Proceedings SOR

Rupnik V. and L. Bogataj (Editors): The 1st Symposium on Operational Research, SOR'93. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1993, 310 pp.

Rupnik V. and M. Bogataj (Editors): The 2nd International Symposium on Operational Research in Slovenia, SOR'94. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1994, 275 pp.

Rupnik V. and M. Bogataj (Editors): The 3rd International Symposium on Operational Research in Slovenia, SOR'95. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1995, 175 pp.

Rupnik V., L. Zadnik Stirn and S. Drobne (Editors.): The 4th International Symposium on Operational Research in Slovenia, SOR'97. Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1997, 366 pp. ISBN 961-6165-05-4.

Rupnik V., L. Zadnik Stirn and S. Drobne (Editors.): The 5th International Symposium on Operational Research SOR '99, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 1999, 300 pp. ISBN 961-6165-08-9.

Lenart L., L. Zadnik Stirn and S. Drobne (Editors.): The 6th International Symposium on Operational Research SOR '01, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2001, 403 pp. ISBN 961-6165-12-7.

Zadnik Stirn L., M. Bastič and S. Drobne (Editors): The 7th International Symposium on Operational Research SOR'03, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2003, 424 pp. ISBN 961-6165-15-1.

Zadnik Stirn L. and S. Drobne (Editors): The 8th International Symposium on Operational Research SOR'05, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2005, 426 pp. ISBN 961-6165-20-8.

Proceedings SOR'07

i

Proceedings of the 9th International Symposium on OPERATIONAL RESEARCH

SOR '07

Nova Gorica, Slovenia
September 26-28, 2007

Edited by:
L. Zadnik Stirn • S. Drobne

SOR '07 Proceedings

The 9th International Symposium on Operational Research in Slovenia
Nova Gorica, SLOVENIA, September 26 - 28, 2007

Edited by:

L. Zadnik Stirn and S. Drobne

Slovenian Society Informatika (SDI)
Section for Operational Research (SOR)

Proceedings of the 9th International Symposium on Operational Research
SOR'07 in Slovenia, Nova Gorica, September 26 - 28, 2007.

Organiser : Slovenian Society Informatika – Section for Operational Research, SI 1000 Ljubljana, Vožarski pot 12, Slovenia (www.drustvo-informatika.si/sekcije/sor/)

Under the auspices of the Slovenian Research Agency

First published in Slovenia in 2007 by Slovenian Society Informatika – Section for Operational Research, SI 1000 Ljubljana, Vožarski pot 12, Slovenia (www.drustvo-informatika.si/sekcije/sor/)

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

519.8(063)(082)

INTERNATIONAL Symposium on Operational Research in Slovenia (9 ; 2007 ; Nova Gorica)
SOR '07 proceedings / The 9th International Symposium on Operational Research in Slovenia, Nova Gorica, Slovenia, September 26-28, 2007 ; [organiser Slovenian Society Informatika, Section for Operational Research] ; edited by L. Zadnik Stirn and S. Drobne. - Ljubljana : Slovenian Society Informatika (SDI), Section for Operational Research (SOR), 2007

ISBN 978-961-6165-25-9
1. Zadnik Stirn, Lidija 2. Slovensko društvo Informatika. Sekcija za operacijske raziskave
234831104

Proceedings of the 9th International Symposium on Operational Research in Slovenia (SOR'07) is cited in: ISI (Index to Scientific & Technical Proceedings on CD-ROM and ISI/ISTP&B online database), Current Mathematical Publications, Mathematical Review, MathSci, Zentralblatt für Mathematic / Mathematics Abstracts, MATH on STN International, CompactMath, INSPEC, Journal of Economic Literature

Technical editor : Samo Drobne
Designed by : Studio LUMINA and Samo Drobne
Printed by : Birografika BORI, Ljubljana, Slovenia

*The 8th International Symposium on Operational Research in Slovenia - **SOR '07***
Nova Gorica, SLOVENIA, September 26 - 28, 2007

Program Committee:

L. Zadnik Stirn, Biotechical Faculty, Ljubljana, Slovenia, Chairwoman
M. Bastič, Faculty of Business and Economics, Maribor, Slovenia
L. Bogataj, Faculty of Economics, Ljubljana, Slovenia
M. Bogataj, Faculty of Maritime Studies and Transport, Portorož, Slovenia
V. Boljunčić, Faculty of Economics and Turism, Pula, Croatia
B. Böhm, Institute for Mathematical Models and Economics, Vienna, Austria
K. Cechlarova, Faculty of Science, Košice, Slovakia
V. Čančer, Faculty of Business and Economics, Maribor, Slovenia
A. Čizman, Faculty of Organising Sciences Kranj, Kranj, Slovenia
S. Drobne, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
L. Ferbar, Faculty of Economics, Ljubljana, Slovenia
J. Grad, Faculty of Public Administration, Ljubljana, Slovenia
S. Indihar, Faculty of Business and Economics, Maribor, Slovenia
P. Köchel, Chemnitz University of Technology, Chemnitz, Germany
J. Jablonsky, University of Economics, Praha, Czech Republic
J. Kušar, Faculty of Mechanical Engineering, Ljubljana, Slovenia
L. Lenart, Institute Jožef Stefan, Ljubljana, Slovenia
M. Marinovič, University of Rijeka, Rijeka, Croatia
L. Neralić, Faculty of Economics, Zagreb, Croatia
J. Povh, Faculty of Logistics, Krško, Slovenija
M. Simončič, Institute for Economic Research, Ljubljana, Slovenia
W. Ukovich, DEEI, University of Trieste, Trieste, Italy
J. Žerovnik, Institute for Mathematics, Physics and Mechanics in Ljubljana, Slovenia, Co-Chairman

Organizing Committee:

B. Nemec, HIT, Nova Gorica, Slovenia, Chairman
M. Bastič, Faculty of Business and Economics, Maribor, Slovenia
S. Drobne, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
A. Lisec, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
B. Peček, Faculty of Public Administration, Ljubljana, Slovenia
D. Škulj, Faculty of Social Sciences, Ljubljana, Slovenia
L. Zadnik Stirn, Biotechical Faculty, Ljubljana, Slovenia
B. Zmazek, Faculty of Mechanical Engineering, Maribor, Slovenia

Chairmen and Chairladies:

M. Bastič, Faculty of Business and Economics, Maribor, Slovenia
M. Bogataj, Faculty of Economics, Ljubljana, Slovenia
V. Boljunčić, Department of Economics and Tourism «Dr.Mijo Mirković», Pula, Croatia
B. Böhm, University of Technology – Vienna, Vienna, Austria
V. Čančer, Faculty of Economics and Business, Maribor, Slovenia
D. Hvalica, Faculty of Economics, Ljubljana, Slovenia
L. Ferbar, Faculty of Economics, Ljubljana, Slovenia
J. Grad, Faculty of Administration, Ljubljana, Slovenia
P. Köchel, Faculty of Informatics, Chemnitz, Germany
J. Kušar, Faculty of Mechanical Engineering, Ljubljana, Slovenia
L. Lenart, Institute Jožef Stefan, Ljubljana, Slovenia
M. Marinović, Faculty of Philosophy, Rijeka, Croatia
L. Neralić, Faculty of Economics, Zagreb, Croatia
J. Povh, School of Business and Management, Novo mesto, Slovenia
L. Zadnik Stirn, Biotechical Faculty, Ljubljana, Slovenia
J. Žerovnik, Institute for Mathematics, Physics and Mechanics in Ljubljana, Slovenia & Faculty of Mechanical Engineering, Maribor, Slovenia

Preface

In order not to lose the experience and the knowledge from The 9th International Symposium on Operations Research, called SOR'07, which was held in Nova Gorica, Slovenia, from September 26 through September 28, 2007, we published this symposium publication. It reflects the scientific and professional activities during SOR'07; the articles presented and discussed at SOR'07 are so permanently stored and available to all those who participated in the symposium and to all those who did not, but are interested in the contents and are considering to participate in the future symposia.

SOR'07 is the premiere scientific event in the area of operations research, one of the traditional series of the biannual international conferences organized by Slovenian Society INFORMATIKA, Section of Operational Research. It represents the continuity of eight previous symposia and has attracted a growing number of national and international audience. At the symposium SOR'07 the scientists, researchers and practitioners from different areas, like mathematics, economics, statistics, computer science engineering, environment and system theory, often working together on common projects, came together, exchanged new developments, opinions, experience, and thus contributed to the quality and reputation of operations research.

The 9th International Symposium on Operations Research SOR'07 stood under the auspices of the Slovenian Research Agency, and was granted by sponsors cited in these Proceedings. The opening address was given by Mr. N. Schlamberger, the President of Slovenian Society INFORMATIKA, Prof. Dr. L. Zadnik Stirn, the President of the Slovenian Section of Operations Research, Mr. B. Nemec, MSc., the representative of HIT, Nova Gorica, and representatives of different professional institutions and Operations Research Societies from other countries.

