t_f$, which is not a fixed one, the system must reach a new state

$$y_f = y(t_f) = \begin{pmatrix} 3 \\ 0 \end{pmatrix}. \tag{18}$$

Introduce the following constraints for the state variable y_1 and the control function u:

$$y_1(t) \le 3, \quad t \in [0; 30.5], \tag{19}$$

$$1 \le u(t) \le 2, \quad t \in [0; 30.5]. \tag{20}$$

We are searching an optimal solution among the class of stepwise functions, given in Fig. 1.

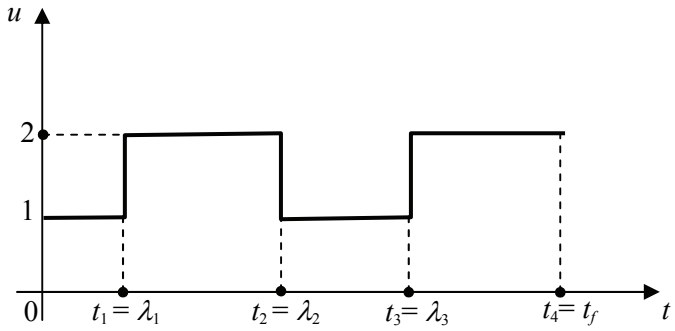


Fig. 1 The class of stepwise functions

This problem with $n = 2$ and $m = 1$ was solved in Tabak and Kuo [2]. The authors reduced the initial dynamic problem to a static nonlinear program with 10 independent variables and 16 constraints given by linear inequalities and nonlinear equalities. They applied the Fiacco and McCormick's method of sequential approximation minimization for finding optimal parameters. An optimal control function was defined by the following parameters: $t_1^* = 11.9\,\text{sec}$, $t_2^* = 14.41\,\text{sec}$, $t_3^* = 15.41\,\text{sec}$, $t_4^* = 21.36\,\text{sec}$. By means of these values the dynamic system reached the final state, defined by the relations (18), in $t_f^* = t_4^* = 21.36\,\text{sec}$.

For solving the optimal control problem (15)-(20) by means of the gravitation centers method, from one hand, it is necessary to consider the following new generalized criterion instead of the optimization functional (16):

$$\widetilde{J}[\lambda_1, \lambda_2, \lambda_3] = 0.5 \int_{t_0}^{t_f} \left(y_1^2 + y_2^2 \right) dt + [y_1(t_f) - 3]^2 + [y_2(t_f) - 0]^2 \to \min. \tag{21}$$

After minimizing the functional (21) we obtain the minimal value of the functional J and the boundary conditions (18) automatically hold with some accuracy as well. From the other hand, we realize the stepwise control functions given in Fig. 1 in a program way by means of the expression

$$u(t) \approx u(\lambda_1, \lambda_2, \lambda_3) = \begin{cases} 1, & t < \lambda_1, \ \lambda_2 \le t < \lambda_3, \\ 2, & \lambda_1 \le t < \lambda_2, \quad t \ge \lambda_3, \end{cases} \tag{22}$$

where the parameters $\lambda_1 = t_1$, $\lambda_2 = t_2$ and $\lambda_3 = t_3$ are randomly selected. By fixing the switching time $t_i = \lambda_i$ the control function (22) is uniquely defined, and then the given optimal control problem has been reduced to the three-dimensional problem for determining parameters λ_1, λ_2, λ_3.

For solving the problem introduce the following values of the program parameters: the number of statistical experiments $N = 100$, the number of Lebesgue levels $\Re = 10$, the integration interval $t \in [0; 30.5]$, the integration step $H = 0.5$, the admissible accuracy value $\varepsilon = 0.1$.

Table 1. The results of the stages

p	ζ	λ_1	λ_2	λ_3
1	76.6833	9.5881	13.6285	17.8128
2	71.7216	9.6380	13.7524	17.7898
3	66.7599	9.7082	13.9460	17.7014
4	61.7982	9.5489	14.4802	17.5250
5	56.8366	9.1184	14.5642	17.6923
6	51.8749	6.9202	13.3518	17.3440
7	46.9132	5.4470	11.6855	15.3373
8	41.9515	5.4470	11.6855	15.3373
9	36.9899	5.4470	11.6855	15.3373
10	32.0281	5.6189	10.8802	14.2979

On the first stage of calculations, based on the preliminary $S=10$ statistical experiments, we define the Lebesgue decomposition levels ς_p, $p=1,2,\ldots,10$, for the functional $\widetilde{J}$.

On the second stage based on the results of main $N=100$ statistical experiments, we get some empirical points of the parameters $\lambda_1,\lambda_2,\lambda_3$. The results of the first and the second stages are given in Table 1 (see above).

On the third stage, based on the parabolic approximation of the empiric points of the functions $\overline{\lambda_i(p)}$, $p=1,2,3$, we determine the following approximation polynomials:

$$\begin{aligned}\lambda_1(p) &= -0.01093\,p^2 - 0.49109\,p + 10.77002,\\ \lambda_2(p) &= -0.06767\,p^2 + 0.38008\,p + 13.48068,\\ \lambda_3(p) &= -0.04868\,p^2 + 0.12622\,p + 17.79764.\end{aligned} \tag{23}$$

For finding optimal values of the parameters λ_1, λ_2, λ_3 we solve a single-dimensional minimization problem $\min\left\{\widetilde{J}[\lambda(p)] \,\middle|\, p\in[8;12]\right\}$ and determine $p^*=9.6683$. Substituting this value into the expressions (23) we get $\lambda_1^*=5.0003$, $\lambda_2^*=10.8303$, $\lambda_3^*=14.4672$. Thus, the optimal control function will be of the following form:

$$u^*(t)\approx u(\lambda_1^*,\lambda_2^*,\lambda_3^*)=\begin{cases}1, & t<5.0003,\quad 10.8303\le t<14.4672,\\ 2, & 5.0003\le t<10.8303,\quad t\ge 14.4672.\end{cases} \tag{24}$$

Substituting (24) into the system (15) and numerically integrating the later one by the Runge-Kutta method we get an optimal trajectory $y^*(t)$, which coordinates are given in Table 2. Besides, we get the minimal value of the generalized functional: $\widetilde{J}^*=\widetilde{J}[\lambda_1^*,\lambda_2^*,\lambda_3^*]=31.5357$.

Table 2. The coordinates of the optimal trajectory

t	0	1	2	3	4	5	6	7	8	9	10
y_1^*	0	0.08	0.30	0.62	0.99	1.35	1.71	2.10	2.47	2.75	2.91
y_2^*	0	0.15	0.28	0.36	0.38	0.34	0.39	0.39	0.33	0.22	0.08

As we can observe from Table 2, the given dynamic system reached the final state in $t_f^*=10$ sec with the accuracy $\varepsilon=0.1$. This result is better than the one achieved by D. Tabak and B. Kuo. As $t_f^*<t_2=10.8303$, we can conclude, that only the first switching moment $t_1=\lambda_1^*=5.0003$ sec of the control function effected on achieving the final state. The mentioned fact stipulated for the necessity to change the structure of the control function, owing to what the same problem was solved for a simpler, single step control function. By the gravitation centers method we determined the following optimal control function:

$$u^*(t)\approx u(\lambda^*)=\begin{cases}1, & t<4.99988,\\ 2, & t\ge 4.99988.\end{cases} \tag{25}$$

By means of the expression (25) we calculated the optimal trajectory's coordinates. Finally, we got again that the system reached the final state in $t_f^*=10$ sec.

References

[1] Jibladze, N., and Topchishvili, A. Static Optimization Numerical Methods, Institute of Control Systems of Georgian Academy of Sciences, Tbilisi, Georgia, 2001, 422pp. (in Georgian).

[2] Tabak, D., and Kuo, B.C. Optimal Control by Mathematical Programming, Prentice-Hall, Inc. Englewood Cliffs, NJ, 1971, 237pp.

Copulas - Risk Measures of Association

K. Vaníček

Charles University
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics
Prague
Czech Republic

Abstract. We introduce an application of the copula theory in the field of integrated risk management (IRM). We first remind some properties of commonly used dependence measures. The axiomatic definition of copulas, Sklar's theorem and some well-known general properties of copulas are presented in the following section. In the third section we define Archimedean copulas with their generators and present a simple way of computing important measures of association–Kendall's tau and Spearman's rho. Examples of parametric families are presented at the end of this section. In the last section we present possible parameter estimation procedures and simulation techniques that can be applied to the individual set of data.

Introduction

A lot of topics associated with measuring risks in finance have already been addressed, but primary focus was dedicated to the one–dimensional problems. Multivariate extensions to higher dimensions, especially for $n \geq 3$, have not almost been discussed and implemented yet, even though there is a strong theoretical background in the probabilistic area, see [10] or [6]. We would like to present in this paper a comprehensive approach for modelling multivariate distribution functions of risks, given their one-dimensional margins and dependence structure. Some of the particular econometric ideas are taken from [2], [3], [5] and [11].

We are especially motivated by the new coming issues in the risk management area associated with aggregating capital requirements for different types of risks (market, credit, operational, financial or volumetric), consult [14].

New trends in the time series analysis are also dealing with copula theory and one of the recent interesting application to FX market can be found in [2]. Structured derivatives such as nth-to-default CDO's are nowadays also priced and evaluated via copula functions and even the famous RiskMetrics group applied such concept into their commercial products for active risk management.

Dependence Measures

Rank Correlation

Kendall's tau and Spearman's rho, both the sample and the population versions, are one of the best alternative to the linear correlation as a measures of dependence, especially for non-elliptical distributions and for non-linear type of dependence. In the modern terminology they are called measures of association. We define these measures in the following. Some of their important properties can be found for instance in [3] and [5].

Definition 1. Kendall's rank correlation
Let $(X_1, X_2), (Y_1, Y_2)$ be independent and identically distributed random vectors with joint distribution function H. Then the population version of Kendall's tau is defined:
$\rho_\tau(X, Y) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$.

Definition 2. Spearman's rank correlation
Let X, Y be random variables with distribution functions F, G. Then the population version of Spearman's rho is defined: $\rho_S(X,Y) = \rho(F(X), G(Y))$.

Copulas

The word copulae is a Latin noun which means "a link, tie or bond" (Cassell's Latin Dictionary). The concept, which can separate the joint distribution function into one dimensional margins and their interdependence structure (copula) seems to us incredibly useful for financial modelling purposes. The central result is formulated in the Sklar's theorem, which is the foundation of many, if not most, of the applications of copulas to statistics. A huge amount of theoretical results, examples and various exercises for 2-dimensional copulas can be found in [12]. Unfortunately possible extensions to higher dimensions are discussed very briefly and practitioners shall consult extensive treatise [10].

Definition 3. Copula
A copula is a function $C : [0,1]^n \to [0,1]$, which has the three properties:

1. *$C(x_1, \ldots, x_n)$ is increasing in each x_i*
2. *$C(1, \ldots, 1, x_i, 1, \ldots, 1,) = x_i \quad \forall i, x_i \in [0,1]$*
3. *$\forall (a_1, \ldots, a_n), (b_1, \ldots, b_n) \in [0,1]^n$ with $a_i \leq b_i$:*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C(x_{1i_1}, \ldots, x_{ni_n}) \geq 0,$$

where $x_{j1} = a_j, x_{j2} = b_j, \forall j = 1, \ldots, n$.

Theorem 4. Sklar's theorem
Let F be an n-dimensional distribution function with margins $F_1, \ldots, F_n$. Then there exists a copula C such that for all $\mathbf{x} = (x_1, \ldots, x_n) \in R^n$,

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

If $F_1, \ldots, F_n$ are all continuous, then C is unique; otherwise C is uniquely determined on $RanF_1 \times \cdots \times RanF_n$. Conversely if C is a copula and $F_1, \ldots, F_n$ are distribution functions, then the function F defined above is an n-dimensional distribution function with margins $F_1, \ldots, F_n$.

Example 5. For independent random variables the copula trivially takes the form

$$C_{ind}(u_1, \cdots, u_n) = u_1 \cdots u_n$$

Example 6. The bivariate Gaussian or normal copula can not be unfortunately expressed in a closed form and is defined as

$$C^{\rho}_{Ga}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{ \frac{-(s^2 - 2\rho st + t^2)}{2(1-\rho^2)} \right\} dsdt,$$

where $-1 < \rho < 1$ and Φ is the univariate standard normal distribution function. Thus the standard bivariate normal distribution function with correlation coefficient ρ is simply given by

$$\Phi^2_{\rho}(x_1, x_2) = C^{\rho}_{Ga}(\Phi(x_1), \Phi(x_2)).$$

Example 7. The copula of the bivariate t-distribution with ν degrees of freedom and correlation ρ can be expressed as

$$C^{\nu,\rho}_{t}(u_1, u_2) = \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} dsdt,$$

where $t_\nu^{-1}(u_1)$ denotes the inverse of the distribution function of the standard univariate t-distribution with ν degrees of freedom.

One of the possible criteria for choosing reasonable and relevant dependence structure is the asymptotic measure of association between the tails of marginal distributions, which is generally called tail dependence. This criterium shall be primary for a multivariate extension of the EVT. Not surprisingly it turns out that tail dependence is a copula based property and as noted and proven in [10], tail dependence is also invariant under strictly increasing transformation of the margins.

Definition 8. Tail dependence
Let X and Y be random variables with distribution functions F, G. The coefficient of upper tail dependence of X and Y is

$$\lim_{u \nearrow 1} P(Y > G^{-1}(u) | X > F^{-1}(u)) = \lambda$$

provided that the limit $\lambda \in [0,1]$ exists. If $\lambda \in (0,1]$, X and Y are said to be asymptotically dependent in the upper tail, if $\lambda = 0$ they are asymptotically independent.

Example 9. The tail dependence basically describes probability of occurrence of two extreme events simultaneously. More precisely, the conditional probability of occurrence of one extreme event conditioned on occurrence of another extreme event. If we assume that the underlying distributions functions F, G are both continuous we can obtain that the limit exists and the following relations in terms of copulas and survival copulas:

$$\lambda = \lim_{u \nearrow 1} \frac{P(Y > G^{-1}(u), X > F^{-1}(u))}{P(X > F^{-1}(u))} = \lim_{u \nearrow 1} \frac{1 - 2u + C(u,u)}{(1-u)}$$

Archimedean Copulas

An essential class of copula families for our modelling intentions are so called Archimedean copulas. The nice properties of such copulas can be summarized as: 1) They can be very easily constructed and implemented; 2) They are able to capture the asymmetry of distributions. The last virtue is essential for our applications in finance, because there are obvious economic evidences of such behaviour.

Definition 10. Archimedean copulas
Let $\varphi : [0,1] \to [0,\infty]$ be a continuous, strictly decreasing function such that $\varphi(0) = \infty$, $\varphi(1) = 0$ and let its inverse φ^{-1} satisfies for all $t \in [0,\infty]$ and $k = 0, 1, 2, \ldots$

$$(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0.$$

The n-dimensional copula, $n \geq 2$

$$C(u_1, \ldots, u_n) = \varphi^{-1}\left(\sum_{i=1}^{n} \varphi(u_i) \right), \quad (u_1, \ldots, u_n) \in [0,1]^n$$

is called Archimedean and the function φ its generator.

Since the families of Archimedean copulas are uniquely defined by their generators, it is possible to study their properties using only the generator function. The computation of Kendall's tau and Spearman's rho becomes now fairly easy and its sample version will turn out to be crucial for estimating the parameters of the underlying copula family.

Corollary 11. Kendall's tau and Spearman's rho for Archimedean copulas
Let X and Y be random variables with an Archimedean copula C generated by φ. The population version of Kendall's tau for X and Y is given by

$$\rho_\tau = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

The population version of Spearman's rho for X and Y is given by

$$\rho_C = 12 \int\int_{\langle 0,1\rangle^2} [C(u,v) - uv] du dv.$$

Similar result, in terms of an application of the generator function, can be obtained for tail dependence, even though we have to assume some additional condition on the generator function φ.

Corollary 12. Tail dependence for Archimedean copulas
Let φ be a generator of an Archimedean copula C. If $(\varphi^{-1})'(0)$ is finite, then C does not have tail dependence, i.e. are asymptotically independent. If $(\varphi^{-1})'(0) = -\infty$, then C has tail dependence and the coefficient of tail dependence is given by

$$\lambda = 2 - 2 \lim_{s \searrow 0} \frac{(\varphi^{-1})'(2s)}{(\varphi^{-1})'(s)}.$$

We have proposed a straightforward way how can we directly compute or estimate the measures of association , it would be quite illustrative to give some examples.