Operations Research comprises a large variety of mathematical, statistical and informational theories and methods to analyze complex situations and to contribute to responsible decision making, planning and the efficient use of the resources. In a world of increasing complexity and scarce natural resources we believe that there will be a growing need for such approaches in many fields of our society.

As traditionally, also SOR'07 was an international forum for scientific exchange at the frontiers of Operations Research in mathematics, statistics, economics, engineering, education, environment and computer science. We believe that the presentations reflected the state of the art in Operations Research as well as the actual challenges. Besides contributions on recent advances in the classical fields, the presentations, on new interactions with related fields as well as an intense dialogue between theory and the numerous applications, were delivered at the symposium. An attention was paid also to simplex method according to its 60th anniversary, as well as to the interior point methods, developed in the last 20 years. Finally, to give the symposium program the final touch, distinguished speakers have been invited to present keynote speeches. Thus, we hope that the division into Invited lectures and 12 sections, reflects on the one hand the variety of fields engaged, on the other hand separating too many subjects which could belong together. The scientific program was divided into the following sections (the number of papers in each section is given in parentheses): Plenary section (7), Networks (5), Stochastic and Combinatorial Optimization (5), Algorithms (3), Multicriteria Decision Making (4), Scheduling and Control (4), Location Theory and Transport (4), Environment and Human

Resource Management (5), Duration Models (5), Finance and Investment (7), Production and Inventory (7), Education and Statistics (5), OR Communications (7).

The first part of the Proceedings includes invited papers, presented by 7 prominent scientists: Valter Boljunčić, Juraj Dobrila University of Pula, Pula, Croatia; Immanuel Bomze, TU Vienna, Vienna, Austria; Martin Gavalec, University Hradec Králové, Hradec Králové, Czech Republic; Juraj Hromković, ETH – Zentrum, Zürich, Switzerland; Leen Stougie, Eindhoven University of Technology, Eindhoven, The Netherlands; Lidija Zadnik Stirn, University of Ljubljana, Ljubljana, Slovenia, and Janez Povh, University of Maribor, Maribor, Slovenia. The second part of the Proceedings includes 61 papers written by 104 authors and co-authors. These papers were accepted among numerous submitted papers after a review process carried out by the members of the Program Committee assisted by a few additional reviewers appointed by the Committee members. Most of the authors of the contributed papers came from Slovenia (41), then from Croatia (30), Bosnia and Herzegovina (7), Czech Republic (4), Austria (3), Algeria (3), Poland (3), Hungary (2), Macedonia (2), Romania (2), Switzerland (1), Canada (1), China (1), Germany (1), Slovak Republic (1), Ukraine (1) and USA (1).

The Proceedings of the previous eight International Symposia on Operations Research organized by Slovenian Section of Operations Research are cited in the following secondary and tertiary publications: Current Mathematical Publications, mathematical Review, MathSci, Zentralblatt fuer Mathematik/Mathematics Abstracts, MATH on STN International, CompactMath, INSPEC. Also the present Proceedings will be submitted and is supposed to be cited in the same publications.

We would not have succeeded in attracting so many distinguished speakers from all over the world without the engagement and the advice of active members of Slovenian Section of Operations Research. Many thanks to them. Further, we would like to express our deepest gratitude to the members of the Program and Organizing Committees, to the reviewers, chairperson, to the sponsors, especially HIT, Nova Gorica and Austrian Science and Research Liaison Office, Department of Ljubljana, and to all the numerous people - far too many to be listed here individually - who helped in carrying out The 9th International Symposium on Operations Research SOR'07 and in putting together these Proceedings. At last, we appreciate the authors' efforts in preparing and presenting the papers, which made The 9th Symposium on Operational Research SOR'07 successful. The success of the appealing scientific events at SOR'07 and the present proceedings should be seen as a result of our joint effort.

Nova Gorica, September 26, 2007

Lidija Zadnik Stirn
Samo Drobne
(Editors)

Foreword

Lionel Terray is one of the climbers that stood on the first of the eight thousand meter high Himalayan peaks. He described climbers as the conquerors of useless land. What they do is discovering new directions just for the fun of it and not asking themselves what will be the use of results of their endeavor. Surprisingly enough, their achievements seem to start to be of interest long after they have accomplished them. Discovering of Alps is one such example. At first the highest peaks were of interest to them just to see if they could be climbed. They could. This done, seemingly forbidding mountain faces attracted interest and stood there as a challenge whether they could be climbed. They could be climbed, too. Followers tried to repeat the undertakings and succeeded, even surpassing the achievements of their predecessors. The result of the process is that in the parts where no practical activity was present in the past, the same parts are now busy with life and action. It all started by answering a simple question: Can it be done?

Researchers in the field of Operations Research, in a way, bear some resemblance to climbers. They, too, discover and prove theorems without asking what will be the use of their work. Having in mind the history we can be confident that even if today there is no use for the large number of theorems that are devised daily, they will surely be indispensable in most unexpected applications in the future. The approach has been most wonderfully described by Nobel laureate in physics (where it must not be forgotten that his great mathematical knowledge and respective contribution played a major role) Richard Feynman. He said that he was motivated by the pleasure of finding things out. The same may be said for findings in operations research/management science. A great deal of them is the result of the pleasure of finding things out.

Mathematics together with operations research is probably the most abstract of all sciences and most formally controlled of them all. It is the abstraction which underlies mathematical work that seems strange to most people which may be the reason why most people are not mathematicians and operation researchers or are at least not fond of mathematics. A process of proving a theorem is based on a small number of axioms that form the foundation of mathematics. They must be observed line after line regardless of length of the proof which may take even hundreds of pages. However, such a rigorous approach could be useful, even recommendable, in other fields as well. Just to think how productive and up to the point the parliamentary debates would be if they were governed by similar principles and not by rhetoric and surprises in a form of an ace from a sleeve.

Operations research as a part of mathematics necessarily shares many common features and approaches of its mother science. Not that the findings are useless; far from it. Theory of games seemed a rather abstract discipline in the beginning whereas today it is virtually unthinkable not to try any serious strategic decision without the apparatus provided by it. Yet the results, however abstract they seem -- or are -- may have unexpected practical significance and importance in a most unforeseen way. Theory of numbers is another such example. Greatest minds of mathematics have been occupied with it and have formulated and proved many theorems that seemed to serve to no purpose at the time they were proposed. A century or so later they were found indispensable in encryption algorithms. History of mathematics and operations research is full of such cases.

So, in a way, mathematicians and operation researchers are not quite unlike alpine climbers. One more feature of similarity between them is that their achievements are completed away from the public attention. It is only after they finish their work that the

outcome of their effort catches public interest and even then it rarely hits headlines. The conquest of highest mountain faces; the proofs of conjectures that were formulated hundreds of years ago are examples of such undertakings where the arena is not a crowded stadium but rather a far away territory or a remote peace of a laboratory. Audience is not needed during the process of finding things out; the recognition will follow after the work is done. However there is a line where the similarity between the both ends. Whereas most climbers do it in their free time most students in mathematics and operation research do it for living which means that the places where they work, notably universities and institutes, must provide for the necessary finance.
Universities and institutes have enjoyed a great deal of autonomy throughout the history and they still do which is good but a shadow lies above their autonomy. The shadow is called the finance most of which comes from the government which in turn needs to have a control over its expenditure. The result is a clash between two autonomies, that of the university/institute and that of the government. It is hard to practice any autonomy without a major autonomous source of finance and that goes for the university/institute as well. That is a situation that will have to be resolved not just in Slovenia but much wider. We have seen attempts that were not very fortunate and that have brought no solutions, only just new problems but that should not discourage us. The Bologna process has been started; it is well under way and possibly some answers will be found there. Also the very concept of the university seems to have begun to change which will necessarily influence its role, its function, and also it's financing in the future. We see more universities emerging and being established recently that we have ever thought possible. We could even ask ourselves whether we are not trying to give an old name to a new entity which could be rather misleading, even confusing and certainly not productive in searching for new solutions.

While it is true that we seem to be facing more questions than we know the answers it is also true that such a situation is not new. It has happened before in various lines of human endeavor and will surely happen again. The common feature of such situations is that a solution was found eventually maybe not such one to satisfy all but undoubtedly such one to make possible further development in the field. We may be confident that also for the autonomy of the universities/institutes an acceptable solution will come into being although we do not know at this time what will it be. Which we hope and trust is that it will make possible the pleasure of finding things out for the generations of mathematicians and operation researchers to come and that they will be able to conquer the useless land for the benefit of future generations.

Nova Gorica, September 26, 2007

Niko Schlamberger
(President
Slovenian Society INFORMATIKA)

Sponsors of the 9th International Symposium on Operational Research in Slovenia (SOR'07)

The following institutions and companies are gratefully acknowledged for their financial support of the symposium:

- *Austrian Science and Research Liaison Office, Department of Ljubljana, Ljubljana*
- *HIT, Nova Gorica*
- *Slovenian Research Agency, The Republic of Slovenia*

Contents

Author index

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Plenary Lectures

ON DUAL MULTIPLIERS IN DEA

Valter Boljunčić[1], Luka Neralić[2]
[1]Juraj Dobrila University of Pula, Department of Economics and Tourism «Dr.Mijo Mirković»
[2]University of Zagreb, Faculty of Economics
vbolj@efpu.hr , lneralic@efzg.hr

Abstract: Data Envelopment Analysis (DEA) is a set of methods and models used to evaluate efficiency of a chosen Decision Making Unit (DMU) compared to set of DMUs. This is done using LP models where usually primal LP program is called envelopment model, while dual is called multiplier model. In the talk we focus on dual multipliers which preserve efficiency of DMU after data perturbations.

Keywords: DEA, linear programming, dual multipliers

1 INTRODUCTION

Data envelopment analysis is a set of methods and models used to evaluate relative efficiency of n decision making units, DMUs, each using m inputs to produce s outputs. There exist different models, see Charnes et al. (1994), and each model is presented as a linear program, with primal LP called envelopment model, and dual LP called multiplier model.

In envelopment model we seek for a linear combination of DMUs that represents, or dominates, DMU under evaluation. These DMUs form peer group for the DMU under evaluation. In multiplier model we seek for weights of each input and output so that efficiency of DMU under evaluation, evaluated as ratio of weighted output and weighted input, is maximized. These weights are called dual multipliers. In this presentation we are interested in dual multipliers, especially in how to obtain optimal dual multipliers for DMU under evaluation given certain criterion, and the role of dual multipliers in sensitivity analysis, i.e. assessment of robustness of DMU considering efficiency classification. In sensitivity analysis we are interested in changes of DMUs inputs and outputs, usually increase of inputs and decrease of outputs, so that DMU preserves its efficiency classification. Applying standard LP based sensitivity analysis we encounter problem of multiple optimal solutions, more rule then exception in DEA. This makes sensitivity analysis more difficult, i.e. in order to obtain sufficient and necessary conditions on the amounts on input and output changes that will not alter the obtained efficiency score, it is not sufficient to consider just the obtained optimal solution. In fact some other optimal solution can possibly allow bigger changes so that DMU remains efficient. In first approaches to sensitivity analysis in DEA change of single input or output was considered, Charnes et al.(1985). After, number of papers followed where multiple simultaneous changes, and different DEA models, were considered, Charnes et Neralić (1989a), (1989b), (1990), (1992a), (1992b) etc. In these approaches envelopment model was used. Multiplier model was considered by Thompson et al. (1994) and after by Gonzales-Lima (1996), Thompson et al.(1996). In this approaches either simplex or interior point method were used, resulting with sufficient, but not also necessary, conditions on amounts on input/output changes. Zhu (1996), Seiford and Zhu(1998), used different approach to sensitivity analysis. Superefficient, or extended DEA model was used, resulting with sufficient and necessary conditions on the amounts on changes, obtained via iterative algorithm. Following this our goal is to obtain such a set of dual multipliers so that efficient DMU remains efficient after applying changes if and only if these dual multipliers are optimal for it.

2 OPTIMAL DUAL MULTIPLIERS

Let us assume that we have n decision making units, DMU for short, each using m (same among DMUs) inputs to produce s (same among DMUs) outputs. We will employ notation

(1) $X_j = (x_{1j},\ldots,x_{mj})^T,\ X_j \geq 0,\ \ Y_j = (y_{1j},\ldots,y_{sj})^T, Y_j \geq 0,\ \ DMU_j = (X_j^T, Y_j^T)^T,\ j=1,\ldots,n$

to represent DMU_j. Based on this data, and assuming constant returns to scale, CRS, production possibility set, PPS, is defined as

(2) $$PPS = \left\{ (X^T, Y^T)^T \middle| Y \leq \sum_{j=1}^{n} \lambda_j Y_j, X \geq \sum_{j=1}^{n} \lambda_j X_j, \lambda_j \geq 0, j = 1,\ldots,n \right\}$$

Efficiency of the DMU_0 is defined as Paretto-Koopmans efficiency, i.e. $\text{DMU}_0 = (X_0^T, Y_0^T)^T$ is efficient if and only if there is no other point $T = (X^T, Y^T)^T$ from PPS where $x_{io} \geq x_i,\ i=1,\ldots,m$ and $y_{r0} \leq y_r,\ r=1,\ldots,s$ with at least one strict inequality in inputs or outputs. Assessment of efficiency can be done using variety of models, see Charnes et al. (1994), for overview of different DEA models. One approach is to use these LP:

Envelopment (primal) — Multiplier (associated dual)

(3)

$$\min_{s.t.} -\sum_{i=1}^{m} s_i^- - \sum_{r=1}^{s} s_r^+$$
$$-\sum_{j=1}^{n} \lambda_j x_{ij} - s_i^- = -x_{io}, i = 1,\ldots,m$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro}, r = 1,\ldots,s$$
$$\lambda_j, s_i^-, s_r^+ \geq 0$$

(4)

$$\max -\sum_{i=1}^{m} v_i x_{io} + \sum_{r=1}^{s} \mu_r y_{ro}$$
$$-\sum_{i=1}^{m} v_i x_{ij} + \sum_{r=1}^{s} \mu_r y_{rj} \leq 0,\ j = 1,\ldots,n$$
$$v_i, \mu_r \geq 1$$

Dual variables in program (4), $v_i, i=1,\ldots,m$ and $\mu_r, r=1,\ldots,s$, are called dual multipliers and are coefficients of PPS's facet that contains the reference point (obtained via (3)). DMU_0 is efficient if and only if optimal value of the objective function is 0, meaning that DMU_0 is not dominated by some point from PPS, and it is on the efficient frontier of PPS. Dual multipliers from (4) can be associated with weights in ratio form as it was originally done in seminal DEA paper, Charnes et al. (1978). In this approach we use intuitive efficiency evaluation, i.e. in single-input single-output case efficiency is evaluated as ratio of output and input value (higher the ratio, more efficient is the DMU). If this ratio equals 1 than DMU is efficient. In applying this approach in multiple input and output settings, we use weights for each input and output to obtain composite single weighted “virtual“ input and single weighted “virtual” output. Efficiency of DMU_0 is evaluated using following objective function

(5) $$\max_{\mu,v} \frac{\sum_{r=1}^{s} \mu_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$$

We define vector of dual multipliers $\omega = (v_1,\ldots,v_m,\mu_1,\ldots,\mu_s)^T, \omega \geq 0$, and also $f_j(\omega) = \mu_1 y_{1j} + \cdots + \mu_s y_{sj}$ and $g_j(\omega) = v_1 x_{1j} + \cdots + v_m x_{mj}$ and DMU_0 is efficient if and only if we can find vector of dual multipliers, such that

(6) $$h_0(\omega) = \frac{f_0(\omega)}{g_0(\omega)} \geq \frac{f_j(\omega)}{g_j(\omega)} = h_j(\omega), j = 1,...,n, j \neq 0$$

This vector is optimal for the DMU_0. Our further discussion is based on relation between these approaches. The idea is to obtain dual multipliers as optimal dual variables from modified LP (3), and use (6) in sensitivity analysis, Thompson et al. (1994). In fact, if DMU_0 is efficient, then optimal value of the objective function in LP (4) is 0. With v_i^* , $i=1,...,m$ and μ_r^* $r=1,...,s$ being optimal values of decision variables in (4), we have

(7) $$-\sum_{i=1}^{m} v_i^* x_{i0} + \sum_{r=1}^{s} \mu_r^* y_{r0} = 0$$

and for vector of dual multipliers $\omega^*=(v_1^*,...,v_m^*,\mu_1^*,...,\mu_s^*)^T$ we have

(8) $$h_j(\omega^*) \leq h_0(\omega^*) = 1$$

So ω^* is optimal vector of dual multipliers for DMU_0, i.e. (6) is fulfilled.

3 ROBUSTNESS OF EFFICIENT DMU

We are interested if the obtained efficiency evaluation is robust, i.e. what is the range of possible changes of inputs and outputs of DMU such that it remains efficient. Changes are such that we 'decrease' efficiency of DMU_0 and 'increase' efficiency of remaining DMUs:
for DMU_0 under evaluation:

absolute changes — proportional changes (as in Thompson et al. (1994)

(9) $$x_{io}^* = x_{i0} + \beta_i, \beta_i \geq 0, i = 1,...,m \qquad x_{io}^* = (1+c)x_{i0}, i = 1,...,m$$
$$y_{ro}^* = y_{r0} - \alpha_r, 0 \leq \alpha_r \leq y_{r0}, r = 1,...,s \qquad y_{ro}^* = (1-c)y_{r0}, r = 1,...,s$$
$$0 \leq c \leq 1$$

for other DMUs:

(10) $$x_{ij}^* = (1-c)x_{ij}, i = 1,...,m$$
$$y_{rj}^* = (1+c)y_{rj}, r = 1,...,s, \qquad j = 1,...,n, j \neq 0, \quad 0 \leq c \leq 1$$

Key question in evaluating robustness of efficient DMU_0 is assessment of maximal changes that will not alter its efficiency, i.e. it will remain efficient. We are looking for a result that states that DMU_0^*, which is DMU_0 after changes applied, is efficient if and only if changes are in certain intervals, i.e. sufficient and necessary conditions on possible data changes are given. Also, we base our analysis on dual multipliers so the above statement can be considered as: does exist vector of dual multipliers such that DMU_0^* is efficient if and only if that vector is optimal for it. In this procedure we can also consider changes of all DMUs, as (11).