Example 13. Consider the bivariate Gumbel family of copulas given by

$$C_\theta^{Gu}(u,v) = \exp\left(- \left((-\ln u)^{1/\theta} + (-\ln v)^{1/\theta} \right)^\theta \right)$$

with generator $\varphi(t) = (-\ln t)^\theta$, where $\theta \geq 1$. Since $\varphi(t)/\varphi'(t) = (t \ln t)/\theta$ and $(\varphi^{-1})'(0) = -s^{1/\theta - 1} \exp(-s^{1/\theta})/\theta|_{s=0} = -\infty$ we can calculate Kednall's tau and the coefficient of tail dependence λ (by the definition or using the generator function φ) as:

$$\begin{aligned}
\rho_\tau &= 1 + 4 \int_0^1 (t \ln t)/\theta = 1 + \frac{4}{\theta}(0 - 1/4) = 1 - 1/\theta \Longrightarrow \theta = \frac{1}{1-\rho_\tau}, \\
\lambda &= 2 - 2 \lim_{s \searrow 0} \frac{-(2s)^{1/\theta - 1} \exp(-(2s)^{1/\theta})/\theta}{-s^{1/\theta - 1} \exp(-s^{1/\theta})/\theta} = 2 - 2^{1/\theta} \lim_{s \searrow 0} \frac{\exp(-(2s)^{1/\theta})}{\exp(-s^{1/\theta})} = 2 - 2^{1/\theta}, \\
\lambda &= \lim_{u \nearrow 1} \frac{1 - 2u + C(u,u)}{1-u} = 2 - \lim_{u \nearrow 1} 2^{1/\theta} u^{2^{1/\theta} - 1} = 2 - 2^{1/\theta}.
\end{aligned}$$

Example 14. Consider the bivariate Clayton family of copulas given by

$$C_\theta^{Cl}(u,v) = \max((u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, 0)$$

with generator $\varphi(t) = (t^{-\theta} - 1)/\theta$, where $\theta \in [-1, \infty) \backslash \{0\}$. Since $\varphi(t)/\varphi'(t) = (t^{\theta+1} - t)/\theta$ and $(\varphi^{-1})'(0) = -(1+\theta s)^{-1/\theta - 1}|_{s=0}$ is finite, we obtain:

$$\begin{aligned}
\rho_\tau &= 1 + 4 \int_0^1 \frac{t^{\theta+1} - t}{\theta} = 1 + \frac{4}{\theta} \left(\frac{1}{\theta + 2} - \frac{1}{2} \right) = \frac{\theta}{\theta + 2} \Longrightarrow \theta = \frac{2\rho_\tau}{1 - \rho_\tau}, \\
\lambda &= 0.
\end{aligned}$$

Estimation procedures and fitting copulas

Step-by-step algorithm

Let consider bivariate problem from stock exchange. Suppose we would like to estimate dependence structure of the daily movements of two stocks (e.g. IBM and Dell). The way how can we solve the problem, is summarized in the following estimation procedure in:

1. Estimate the marginal distribution functions of daily movements of both stocks using appropriate techniques, consult [15].

2. Choose some suitable parametric family of Archimedean copulas (will be developed in more details)
3. Estimate the sample version of Kendall's tau for the bivariate historical daily movements.
4. For a given generator function φ evaluate one dimensional integral and determine direct link between Kendall's tau and the parameter θ. Consult examples above.
5. Bivariate distribution function of the daily movements is uniquely determined by the two marginal distribution functions and its copula function.

Simulation procedure

We have already presented the relationship between the parameter θ and the Kendall's tau ρ_τ for some families of copulas. There is indeed another simulation problem, which can be solved by numerical root finding. Consider that we have in hands the parameter θ and we would like to simulate the random vector from predetermined family of copulas. The following theorem gives us the general solution tool.

Theorem 15. Distribution function of Archimedean copulas
Let U, V be uniform $(0, 1)$ random variables whose joint distribution function is an Archimedean copula C generated by φ in Ω. Then $K_C(t)$ given by

$$K_C(t) = t - \frac{\varphi(t)}{\varphi'(t^+)}$$

is the distribution function of the random variable C(U,V).

If we employ this result we can obtain following procedure for generating random vector from particular copula.

1. Simulate two independent uniform $(0, 1)$ random variables u and v.
2. Set $t = K_C^{-1}(v)$, where K_C is the distribution function of $C(U, V)$.
3. Set $q = \varphi^{-1}(u\varphi(t))$ and $s = \varphi^{-1}((1 - u)\varphi(t))$.

Sometimes it is impossible to compute the inverse of K_C so the result of step 2 can be obtained through the equation

$$(t - \frac{\varphi(t)}{\varphi'(t^+)}) - v = 0$$

by numerical root finding. The 2-dimensional random vector (q, s) is generated from given copula family with parameter θ.

Discussion

When studying the Archimedean copulas and their implementation in practice, we have identified at least two basic problems we will have to deal with. As already indicated, we would like to extend the application of Archimedean copulas for the 3-dimensional distribution function. While it is simple to generate 3-dimensional Archimedean copulas of the form given in the basic definition, they suffer from a limited dependence structure since all 2-margins are identical.

We would like to design a multivariate extension of the basic definition with different generator functions for each 2-marginal copulas. This could provide us a dependence structure with nonexchangeable and different margins which shall be more appropriate for practical applications. But we shall still keep in mind conditions in the basic definition since any multivariate extension should contain the basic definition as a special case. We haven't gone into more details yet, but several references for such extensions are discussed in [10].

Another problem, which immediately arises already for bivariate distributions is the choice of the parametric family of Archimedean copulas. Which family of Archimedean copulas would fit our data best? We are able to compare predetermined families of Archimedean copulas employing

Q-Q plots, statistical tests or by minimizing different reasonable metrics as discussed in [7], [8], [9] and [1] with practical application in MS Office (VBA code) presented in [13]. However there is an unbounded variaty of possible choice of the generator functions (parametric families), we haven't found by now any procedure which would indicate the most "appropriate" family of copulas. We would like to drive our attention to this problem in order to design both parametric and non-parametric tests for identifying such family of Archimedean copulas.

Conclusion

We have presented the main ideas standing behind the theory of copulas and dependence in general, but there is still an extensive list of properties, outcomes, generalizations and applications which have remained undiscussed. Interested reader shall consult references for further theoretical and practical results. Even we have described a straightforward and natural way of implementation of Archimedean copulas in practice, we have also pointed out some deficiencies that shall be studied in more details. Especially we would like to focus on the multivariate extension of nonexchangeable Archimedean copulas in our further research.

References

1. Caperaa P., Fougeres A., Genest C. (1997): *A Nonparametric Estimation Procedure for Bivariate Extreme Value Copulas.* Biometrika 84: 567-577.
2. Dias A., Embrechts P. (2003): *Dynamic copula models for multivariate high-frequency data in finance.* ETH Zurich, Preprint, http://www.math.ethz.ch/%7Ebaltes/ftp/fx2004.pdf.
3. Embrechts P., McNeil A., Straumann D. (2002a): *Correlation and Dependence in Risk Management: Properties and Pitfalls.* In: Risk Management: Value at Risk and Beyond (Ed.M. Dempster), Cambridge University Press, Cambridge, 176-223.
4. Embrechts P., Lindskog F., McNeil A. (2003): *Modelling Dependence with Copulas and Applications to Risk Management.* In: Handbook of Heavy Tailed Distributions in Finance, ed. S. Rachev, Elsevier, Chapter 8, pp. 329-384.
5. Embrechts P., Höing A., Juri A. (2002b): *Using copulae to bound Value-at-Risk for functions of dependent risks.* Finance & Stochastics 7(2), 145-167.
6. Fang K.-T., Kotz S., Ng K.-W. (1990): *Symmetric Multivariate and Related Distributions.* Chapman & Hall, London.
7. Frees E. W., Valdez E. A. (1998): *Understanding relationships using copulas.* North American Actuarial J. 2, 1-25.
8. Genest C., Rivest L. (1993): *Statistical Inference Procedures for Bivariate Archimedean Copulas.* Journal of the American Statistical Association 88: 1034-1043.
9. Genest C., Ghoudi K., Rivest L. (1995): *A Semiparametric Estimation Procedures of Dependence Parameters in Multivariate Families of Distributions.* Biometrika 82: 543-552.
10. Joe H. (1997): *Multivariate Models and Dependence Concepts.* Chapman & Hall, London. Chapman & Hall, London.
11. Juri A. (2002): *Applications of dependence concepts in insurance and finance.* ETH Zurich, PhD Thesis, http://e-collection.ethbib.ethz.ch/ecol-pool/diss/fulltext/eth14499.pdf.
12. Nelsen R. (1999): *An Introduction to Copulas.* Springer, New York.
13. Melchiori M. R. (2003): *Which Archimedean Copula is the right one?* Universidad Nacional del Litoral, Santa Fe, http://www.riskglossary.com/papers/Copula_carta.pdf.
14. Rosenberg J.V., Schuermann T. (2004): *A General Approach to Integrated Risk Management with Skewed, Fat-Tailed Risk.* Federal Reserve Bank of New York, Staff Reports.
15. Vaníček K. (2003): *Securing Market Risk Extreme Value Theory.* MFF UK, Prague, Diploma Thesis.

A Model Interpretation of the Czech Inflation Targeting and the Monetary Policy*

Osvald Vašíček and Karel Musil

Faculty of Economics and Administration of Masaryk University in Brno
Department of Applied Mathematic and Computer Science, Lipová 41a, 602 00 Brno
email: osvald@econ.muni.cz, musilk@econ.muni.cz

Abstract

The paper introduces a New Keynesian DSGE model that describes the inflation targeting policy in the Czech economy. This model is based strictly on microfoundations and consists of representative finished and intermediate goods–producing firms, representative households and a central bank. The central bank implements its policy according to the generalized Taylor rule. A suitable method for solving the model is the Kalman filter evaluating a likelihood function and the Kalman smoother evaluating a time series of a smoothed estimate of the unobserved state variable (target inflation). The model seems to give an approximation of the behavior of the Czech economy.

Keywords

DSGE model, New Keynesian model, monetary policy, Taylor rule, inflation targeting, rational expectations, Kalman filter with likelihood function, Kalman filter smoother

1 Model

The basic goal of this paper is to interpret the implementation of the monetary policy of the Czech National Bank. That means to use a model approach to clarify the inflation targeting in the Czech economy[1].

In the first part the paper introduces a suitable model of the economy that allows to use it as an adequate tool for the following interpretation. In the next part there is an estimation of this model with respect to the condition of the Czech economy. In the end we assess and analyze estimated results and try to interpret them.

The model[2] is a dynamic stochastic general equilibrium model with microfoundations. Its structure comes under the New Keynesian approach. For our purposes we use a model of Peter N. Ireland (2005*a*) and the whole next part is based on his paper.

The model consists of four representative agents: representative households, intermediate goods–producing firms, finished goods–producing firms and a central monetary authority.

1.1 The Representative Households

The budget constraint of a representative household is following:

$$M_{t-1} + T_t + B_{t-1} + W_t h_t + D_t \geq P_t C_t + M_t + B_t/R_t \qquad (1)$$

*This paper has been worked as a part of a project at the Center of Czech Economy Competitiveness Research (identification code 1M0021622405).

[1]For some practical experience or more information about the inflation targeting in the Czech economy see Czech National Bank (2005).

[2]Authors thank to Peter N. Ireland for his permission to use his model for our purpose and applications on the conditions in the Czech economy.

for all t = 0, 1, 2, ..., where money in the period t is M_t, bonds B_t, a lump–sum nominal transfer T_t, nominal profits D_t in the form of dividends, R_t denotes the gross nominal interest rate, h_t a supply of labor, W_t a nominal wage, C_t a consumption and P_t a price of goods.

The household's preferences are described by the expected utility function, where the discount factor is β and the habit formation parameter is γ: $1 > \beta > 0$ and $1 > \gamma \geq 0$: $E_0 \sum_{t=0}^{\infty} \beta^t a_t [ln(C_t - \gamma C_{t-1}) + ln(M_t/P_t) - h_t]$. The preference shock a_t follows the stationary autoregressive process: $ln(a_t) = \rho_a ln(a_{t-1}) + \sigma_a \epsilon_{at}$ (2) for all t = 0, 1, 2, ..., with $1 > \rho_a \geq 0$ and $\sigma_a \geq 0$, where the serially uncorrelated innovation ϵ_t has the standard normal distribution.

The first order conditions of this problem are (for t = 1, 2, 3, ...):

$$\Lambda_t = \frac{a_t}{C_t - \gamma C_{t-1}} - \beta\gamma E_t \left(\frac{a_{t+1}}{C_{t+1} - \gamma C_t} \right) \quad (3)$$

$$\Lambda_t = \beta R_t E_t \left(\frac{\Lambda_{t+1}}{P_t / P_{t-1}} \right) \quad (4)$$

$$\frac{M_t}{P_t} = \left(\frac{W_t}{P_t} \right) \left(\frac{R_t}{R_t - 1} \right) \quad (5)$$

1.2 The Representative Intermediate and Finished Goods–Producing Firm

The aggregate technology shock to production is $ln(Z_t) = ln(z) + ln(Z_{t-1}) + \sigma_z \epsilon_{zt}$ (6) for all t = 0, 1, 2, ..., with $z > 1$ and $\sigma_z \geq 0$, where the serially uncorrelated innovation ϵ_{zt} has the standard normal distribution.

Every firm tries to maximize its real profits during period t as a real price of the whole production reduced by a real cost and a cost of price adjustment:

$$\frac{D_t(i)}{P_t} = \left[\frac{P_t(i)}{P_t} \right]^{1-\theta_t} Y_t - \left[\frac{P_t(i)}{P_t} \right]^{-\theta_t} \left(\frac{W_t}{P_t} \right) \left(\frac{Y_t}{Z_t} \right) - \frac{\phi}{2} \left[\frac{P_t(i)}{\Pi_{t-1}^{\alpha} (\Pi_t^*)^{1-\alpha} P_{t-1}(i)} - 1 \right]^2 Y_t \quad (7)$$

There is a quadratic cost of adjusting the price between periods, measured in real terms (expressed as the last term of the previous equation), where $\phi \geq 0$ governs the magnitude of the adjustment cost, Π_t^* denotes the central bank's inflation target for period t, and the parameter α lies between zero and one: $1 \geq \alpha \geq 0$.

The first–order condition of the dynamic problem is (for all t = 0, 1, 2, ...):

$$(1-\theta_t) \left[\frac{P_t(i)}{P_t} \right]^{-\theta_t} + \theta_t \left[\frac{P_t(i)}{P_t} \right]^{-\theta_t - 1} \left(\frac{W_t}{P_t Z_t} \right) - \phi \left[\frac{P_t(i)}{\Pi_{t-1}^{\alpha} (\Pi_t^*)^{1-\alpha} P_{t-1}(i)} - 1 \right] \left[\frac{P_t(i)}{\Pi_{t-1}^{\alpha} (\Pi_t^*)^{1-\alpha} P_{t-1}(i)} \right]$$

$$+ \beta\phi E_t \left\{ \left(\frac{\Lambda_{t+1}}{\Lambda_t} \right) \left[\frac{P_{t+1}(i)}{\Pi_t^{\alpha} (\Pi_{t+1}^*)^{1-\alpha} P_t(i)} - 1 \right] \left[\frac{P_{t+1}(i)}{\Pi_t^{\alpha} (\Pi_{t+1}^*)^{1-\alpha} P_t(i)} \right] \left[\frac{P_t}{P_t(i)} \right] \left(\frac{Y_{t+1}}{Y_t} \right) \right\} = 0 \quad (8)$$

When $\phi = 0$, the firm sets its price $P_t(i)$ as a markup $\theta_t/(\theta_t - 1)$ over marginal cost W_t/Z_t and $\theta_t/(\theta_t - 1)$ can be interpreted as the firm's desired markup, and random fluctuations in θ_t act like shocks to the firm's desired markup.

The production of the intermediate goods–producing firms $Y_t(i)$ for $i \in [0, 1]$ can be described by the constant–returns–to–scale technology: $\left[\int_0^1 Y_t(i)^{\theta_t - 1/\theta_t} di \right]^{\theta_t/\theta_t - 1} \geq Y_t$ with

the autoregressive process for θ_t: $ln(\theta_t) = (1-\rho_\theta)ln(\theta) + \rho_\theta ln(\theta_{t-1}) + \sigma_\theta \epsilon_{\theta t}$ (9) for all t = 0, 1, 2, ..., $1 > \rho_\theta \geq 0$, $\sigma_\theta \geq 0$, and the serially uncorrelated innovation $\epsilon_{\theta t}$ with the standard normal distribution.

1.3 The Central Bank

The central bank implements monetary policy according to the Taylor rule that can be adjusted to the following form[3]. We use log–linearized form of this rule:

$$ln(R_t) - ln(R_{t-1}) = \rho_\pi ln(\Pi_t/\Pi_t^*) + \rho_{gy} ln(g_t^y/g^y) + ln(v_t) \qquad (10)$$

where ρ_π and ρ_{gy} are positive constant, for all t = 0, 1, 2, ...

According to this equation the central authority increases short-run nominal interest rate if: a) the current inflation (Π_t) is higher than the inflation target (Π_t^*); b) the output growth $g_t^y = \frac{y_t}{y_{t-1}}$ (11) is higher than the long–run equilibrium of the output[4] and c) if there is a positive transitory monetary policy shock: $ln(v_t) = \rho_v ln(v_{t-1}) + \sigma_v \epsilon_{vt}$ (12) for all t = 0, 1, 2, ..., with $1 > \rho_v \geq 0$ and $\sigma_v \geq 0$, where the serially uncorrelated innovation ϵ_{vt} has the standard normal distribution.