We base our procedure on comparison between different approaches to sensitivity analysis in DEA. We will aplly superefficient DEA model, Seiford and Zhu (1998), since this procedure results in necessary conditions on data changes, i.e. DMU_0 is projected on the frontier of the production set spanned with remaining DMUs. Any further change will move it in the interior or on the inefficient part of the frontier, thus it will be inefficient.. We will use envelopment model and results as in series of paper by Charnes and Neralić , i.e. we will consider optimal basis matrix and its inverse obtained using simplex method. Also primal-dual relationship as well as parametric programming will be applied. The obtained optimal dual variables will be considered as dual multipliers and these variables are coefficients of facet of production set spanned with remaining DMUs .Third, in assessing data changes we will consider (6), as in Thompson and Thrall (1994).

In figure 1. we can visualize the idea of the above procedure

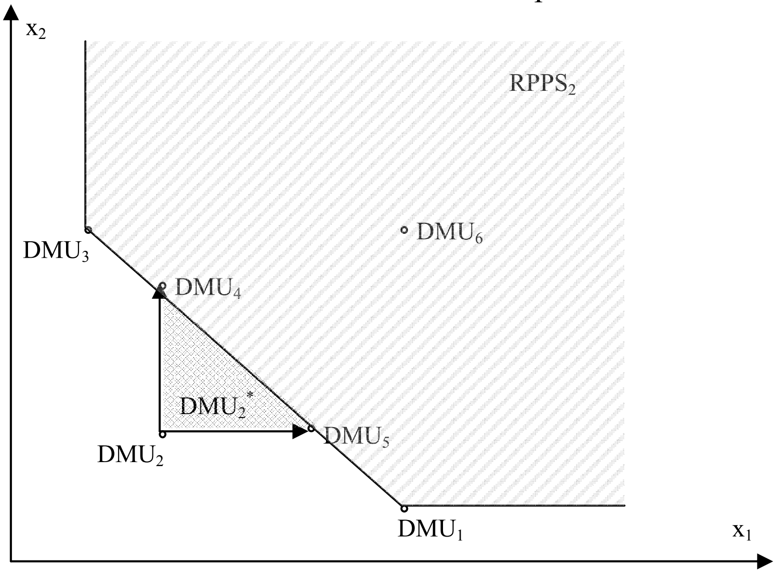


Figure 1.

In figure 1. DMU_2 is under evaluation. $RPPS_2$ represent the production set spanned with remaining DMUs. Possible data changes are such that DMU_2^* can be in the shaded triangle spanned with DMU_2 , DMU_4 and DMU_5 . The optimal dual multipliers are coefficients of the facet spaned with DMU_1 and DMU_3 . These coefficients can be obtained via appropriate superefficient DEA model.

In order to obtain equation of that facet, and with this dual multipliers we project DMU under evaluation to this facet, choosing just one input to increase or just one output to decrease. We can use the following LP, as in Seiford and Zhu (1998)

Primal associated dual

$$\min_{s.t.} \beta_1 \qquad \max \sum_{r=1}^{s} \mu_r y_{ro} - \sum_{i=1}^{m} \nu_i x_{io}$$

$$\beta_1 - \sum_{\substack{j=1\\ j\neq 0}}^{n} \lambda_j x_{1j} \geq -x_{1o} \qquad \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} \leq 0, \forall j, j \neq o$$

(11) $$-\sum_{\substack{j=1\\ j\neq 0}}^{n} \lambda_j x_{ij} \geq -x_{io}, i = 2,\ldots,m \qquad \nu_1 \leq 1$$

$$\sum_{\substack{j=1\\ j\neq 0}}^{n} \lambda_j y_{rj} \geq y_{ro}, r = 1,\ldots,s \qquad \mu_r, \nu_i \geq 0$$

$$\lambda_j, \beta_1 \geq 0$$

When solving (11), in some cases it can be infeseable, we obtain β_1 as increase in input 1 and also optimal dual variables, which can be easily red from the obtained simplex tableau. These variables are coefficients of the facet and thus optimal dual multipliers for the DMU under evaluation.

We have, for the facet which is border to possible changes, for each DMU besides DMU_0 $-\sum_{i=1}^{m} \nu_i x_{i0} + \sum_{r=1}^{s} \mu_r y_r \leq 0$ and for DMU_0 we have $-\sum_{i=1}^{m} \nu_i x_{i0} + \sum_{r=1}^{s} \mu_r y_{r0} > 0$ (since optimal solution of (11) is greater then 0).

It follows, after applying changes (9) until DMU_0^* is on the frontier, $-\sum_{i=1}^{m} v_i x_{io}^* + \sum_{r=1}^{s} \mu_r y_{ro}^* = 0 = -\sum_{i=1}^{m} v_i (x_{io} + \beta_i) + \sum_{r=1}^{s} \mu_r (y_{ro} - \alpha_r) = -\sum_{i=1}^{m} v_i x_{io} + \sum_{r=1}^{s} \mu_r y_{ro} - \sum_{i=1}^{m} v_i \beta_i - \sum_{r=1}^{s} \mu_r \alpha_r = 0$

implying $-\sum_{i=1}^{m} v_i x_{io} + \sum_{r=1}^{s} \mu_r y_{ro} = \sum_{i=1}^{m} v_i \beta_i + \sum_{r=1}^{s} \mu_r \alpha_r > 0$ since at least one of the α_r and β_i is greater then 0, and all v_i and μ_r in our case are greater then 0. This gives us sufficient and necessary conditions on amount of changes in inputs and outputs. Any further change will make DMU under evaluation inefficient. We can show results using following example, Thompson et al. (1994) with 6 DMUs, each using two inputs to produce one output, example shown in figure 1.

	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5	DMU_6
x_1- first input	4	2	1	2	3	4
x_2-second input	1	2	4	3	2	4
y- output	1	1	1	1	1	1

Table 1

Facet spanned with DMU_1 and DMU_3, more precisely only part of that facet spanned with DMU_4 and DMU_5, which is border to possible changes, figure 1, has equation $-x_1 - x_2 + 5y = 0$. We use coefficients to obtain dual multipliers, i.e. v_1=1, v_2=1 and μ=5, ω=(1,1,5)^T. Using previous definitions, $h(\omega) = \frac{5y}{x_1 + x_2}$, we have for each of six DMUs

$$
\begin{aligned}
&h_1(\omega) = \frac{5}{4+1} = 1, \quad h_2(\omega) = \frac{5}{2+2} = 1.25, \quad h_3(\omega) = \frac{5}{1+4} = 1 \\
&h_4(\omega) = \frac{5}{2+3} = 1, \quad h_5(\omega) = \frac{5}{3+2} = 1, \quad h_6(\omega) = \frac{5}{4+4} = 0.625
\end{aligned} \tag{12}
$$

So ω is optimal for DMU_2, (6), and we will assess possible data changes of DMU_2 with proportional changes, as in (9) and (10). In general case we have

$$
\begin{aligned}
&h_0(\omega) = d > 1 \\
&h_j(\omega) \le 1, j = 1,\dots,n, j \ne 0
\end{aligned} \tag{13}
$$

with at least one equation for some j in (13). We can apply changes until

$$h_0^*(\omega) = \frac{f_0^*(\omega)}{g_0^*(\omega)} = \frac{\mu_1 y_{10}^* + \cdots + \mu_s y_{s0}^*}{v_1 x_{10}^* + \cdots + v_m x_{m0}^*} = \frac{(1-c) f_0(\omega)}{(1+c) g_0(\omega)} = \frac{(1-c)}{(1+c)} d = 1 = h_j(\omega) \quad \text{for some } j = 1,\dots,n, j \ne o$$

thus $\frac{(1-c)}{(1+c)} d = 1 \Rightarrow c = \frac{d-1}{d+1}$

giving same results as in Thompson et al. (1994) and Charnes et Neralic (1990). Thus we obtained sufficient but also necessary conditions on amount of data changes so that DMU under evaluation remains efficient.

This simple example will guide us to the more complex situation which can be represented on figure 2.

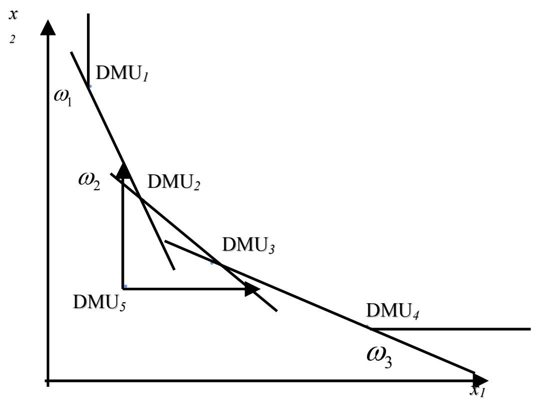


Figure 2

This is the situation when there are more facets that can be reached by DMU_0^* after changes are applied. The approach consists in finding equations of all these facets. In order to do this we will find equation of one facet using LP (11) using previous procedure. Exploring all possible changes, using parametric programming where only RHS is changed, we will either obtain part of the facet that is border to possible changes, or we will have to move from one facet to the neighboring ones continuing the process. In figure 2 there are three facets, resulting with three vectors of dual multipliers, ω_1, ω_2 and ω_3. The result is that DMU under evaluation is efficient when applying changes if and only if at least one of the obtained vector of dual multipliers is optimal for it.

4 CONCLUSION

Sensitivity analysis in DEA is one of the key topics. We are interested in robustness of the efficient DMUs, seeking for the possible input/output changes that will not alter its efficiency status. In this approach we concentrate on dual multipliers, use primal-dual relationship to obtain dual multipliers and finally obtain results with sufficient and necessary conditions on range of possible data changes so that DMU remains efficient.

References

Ali, A.I., Lerme, C.S., Seiford, L.M.,"Components of efficiency evaluation in data envelopment analysis", European Journal of Operational Research 80, (1995), 462-473

Andersen, P.,Petersen, N.C.,"A procedure for ranking efficient units in data envelopment analysis", Management Science 39, (1993), 1261-1264

Charnes, A., Cooper, W.W., Rhodes, E.,"Measuring the efficiency of decision making units", European Journal of Operational Research 2, (1978), 429-444

Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M., Stutz, J.,"Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions", Journal of Econometrics 30, (1985), 91-109

Charnes, A., Neralić, L., "Sensitivity analysis in data envelopment analysis 1", Glasnik matematički Ser.III 24(44), (1989), 211-226

Charnes, A., Neralić, L., "Sensitivity analysis in data envelopment analysis 2", Glasnik matematički Ser.III 24(44), (1989), 449-463

Charnes, A., Neralić, L., "Sensitivity analysis in data envelopment analysis 3", Glasnik matematički Ser.III 27(47), (19929), 191-201

Charnes, A., Cooper, W.W., Thrall, R.M., "A Structure for Classifying and Characterizing Efficiency and Inefficiency in Data Envelopment Analysis", The Journal of Productivity Analysis 2 (1991) 197-237

Charnes, A., Cooper, W.W., Thrall, R.M., "A Structure for Classifying and Characterizing Efficiency and Inefficiency in Data Envelopment Analysis", The Journal of Productivity Analysis 2 (1991) 197-237

Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M. ed.,:"Data Envelopment Analysis: Theory , Methodology, and Application" Kluwer Academic Publishers, (1994)

Charnes, A., Rousseau, J.J., Semple, J.H., "Sensitivity and Stability of Efficiency Classifications in Data Envelopment Analysis", The Journal of Productivity Analysis 7, (1996), 5-18

Gonzalez-Lima, M.D., Tapia, R.A., Thrall, R.M., "On the construction of strong complementarity slackness solutions for DEA linear programming problems using a primal-dual interior-point method", Annals of Operations Research 66, (1996),139-162

Seiford, L.M., Zhu, J.,"Stability regions for maintaining efficiency in data envelopment analysis", European Journal of Operational Research 108(1), (1998), 127-139

Seiford, L.M., Zhu, J.,"Sensitivity analysis of DEA models for simultaneous changes in all data", Journal of the Operational Research Society Vol. 49, No. 10 (1998) 1060-1071

Thompson, R., Dharmapala, P. S., Thrall, R. M., "Sensitivity analysis of efficiency measures with application to Kansas farming and Illinois coal mining", In Charnes et al. (editors): Data Envelopment Analysis: Theory, Methodology and Applications, Kluwer Academic Publishers, (1994)

Thompson, R.G., Dharmapala, Diaz, J., Gonzalez-Lima, M.D., Thrall, R. M., "DEA multiplier analytic center sensitivity with an illustrative application to independent oil companies", Annals of Operations Research 66, (1996),163-177

Zhu, J.,"Robustness of the efficient DMUs in data envelopment analysis",European Journal of Operational Research 90(3), (1996), 451-460

Zhu, J.,"Super-efficiency and DEA sensitivity analysis",European Journal of Operational Research 129, (2001), 443-455

RECENT DEVELOPMENTS IN COPOSITIVE PROGRAMMING

Immanuel M. Bomze
Dept.of Statistics and Decision Support Systems, University of Vienna
Bruenner Strasse 72, A-1210 Wien, Austria
immanuel.bomze@univie.ac.at

Abstract: A symmetric matrix is called copositive, if it generates a quadratic form taking no negative values over the positive orthant. Contrasting to positive-semidefiniteness, checking copositivity is NP-hard. In a copositive program, we have to minimize a linear function of a symmetric matrix over the copositive cone subject to linear constraints. This convex program has no non-global local solutions. On the other hand, there are several hard non-convex programs which can be formulated as copositive programs, thus shifting the complexity from global optimization towards sheer feasibility questions.

Key words. Complete positivity; quadratic optimization; conic programs; interior point methods

Extended abstract.

The set of all symmetric *nxn* matrices which generate a quadratic form taking no negative values over the positive orthant forms a closed convex cone ***C*** in the space of all symmetric *nxn* matrices $A=A'$, where ´ denotes transposition.

Under the Frobenius inner product $<A,B> = trace\ (AB)$, this cone ***C*** is not self-dual like the cone ***P*** of all positive-semidefinite symmetric *nxn* matrices, or the cone ***N*** of all symmetric *nxn* matrices with no negative entries. Rather, the dual cone of ***C***,

$$\boldsymbol{C^*} = \{ B = B' : \text{for all } A \text{ in } \boldsymbol{C},\ <A,B> > 0 \text{ or } <A,B> = 0 \}$$

is the cone of all completely positive matrices, which is given by all products FF', for rectangular *nxk* matrices F with no negative entries, where k may exceed n. Alternatively, one may write

$$\boldsymbol{C^*} = \text{convex hull } \{ xx' : x \text{ in } \boldsymbol{R}^n \text{ has no negative coordinates } \} .$$

If n does not exceed 4, these cones are relatively simple: ***C**** is the intersection of ***P*** with ***N***, and ***C*** is the Minkowski sum ***P*** +***N***. However, if n exceeds 4, then both ***C*** and ***C**** are much more complicated: the former is strictly larger than ***P+N*** while the latter is strictly included in the intersection of ***P*** with ***N***.

This perfectly corresponds to the fact that linear optimization over ***P*** and ***N*** can be accomplished in polynomial time to arbitrary accuracy via interior-point methods (over ***P*** we speak of semidefinite programming while over ***N*** we get the familiar linear programming problem).

By contrast, Burer recently showed [21] that every (possibly non-convex) quadratic optimization problem, even if some variables are restricted to be binary, can be represented as a linear optimization problem over ***C****.

Speaking more generally, copositivity plays a central role in non-convex quadratic optimization, since also conditions characterizing local and global optimization of critical points, involve copositivity in a natural manner [4], [24].

It is easy to write down primal-dual pairs of copositive programs: if the primal program is

$$min \{ <C,X> : <A_i, X> = b_i\ , i=1..m,\ X \text{ in } \boldsymbol{C^*} \}$$

then the dual program is

$$sup \{ b'y : C- (y_1 A_1 + \ldots + y_m A_m) \text{ in } \boldsymbol{C} \} .$$

As in semidefinite programming, for this primal-dual pair, duality theory is slightly more complex than in the linear programming case. For instance, there can be a positive duality gap, or the dual program may not attain the optimal value although it is feasible and bounded [39].

This may happen for an important class of copositive representations, namely those of so-called multi-standard QPs, which consist of optimizing a quadratic form over the direct product of several standard simplices. Recently, different convergent monotone interior-point methods for multi-StQPs have been proposed, some of which extend so-called Relaxation Labelling Processes arising in Pattern Recognition and Image Processing [18], [41], [42].

These multi-StQPs form a generalization of Standard QPs (optimizing a quadratic form over a single standard simplex) which in turn form a central class in quadratic optimization [6], [7], [9], [10], [19]. These non-convex quadratic problems may have up to $1.25*2^n/n^{1/2}$ inefficient local solutions and are NP-hard as they encode, among others, also the Maximum-Clique Problem, according to the famous Motzkin-Straus theorem [33] – as an aside, Motzkin also coined the term copositivity, which apparently is an abbreviation of **co**nditional **positiv**e-semidefiniteness.