An important feature of the model is the time–varying inflation target Π_t^*:

$$ln(\Pi_t^*) = ln(\Pi_{t-1}^*) - \delta_\theta \epsilon_{\theta t} - \delta_z \epsilon_{zt} + \delta_\pi \epsilon_{\pi t} \qquad (13)$$

for all t = 0, 1, 2, ..., where the response coefficients $\delta_\theta \geq 0$ and $\delta_z \geq 0$ are again chosen by the central bank, where $\sigma_\pi \geq 0$, and where the serially uncorrelated innovation $\epsilon_{\pi t}$ has the standard normal distribution.

1.4 The Model Equilibrium and The Steady State

The aggegate households' budget constraint takes this form:

$$Y_t = C_t + \frac{\phi}{2}\left[\frac{\Pi_t}{\Pi_{t-1}^\alpha (\Pi_t^*)^{1-\alpha}} - 1\right]^2 Y_t$$

The rule for price adjustment (9) is simplified to the following form (for all t = 0, 1, 2,...):

$$\theta_t - 1 = \theta_t \left(\frac{a_t}{\Lambda_t Z_t}\right) - \phi\left[\frac{\Pi_t}{\Pi_{t-1}^\alpha (\Pi_t^*)^{1-\alpha}} - 1\right]\left[\frac{\Pi_t}{\Pi_{t-1}^\alpha (\Pi_t^*)^{1-\alpha}}\right]$$
$$+\beta\phi E_t \left\{\left(\frac{\Lambda_{t+1}}{\Lambda_t}\right)\left[\frac{\Pi_{t+1}}{\Pi_t^\alpha (\Pi_{t+1}^*)^{1-\alpha}} - 1\right]\left[\frac{\Pi_{t+1}}{\Pi_t^\alpha (\Pi_{t+1}^*)^{1-\alpha}}\right]\left(\frac{Y_{t+1}}{Y_t}\right)\right\}$$

The rest equations (2) – (4), (6), and (9) – (13) are unchanged.

The previous system of equilibrium equtations consists of some variables that include the unit root either from (6) or (13). We must transform them in this respect[5].

These conditions hold in the steady state: $a = 1, \pi^* = 1, v = 1, \pi = 1, g^\pi = 1, g^r = 1, a_t = \theta, z_t = z$ and after next calculations we get: $g^y = z, \lambda = \frac{\theta}{\theta-1}, y = \left(\frac{\theta-1}{\theta}\right)\left(\frac{z-\beta\gamma}{z-\gamma}\right), r = \frac{z}{\beta}$ and $r^{r\pi} = \frac{z}{\beta}$ for t = 1, 2, 3, ...

[3] Woodford(2001) describes this modification of the rule.

[4] In this context we introduce $g_t^\pi = \pi_t/\pi_{t-1}$ and $r_t^{r\pi} = r_t/\pi_t$.

[5] The new variables are: $y_t = Y_t/Z_t$, $c_t = C_t/Z_t$, $\lambda_t = Z_t\Lambda_t$, $z_t = Z_t/Z_{t-1}$ (with a reference to Z_t) and $\pi_t = \Pi_t/\Pi_t^*$, $r_t = R_t/\Pi_t^*$ and $\pi_t^* = \Pi_t^*/\Pi_{t-1}^*$ (with a reference to Π_t^*).

1.5 The Linearized System

The system of the stationary equations can be log–linearized around the steady state[6]. A first–order approximation to the aggregate resource constraint reveals that $\hat{c}_t = \hat{y}_t$ and for the remaining equations imply:

$$(z-\gamma)(z-\beta\gamma) = \gamma z\hat{y}_{t-1} + \beta\gamma z E_t\hat{y}_{t+1} - (z^2+\beta\gamma^2)\hat{y}_t + (z-\gamma)(z-\beta\gamma\rho_a)\hat{a}_t - \gamma z\hat{z}_t \quad (14)$$

$$\hat{\lambda}_t = E_t\hat{\lambda}_{t+1} + \hat{r}_t - E_t\hat{\pi}_{t+1} \quad (15)$$

$$(1+\beta\alpha)\hat{\pi}_t = \alpha\hat{\pi}_{t+1} + \beta E_t\hat{\pi}_{t+1} + \psi(\hat{a}_t - \hat{\lambda}_t) - \hat{e}_t - \alpha\hat{\pi}^*_t \quad (16)$$

$$\hat{r}_t - \hat{r}_{t-1} = \rho_\pi\hat{\pi}_t + \rho_{gy}\hat{g}^y_t - \pi^*_t + \hat{v}_t \quad (17)$$

$$\hat{\pi}^*_t = \sigma_\pi\epsilon_{\pi t} - \delta_e\epsilon_{et} - \delta_z\epsilon_{zt} \quad (18)$$

$$\hat{g}^y_t = \hat{y}_t - \hat{y}_{t-1} + \hat{z}_t \quad (19)$$

$$\hat{g}^\pi_t = \hat{\pi}_t - \hat{\pi}_{t-1} + \hat{\pi}^*_t \quad (20)$$

$$\hat{r}^{r\pi}_t = \hat{r}_t - \hat{\pi}_t \quad (21)$$

$$\hat{a}_t = \rho_a\hat{a}_{t-1} + \sigma_a\epsilon_{at} \quad (22)$$

$$\hat{e}_t = \rho_e\hat{e}_{t-1} + \sigma_e\epsilon_{et} \quad (23)$$

$$\hat{z}_t = \sigma_z\epsilon_{zt} \quad (24)$$

$$\hat{v}_t = \rho_v\hat{v}_{t-1} + \sigma_v\epsilon_{vt} \quad (25)$$

for all t = 1, 2, ..., where $\hat{e}_t = (1/\phi)\hat{\theta}_t, \psi = (\theta-1)/\phi, \delta_e = \delta_\theta$ and $\sigma_e = \sigma_\theta/\phi$.

The first five equations are the ground of the model: a marginal utility of households' consumption (14), a New Keynesian IS curve (15), a New Keynesian Phillips curve (16), modified Taylor equation (17) and inflation targeting equation (18). The subsequent three equations state the definitions of the growth rate for the output (19), the inflation (20) and the nominal interest rate to the inflation (21). The rest equation describes the households' preference (22), cost–push (23), technology (24) and monetary shock (25).

2 Solving the Model and Results

There are 17 parameters[7] and 3 observable variables in the model: the growth rate of the output ($\hat{g}^y_t$), of the inflation ($\hat{g}^\pi_t$) and of the nominal interest rate to the inflation ($\hat{r}^{y\pi}_t$). For the solving the model we use the Czech economy quarterly data.

In the steady state $g^y = z$ and $r^{r\pi} = \beta$. The level of output growth rate is 1.0057 and the level of the nominal interest rate to the inflation divided by the value of parameter z is 0.99778. The coefficient on real marginal cost in Phillips curve is 0.1 (see Ireland (2005)). The others parameters are estimated via maximum likelihood – more precisely: the method is the Kalman filter evaluating a likelihood function and the Kalman smoother evaluating time series of a smoothed estimate of the unobserved state variable (target inflation). This technique is based on Ireland (2005*b*).

[6] A percentage deviation of the variable from its steady state are: $\hat{y}_t = ln(y_t/y)$, $\hat{c}_t = ln(c_t/c)$, $\hat{\pi}_t = ln(\pi_t)$, $\hat{r}_t = ln(r_t/r)$, $\hat{g}^y_t = ln(g^y_t/g^y)$, $\hat{g}^\pi_t = ln(g^\pi_t)$, $\hat{y}_t = ln(y_t/y)$, $\hat{g}^r_t = ln(g^r_t)$, $\hat{r}^{r\pi}_t = ln(r^{r\pi}_t/r^{r\pi})$, $\hat{\lambda}_t = ln(\lambda_t/\lambda)$, $\hat{a}_t = ln(a_t)$, $\hat{\theta}_t = ln(\theta_t/\theta)$, $\hat{z}_t = ln(z_t/z)$, $\hat{v}_t = ln(v_t)$, and $\hat{\pi}^*_t = ln(\pi^*_t)$.

[7] The parameters are: $z, \beta, \psi, \gamma, \alpha, \rho_\pi, \rho_{gy}, \rho_a, \rho_e, \rho_v, \sigma_a, \sigma_e, \sigma_z, \sigma_v, \sigma_\pi, \delta_e$ and δ_z. We introduce an unconstrained endogenous target model but there is a possibility to use a constrained exogenous version too.

In the model there is a quite high consumption habit of households ($\gamma = 0.91$). There is only a forward looking behavior in the price setting of firms ($\alpha = 0$) that is consistent with a rational expectations approach. The result reveals that both output as well as inflation growth enter the Taylor rule. There is no influence of the last value of the preference shock, cost-push and monetary shock to its present value – there is no persistence. The influence of technology and cost–push shock into the inflation target is not so high but there exists (coefficients δ_e and δ_z are non–zero). Except ρ_{gy} all parameters are statistically significnat.

Table 1 contains the maximum likelihood parameters estimates and standard deviationds.

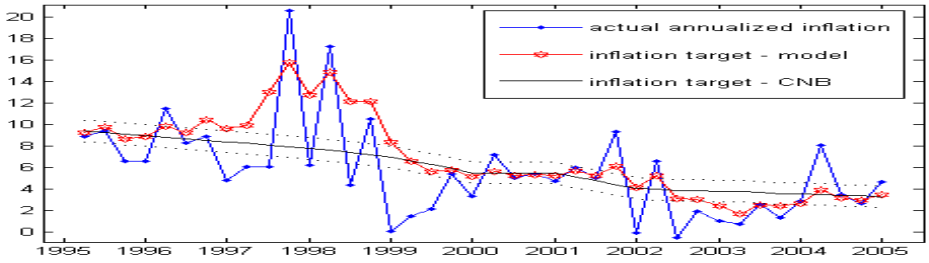
Table 1: Estimates of the Model with Endogenous Target

Parameter	Estimate	Standard Deviation	Parameter	Estimate	Standard Deviation
z	1.0057	0	ρ_e	0	*
β	0.99778	0	ρ_v	0	*
ψ	0.1	0	σ_a	0.024039	0.0062766
			σ_e	0.0041403	0.0018223
γ	0.91272	0.025536	σ_z	0.031182	0.0052965
α	0	*	σ_v	0.0049429	0.00062696
ρ_π	0.30946	0.080288	σ_π	0.0024665	**
ρ_{gy}	0.023574	0.069327	δ_e	0.0023736	0.00062486
ρ_a	0	*	δ_z	0.0020351	0.000914

* the estimate lies up the boundary of the parameter space, ** the calibrated value
The maximized value of the log likelihood function is 456.8707.

The overall fit of the model is presented by Figure 1. The basic workings of the model is illustrated by impulse responses in Figure 2.

Figure 1: Consumer price index inflation and inflation target (annualized)

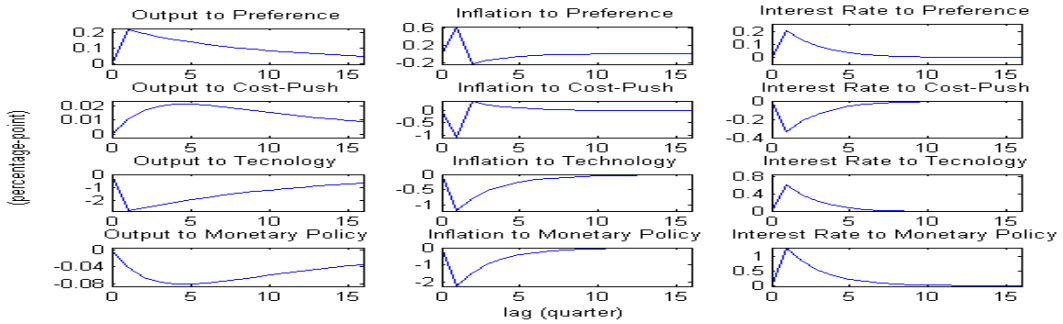


A preference shock seems to be a demand shock. It increases the output of the economy The higher level of the output is accompanied by an increase in inflation by 0.6 percentage point and by an increase of the interest rate.

A cost–push shock acts as a supply side shock on the other hand. It causes a rise in the output and a fall of the inflation and interest rate. Because there is a rational expectations it influences in the first period the inflation rate and cause an impulse to the monetary policy.

The shock in technologies proves negatively in the output and inflation and positively in the interest rate. This kind of shocks are permanent which means a higher output growth rate in steady state but this long run level of output is not reached immediately. If the agents behave rationally, they must change their behavior to adapt to a new conditions and these costs expressed in real terms mean some loss in output (more than two percentage points). When costs disappear the output goes back to its steady state at the higher growth rate.

Figure 2: Impulse responses (percentage–point) to a one standard deviation shock of preference, cost–push, technology, monetary policy and inflation targeting (the annualized inflation and the interes rate)



One standard deviation monetary shock induce insignificant drop in output (0.08 percentage points) connected with an important fall in the inflation by two percentage points. An adequate reaction of the central bank means to incerase the interest rate.

A change in inflation targets invokes no change due to rational expectations.

3 Conclusions

The model seems to give quite suitable approximation of the behavior of the Czech economy with respect to the results. From the results it is evident that every shock (except for the shock to the inflation target) has very different impact on the output and often important impact on the inflation and interest rate. All changes influence the inflation and if the central bank is obligated to the goal of the price stability, it must react to this situation according to the Taylor rule: adjust the interest rate. The model is influenced by the rational expectations.

This model is a closed model of a complex economy. Our first task in futher research is to modify it to our conditions: small open economy model. Next steps will also lead to a better and more precise description of the behavior of agents in economy: an introduction a full version of a market of inputs and an inspection of possible influence of money[8].

References

[1] CZECH NATIONAL BANK (2005): Cílování inflace v ČR, CNB. Access from internet www.cnb.cz/c.php?shortcut=yes (cited to date: 2. 5. 2005)

[2] DAVID, S. – VAŠÍČEK, O. (2005): Money's Function in the Czech Monetary Business Cycle. So far not publicated

[3] IRELAND, P. N. (2005*a*): Changes in the Federal Reserve's Inflation Target: Causes and Consequences, Boston College Personal Web Server. Access from internet www2.bc.edu/˜ irelandp/targets.pdf (cited to date: 12. 1. 2005)

[4] IRELAND, P. N. (2005*b*): Data and Programs (zip): Changes in the Federal Reserve's Inflation Target: Causes and Consequences, Boston College Personal Web Server. Access from internet www2.bc.edu/˜ irelandp/#programs (cited to date: 12. 1. 2005)

[5] WOODFORD M. (2001): The Taylor Rule and Optimal Monetary Policy, Princeton University - M. Woodford page. Access from internet www.princeton.edu/˜ woodford/taylor.pdf (cited to date: 2. 5. 2005)

[8] In this respect some important results gives the work of David and Vašíček (2005).

Credit Scoring Methodologies*

Martin Vojtek and Evžen Kočenda

CERGE-EI, P.O.BOX 882, Politických věznů 7, 111 21 Prague 1, Czech Republic;
telephone: +420 224005167; fax: +420 22421374; martin.vojtek@cerge-ei.cz

Abstract. In this paper we identify the methods used in the process of credit scoring of applicants for a commercial loan and perform their theoretical analysis. We are interested in their possible application in the framework of post-transition countries. We concentrate on retail loans as the sharp increase in the amount of loans for this clientele has been recorded in the last few years. The empirical study based on the above specified models can answer the question what are the main determinants (what variables influenced the decision) of approving the loan application in the existing practices of local banks. It also will clarify what are the determinants for defaulting loan by debtor.
Keywords: credit scoring, discrimination analysis, post-transition countries, banking

1 Introduction

Despite the proliferation of banking services, the basic commercial and industrial lending is still the core of the incomes of commercial banks and other lending institution in developed countries. Also in transition countries lending is the leading business of commercial banks, at least in the sector of medium- and large-size firms. The goal of this paper is to review the credit-scoring methods and to identify factors in these methods that are specific to post-transition countries in order to improve the credit-scoring performance.

In the last few years, banks in the Czech and Slovak Republics have moved in their commercial lending to a retail clientele and a sharp increase in business activity regarding commercial loans to retail clientele has been recorded.[1] The amount of commercial loans to retail clientele will continue to increase sharply. [2] From the technical perspective, the lending process in general is a relatively straightforward series of actions involving two principal parties. These activities range from the initial loan application to the successful or unsuccessful repayment of the loan. Although consumer lending belongs to the most profitable investments in lenders' asset portfolios (at least in the developed countries), the increases in the amounts of loans bring in also increases in the number of defaulted loans, i.e. loans that either are not repaid at all or the borrower has problems with paying his/her debts. This paper addresses the primary problem of the lenders: credit scoring, i.e. how to determine between good and bad debtors prior to granting credit. Our approach aims at addressing the specifics in the emerging post-transition economies.

* We acknowledge the support by grant of the Grant Agency of the Czech Republic.

[1] According to Czech National Bank monetary database, the amount commercial loans to retail clientele in the Czech Republic was 296 billions of CZK by the end of March 2005. In March 2002 this amount was only 124 billions of CZK. Over the last year, this amount jumped by 33.8%.