The copositive formulation of Standard QPs was the first application of copositive programming [13], and has been extensively used for deriving SDP-based bounds on the clique number [25], [15], [28] or to improve bounds on the crossing numbers of graph classes towards Zarankiewicz' conjecture [26]. To be more specific, if S denotes the standard simplex in $\boldsymbol{R}^n$, then the StQP with symmetric nxn data matrix Q reads $min \{ x'Qx : x \text{ in } S \}$ and can be written as the copositive programming problem

$$min \{ x'Qx : x \text{ in } S \} = min \{ <Q,X> : <J,X> = 1, X \text{ in } \boldsymbol{C^*} \} = max \{ y : Q - yJ \text{ in } \boldsymbol{C} \} ,$$

where J denotes the nxn matrix with all unit entries. Hence, if checking membership in $\boldsymbol{C}$ or $\boldsymbol{C^*}$ would be easy, the NP-hard StQP would be reduced to a line search problem. So detecting
whether or not a given matrix belongs to $\boldsymbol{C}$ is also NP-hard. Nevertheless there are many copositivity detection procedures [3], [5], [8], [11], [20], [22], [23], [29], [32] and much less methods for detecting complete positivity. For a recent survey of complete positivity see [2].

Other applications of copositive programming include further combinatorial problems like Graph Partitioning [37], Maximum-Cut, Quadratic Assignment [37], or, in a statistical context, supervised (classification by support vector machines) and unsupervised learning (SQP strategies in k-means MSE clustering) [18]. From an algorithmic point of view, cheap but efficient bounds for quadratic optimization problems can be derived from copositive programming [1], [12], [16], [17].

All these approaches use, in some way or another, approximations of the intractable cones $\boldsymbol{C}$ or $\boldsymbol{C^*}$, by cones which occur in (albeit larger) SDPs. A rapidly evolving research topic is the construction of approximation hierarchies [12], [28], [30], [31], [34], [35], [36]. These consist of a sequence of nested tractable cones which eventually include every matrix in the interior of $\boldsymbol{C}$ (the dual cones would then shrink towards $\boldsymbol{C^*}$).

The starting point of this cone sequence approximating $\boldsymbol{C}$, the approximation of order zero, is the smallest cone, $\boldsymbol{P} + \boldsymbol{N}$, which we already encountered above. Applied to the above-sketched bounding of the clique number, one would arrive by this procedure at the well-known Lovász-Schrijver bound [40]. However, an approximation of larger order r typically involves SDPs of size n^{r+1}. So even for moderate problem dimensions n, cheap but efficient bounds are basically restricted to small approximation order. For instance, recent publications suggest to add only one cut (copositivity constraint) [14], or a limited number of these cuts (triangle inequalities) [27] to improve the Lovász-Schrijver bound, and rather employ data-driven improvements like exploiting symmetry and/or decompositions of the given problem instances.

References

[1] Anstreicher, K., and S. Burer (2005), ``D.C. Versus Copositive Bounds for Standard QP,'' *J. Global Optimiz* **33**, 299-312.

[2] A. Berman, and N. Shaked-Monderer (2003), *Completely Positive Matrices*, World Scientific Publ., London.

[3] Bomze, I.M. (1987), ``Remarks on the recursive structure of copositivity,'' *J.Inf.& Optimiz.Sciences* **8**, 243-260.

[4] Bomze, I.M. (1992), ``Copositivity conditions for global optimality in indefinite quadratic programming problems,'' *Czechoslovak J. Operations Research* **1**, 7-19.

[5] Bomze, I.M. (1996), ``Block pivoting and shortcut strategies for detecting copositivity,'' *Linear Alg.Appl.* **248**, 161-184.

[6] Bomze, I.M. (1997), ``Evolution towards the maximum clique,'' *J. Global Optimiz.* **10**, 143-164.

[7] Bomze, I.M. (1998), ``On standard quadratic optimization problems,'' *J. Global Optimiz*.**13**, 369-387.

[8] Bomze, I.M. (2000), ``Linear-time detection of copositivity for tridiagonal matrices and extension to block-tridiagonality,'' *SIAM J.Matrix Anal.Appl.* **21**, 840-848.

[9] Bomze, I.M. (2002), ``Branch-and-bound approaches to standard quadratic optimization problems,'' *J. Global Optimiz.* **22**, 17-37.

[10] Bomze, I.M. (2005), ``Portfolio selection via replicator dynamics and projection of indefinite estimated covariances,'' *Dynamics of Continuous, Discrete and Impulsive Systems B* **12**, 527-564.

[11] Bomze, I.M., and G. Danninger (1993), ``A global optimization algorithm for concave quadratic problems,'' *SIAM J. Optimiz*. **3**, 836-842.

[12] Bomze, I.M., and E. de Klerk (2002), ``Solving standard quadratic optimization problems via linear, semidefinite and copositive programming,'' *J. Global Optimiz*. **24**, 163-185.

[13] Bomze, I.M., M. Dür, E. de Klerk, A. Quist, C. Roos, and T. Terlaky (2000), ``On copositive programming and standard quadratic optimization problems,'' *J. Global Optimiz*.**18**, 301-320.

[14] Bomze, I.M., F. Frommlet, and M. Locatelli (2007), ``The first cut is the cheapest: improving SDP bounds for the clique number via copositivity,'' submitted.

[15] Bomze, I.M., F. Frommlet, and M. Locatelli (2007), ``Gap, cosum, and product properties of the Lovász-Schrijver bound on the clique number,'' submitted.

[16] Bomze, I.M., F. Frommlet, and M. Rubey (2007), ``Improved SDP bounds for minimizing quadratic functions over the l^1 ball,'' *Optimiz. Letters* **1**, 49-59.

[17] Bomze, I.M., M. Locatelli, and F. Tardella (2007), ``New and old bounds for standard quadratic optimization: dominance, equivalence and incomparability,'' to appear in *Math. Programming*.

[18] Bomze, I.M., and W.Schachinger (2007), ``Multi-Standard Quadratic Optimization Problems,” submitted.

[19] Bomze, I.M., and V. Stix (1999), ``Genetical engineering via negative fitness: evolutionary dynamics for global optimization, '' *Annals of O.R* . **89**, 279-318.

[20] Bundfuss, S., and M. Dür (2006), ``Criteria for copositivity and approximations of the copositive cone,“ preprint, Techn. Univ. Darmstadt.

[21] Burer, S. (2006), ``On the copositive representation of binary and continuous nonconvex quadratic programs," preprint, Univ. of Iowa.

[22] Cottle, R.W., G.J. Habetler, and C.E. Lemke (1970), ``Quadratic forms semi-definite over convex cones," in: H.W. Kuhn (ed.), Proc. Princeton Sympos.\ Math. Programming, 551-565. Princeton University Press.

[23] Danninger, G. (1990), ``A recursive algorithm for determining (strict) copositivity of a symmetric matrix," in: U. Rieder et al. (eds.), Methods of Operations Research **62**, 45-52. Hain, Meisenheim.

[24] Danninger, G. (1992), ``Role of copositivity in optimality criteria for nonconvex optimization problems, " *J.Optimiz.Theo.Appl*. **75**, 535-558.

[25] de Klerk, E., and D.V. Pasechnik (2002), ``Approximation of the stability number of a graph via copositive programming," *SIAM J. Optimiz.* **12**, 875-892.

[26] de Klerk, E., J. Maharry, D.V. Pasechnik, R.B. Richter, and G. Salazar (2006), ``Improved bounds for the crossing numbers of $K_{m,n}$ and K_n," *SIAM J. Discrete Mathematics* **20**, 189-202.

[27] Dukanovic, I., and F. Rendl (2007), ``Semidefinite programming relaxations for graph coloring and maximal clique problems," *Math. Programming* **109**, 345-365.

[28] Gvozdenovic, N., and M. Laurent (2005), ``Semidefinite bounds for the stability number of a graph via sums of squares of polynomials," Lecture Notes in Computer Science, vol 3509/2005, pages 136-151. Integer Programming and Combinatorial Optimization: 11th International IPCO Conference.

[29] Ikramov, Kh.D. (2002), ``Linear-time algorithm for verifying the copositivity of an acyclic matrix," *Comput.Math.Math.Phys*.**42**, 1701-1703.

[30] Lasserre, J.B. (2001), ``Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization* **11**, 796-817.

[31] Lasserre, J.B. (2001), ``An explicit exact SDP relaxation for nonlinear 0-1 programming," In: K.Aardal and A.H.M. Gerards, eds., Lecture Notes in Computer Science 2081, 293-303.

[32] Martin, D.H., and D.H. Jacobson (1981), ``Copositive matrices and definiteness of quadratic forms subject to homogeneous linear inequality constraints," *Linear Alg. Appl*. **35**, 227-258.

[33] Motzkin, T.S., and E.G. Straus (1965), ``Maxima for graphs and a new proof of a theorem of Turán," *Canadian J. Math.* **17**, 533-540.

[34] Parrilo, P.A. (2000), Structured Semidefinite Programs and Semi-algebraic Geometry Methods in Robustness and Optimization, PhD thesis, California Institute of Technology, Pasadena, USA.

[35] Parrilo, P.A. (2003), ``Semidefinite programming relaxations for semi-algebraic problems," *Math. Programming B* **696**, 293-320.