[2] Financial liabilities of households per capita were around 14 500 euro in the 15 older members of European Union and they account for 59% of GDP in the 2003.

In the Czech Republic the financial liabilities of households were around 900 euro per capita and they formed 11% of GDP in 2004. These numbers cover only banking sectors. Hence, there is a large space for expansion in the financial liabilities.

2 Basic Literature Review

To our best knowledge, no empirical study regarding commercial loans in transition countries exists in the relevant literature and similar literature is only rare for developed markets. Most of the credit scoring literature deals with the industry loans, i.e. loans received by firms. This part is different from the previous case, as not only there are different decision variables (various ratios of financial indicators for example) used in the decision process, but also the amounts lent are much lower in the case of consumer lending.

Credit scoring as a quantitative method relies on mathematical and statistical techniques, and thus development of credit screening goes hand-in-hand with progress in the mentioned techniques. First major use of the credit-scoring dates to the 60's, where credit card business grew up and the automatized decision process was a must. Credit scoring was fully recognized in the USA by 1975 Equal Opportunity Act, where it was stated, that any discrimination can be based only on statistical assessments.

All these approaches focused both on prediction of failure as well as on the classification of credit quality. This distinction is important as it is often not clear what is the aspect to focus on (whether classification or prediction). [1] described basic bank-lending process as an integrated system and analyzes conceptual approach to setting criteria for commercial loans and reviews previous results. As an example of methodologies concerning retail loans, [10] studied the selection of the best of alternative systems, instead of discussing the use of particular system or the development of new system. Since any system is loosing its efficiency over time he proposed a criteria for optimal updating cycle of credit scoring system. [6] examined how personal bankruptcy and personal bankruptcy exemptions affect the supply and demand for credit. They found that bankruptcy exemptions redistribute credit towards borrowers with high assets.

3 Techniques Used in Credit Scoring

In this section we will review the classic methods of the credit scoring and discuss the possibility of their use for the transition countries. We will discuss problems connected with implementing theses approaches as well. However to discuss the issue that applicants are subject to influences (such as external social or financial changes), which can change their probability of default, is beyond the scope of this paper.

First, we introduce a concept of incurred costs associated with probabilities of loans' repayment. For the simplicity let us assume that the population of loans consists of two groups or classes G and B that denote good and bad loans, respectively. Usually the class sizes are very different, so that for the probability that a randomly chosen candidate (customer) belongs to group G, denoted as p_G, we have $p_G \gg p_B$. Let x be a vector of independent variables (called measurement vector) used in the process of deciding whether an applicant belongs to the group G or B. Let the probability that an applicant with measurement vector x belongs to the group G (B) is $p(G|x)$ ($p(B|x)$). Let the probability that a good (bad) applicant has measurement vector x is $p(x|G)$ ($p(X|B)$). The task is to estimate probabilities $p(\cdot|x)$ from the set of given data about applicants which turns out to be good or bad and to find a rule how to partition the space X of all measurement vectors into two groups A_G and A_B based on these probabilities such that in A_G would be the measurement vectors of applicants who turn out to be good and vice versa. It is necessary to find a rule which will minimize the costs of the agency giving a credit connected with the misclassification of applicants. Let us denote c_G the costs connected with misclassifying a good applicant as bad and c_B as the costs connected with classifying a bad applicant as good. Usually $c_B > c_G$, because costs caused by misclassifying bad customer are financially more damaging costs. If applicants with

x are assigned to class G the expected costs are $c_B p(B|x)$ and the expected loss for the whole sample is $c_B \sum_{x \in A_G} p(B|x)p(x) + c_G \sum_{x \in A_B} p(G|x)p(x)$, where $p(x)$ is a probability that the measurement vector is equal to x. This is minimized when into group G are assigned such applicants which have

$$A_G = \{x | c_B p(B|x) \leq c_G p(G|x)\}, \tag{1}$$

what is equivalent with

$$A_G = \{x | p(G|x) \geq \frac{c_B}{c_B + c_G}\}. \tag{2}$$

Without loss of generality we can assume that $c_B + c_G = 1$. In that case the rule for classification is to assign applicant with x to class G if $p(G|x) > c_B$ and otherwise to class B.

3.1 Linear Discriminant Analysis

The aim of the Linear Discriminant Analysis is to classify a heterogeneous population into homogeneous subsets. We can assume that for each applicant there is a specific number of explanatory variables available. In a simple case of two subsets, the goal is to find the linear combination of explanatory variables which leaves the maximum distance between means of the two subsets.

In a general case we consider the distributions $p(x|G)$ and $P(x|B)$ that are multivariate normal distributions with the common variance. Then equation (1) reduces to

$$A_G = \{x | \sum w_i x_i > c\}.$$

Here x_i are explanatory variables, w_i are associated coefficients (weights) in the linear combination of explanatory variables. If one takes $s(x) = \sum w_i x_i$ then it is possible to discriminate according to this *score* and thus to reduce problem to only one dimension.

The discriminant analysis was introduced by [5] who was looking for the best way to separate the 2 groups using linear combination of the variables. [3] criticized this method that the rule is optimal only for a small class of distributions. However, [7] claims that 'If the variables follow a multivariate ellipsoidal distribution (of which the normal distribution is a special case), then the linear discriminant rule is optimal.'

As we have mentioned the advantages of the Linear Discriminant Analysis method are that it is simple, it can be very easily estimated and it actually works very well as it is often used in the credit scoring. The disadvantage is that the LDA requires normally distributed data but the credit data are often nonnormal. The data from post-transition economies are no different and the evidences for the Czech Republic show that for example income data, house ownership data or employment status indeed are nonnormal. Also data are often categorized so that we do not have full information. (Banks often ask to choose the category where the applicant belongs, for example income more than 10 000 CZK and less than 15 000 CZK). These facts lead us to an assumption that also a parametric approach may be adopted in the case of post-transition countries.

3.2 Logit Analysis

One way to overcome the problems with nonnormality of data is to use an extension of the linear discriminant analysis model that allows for some parametric distribution. In this case a suitable extension is generalized linear model known as logit model. Given a vector

of application characteristics x, the probability of default p is related to vector x by the relationship

$$\log\left(\frac{p}{1-p}\right) = w_0 + \sum w_i \log x_i. \tag{3}$$

One of the advantages over the linear discriminant analysis is the use of maximum likelihood method for the estimation of parameters w_i. Another advantage is that one can provide probabilities of being in some concrete class. Several studies found that logit model outperforms discriminant analysis, however this approach has not been so widely used as the linear discriminant analysis so far. More recently the logit belongs to the main approaches ofthe classification in credit scoring. The coefficients obtained have the same values as in the previous decision rule. However, they are obtained under much weaker assumptions. Actual classification results are similar for both types of regression and because both are sensitive to high correlation among explanatory variables, one should ensure that there are no such variables left in the training set.

To answer the question what are the main determinants of approving the loan application in the existing practices of the local banks it is necessary to design a model specification containing the right variables that influence the decision process of the loan approval.[3]

3.3 k-nearest Neighbor Classifier

The k-nearest neighbors classifier serves as an example of the non-parametric statistical approach. This technique assesses the similarities between the pattern identified in training set and the input pattern. One chooses a metric on the space of applicants and takes k-nearest neighbors (k-NN hereafter) of the input pattern that is nearest in some metric sense and new applicant will be classified to the class to which majority of the neighbors belong (in the case when the costs of misclassification are equal) or according to the rule expressed by equation (2). This means that this method estimates the $p(G|x)$ ($p(B|x)$) probability by the proportion of G (B) class points among the k nearest neighbors to the point x to be classified.

The first to use this method were [2]. Identifying the advantages of this method, [9] stated that the non-parametric nature of the method enables modelling of irregularities in the risk function over the feature space. The k-NN method is a fairly intuitive procedure and as such could be easily explained to business managers who would need to approve its implementation. It can be also used dynamically by adding applicants, when their class becomes known and deleting old applicants to overcome problems with the changes in population over time. The method is often superior over linear discriminant analysis (it is shown by [11], who compares linear discriminant, k-NN and various other classification methodologies). When performing the k-NN methodology, the very important step is the choice of a metric used. In their paper [9] described the choice of metric and the choice of the number of nearest neighbors to consider. A commonly taken metric is used the standard Euclidean norm.

However, when the variables are in different units or categorized, it is necessary to use some appropriate standardization of variables as well as to select some data-dependent version of the Euclidean metric such as

$$\rho_2(x,y) = \left[(x-y)^{\mathrm{T}} A (x-y)\right]^{1/2}, \tag{4}$$

where x and y are two measurement vectors and A is a $n \times n$ matrix where n is a number of variables. As matrix A can depend on x we can define 2 types of metrics according to how A is selected: local metrics are those where A depends on x and global metrics those where A is

[3] Training set is the data sample with known properties that serves to calibrate and verify the model performance before the model is applied.

independent on x. [9] proposed global metric given by the parametrization $A = I + Dww^{\mathrm{T}}$, where w is the direction orthogonal to equiprobability contours for p(G—x), D is a distance parameter. The direction w can be given using linear regression weights.

The choice of k (the number of nearest neighbors chosen) determines the bias/variance trade-off in the estimator. k has to be much smaller than the smallest class. A simulation study by [4] suggested that $k \approx n^{2/8}$ or $n^{3/8}$ is reasonable. On the other way a choice of k via cross-validation on the misclassification rate is often adopted in empirical literature. T

Recently, [8] observed that in problems, where there are 2 unbalanced classes, the fact that k is finite (and thus asymptotic properties do not hold) results in non-monotonic relationship between the k and the proportion of each class correctly classified. That means, in general, that larger k may not yield better performance than smaller k. For example if the number of points from the smaller class is less than $(1 - c_B)^{-1}$, then the best classification rule for predicting class G membership is to use $k = 1$.

When using k-NN method we do not need to identify variables which are distinctive for the transition countries. The specifics of emerging markets will be expressed in the partition of the set of all measurement vectors, where this partition can be different than it is in the developed markets. However this approach may serve as interesting comparison to the results from the discriminant analysis.

4 General and Specific Explanatory Variables

In the following section we identify the variables which may have explanatory power for the credit scoring in transition markets. Of course, the legal system may prohibit the use of some of these variables, but we will not deal with these issues in this paper. It is of our interest to identify such variables that are able to help to estimate the probability of default of an applicant.

- Demographic data - Here we identified variables as *Age of borrower*, *Sex of borrower*, *Marital status of borrower*, *Number of dependent persons*, *Home status (own house, mortgage, renting and so on)* and *District of address of borrower*
- Financial situation - Among others, these variables are important determinants: *Total assets of borrower, Gross income, Gross income of whole family, Number of dependents, Checking account(yes/no), Average balance on checking account, Monthly costs of family* and *Loans outstanding in the time of application*
- Employment status - As we discuss below, employment variables are very different from the situation in the western markets. Here are variables *Type of employment, Length of current employment, Number of employments over the last x years*
- Loan situation - These variables identifies the loan and potential problems with the repayment of loan (in the case of earlier default). The loans can be characterized by *Amount of loan, Maturity of loan, Purpose of loan, Assurance of loan, Number of payment/year, Total monthly costs of loan (or payment), Total loans outstanding/repaid, Problems with repaying debt (1=yes,0=no), If Problems=1 then Defaulted (1=yes,0=no), If Problems=1 then Renegotiated debt (1=yes,0=no) (specification of renegotiations - the change of payments, postponement of payments and so on), Total loans outstanding in the time of problems*

We believe that these several variables may be good determinants of the probability of default. Although they are in many way similar to the variables used in the western markets, the situation is often different. Good example is an employment status. In the developed countries, the big proportion of people are self-employed and this category frequently receives a smaller score in the assessment of loan application than employed people. However in the

Czech or Slovak Republic this category is not so widespread and the situation is biased by the fact that self-employment is often the second source of the income (so that people are employed as well as self-employed). The another difference is that banks are ready to accept higher risk as they do not have many good investment opportunities and the yields of default-free government bonds dropped to very low levels. The amounts of loans are much lower than in the western countries, what is other fact why banks are willing to accept higher risk the loans where they can charge higher interest rates. On the other hand the households are still ready to absorb a huge increase in the loans.

Another difference in comparison with the developed markets is situation of persons who are living of the social transfers. These transfers are still high and they are considered as "safe" income. That means that these persons can afford a small consumer loan, what is not possible in the USA for example. However, a negative determinant of the default probability is different expenditure structure. The "mandatory" expenditure (as costs of living and housing) creates higher percentage of income than in the western markets and therefore the household has smaller disposable income.

5 Conclusions

In this paper we identify the methods used in the process of credit scoring of applicants for a commercial loan. We are interested in their possible application in the framework of post-transition countries. We concentrate on retail loans as the sharp increase in the amount of loans for this clientele has been recorded in the last few years. As very interesting further work we plan to conduct empirical study based on the above specified models. It can answer the question what are the main determinants (what variables influenced the decision) of approving the loan application in the existing practices of local banks. It also will clarify what are the determinants for defaulting loan by debtor as well as confirm our selection of variables useful in the screening of potential customer.

Another goal is to specify the cost of lending errors for commercial bank loans and, in doing so, more accurately specify the optimal cutoff-score approach to credit scoring. Namely the bank has to choose the optimal tradeoff between profitability and risk. Too restrictive loan policies may insure minimal costs in the terms of defaulted loans, but the opportunity costs of rejected loans may exceed potential bad debt costs and thus the profit is not maximized. Conversely, too liberal policies may result in high losses from bad debts.

References

1. E. Altman. Commercial bank lendings: Process, credit scoring and costs of errors in lending. Journal of Financial and Quantitative Analysis, 15:813–832, 1980.
2. S. Chatterjee and S. Barcun. A nonparametric approach to credit screening. Journal of American Statistical Association, 65:150–154, 1970.
3. R.A. Eisenbeis. Pitfalls in the application of discrimant analysis in business, finance and economics. Journal of Finance, 32:875–900, 1977.
4. G.G. Enas and S.C. Choi. Choice of the smoothing parameter and efficiency of k-nearest neighbor classification. Computers and Mathematics with Applications, 12:235–244, 1986.
5. R. A. Fisher. The use of multiple measurements in taxonomic problems. Annals of Eugenic, 7:179–188, 1936.
6. R. Gropp, J. Scholz, and M. White. Personal bankruptcy and credit supply and demand. Quarterly Journal of Economics, 112:217–251, 1997.
7. D.J. Hand and W.E. Henley. Statistical classification methods in consumer credit scoring. Journal of the Royal Statistical Society. Series A (Statistics in Society), 160:523–541, 1997.

8. D.J. Hand and V. Vinciotti. Choosing k for two-class nearest neighbour classifiers with unbalanced classes. Pattern Recognition Letters, 24:1555–1562, 2003.
9. W.E. Henley and D.J. Hand. A k-nearest neighbour classifier for assessing consumer credit risk. Statistician, 45:77–95, 1996.
10. M. Long. Credit screening system selection. Journal of Financial and Quantitative Analysis, 11:313–328, 1976.
11. B.D. Ripley. Neural networks and related method for classification. Journal of the Royal Statistical Society. Series B, 56:409–456, 1994.

Bayes analysis of time series with covariates

Petr Volf
ÚTIA AV ČR
Pod vodárenskou věží 4, Praha 8
volf@utia.cas.cz

Abstract

In the last decade, the Bayes methods (often supported by MCMC computations) lead to the use of enhanced statistical models, hardly tractable by standard approaches. It concerns also to time series analysis, where the autoregressive character can be incorporated already to Bayes prior model (again, a richer notion than mere prior distribution) and one can consider simultaneously a similar time development of other parameters. In present contribution the methodology is used to the analysis of time series of aggregated unemployment data, a part of proposed model is the regression on covariates (age, gender, region) and time-dependent variance.

Key words

Bayes analysis, time series, regression, MCMC, unemployment data

1 Time series and Bayes method

Let us consider a time series model with regression, namely

$$X_j(t) = a(t) + \boldsymbol{Z}_j(t) \cdot \boldsymbol{b} + \varepsilon_j(t), \quad t = 1, 2, \ldots, T, \;\; j = 1, \ldots, J, \tag{1}$$

where $X_j(t)$ denotes the time development of followed random variable for object j (or in group j), $a(t)$ is a trend function, $\boldsymbol{Z}_j(t)$ are observed covariates from R^K, $\boldsymbol{b}$ is a K-dimensional vector of parameters, $\varepsilon_j(t)$ are random variables; we assume them mutually independent and with distribution $\mathcal{N}(0, \sigma^2(t))$, i. e. with variance depending on t. Hence, the estimation problem consists in estimation of two functions $a(t)$, $\sigma(t)$ (we take them nonparametrized) and of parameters b (in more general cases, not considered there, even parameters b can be time-dependent). For each t, $\boldsymbol{Z}(t)$ is a matrix $J \times K$.

One of standard solutions could consists of re-parametrization of unknown functions (e. g. expressing them as regression splines) the other in considering kernel-like estimators for them. Both approaches will, after some complicated computation, lead to a solution. An interest in the Bayes formulation of problems of similar type (i. e. with rather rich structure) has arisen recently in connection with the wide use of random generation methods called Markov chain Monte Carlo (MCMC). The main problem with Bayes methods always was to obtain posterior distribution. This is now solved by simulation of a sample from it (at least approximately). Thus, non-trivial numerical optimizations in standard formulation of models were replaced by rather intensive (but essentially easy) random generation. Even now, it is better to use conjugated families of distribution (i. e. yielding the same type of both prior and posterior distribution) because it leads to a simpler (and more effective) generating method – namely the Gibbs sampler instead of Metropolis–Hastings algorithm. Another advantage of Bayes approach is that, with the aid of prior model structure (i. e. formulation of model before the data are available), we can express our assumptions and "expectations" on data mechanism. We shall try to demonstrate it on a simple example.

2 Specification of model and solution

The Bayes formulation consist of two parts – we have to specify the model of data, given the parameters (i. e. the likelihood), and then the prior of parameters. From (1) it follows that the likelihood is proportional to

$$f(\boldsymbol{X}|a,b,\rho) \sim \prod_{t=1}^{T}\prod_{j=1}^{J}\sqrt{\rho(t)}\exp\left\{-\frac{\rho(t)}{2}(X_j(t)-a(t)-\boldsymbol{Z}_j(t)\,\boldsymbol{b})^2\right\}, \tag{2}$$

where we denoted $\rho(t) = 1/\sigma^2(t)$. (It is typical that when formulating problems for MCMC solutions we need not care on normalizing constants. Their exact computations for posterior distribution is not, as a rule, easy, We overcome this problem with the aid of a convenient sampling method).

As regards the prior distribution, we shall traditionally use normal distributions, and gamma distribution for variables $\rho(t)$. However, we shall also impose certain autoregressive development to both $a(t)$ and $\rho(t)$, similarly as Gamerman and West in [4].

2.1 Prior distributions

Let us assume the following structure of model components:

1. $a(t) = a(t-1) + e(t)$, $t = 1, 2, \ldots, T$, where $e(t) \sim \mathcal{N}(0, \delta_1^2)$ and $a(0) \sim \mathcal{N}(\mu_0, \delta_0^2)$.
2. $\rho(t) \sim g(\cdot; \alpha, \alpha/\rho(t-1))$, where $g(\cdot)$ is the gamma density, so that conditionally $\mathsf{E}\,\rho(t) = \rho(t-1)$, $\mathsf{var}\,\rho(t) = \rho(t-1)^2/\alpha$.
   Further, let $\rho(0) \sim g(\cdot; \alpha, \beta)$, i. e. $\mathsf{E}\,\rho(0) = \alpha/\beta$, $\mathsf{var}\,\rho(0) = \alpha/\beta^2$.
3. $\boldsymbol{b} \sim \mathcal{N}(0, V)$, where $V = \delta_2^2 \cdot I_K$, We thus have prior parameters $\boldsymbol{b}$ independent, however their posterior distribution will yield their mutual dependence.

Initial parameters μ_0, δ_0, δ_1, δ_2, α, β are chosen as fixed and their influence on results of analysis can be studied, too.

2.2 Conditional posteriors

The posterior distribution is proportional to the product of likelihood and priors, in our case of (1) and all prior densities formulated above. Hence, the conditional posterior distribution of a certain parameter of model consists only of components containing it. Namely, for $t = 1, \ldots, T$, we have that

$$\begin{aligned} p(a(t)|\cdots) &\sim \exp\left\{-\frac{1}{2}\left[\rho(t)\sum_{j=1}^{J}(X_j(t)-a(t)-\boldsymbol{Z}_j(t)\cdot\boldsymbol{b})^2 + \right.\right. \\ &\qquad \left.\left. + \frac{(a(t)-a(t-1))^2+(a(t+1)-a(t))^2}{\sigma_1^2}\right]\right\} \\ &\sim \exp\left\{-\frac{1}{2}\left[a(t)^2\cdot B_1 - 2a_t\cdot B_2\right]\right\} \sim N\left(\mu=\frac{B_2}{B_1},\, var=\frac{1}{B_1}\right), \end{aligned}$$

where $B_1 = \left(\frac{J}{\sigma_t^2}+\frac{2}{\delta_1^2}\right)$ and $B_2 = \left(\sum_j \frac{(X_j(t)-\boldsymbol{Z}_j(t)\cdot\boldsymbol{b})}{\sigma(t)^2} + \frac{(a(t-1)+a(t+1))}{\delta_1^2}\right)$.

Further

$$p(\rho(t)|\cdots) \sim \rho(t)^{J/2}\cdot\exp\left[-\frac{\rho(t)}{2}\sum_{j=1}^{J}(X_j(t)-a(t)-\boldsymbol{Z}_j(t)\cdot\boldsymbol{b})^2\right]\cdot$$

$$\cdot\rho(t)^{\alpha-1}\cdot\exp\left(-\frac{\rho(t)\cdot\alpha}{\rho(t-1)}\right)\cdot\exp\left(\frac{-\rho(t+1)\cdot\alpha}{\rho(t)}\right)\cdot\left(\frac{1}{\rho(t)}\right)^{\alpha} =$$

$$= \rho(t)^{A+J/2-1}\exp(-\rho(t)\cdot B)\cdot\exp\left(-\frac{C}{\rho(t)}\right)\cdot\left(\frac{1}{\rho(t)}\right)^{A},$$

where

$$B=\frac{1}{2}\sum_j(X_j(t)-a(t)-\boldsymbol{Z}_j(t)\cdot\boldsymbol{b})^2+\frac{\alpha}{\rho(t-1)},\quad C=\rho(t+1)\cdot\alpha$$

and $A>0$ is arbitrary. In this detail, convenient for effectiveness of Gibbs sampling, we follow Arjas and Gasbarra [2]. Posteriors for $t=0$ and $t=T$ are slightly simpler:

$$P(a(0)|\cdots)\sim\exp\left\{-\frac{1}{2}\left[\frac{(a(1)-a(0))^2}{\delta_1^2}+\frac{(a(0)-\mu_0)^2}{\delta_0^2}\right]\right\}$$

$$\sim\exp\left\{-\frac{1}{2}[a^2(0)\cdot B_1-2a(0)\cdot B_2]\right\},$$

so that $\sim N(\mu=\frac{B_2}{B_1}, var=\frac{1}{B_1})$, where $B_1=\frac{1}{\delta_0^2}+\frac{1}{\delta_1^2}$, $B_2=\frac{a(1)}{\delta_1^2}+\frac{\mu_0}{\delta_0^2}$.

$$p(a(T)|\cdots) \sim \exp\left\{-\frac{1}{2}\left[\rho(T)\sum_{j=1}^{J}(X_j(T)-a(T)-\boldsymbol{Z}_j(T)\,\boldsymbol{b})^2+\frac{(a(T)-a(T-1))^2}{\delta_1^2}\right]\right\}$$

$$\sim \exp\left\{-\frac{1}{2}[a(T)\cdot B_1-2a(T)\,B_2]\right\}\sim N\left(\mu=\frac{B_2}{B_1},\, var=\frac{1}{B_1}\right)$$

with $B_1=\dfrac{J}{\sigma(T)^2}+\dfrac{1}{\delta_1^2}$, $B_2=\displaystyle\sum_j\frac{(X_j(T)-\boldsymbol{Z}_j(T)\cdot\boldsymbol{b})}{\sigma(T)^2}+\frac{a(T-1)}{\delta_1^2}$.

$$P(\rho(0)|\cdots) \sim \rho(0)^{\alpha-1}\exp(-\beta\cdot\rho(0))\cdot\exp\left(-\frac{\alpha(\rho(1))}{\rho(0)}\right)\cdot\frac{1}{\rho(0)^{\alpha}} =$$

$$= \rho(0)^{A-1}\exp(-\beta\rho(0))\exp\left(-\frac{C}{\rho(0)}\right)\cdot\frac{1}{\rho(0)^{A}},$$

for arbitrary $A>0$, $C=\alpha\cdot\rho(1)$.

$$p(\rho(T)|\cdots)\sim\rho(T)^{J/2}\cdot\exp\left[-\frac{\rho(T)}{2}\sum_{j=1}^{J}(X_j(T)-a(T)-\boldsymbol{Z}_j(T)\cdot\boldsymbol{b})^2\right]\cdot$$

$$\cdot\rho(T)^{\alpha-1}\exp\left(-\frac{\rho(T)\cdot\alpha}{\rho(T-1)}\right)\sim g(\rho(T)\,;\,\alpha+\frac{J}{2},\,B),$$

where $g(\cdot)$ is a gamma density, $B=\dfrac{1}{2}\displaystyle\sum_{j=1}^{J}(X_j(T)-a(T)-\boldsymbol{Z}_j(T)\cdot\boldsymbol{b})^2+\frac{\alpha}{\rho(T-1)}$.

As regards parameter $\boldsymbol{b}$, we can easily find its joint posterior normal distribution

$$p(\boldsymbol{b}|\cdots)\sim\exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}\sum_{j=1}^{J}\rho(t)(X_j(t)-a(t)-\boldsymbol{Z}_j(t)\,\boldsymbol{b})^2+\boldsymbol{b}'V^{-1}\boldsymbol{b}\right]\right\}.$$

Let us denote $\boldsymbol{Y}$ the matrix composed from rows of matrices $\sqrt{\rho(t)}\boldsymbol{Z}_j(t)$, $t = 1, \ldots, T$, $j = 1, \ldots, J$, so that $\boldsymbol{Y}$ is $N \times K$, $N = T \times J$, further a $N \times 1$ dimensional vector $\boldsymbol{d}$ composed from elements $\sqrt{\rho(t)}(X_j(t) - a(t))$, then

$$\begin{aligned} p(\boldsymbol{b}|\cdots) &\sim \exp\left\{-\frac{1}{2}(\boldsymbol{b} - (\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{d})'\boldsymbol{Y}'\boldsymbol{Y}(\boldsymbol{b} - (\boldsymbol{Y}'\boldsymbol{Y})^{-1}\boldsymbol{Y}'\boldsymbol{d}) - \frac{1}{2}\boldsymbol{b}\,V^{-1}\boldsymbol{b}\right\} \\ &\sim \mathcal{N}\left\{\mu = (\boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{V}^{-1})^{-1} \cdot \boldsymbol{Y}'\boldsymbol{d},\ \mathsf{var} = (\boldsymbol{Y}'\boldsymbol{Y} + \boldsymbol{V}^{-1})^{-1}\right\}. \end{aligned}$$

From this, the marginal conditional posterior distributions can be derived for each component $p_k(b_k|\cdots)$, given the other parameters and data.

2.3 Random generation using Gibbs sampler

We have just formulated conditional posterior distributions of all unknown components of model. Now, the problem is to "evaluate" these distribution, to select a "point estimates" of unknown quantities and also to use the shape of posterior distribution to additional analysis, e. g. construction of credibility intervals – Bayes version of confidence intervals – and statistical testing of certain hypotheses, for instance on significance of parameters $\boldsymbol{b}$ or on constancy of functions $a(t)$, $b(t)$.

The Gibbs sampler is described elsewhere, for instance in [3]. It proceeds in the following way: It starts from an initial (chosen by the analyst) configuration of parameters and at each step generates randomly one new value of selected parameter. The order is selected randomly or systematically. Resulting Markov chain of values has the property that its distribution converges to the posterior distribution (it is the stationary distribution of created Markov chain). Hence, when we cut off sufficiently long initial part of the chain, we obtain the posterior represented by a sample, which (approximately) corresponds to it. From it, we can compute the averages (or sample medians, as point estimates of parameters) and also sample quantiles as borders of credibility intervals for them. These sample characteristics also converge, when the simulation is prolonged, to corresponding characteristics of posterior distribution.

As regards our concrete formulation, we see that in most cases we can sample a new value directly, mostly from normal distribution, $\rho(T)$ from gamma distribution. New values of $\rho(t)$, $t = 1, \ldots, T-1$, will be generated with the help of Gibbs sampler combined with "rejection sampling" method, see for instance [1]. In one step we wish to sample a value from the distribution

$$p(x) \sim x^{A+J/2-1} \cdot e^{-Bx} \cdot e^{-\frac{C}{x}} \cdot \left(\frac{1}{x}\right)^A = h_1(x) \cdot h_2(x),$$

where B, C are specified in part 2.2 and $A > 0$ is arbitrary, $h_1(x)$ (two first factors) is a part of gamma $g(x; A + \frac{J}{2}, B)$ distribution, while $h_2(x)$ is a part of "inverse gamma" distribution. We can set A in such a way that both parts have maximum at the same point, namely $\tilde{x} = C/A = (A + \frac{J}{2} - 1)/B$, i. e. for $A = \sqrt{BC + (\frac{J-2}{4})^2} - \frac{J-2}{4}$. Let us denote $M = h_2(\tilde{x})$. The rejection-sampling method now recommends to sample new candidate, say r, from $g(x; A + \frac{J}{2}, B)$ distribution and accept it as a new member of our chain $\rho^*(t)$, with probability $h_2(r)/M$ (again compare with Arjas and Gasbarra, 1994). If r is not accepted, we may try another r, till first accepted. Sampling of $\rho(0)$ is quite similar, except that we have in function $h_1(x)$ exponent $A - 1$ instead $A + \frac{J}{2} - 1$, so that optimal $A = \sqrt{BC + \frac{1}{4}} + \frac{1}{2}$.

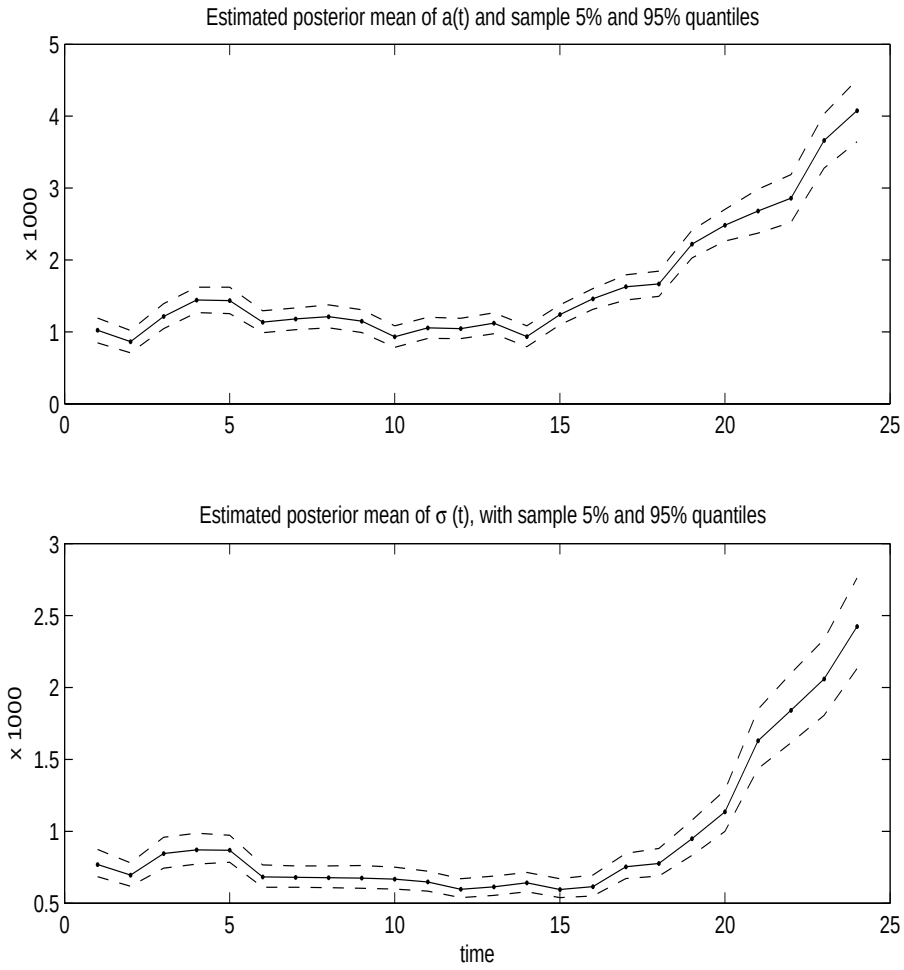


Figure 1: Posterior estimates of development of $a(t)$ and $\sigma(t)$

3 Numerical example with unemployment data

The data of our example are $X_j(t)$ = numbers of registered unemployed people in ČR, $t = 1, \ldots, 24$ are quarters 1993 – 1998 (seasonal variability is not taken into account, though it is possible to enlarge model (1) and add seasons as other factors to regression part). The data are grouped along 3 covariates, namely region (in those years there were 8 regions), age groups (5) and 2 genders. Then, matrix Z is a matrix of indicators of combinations of groups, $j = 1, \ldots, J = 80 = 8 \cdot 2 \cdot 5$, while, in order to obtain unique results, we have only $K = 12$ parameters b_k corresponding to regions 1,3 – 8, men, age group 2 – 5. It means that series of number of unemployed women from region 2 (Central Bohemia) and of first age class is modeled directly by $a(t) + \varepsilon(t)$ (their b's are zero). Interactions of covariates are not included. The same data has been analyzed under another model framework (Poisson regression) in [5].