[36] Pena J., J. Vera, and L. Zuluaga (2007), ``Computing the stability number of a graph via linear and semidefinite programming," *SIAM J. Optimization* **18**, 87-105.

[37] Povh, J., and F. Rendl (2006), ``Copositive and Semidefinite Relaxations of the Quadratic Assignment Problem,” submitted.

[37] Povh, J., and F. Rendl (2007), ``A copositive programming approach to graph partitioning,” *SIAM Journal on Optimization* **18**, 223-241.

[39] Schachinger, W., and I.M. Bomze (2007), ``A conic duality Frank-Wolfe type theorem via exact penalization in quadratic optimization,” submitted.

[40] Schrijver, A. (1979), ``A comparison of the Delsarte and Lovász bounds," *IEEE Trans. Inform.Theory* **25**, 425-429.

[41] Tseng, P. (2007), ``A scaled projected reduced-gradient method for linearly constrained smooth optimization," preprint, Univ.of Washington.

[42] Tseng, P., I.M. Bomze, and W. Schachinger (2007), ``A first-order interior-point method for Linearly Constrained Smooth Optimization," preprint, Univ.of Washington.

EIGENPROBLEM IN EXTREMAL ALGEBRAS

MARTIN GAVALEC, JÁN PLAVKA

Abstract. Extremal algebra deals with extremal operations: maximum and minimum, which are used in place of operations of addition and multiplication used in the linear algebra. For a given $n \times n$ matrix A in an extremal algebra, the eigenvalue-eigenvector problem is studied. The properties of eigenvectors and the structure of the eigenspace $\mathcal{F}(A)$, from various points of view are described. The computational complexity of the presented algorithms, for general case and also for special types of matrices, is evaluated.

1. Introduction

Extremal algebras deal with the operations of maximum and minimum which are involved in many optimization problems. Matrix computations using these operations were considered by a number of authors, e.g. in [1, 6, 8, 18], and analogies of various notions from the classical linear algebra were studied.

Max-plus algebras are important in the study of discrete events systems (DES, in short). The steady states of DES correspond to eigenvectors of max-plus matrices, see [7, 23], hence investigation of the properties of eigenvectors and the characterization of the eigenspace structure is important for the applications. In some cases, the investigation is more efficient, if the considered matrix has special properties. Many efficient solution of problems concerning Monge matrices were described in [2]. Problems connected with eigenvectors of Monge matrices were studied in papers [9, 11, 13], in which efficient algorithms for various questions were presented.

Max-min algebras have wide applications in the fuzzy set theory (the max-min algebra on the unit real interval is one of the most important fuzzy algebras). The eigenvectors of max-min matrices are useful in cluster analysis (see [17]), or in fuzzy reasoning (see [26]). The eigenproblem in max-min algebra and its connections to paths in digraphs were investigated in [4, 17, 18, 19]. A procedure for computing the greatest eigenvector of a given max-min matrix was proposed in [26] and an efficient algorithm was described in [5]. The eigenproblem in distributive lattices was studied in [27].

The first part of this paper deals with the eigenvalue-eigenvector problem in max-plus algebra. The problem is studied for general matrices and also for special types such as circulant or Monge matrices. As a generalization, the multiparametric version of the eigenproblem is investigated.

The second part is concentrated to max-min algebra. We discuss several questions related to the structure of the eigenspace $\mathcal{F}(A)$, the robustness and to the simple image set of a given max-min square matrix A.

2. Eigenproblem in max-plus algebra

By a max-plus algebra we understand the algebraic structure $(G, \oplus, \otimes) = (\mathcal{R}^\star, \max, +)$, where $G = \mathcal{R}^\star$ is the set of all real numbers $\mathcal{R}$ extended by an infinite element $\varepsilon = -\infty$, and

Date: July 15, 2007.

1991 *Mathematics Subject Classification.* Primary 04A72; Secondary 05C50, 15A33.

Key words and phrases. eigenproblem, max-plus algebra, max-min algebra.

This work was supported by Czech Science Foundation #402/06/1071 and VEGA #1/2168/05.

$\oplus, \otimes$ are the binary operations on $\mathcal{R}$: $\oplus = \max$ and $\otimes = +$. The infinite element is neutral with respect to the maximum operation and absorbing with respect to addition.

The results presented in this paper for the max-plus algebra $(\mathcal{R}^\star, \max, +)$ are valid also for the general notion of max-plus algebra, in which $(G, \oplus, \otimes)$ is derived in a similar way from an arbitrary divisible commutative linearly ordered group in additive notation. In the general case, the neutral element $e \in G$ in the additive group must be used instead of $0 \in \mathcal{R}$.

For any natural $n > 0$, we denote $N = \{1, 2, \dots, n\}$. Further, we denote by $G(m, n)$ the set of all $m \times n$ matrices over G. The matrix operations over the max-plus algebra G are defined with respect to $\oplus, \otimes$, formally in the same manner as the matrix operations over any field. The operation $\otimes$ for matrices denotes the formal matrix product with operations $\oplus = \max$ and $\otimes = +$ replacing the usual operations $+, \cdot$, while the operation $\oplus$ for matrices is performed componentwise.

The problem of finding a vector $x \in G(n, 1)$ and a value $\lambda \in G$ satisfying

$$A \otimes x = \lambda \otimes x \tag{2.1}$$

is called an *max-plus eigenproblem* corresponding to the matrix A, the value λ is called eigenvalue, and x is called eigenvector of A.

The *associated digraph* D_A of a matrix $A \in G(n, n)$ is defined as a complete arc-weighted digraph with the node set $V = N$, and with the arc weights $w(i, j) = a_{ij}$ for every $(i, j) \in N \times N$. If p is a path or a cycle in D_A, of length $r = |p|$, then the weight $w(p)$ is defined as the sum of all weights of the arcs in p. If $r > 0$, then the mean weight of p is defined as $w(p)/r$. Of all the mean weights of cycles in D_A, the maximal one is denoted by $\lambda(A)$. By Cuninghame-Green in [8], the maximal cycle mean $\lambda(A)$ is the unique eigenvalue of A. The problem of finding the eigenvalue $\lambda(A)$ has been studied by a number of authors and several algorithms are known for solving this problem. The algorithm described by Karp in [21] has the worst-case performance $O(n^3)$. The iterative algorithm by Howard has been reported to have on average almost linear computational complexity, though a tight upper bound has not yet been found (see [20]).

Theorem 2.1. [8] *Let $A \in G(n, n)$. Then $\lambda(A)$ is the unique eigenvalue of A.*

For $B \in G(n, n)$ we denote by $\Delta(B)$ the matrix $B \oplus B^{(2)} \oplus \cdots \oplus B^n$ where $B^{(s)}$ stands for the s-fold iterated product $B \otimes B \otimes \cdots \otimes B$. Further, we denote $A_\lambda = -\lambda(A) \otimes A$ (here we have a formal product of a scalar value $-\lambda(A)$ and a matrix A, i.e. $[A_\lambda]_{ij} = -\lambda(A) + a_{ij}$ for any $(i, j) \in N \times N$). It is shown in [8] that the matrix $\Delta(A_\lambda)$ contains at least one column, the diagonal element of which is 0 and every such a column is an eigenvector (so called: fundamental eigenvector) of the matrix A.

Theorem 2.2. [8] *Let $A \in G(n, n)$. Every eigenvector of A can be expressed as a linear combination of fundamental eigenvectors.*

Let $\Delta(A_\lambda) = (\delta_{ij})$. It follows from the definition of $\Delta(A_\lambda)$ that δ_{ij} is the maximal weight of a path from i to j in D_{A_λ}. Hence, $\Delta(A_\lambda)$ can be computed in $O(n^3)$ time, using the Floyd-Warshall algorithm [22]. In this way, a complete set of fundamental eigenvectors can be found by at most $O(n^3)$ operations. However, if we wish to compute only one single eigenvector of A, no better algorithm than $O(n^3)$ is known for matrices of a general type.

3. Special matrices

In special cases, when the matrix A is circulant or Monge, the above computations can be performed in a more efficient way.

Let $a_0, a_1, \dots, a_{n-1} \in \mathcal{G}$. We say that $A \in G(n,n)$ is a circulant matrix (generated by elements $a_0, a_1, \dots, a_{n-1}$), if A has the form

$$A = A(a_0, a_1, \dots, a_{n-1}) = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix}$$

Theorem 3.1. [24] *Let $A \in G(n,n)$ be circulant matrices. Then $\lambda(A) = \max_{i \in N} a_i$.*

Theorem 3.2. [24] *Let A, B be circulant matrices. Then also $A \oplus B$, $A \otimes B$ are circulant matrices.*

The last theorem allows compute the $\Delta(A_\lambda)$ by using Dijkstra algorithm for computing 1-to-all heaviest paths. Then, after the reconstruction of matrix $\Delta(A_\lambda)$ (which also is circulant), we obtain all eigenvectors of A.