We generated the Markov chain containing 5 000 values of all model components, first 2 000 were omitted and from remaining 3 000 we obtained the sample approximately representing the posterior distributions. From them, the following results can be displayed:

Figure 1 above shows mean development of $a(t)$, together with connected pointwise 5 % and 95 % sample quantiles, lower subplot contains the same characteristics of posterior

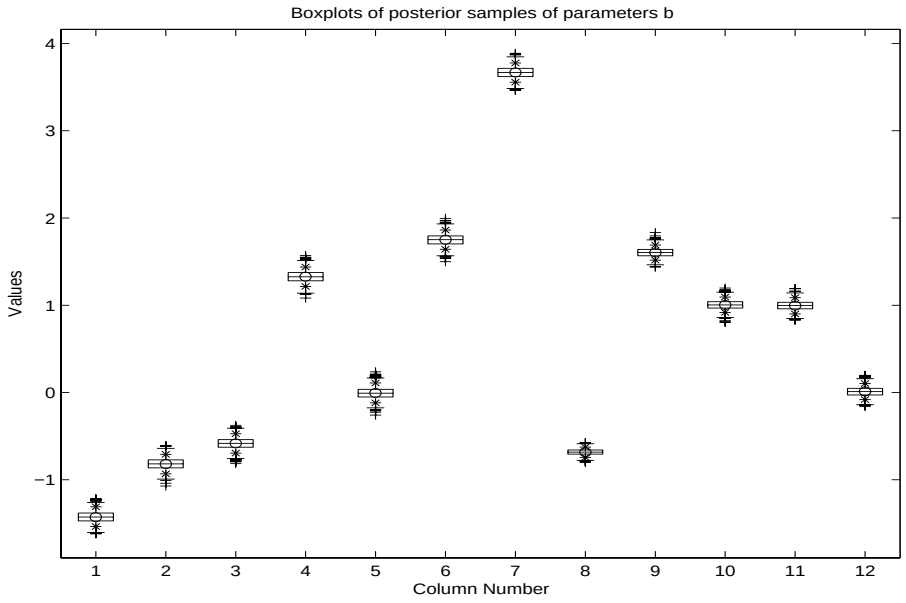


Figure 2: Boxplots from approximated posterior distribution of parameters $b(1), ..., b(12)$, stars denote 5% and95% sample quantiles

of $\sigma(t)$ (i.e. $\sqrt{(1/\rho(t))}$. Figure 2 then shows boxplots of samples of 12 parameters $\boldsymbol{b}$. It is seen that the situation was better in Prague (parameter b_1), negative b_8 shows that situation of men is still better than of women ($b_{women} = 0$. The regions (parameters 1 to 7) are in order Prague, South, West, North, East Bohemia, South,North Moravia, age groups corresponding to parameters $b_9 - b_{12}$ are 21-30, 31-40, 41-50, over 50 (the youngest group, up to 20, has by definition its parameter zero). Analysis of credibility (Bayes analog to significance) revealed that parameters b_5 and b_{12} were not significant, i.e. the range between 5% and 95% sample quantile contained zero.

Acknowledgement: The research is supported by the grant of GA ČR 402/04/1294.

References

[1] Antoch J. and Vorlíčková D.: Vybrané metody statistické analýzy dat. Academia, Praha 1992.

[2] Arjas E. and Gasbarra D.: Nonparametric Bayesian inference from right censpred survivak data, using the Gibbs sampler. Statistica Sinica 4 (1994), 505–524.

[3] Gamerman D.: Markov Chain Monte Carlo. Chapman and Hall, 1997.

[4] Gamerman D. and West M.: An application of dynamic survival models in unemployment studies. The Statistician 36 (1987), 269–274.

[5] Volf P.: Cox's regression model for dynamics of grouped unemployment data. Bulletin of the Czech Econometric Society, 10 (2003), 151–162.

Metaheuristics in Automated Storage and Retrieval Systems

Blaž Zmazek[1,2,3] and Janez Žerovnik[1,2,4]

[1] University of Maribor, FS, Smetanova 17,2000 Maribor, Slovenia
[2] Institute of Mathematics, Physics and Mechanics, Department of Theoretical Computer Science, Jadranska 19, 1111 Ljubljana, Slovenia
[3] blaz.zmazek@uni-mb.si
[4] janez.zerovnik@uni-lj.si

Abstract. This paper proposes the use of some metaheuristics to improve the performance of AS/RS; focusing attention on storage and retrieval policies, on-board storage capacity, S/R machine operation modes, and dwell point determination.

1 introduction

In today's manufacturing environments, inventories are maintained at lower levels than in the past. These reduced inventories have led to smaller storage systems, which, in turn, have created the need for quick access to the material being held in storage. Hence, automated storage/retrieval systems used in manufacturing, warehousing, and distribution applications must be designed to provide quick response times to service requests in order to keep the system operating efficiently. One important operational aspect of the AS/RS, which contributes to the system response time, is the dwell point location of the S/R machine. The dwell point is the location where the S/R machine is positioned when the AS/RS is idle and awaiting the next service request. Authors in [8] have shown the choice of the dwell point location can have significant impact on the response time of the AS/RS. The basic components of AS/RS are storage racks (SR), storage and retrieval machines (S/R machines), input/output (I/O) locations and accumulating conveyors (AC). Advantages by the application of AS/RS are: efficient utilization of warehouse space, reducing of damages and of loose of goods, increased control upon storage and retrieval of goods and decreasing the number of warehouse workers.

2 Literature Survey

Automated storage/retrieval systems have been the subject of much research over the past several years. Many of the papers deal with performance modeling of the AS/RS. Bozer and White (1984) developed expressions for the expected cycle time of an AS/RS performing single and dual command cycles. They used a continuous rack approximation to develop analytical models of the expected

cycle time. In addition, they suggested several static dwell point rules for AS/RS, although they provide no quantitative comparison of their performance. Han et al. [13] have improved the performance of an AS/RS by sequencing the order of servicing the retrieval requests. In this case, the AS/RS is assumed to be throughput-bound and therefore is never idle. Many other papers have looked at various aspects of AS/RS design and operation [12, 14, 18].

Egbelu [7] studied the AS/RS dwell point location problem. He developed formulations for minimizing the expected response time and minimizing the maximum response time for an AS/RS dedicated to a single aisle or shared between two aisles. He then transformed these constrained nonlinear programming formulations into linear programming problems. He discussed a framework for the dynamic selection of the dwell point, where updated information is used to generate a new linear programming problem each time a decision is to be made. Hwang and Lim [16] showed that the formulations from Egbelu [7] could be transformed to facility location problems in order to improve the solution time. Egbelu and Wu [8] used simulation to compare several dwell point rules. In particular, they compared the two formulations of Egbelu [7] and the four static rules proposed by Bozer and White [3]. They found that the solution from the minimum expected response time formulation performed well as did the always dwell at the input point rule of Bozer and White. These rules dominated the other rules studied, but one was not found to always dominate the other.

Designers and managers now view factory automation in a different perspective with AS/RS playing an important role. AS/RS also plays an important role in material handling and inventory control in warehouses. With on-board storage in AS/RS, parts are stored or retrieved in bulk. The major operational problems are storage or retrieval orders (a batching problem) and sequencing of items in each batch (a sequencing problem). Elsayed [9] proposed a heuristic algorithm to solve the order-picking problem in the case of a person on-board. He presents four heuristic algorithms for the grouping (batching) of orders and then utilizes Traveling Salesman Problem (TSP) algorithm to sequence the items of a single batch. The main objective of his study is to minimize the total travel time for picking orders. They found that the performance of these algorithms depend on the specific structure of the orders and the capacity of the S/R machine. During the last two decades, the order-picking problem has been extensively studied [10, 19]. Hwang and Lee [15] investigated a system where customer orders are batched using the heuristic batching algorithm based on cluster analysis. Various tour construction procedures [2] for multi command picking, as well as different rack shape configurations, in case of shared storage, have been investigated.

3 Definitions

The storage space is a rectangular structure, consisting of bins in m aisles, n rows and k columns. Each storage bin is capable of holding one part type only. In the generalization of the problem, each bin can have (fixed) capacity w. The storage space can be represented as a graph $G(V, E)$ with set of vertices - bins

V and edge set E of all direct paths between corresponding bins. In this way, in S/R machine the Euclidian distance is defined, therefore each edge uv in G has distance $d(u, v)$, the Euclidian distance between endpoints (bins) of the edge.

A *weighted graph* $G = (V, E, w, \lambda)$ is a combinatorial object consisting of an arbitrary set $V = V(G)$ of *vertices*, a set $E = E(G)$ of unordered pairs $\{x, y\} = xy$ of distinct vertices of G called *edges*, and two *weighting functions*, w and λ. $w : V(G) \mapsto R$ assigns positive real numbers (weights) to vertices and $\lambda : E(G) \mapsto R$ assigns positive real numbers (lengths) to edges.

A *simple path* from x to y is a finite sequence of distinct vertices $P = x_0, x_1, \ldots, x_\ell$ such that each pair x_{i-1}, x_i is connected by an edge and $x_0 = x$ and $x_\ell = y$. The *length* of the path is the sum of lengths of its edges, $l(P) = \sum_{i=1}^{\ell} \lambda(x_{i-1}x_i)$. For any pair of vertices x, y we define the *distance* $d(x, y)$ to be the minimum of lengths over all paths between x and y. If there is no such path, we set $d(x, y) = \infty$.

3.1 AS/RS as facility location problem

As mentioned above the position of dwell points can be reduced to a facility location which can, for example, be a variant of the center problem, if the maximal distance is to be minimized, or a median type problem if the average distance is to be minimized.

The *center problem* on a graph G is to minimize $f(z) = \max_{x \in V} w(x)d(z, x)$ over all points z on the edges of G. This objective function reflects the goal to locate a facility (center) z as close to the clients x as possible, so that the clients can quickly get services from the center in case of emergency. In the case of an *obnoxious* facility one wants to maximize $g(z) = \min_{x \in V} w(x)d(z, x)$. This objective function places the new location as far as possible from the sites (vertices) x of G. In this case important and highly sensitive sites (or highly obnoxious sites) receive small weights. In measuring the sensitivity of the sites we usually dispose of the finite set of marks. In this case each vertex of a graph can be valued with one of the c different marks (weights) with respect to its sensitivity. Finally, *median problem* on a graph G is to minimize $f(z) = \sum_{x \in V} w(x)d(z, x)$ over all points z on the edges of G. This objective function reflects the goal to locate a facility (median) z as close to all clients x as possible, so that the overall cost of distribution is minimal. This aspect arises directly from the AS/RS as a dwell point determination.

3.2 AS/RS as VRP

In the field of logistics many techniques have been already developed, from the travelling salesman problem to complex dynamic routing problems. One of the prominent problems in logistics is the *vehicle routing problem* (VRP). VRP is a generic name given to a whole class of problems in which a set of routes for a fleet of vehicles based at one or several depots must be determined for a number

of geographically dispersed cities or customers. The objective of the VRP is to deliver a set of customers with known demands on minimum-cost vehicle routes originating and terminating at a depot.

The VRP is a well known integer programming problem which falls into the category of NP Hard problems, meaning that the computational effort required to solve this problem increases exponentially with the problem size. For such problems it is often desirable to obtain approximate solutions, so they can be found fast enough and are sufficiently accurate for the purpose. Usually this task is accomplished by using various heuristic methods, which rely on some insight into the problem nature. This difficult combinatorial problem conceptually lies at the intersection of these two well-studied problems:

1. The Traveling Salesman Problem (TSP) (If the capacity of the vehicles C is infinite).
2. The Bin Packing Problem (BPP): (all edge costs in VRP are taken to be zero).

Because of the interplay between the two underlying models (both of them are NP Hard problems), instances of the Vehicle Routing Problem can be extremely difficult to solve in practice.

The VRP arises naturally as a central problem in the fields of transportation, distribution and logistics.

A feasible solution of VRP is composed of a partition $R_1, \ldots, R_m$ of V and a permutation σ_j of $R_j \cup 0$ specifying the order of the customers on route j.

The cost of a given route ($R_j = \{v_{j_0}, v_{j_1}, \ldots, v_{j_{(m_j+1)}}\}$), where $v_{j_i} \in V$ and $v_{j_0} = v_{j_{(m_j+1)}} = 0$ (0 denotes the dwel point), is given by:

$$C(R_j) = \sum_{i=0}^{m_j} d(v_{j_i}, v_{j_{i+1}}).$$

A route R_j is feasible if the vehicle stops exactly once in each customer and the total duration of the route does not exceed a prespecified bound D: $C(R_j) \leq D$. Finally, the cost function of the problem solution S is:

$$f(S) = \sum_{j=1}^{m} C(R_j).$$

4 Metaheuristics

It is well-known that many optimization problems including, for example, the traveling salesman problem [17] are NP-hard [11]. Roughly speaking, this means that there is no known efficient, i.e. polynomial time algorithm known and it is very unlikely that it is not possible to design efficient algorithms for such problems. More formally, such algorithms are not possible if the famous conjecture NP≠P holds [6]. For practical problems, this implies that we have to use heuristic methods. Among the most well known methods, also known metaheuristics,

are local search based heuristics [1] including, besides the basic versions of local improvement, tabu search, simulated annealing, genetic algorithms, and others. In order to apply these methods, we assume to have defined a set of feasible solutions for which the value of appropriate fittness function can be computed. Construction of a starting feasible solution or solutions is assumed to be possible. Furthermore, a neighborhood structure has to be defined on the set of feasible solutions, which enables relatively fast random walk in the space of feasible solutions.

Depending on particular definition of the feasible solutions, the fitness function and the 1 neighborhood structure, the behaviour of the same metaheuristic can vary considerable on the same problem instance. For further references on the literature on both theoretical and practical results on metaheuristics see, for example [1, 17]. Without going into details, let us only mention here that on the problems relevant here, our experience includes an example where already the generation of feasible solutions is very hard (a variant of the vehicle routing problem, [20]), and a very simple and successful tour generation and local improvement of the traveling salesman problem and some generalizations [4, 5]. Finally, there are usually some parameters for which values are to be chosen in order to implement the heuristics. The choice of optimal parameters is normally again a difficult problem, and from practical point of view it is reasonable to use a heuristics which does not have too many parameters. For example, considering constant time simulated annealing may be an interesting decision [21].

5 Conclusions and Scope for Future Work

The proposed methodology can be used to study the influences of the combinations of on-board storage capacity, S/R machine operation modes and dwell points and determine the best combination. AS/RS is one of the important materials handling facility that offers flexibility to keep pace with the rapidly changing demands of manufacturing. Considering the limitations of the existing methods, it is necessary to find techniques that will improve the performance of the system. In this context, this paper advocates the use of metaheuristic methods.

References

1. Aarts, E., and Lenstra, E. (eds.), Local Search in Combinatorial Optimization, John Wiley & sons, New York, 1997.
2. Bozer, Y.A., Schorn, E.C., and Sharp, G.P., Geometric approaches to solve the Chebyshev traveling salesman problem, IIE Transactions, (22) 3, 1990, 238-254.
3. Bozer, Y. A. and White, J. A., Travel-Time Models for Automated Storage/Retrieval Systems, IIE Transactions, Vol. 16, No. 4, 1984, pp. 329-338.
4. Brest, J., and Žerovnik, J., An approximation algorithm for the asymmetric traveling salesman problem. Ric. oper., 1999, vol. 28, pp. 59-67.

5. Brest, J., and Žerovnik, J., On the probabilistic traveling salesman problem. In: Plešingr, J. (ed.). Proceedings of the 17th International Conference on Mathematical methods in Economics '99. Jindrichuv Hradec: Faculty of Management: Czech Society for Operations Research: University of Economics, Prague, 1999, pp. 289-295.
6. Clay Mathematics Institute offers one million US dollars for proof or disproof of the P≠NP conjecture, http://www.claymath.org/millennium/P_vs_NP/
7. Egbelu, P. J., Framework for Dynamic Positioning of Storage/Retrieval Machines in an Automated Storage and Retrieval System, International Journal of Production Research, Vol. 29, No. 1, 1991, pp. 17-37.
8. Egbelu, P. J. and Wu, C. T., A Comparison of Dwell Point Rules in an Automated Storage/Retrieval System, International Journal of Production Research, Vol. 31, No. 11, 1993, pp. 2515-2530.
9. Elsayed, E.A., and Stren, R.G., Computerized algorithms for processing in automated warehousing systems, International Journal of Production Research, 21, 1983, 559-586.
10. Elsayed, E.A., and Unal, O.I., Order batching algorithms and travel time estimation for Automated Storage/Retrieval Systems, International Journal of Production Research, (27) 7, 1989, 1097-1114.
11. Garey, M.R., Johnson, D.S., Computers and Intractability, W.H.Freeman and Co., San Francisco, 1979.
12. Graves, S. C., Hausman, W. H., and Schwarz, L. B., Storage-retrieval Interleaving in Automatic Warehousing Systems, Management Science, Vol. 23, No. 9, 1977, pp. 935-945.
13. Han, M. H., McGinnis, L. F., Shieh, J. S., and White, J. A., On Sequencing Retrievals in an Automated Storage/Retrieval System, IIE Transactions, Vol. 19, No. 1, 1987, pp. 56-66.
14. Hausman, W. H., Schwarz, L. B., and Graves, S. C., Optimal Storage Assignment in Automatic Warehousing Systems, Management Science, Vol. 22, No. 6, 1976, pp. 629-638.
15. Hwang H., and Lee, M-K, Order batching algorithm for a man on-board AS/RS, Engineering Costs and Production Economics, 13, 1988, 285-294.
16. Hwang, H. and Lim, J. M., Deriving an Optimal Dwell Point of the Storage/Retrieval Machine in an Automated Storage/Retrieval System, International Journal of Production Research, Vol. 31, No. 11, 1993, pp. 2591-2602.
17. Lawler, E.L., Lenstra, J.K., Rinnooy Kan A.H.G., and Shmoys, D.B., The Traveling Salesman Problem, John Wiley & Sons, New York, 1985.
18. Linn, R. J. and Wysk, R. A., An Expert System Framework for Automated Storage and Retrieval System Control, Computers and Industrial Engineering, Vol. 18, No. 1, 1990, pp. 37-48.
19. Rosenwein. M. B., A comparison of heuristics for the problem of batching orders for warehouse selection, International Journal of Production Research, 34(3), 1996, 657-664
20. Zmazek, B., O problemu cestnega razvoza. In: Dnevi slovenske informatike, Portoroz, 14.-16. april 2004. Management in informatika : zbornik posvetovanja. Ljubljana: Slovenian Society Informatika, 2004, pp. 585-588 (in Slovene).
21. Žerovnik, J., Simulated annealing type metaheuristics - to cool or not to cool. In: 7th International Symposium on Operational Research in Slovenia, Podcetrtek, Slovenia, September 24-26, 2003. SOR '03 proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2003, [6] pages.

Approach to Real Option Model Application on Soft Binomial Basis
Fuzzy - stochastic approach

Zdeněk Zmeškal[1]

VŠB - Technical University of Ostrava, Faculty of Economics, department of Finance, Sokolská 33, 701 21 Ostrava 1, The Czech Republic, E-mail: Zdenek.Zmeskal@vsb.cz

Abstract

The real option valuation and decision-making is relatively new concept in financial decision-making. The stochastic discrete binomial models and continuous Black-Scholes-Meton models are usually applied. However, there is not in several situations in real option methodology application to have at to disposal input data of required quality. Traditionally, input data are in a form of real (crisp) numbers or crisp-stochastic distribution function. However, in several cases, input data is possible introduce only vaguely, by fuzzy numbers or fuzzy-stochastic distribution function. Hybrid fuzzy-stochastic binomial model under fuzzy numbers (T-numbers) and Decomposition principle is proposed and described. Input data are in a form of fuzzy numbers and result, possibility-expected real option value is also determined vaguely as a fuzzy set. Illustrative example of valuation firm equity with dividends is presented.

Keywords*:* Real option; Call option; Discrete Binomial Model; Black-Scholes-Merton model; Decision support system; Finance; Fuzzy sets; Pricing; Stochastic processes

1 Introduction

Applications of the option methodology in corporate finance decision-making, which have some special features, are called as *real options*. A survey of the real option literature is available in (Dixit (1994), Sick (1995), Trigeorgis (1998), Brennan (1999)). One possible and tractable usage of this methodology is stating the firm value as an option, investment project etc. Deterministic binomial (trinomial) models and analytical continuous Black–Scholes-Merton are often used, see (Black (1973), Boyle (1995)). In the case of application an option methodology the estimation of an option value is determined by input data precision, mainly concerning of validity, quality and availability of input data. There are also problems with data frequency and stochastic error as well.

We can distinguish and see two basic approaches in dealing with soft (fuzzy-stochastic) financial modeling. A survey of input data vagueness problems in finance and accounting decision-making is in (Zebda (1995)). The first one concerns using fuzzy measure in contrast with probability measure, a sub-additive property is not fulfilled, in decision-making and apprising, see e.g. Simonelli, (2001), Yoshida (2003). The second approach is based on assumption that input data (parameters, distribution functions) is possible introduce only vaguely. The approach might be seen as generalized sensitivity analysis; see e.g. Inuiguchi et.al. (2000), Tanaka et. al. (2000), Zmeskal (2001), Carlsson et.al. (2002), Sakawa et .al. (2003), Carlsson et.al. (2003), Yoshida (2003), Zmeskal (2005).

One of suitable approaches for solving the problem is to apply a fuzzy-stochastic methodology and create fuzzy-stochastic model. There are several references combining the fuzzy and stochastic processes, see (Dubois (1980), Kacprzyk (1988), Luhandjula (1996), Wang (1993), Viertl (1996)). Further the Wang (1993) approach will be mainly followed. Fuzzy-stochastic option valuation models are presented in Zmeskal (2001), Simonelli (2001), Carlsson et.al. (2003), Yoshida (2003). In the paper the approach of in Zmeskal (2001) is followed.

---

[1]The research was supported by Grant Agency of the Czech republic (GAČR) CEZ: 402/04/1357 this support is gratefully acknowledged.

2 Fuzzy-stochastic model development under normal fuzzy sets of T-number type

Following basic fuzzy-stochastic elements are very useful from an application point of view, (1) normal fuzzy set and T-number sets, (2) fuzzy-random variable, (3) ε-cut, (4) decomposition principle.

Many fuzzy sets could be created. The most common type of the fuzzy set meeting the specified preconditions of normality, convexity and continuity with upper semi-continuous membership function is very well known normal *fuzzy set*, see (e.g. Dubois et. al., 1980, Ramík et. al., 2001). One of the most widely applied normal fuzzy set types is the *T-number*.

Definition 1. A fuzzy set meeting preconditions of normality, convexity, continuity with upper semi-continuous membership function and closeness called the *T-number* and being defined as quadruple $\tilde{s} = (s^L, s^U, s^\alpha, s^\beta)$, as follows,

$$\tilde{s} \equiv \mu_{\tilde{s}}(x) = \left\{ \begin{array}{llll} 0 & for & x \le s^L - s^\alpha; & \phi(x) \quad for \quad s^L - s^\alpha < x < s^L; \\ \psi(x) & for & s^U < x < s^U + s^\beta; & 0 \quad for \quad x \ge s^U + s^\beta \end{array} \right\},$$

where $\phi(x)$ is a non-decreasing function and $\psi(x)$ is a non-increasing function.

Definition 2. The *ε-cut* of the fuzzy set $\tilde{s}$, depicted $\tilde{s}^\varepsilon$, is defined as follows,

$\tilde{s}^\varepsilon = \{x \in E^n; \mu_{\tilde{s}}(x) \ge \varepsilon\} = [{}^-s^\varepsilon, {}^+s^\varepsilon]$, where,

${}^-s^\varepsilon = \inf\{x \in E^n; \mu_{\tilde{s}}(x) \ge \varepsilon\}$, ${}^+s^\varepsilon = \sup\{x \in E^n; \mu_{\tilde{s}}(x) \ge \varepsilon\}$.

The crucial category in the fuzzy-stochastic modelling is the fuzzy-random variable.

Definition 3. It is said, $\tilde{\tilde{s}} : \Omega \to F_T(E)$ is the *fuzzy-random variable* (depicted with tilde and line), if for every $w \in \Omega$ and $\varepsilon \in]0,1]$, $\tilde{\tilde{s}}_w^\varepsilon = \{x : x \in E^n, \tilde{\tilde{s}}_w \ge \varepsilon\} = [{}^-s_w^\varepsilon, {}^+s_w^\varepsilon]$, is random interval (random ε-cut), ${}^-s_w^\varepsilon$, ${}^+s_w^\varepsilon$ are two random variables (or finite measurable functions). Let us denote the set of the fuzzy-random variables $FR(\Omega, P)$ where $P : \Omega \to [0,1]$ and thus $\tilde{\tilde{s}} \in FR(\Omega, P)$.

From definition implies that, (a) $\tilde{\tilde{s}} = \bigcup_\varepsilon \tilde{\tilde{s}}^\varepsilon$, because $\forall w \in \Omega, \tilde{\tilde{s}}_w = \bigcup_\varepsilon \tilde{\tilde{s}}_w^\varepsilon$. Where $\tilde{\tilde{s}}_w$ is fuzzy set and $\tilde{\tilde{s}}_w^\varepsilon$ is random interval, (b) $\tilde{\tilde{s}}$ is the fuzzy-random variable iff $\tilde{\tilde{s}}_w^\varepsilon$ is a random interval, it means $\tilde{\tilde{s}} = \bigcup_{\varepsilon, w} \tilde{\tilde{s}}_w^\varepsilon$ for every $w \in \Omega$ *and* $\varepsilon \in]0,1]$.

Very useful and powerful instrument that might be used for calculating a function of fuzzy sets is the extension principle, see (Zadeh, 1965). In general conditions an analytic solution according to extension principle is not available. Assuming a fuzzy set is of fuzzy number type (the T-number type as well) there is possible to solve function of fuzzy numbers $\tilde{s} = f(\tilde{r}_1 ... \tilde{r}_n)$ in accordance with the extension principle by decomposition principle as the approximate procedure of ε-cuts.

Definition 4. *Decomposition principle (Resolution identity)* is defined as follows,

$\mu_{\tilde{s}}(y) = sup_\varepsilon \{\varepsilon \cdot I_{\tilde{s}^\varepsilon}; y \in \tilde{s}^\varepsilon\}$ for any $y \in E^n$ and $\varepsilon \in [0;1]$, where $\tilde{s}^\varepsilon = [{}^-s^\varepsilon, {}^+s^\varepsilon]$ is ε - cut,

$y = f(x)$, ${}^-s^\varepsilon(x) = \min_{x \in \tilde{X}^\varepsilon \subset E^n} f(x)$, ${}^+s^\varepsilon(x) = \max_{x \in \tilde{X}^\varepsilon \subset E^n} f(x)$. Here $I_{\tilde{s}^\varepsilon}$ is characterisation function,

$I_{\tilde{s}^\varepsilon} = \{1 \; if \; y \in [{}^-s^\varepsilon, {}^+s^\varepsilon], 0 \; if \; y \notin [{}^-s^\varepsilon, {}^+s^\varepsilon]\}$.

It is apparent that applying the Definition 4 the function of fuzzy numbers $\tilde{s} = f(\tilde{r}_1 ... \tilde{r}_n)$ could be solved as several mathematical programming problems for ε in this way.

Problem 1 max (min) $s = {}^+s^\varepsilon, ({}^-s^\varepsilon)$,

where $s = f(x_1 .... x_n)$,

s.t. $x_i \in [{}^-x_i^\varepsilon, {}^+x_i^\varepsilon]$ for $i \in \{1;2.....n\}$, and $\varepsilon \in [0;1]$.

Advantage of procedure can be seen in generalized application possibility; disadvantage consists in a computation difficulty

Definition 5. Application of the Decomposition principle for function of fuzzy numbers allows expressing selected fuzzy operations $\tilde{*}$ among fuzzy numbers directly, as follows:

$$\tilde{w} = \tilde{s}\,\tilde{*}\,\tilde{r} = \bigcup_{\varepsilon} \varepsilon\left(w^{\varepsilon}\right) = \bigcup_{\varepsilon} \varepsilon\left(s^{\varepsilon} * r^{\varepsilon}\right).$$

Fuzzy addition $s^{\varepsilon} + r^{\varepsilon} = \left[{}^{-}s^{\varepsilon} + {}^{-}r^{\varepsilon}; {}^{+}s^{\varepsilon} + {}^{+}r^{\varepsilon}\right]$,

Fuzzy subtract $s^{\varepsilon} - r^{\varepsilon} = \left[{}^{-}s^{\varepsilon} - {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} - {}^{-}r^{\varepsilon}\right]$.

Fuzzy scalar product $k \cdot s^{\varepsilon} = \left[k \cdot {}^{-}s^{\varepsilon}; k \cdot {}^{+}s^{\varepsilon}\right]$ for $k \geq 0$, $k \cdot s^{\varepsilon} = \left[k \cdot {}^{+}s^{\varepsilon}; k \cdot {}^{-}s^{\varepsilon}\right]$ for $k < 0$.

Fuzzy multiplication $s^{\varepsilon} \cdot r^{\varepsilon} = \left[{}^{-}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon}; {}^{+}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon}\right]$ for $\tilde{s} > 0,\ \tilde{r} > 0$,

$s^{\varepsilon} \cdot r^{\varepsilon} = \left[{}^{-}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon}\right]$ for $\tilde{s} < 0,\ \tilde{r} > 0$, $s^{\varepsilon} \cdot r^{\varepsilon} = \left[{}^{+}s^{\varepsilon} \cdot {}^{+}r^{\varepsilon}; {}^{-}s^{\varepsilon} \cdot {}^{-}r^{\varepsilon}\right]$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy division $s^{\varepsilon} : r^{\varepsilon} = \left[{}^{-}s^{\varepsilon} : {}^{+}r^{\varepsilon}; {}^{+}s^{\varepsilon} : {}^{-}r^{\varepsilon}\right]$ for $\tilde{s} > 0,\ \tilde{r} > 0$, $s^{\varepsilon} : r^{\varepsilon} = \left[{}^{+}s^{\varepsilon} : {}^{+}r^{\varepsilon}; {}^{-}s^{\varepsilon} : {}^{-}r^{\varepsilon}\right]$ for $\tilde{s} < 0,\ \tilde{r} > 0$, $s^{\varepsilon} \cdot r^{\varepsilon} = \left[{}^{+}s^{\varepsilon} : {}^{-}r^{\varepsilon}; {}^{-}s^{\varepsilon} : {}^{+}r^{\varepsilon}\right]$ for $\tilde{s} < 0,\ \tilde{r} < 0$.

Fuzzy max $\max\left(s^{\varepsilon}\right) = \left[\max {}^{-}s^{\varepsilon}; \max {}^{+}s^{\varepsilon}\right]$.

Here $\tilde{s} > 0$ is positive fuzzy number, positive $\tilde{s} : \{x;$ for which $\mu_{\tilde{s}} \geq 0\}$ and simultaneously $x \in R^{+}$ (set of positive numbers), negative $\tilde{s} : \{x;$ for which $\mu_{\tilde{s}} \geq 0\}$ and simultaneously $x \in R^{-}$ (set of negative numbers).

We are able to show, that result of fuzzy operation of addition, subtract, and scalar product for linear T-numbers is also a linear T-number. For other operations it is not valid, but sometimes the approximation is applied such a way, that the result is also a linear T-number.

3 Binomial model of replication strategy – crisp stochastic approach

Replication strategy is based on creation a portfolio from underlying asset *S* and risk-free asset *B* so, for every situation to be replicated derivative value, it means a derivative value equals a portfolio value. Portfolio value in apprising moment *t*, $a \cdot S_t + B_t = f_t$,

the portfolio value in moment $t + dt$ for growing price, $a \cdot S^{u}_{t+dt} + B_t \cdot (1+R)^{dt} = f^{u}_{t+dt}$,

the portfolio value in moment $t + dt$ for declining price, $a \cdot S^{d}_{t+dt} + B_t (1+R)^{dt} = f^{d}_{t+dt}$,

where *S* is underlying asset value, *a* is amount of underlying asset, *B* is risk-free asset value, *f* is derivative value, *R* is risk-free rate, *u* (*d*) are indexes of growth (fall) of underlying asset. By solution of three equations for variables *a, B,* f_t, we can get a general formula for derivative price,

$$f_t (1+R)^{dt} = f^{u}_{t+dt} \cdot \left[\frac{(1+R)^{dt} \cdot S_t - S^{d}_{t+dt}}{S^{u}_{t+dt} - S^{d}_{t+dt}}\right] + f^{d}_{t+dt} \cdot \left[\frac{S^{u}_{t+dt} - (1+R)^{dt} \cdot S_t}{S^{u}_{t+dt} - S^{d}_{t+dt}}\right].$$

This is general formula for derivative price by replication strategy, which should be written as follows, $f_t = (1+R)^{-dt} \cdot \left[f^{u}_{t+dt} \cdot (p) + f^{d}_{t+dt} \cdot (1-p)\right]$ or $f_t = (1+R)^{-dt} \cdot E(f_{t+dt})$, where *p* is synthetic probability of growth (risk-neutral probability) and $E(f_{t+dt})$ is risk-neutral expected value. Derivative price is determined as present value of expected value in following period.

In a case of geometric Brown process, expressing $S^{u}_{t+dt} = S_t \cdot u$, $S^{d}_{t+dt} = S_t \cdot d$, then

$$p = \left[\frac{(1+R)^{dt} - d}{u - d}\right] \text{ and } f_t (1+R)^{dt} = f^{u}_{t+dt} \cdot \left[\frac{(1+R)^{dt} - d}{u - d}\right] + f^{d}_{t+dtt} \cdot \left[\frac{u - (1+R)^{dt}}{u - d}\right].$$

Stochastic option valuation procedure

- Modelling an evolution of underlying asset in accordance with the observed volatility as follows, for geometric Brown process, $S^{u}_{t+dt} = S_t \cdot u$; $S^{d}_{t+dt} = S_t \cdot d$.

- Computation of the intrinsic value (payoff function), *g;*, e.g. call option, $g_t^u = \max\left(S_t^u - X;0\right)$, and a put option, $g_t^u = \max\left(X - S_t^u;0\right)$, where *X* is exercise price.
- At the maturity day, *T*, option price is equal to intrinsic value, $f_T^u = g_T^u$, or $f_T^d = g_T^d$.
- Working backwards from the end of the binomial tree to the beginning, price of an option is calculated at the initial node. American option can be exercised whenever during pre-specified period and for its price can be written,
  $f_t = \max\left(g_t;F_t\right)$; where $F_t = \left(1+R\right)^{-dt} \cdot \left[f_{t+dt}^u \cdot \left(p\right) + f_{t+dt}^d \cdot \left(1-p\right)\right]$.
- At the beginning of the period f_0 is the price of an option.

4 Soft binomial model of replication strategy – fuzzy stochastic approach

We suppose a binomial model based on replication strategy with fuzzy parameter $\widetilde{S}$ $\widetilde{X}$ and fuzzy-probability function $\widetilde{p}$ derived from fuzzy volatility $\widetilde{u}, \widetilde{d}$, other parameters are introduced as crisp (real) numbers, risk-free rate, *R, dividend rate q*. Fuzzy sets are of fuzzy number type, Decomposition principle on ε*-cut* of function is applied and operations between fuzzy numbers are used, as well. Backward procedure is applied according to formula, $\widetilde{f}_t = \max\left(\widetilde{g}_t;\widetilde{F}_t\right)$; where $\widetilde{F}_t = \left(1+R\right)^{-dt} \cdot \left[\widetilde{f}_{t+dt}^u \cdot \left(p\right) + \widetilde{f}_{t+dt}^d \cdot \left(1-p\right)\right]$.

Soft valuation procedure

- Modelling an evolution of soft underlying asset in accordance with observed volatility applying fuzzy operations as follows, ${}^{-}S_{t+dt}^{\varepsilon u} = {}^{-}S_t^{\varepsilon} \cdot {}^{-}u^{\varepsilon}$; ${}^{+}S_{t+dt}^{\varepsilon u} = {}^{+}S_t^{\varepsilon} \cdot {}^{+}u^{\varepsilon}$, ${}^{-}S_{t+dt}^{\varepsilon d} = {}^{-}S_t^{\varepsilon} \cdot {}^{-}d^{\varepsilon}$, ${}^{+}S_{t+dt}^{\varepsilon d} = {}^{+}S_t^{\varepsilon} \cdot {}^{+}d^{\varepsilon}$.
- Computation of the intrinsic value (payoff function), *g*. For example, in the case of a call option, ${}^{-}g_t^{u\varepsilon} = \max\left({}^{-}S_t^{u\varepsilon} - {}^{+}X^{\varepsilon};0\right)$, ${}^{+}g_t^{u\varepsilon} = \max\left({}^{+}S_t^{u\varepsilon} - {}^{-}X^{\varepsilon};0\right)$, ${}^{-}g_t^{d\varepsilon} = \max\left({}^{-}S_t^{d\varepsilon} - {}^{+}X^{\varepsilon};0\right)$, ${}^{+}g_t^{d\varepsilon} = \max\left({}^{+}S_t^{d\varepsilon} - {}^{-}X^{\varepsilon};0\right)$ and where *X* is exercise price.
- At the maturity day, *T*, option price is equal to intrinsic value, ${}^{-}f_T^{u\varepsilon} = {}^{-}g_T^{u\varepsilon}$, ${}^{+}f_T^{u\varepsilon} = {}^{+}g_T^{u\varepsilon}$ or ${}^{-}f_T^{d\varepsilon} = {}^{-}g_T^{d\varepsilon}$, ${}^{+}f_T^{d\varepsilon} = {}^{+}g_T^{d\varepsilon}$.
- Working backwards from the end of the binomial tree to the beginning, price of an option is calculated at the initial node. American option can be exercised whenever during pre-specified period and for its price can be written,
  ${}^{-}f_t^{\varepsilon} = \max\left({}^{-}g_t^{\varepsilon};{}^{-}F_t^{\varepsilon}\right)$; ${}^{+}f_t^{\varepsilon} = \max\left({}^{+}g_t^{\varepsilon};{}^{+}F_t^{\varepsilon}\right)$,

where Decomposition Principle is realized by following mathematical programming problem.

Problem 2 max (min) $F_t = {}^{+}F_t^{\varepsilon}, \left({}^{-}F_t^{\varepsilon}\right)$, for $\varepsilon \in [0;1]$,

where $F_t = \left(1+R\right)^{-dt}\left\{f_{t+dt}^u \cdot [p] + f_{t+dt}^d \cdot [1-p]\right\}$, $p = \left[\dfrac{\left(1+R-q\right)^{dt} \cdot S_t - S_{t+dt}^d}{S_{t+dt}^u - S_{t+dt}^d}\right]$,

s.t. $p \in [0;1]$, $S_t \in \left[{}^{-}S_t^{\varepsilon}, {}^{+}S_t^{\varepsilon}\right]$, $S_{t+d}^d \in \left[{}^{-}S_{t+d}^{d\varepsilon}, {}^{+}S_{t+d}^{d\varepsilon}\right]$, $S_{t+d}^u \in \left[{}^{-}S_{t+d}^{u\varepsilon}, {}^{+}S_{t+d}^{u\varepsilon}\right]$,

$f_{t+d}^d \in \left[{}^{-}f_{t+d}^{d\varepsilon}, {}^{+}f_{t+d}^{d\varepsilon}\right]$, $f_{t+d}^u \in \left[{}^{-}f_{t+d}^{u\varepsilon}, {}^{+}f_{t+d}^{u\varepsilon}\right]$.

- Soft value of an option at the beginning of the period *is following* $\widetilde{f}_0^{\varepsilon} = \left[{}^{-}f_0^{\varepsilon};{}^{+}f_0^{\varepsilon}\right]$.

Example of company valuation with dividends payment, fuzzy stochastic approach

We are appraising equity of company with dividend payments on the real option binomial replication strategy under soft condition as the fuzzy-stochastic model. We are using methodology described in chapter 4. Fuzzy input data are in the form of triangular fuzzy numbers, defined by this way $\widetilde{x} = \left[{}^{-}x^0; x^1, {}^{+}x^0\right]$.

Solution: Computation of the underlying asset and payoff function is shown in Appendix 1. There is illustrated in Appendix 2 development of the option value - equity value under flexible

conditions. Option value under above conditions is $\tilde{x} = [100{,}6314;\ 148{,}95;\ 208{,}7292]$. The result should be considered to be soft real option value respecting input data preciousness.

Input data

Triangular number	Actual asset value	Debt	Up index	Down index	Risk-free rate	Continual dividend rate
Symbol	S	X	u	d	r	q
$^{-}x^{0}$	990	1095	1,198	0,832639	0,12	0,02
x^{1}	1010	1100	1,2	0,833	0,12	0,02
$^{+}x^{0}$	1000	1105	1,201	0,834725	0,12	0,02

5 Conclusion

The purpose of the paper was to propose and verify one approach for dealing with the problem of real option valuation under vagueness conditions on the fuzzy-stochastic basis. Hybrid soft binomial real option on replication strategy basis was explained, methodology described and illustrative example presented. Model of application of the hybrid fuzzy-stochastic apparatus might be seen as one suitable means for computing and estimates real options. The terms of described model application are characteristic for financial long-term decision-making and real options methodology and thus development and verifying the models of fuzzy-stochastic type might be useful. The described model might be considered to be an advanced method and a generalised sensitivity analysis of estimating real option value under contingent claim conditions and non-precise data. Consequently, we cannot expect better and precise estimation, however, model better coincide with soft input data and decision-making conditions. Impreciseness of results reflects impreciseness of objectives and input data of model. Therefore, successful application of the particular models depends on decision-making circumstances, financial analyst objectives and input data preciousness.

Acknowledgements
The research was supported by Grant Agency of the Czech republic (GAČR) CEZ: 402/04/1357. This support is gratefully acknowledged.

References

[1] Black, F., Scholes, M. (1973), „The Pricing of Options and Corporate Liabilities“, Journal of Political Economy, Vol. 81., 637-659.

[2] Boyle, P., Longstaff, F.A., Ritchken, P. (1995), Advances in Futures and Options Research, JAI Press, Vol. 8.

[3] Brennan, M. J., Tigeorgis, L. (1999), Project, Flexibility, Agency and Product Market Competition: New Development in the Theory and Application of real Options Analysis, Oxford university Press.

[4] Carlsson, C, Fuller, R. (2003) A fuzzy approach to real option valuation, Fuzzy Sets an Systems, 139, pp. 297-312

[5] Carlsson, Ch., Fuller, R., Majlender, P. (2002) A possibilistic approach to selecting portfolios with highest utility score. Fuzzy Sets and Systems, 13, 2002, pp 13-21

[6] Dixit, A. K., Pindyck, R.S. (1994), Investment under Uncertainty, University Press.

[7] Dubois, D., Prade, H. (1980), Fuzzy Sets and Systems, Academic Press, New York.

[8] Hull, J. C. (2000), Options, Futures, and other Derivatives. Prentice Hall.

[9] Inuiguchi, M. Ramik, J. (2000) Possibilistic linear programming: a brief review of fuzzy mathematical programming and comparison with stochastic programming in portfolio selection problem, Fuzzy Sets and Systems, 111, pp 3-28

[10] Kacprzyk, J., Fedrizzi, M.(1988), Combining Fuzzy Imprecision with Probabilistic Uncertainty in Decision-making. Springer-Verlag, Heidelberg.

[11] Luhandjula, M.K.(1996), „ Fuzziness and Randomness in an Optimization Framework“, Fuzzy Sets and Systems, No. 77, 291-297.

[12] Ramík, J. (1996), „New Interpretation of the Inequality Relations in Fuzzy Goal Programming Problems“, Central European Journal for Operation Research and Economics, No. 4.

[13] Ramík, J., Vlach, M. (2001) Generalized concavity in fuzzy goal and decision analysis. Kluwer Academic Publishing, Boston – Dodrecht – London

[14] Rommelfanger, H. J. (1999): "Fuzzy Logic Systems for Checking Credit Solvency of Small Business Firm". In: Riberio, R. A., Zimmermann, H. J., Yager, R. R., Kacprzyk, J. (1999): Soft Computing in Financial Engineering, Springer Verlag. 371-387.

[15] Sakawa, M., Kato, K., Nishizaki, I. (2003) An interactive fuzzy satisfying method for multiobjective stochastic linear programming problems through an expectation model. European Journal of Operational Research, Vol. 145/3, pp 665-672

[16] Sick, G.(1995), „Real Options“. In Jarrow, R et all, Handbooks in OR and MS, Vol. 9, Elsevier Science B.V., 631-691.

[17] Simonelli, M. R. (2001) Fuzziness in valuing financial instruments by certainty equivalents, European Journal of Operational Research, 135/2, 296-302

[18] Trigeorgis, L. (1998), Real Options - Managerial Flexibility and Strategy in Resource Allocation, Harvard University.

[19] Viertl, R. (1996), Statistical Methods for Non-Precise Data, CRC Press, Boca Raton, Florida.

[20] Wang, G. Y., Qiao, Z. (1993), „Linear Programming with Fuzzy-random Variable Coefficients“, Fuzzy Sets and Systems, 57, 295-311.

[21] Yoshida. Y. (2003) The valuation of European options in uncertain environment. European Journal of Operational Research, 145, 2003, pp 221-229

[22] Zadeh, L. A. (1965), „Fuzzy sets“, Information Control, 8 (3), 338-353.

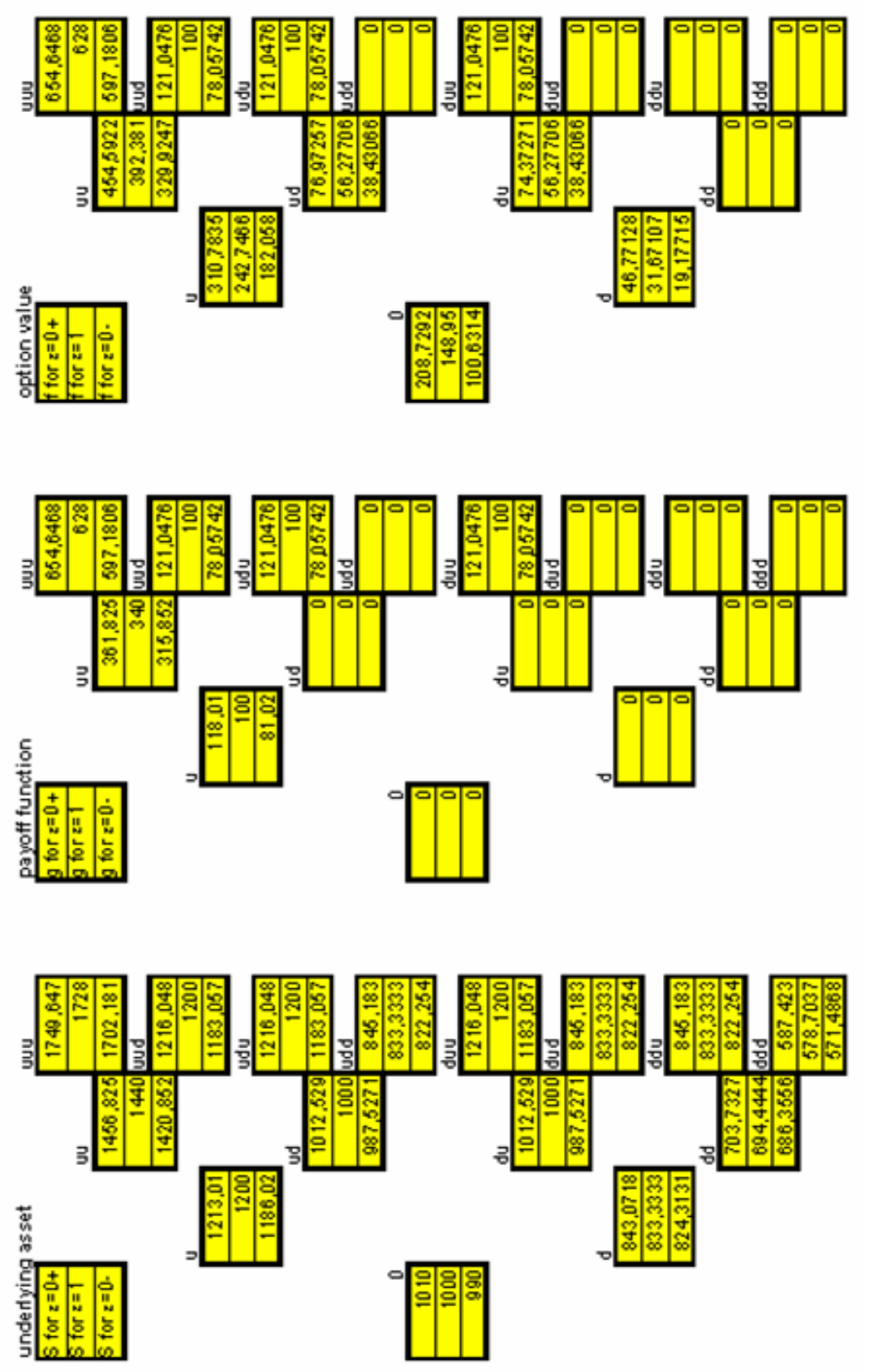
[23] Zebda, E. (1995), "The Problem of Ambiguity and the Use of Fuzzy Set Theory in Accounting, a Perspective and Opportunities for Research", In: Siegel, H.P., de Korvin, A., Omer, K. (1995): Application of Fuzzy Sets and Theory of Evidence to Accounting, JAI Press Inc., 19-35.

[24] Zmeskal, Z. (2001) Application of the fuzzy - stochastic Methodology to Appraising the Firm Value as a European Call Option, European Journal of Operational Research, Vol. 135/2, pp 303-310

[25] Zmeskal, Z. (2005) Value at Risk Methodology under Soft Conditions (fuzzy-stochastic approach). European Journal of Operational Research, 2005, 161, 337-347, ISSN 0377-2217.

[26] Zmeskal, Z. (2005) Value at Risk Methodology of International Index Portfolio under Soft Conditions (fuzzy-stochastic approach), International Review of Financial Analysis, 2005, 14/2, pp. 263-275, ISSN-1057-5219

Appendix 1 Development of the underlying asset, payoff function and option value (value of company equity)



Mathematical Methods in Economics 2005

Proceedings of the 23rd International Conference
Hradec Králové, 14th – 16th September 2005

Editor Hana Skalská

Vydalo nakladatelství GAUDEAMUS Univerzity Hradec Králové jako svou 583. publikaci
Hradec Králové 2005, 1. vydání 130 výtisků
Tisk: Libor Dvořák, tiskárna, E. Beneše 1562, Hradec Králové

ISBN 80-7041-535-5