Theorem 3.3. [24] *There exists an algorithm $\mathcal{A}$ which for a given circulant matrix $A \in G(n,n)$ computes the eigenvalue and the eigenvectors in $O(n^2)$ time.*

We say that a matrix $A = (a_{ij}) \in G(n,n)$ is Monge if

$$a_{ij} + a_{kl} \leq a_{il} + a_{kj} \quad \text{for all} \quad i < k,\ j < l$$

Similarly, we say that a matrix $A = (a_{ij}) \in G(n,n)$ is inverse Monge if

$$a_{ij} + a_{kl} \geq a_{il} + a_{kj} \quad \text{for all} \quad i < k,\ j < l$$

The following theorems show that, in computing the eigenvalue of a given matrix with Monge (inverse Monge) property, the computation may be restricted to cycles of lengths 1 and 2 (to cycles of length 1).

Theorem 3.4. [11] *If $A = (a_{ij})$ has the Monge property, then*

$$\lambda(A) = \max_{i,j \in N} \left\{ a_{ii},\ \frac{a_{ij} + a_{ji}}{2} \right\}$$

Theorem 3.5. [11] *If $A = (a_{ij})$ has the inverse Monge property, then*

$$\lambda(A) = \max_{i \in N} \{ a_{ii} \}$$

As a consequence, the eigenvalue $\lambda(A)$ of a Monge (inverse Monge) matrix can be found in $O(n^2)$ time (in $O(n)$ time). Next theorem shows that the computation of a single eigenvector of a Monge (inverse Monge) matrix can also be performed in $O(n^2)$ time.

Theorem 3.6. [13] *There is an algorithm $\mathcal{A}$ which, for a given Monge matrix $A \in G(n,n)$ over a max-plus algebra G, computes an eigenvector of A in $O(n^2)$ time.*

The eigenspace for Monge (inverse Monge) matrices has been described in papers [15] and [16].

4. Multiparametric eigenproblem

For p arbitrary parameters $\alpha_1, \dots, \alpha_p \in G$ and for given $A \in G(n,n)$, find vector $x_{\alpha_1,\dots,\alpha_p} \in G(n,1)$ and value $\lambda(A(\alpha_1, \dots, \alpha_p)) \in G$ satisfying

$$A(\alpha_1, \dots, \alpha_p) \otimes x_{\alpha_1,\dots,\alpha_p} = \lambda(A(\alpha_1, \dots, \alpha_p)) \otimes x_{\alpha_1,\dots,\alpha_p}.$$

For given matrix $A = (a_{kl}) \in G(n,n)$, for $i \in N$ and for cyclic permutation $\sigma = (i_1, \dots, i_s)$ denote $J = \{i_1, i_2, \dots, i_s\} \bigcap \{1, 2, \dots, p\}$. Indices in J will be denoted as $j_1, j_2, \dots, j_k$. Define

$$m_s^J = \max_{\sigma \in C_n^k} \left\{ \frac{a_{i_1 i_2} + a_{i_2 i_3} + \cdots + a_{i_s i_1}}{s} \right\},$$

where $C_n^k \subset C_n$ is the set of all cyclic permutations on subsets of N containing elements $j_1, j_2, \dots, j_k$, and

$$M_s^J = m_s^J + \frac{\alpha_{j_1} + \cdots + \alpha_{j_k}}{s}$$

$$P_>^J(v) = \left\{ (\alpha_{j_1}, \dots, \alpha_{j_k}) \in R^k;\ M_v^J > \max_{v \neq s \in N,\ E \subseteq L} \left\{ M_s^E;\ \lambda(C) \right\} \right\}.$$

Denote by $B_A = (b_{ij})$ the $n \times n$ matrix which arose from the matrix A by replacing all entries of first p rows and first p columns by $-\infty$. If $b_j^1 = (b_{1j}^1, \dots, b_{nj}^1)$ is j-th column of A, $j = 1, \dots, n$, then define

$$b_j^{k+1} = B_A \otimes b_j^k, \quad \text{for} \quad k = 1, \dots, n-1.$$

Theorem 4.1. [25] *Let* $J = \{j_1, j_2, \dots, j_k\}$. *Then*

$$m_s^J = \max_{r_1 + \cdots + r_k + k + 1 = s} \frac{a_{j_1 v_1} + b_{v_1 s_1}^{r_1} + a_{s_1 j_2} + \cdots + b_{j_{k-1} v_k}^{r_k} + a_{v_k j_1}}{s}.$$

Theorem 4.2. [25] *Let* $(\alpha_{j_1}, \dots, \alpha_{j_k}) \in P_>^J(v)$. *Then* $|F_A'(\alpha_1, \dots, \alpha_p)| = 1$.

Theorem 4.3. [25] *Let*

$$\lambda(A(\alpha_1, \dots, \alpha_p)) = M_v^J = m_v^J + \frac{\alpha_{j_1} + \cdots + \alpha_{j_k}}{v}.$$

Then

$$\xi_{ij_\ell}(\alpha_1, \dots, \alpha_p) = \max_k \left\{ b_{ij_\ell}^k - k \left(m_v^J + \frac{\alpha_{j_1} + \cdots + \alpha_{j_k}}{v} \right) \right\},\ j_\ell \in J.$$

Theorem 4.4. [25] *There is an algorithm that computes values* m_s^J *for all* J *and the coordinates of the eigenvectors* $\xi_{ij_\ell}(\alpha_1, \dots, \alpha_p)$ *in* $O(2^p n^4)$ *time, i.e. in* $O(n^4)$ *time for any fixed* p.

5. Eigenproblem in max-min algebra

By a max-min algebra we understand a linearly ordered set $(\mathcal{B}, \leq)$ with the binary operations of maximum and minimum, denoted by $\oplus$ and $\otimes$. The matrix operations over $\mathcal{B}$ are defined with respect to $\oplus, \otimes$, formally in the same manner as matrix operations over any field. For a given natural $m, n > 0$, we denote by $\mathcal{B}(m, n)$ the set of all $m \times n$ matrices over $\mathcal{B}$. Similarly as in the max-plus case, the matrix operations over the max-min algebra $\mathcal{B}$ are defined with respect to $\oplus, \otimes$.

We say that a vector $b \in \mathcal{B}(n, 1)$ is increasing, if $b_i \leq b_j$ holds for any $i, j \in N$, $i \leq j$. Vector b is strictly increasing, if $b_i < b_j$ whenever $i < j$. The set of all increasing (strictly increasing) vectors in $\mathcal{B}(n, 1)$ is denoted by $\mathcal{B}^{\leq}(n, 1)$ (by $\mathcal{B}^{<}(n, 1)$). For $x, y \in \mathcal{B}(n, 1)$, we write $x \leq y$, if $x_i \leq y_i$ holds for all $i \in N$, and we write $x < y$, if $x \leq y$ and $x \neq y$. In other words, $x < y$ if $x_i \leq y_i$ for all $i \in N$, but the strong inequality $x_i < y_i$ holds true for at least one $i \in N$.

For given $A \in \mathcal{B}(n, n), h \in \mathcal{B}$, the *threshold digraph* $\mathcal{G}(A, h)$ is the digraph $\mathcal{G} = (N, E)$, with the vertex set N and with the arc set $E = \{(i, j);\ i, j \in N,\ a_{ij} \geq h\}$. The *strict threshold digraph* $\mathcal{G}(A, h^+)$ has the arc set $\{(i, j);\ i, j \in N,\ a_{ij} > h\}$. The set of all permutations on N will be denoted by P_n. If $A \in \mathcal{B}(n, n)$ and $b \in \mathcal{B}(n, 1)$, $\varphi \in P_n$, then we denote by $A_{\varphi\psi}$ the matrix created by applying permutation φ to the rows and permutation ψ to the columns of A, and by b_φ we denote the vector created by applying the permutation φ to vector b.

For any square matrix $A \in \mathcal{B}(n, n)$, the eigenspace of A is defined by

$$\mathcal{F}(A) := \{\, b \in \mathcal{B}(n, 1);\ A \otimes b = b \,\}$$

The vectors in $\mathcal{F}(A)$ are called eigenvectors of matrix A. The set of all increasing eigenvectors is denoted by $\mathcal{F}^{\leq}(A)$, and the set of all strictly increasing eigenvectors of A is denoted by

$\mathcal{F}^{<}(A)$. As any vector $b \in \mathcal{B}(n,1)$ can be permuted to an increasing vector, the next theorem says that the structure of the eigenspace $\mathcal{F}(A)$ of a given $n \times n$ max-min matrix A can be described by investigating the structure of monotone eigenspaces $\mathcal{F}^{\leq}(A)$ and $\mathcal{F}^{<}(A)$.

Theorem 5.1. [10] *Let $A \in \mathcal{B}(n,n)$, $b \in \mathcal{B}(n,1)$ and $\varphi \in P_n$. Then $b \in \mathcal{F}(A)$ if and only if $b_\varphi \in \mathcal{F}(A_{\varphi\varphi})$.*

For $A \in \mathcal{B}(n,n)$, we define vectors $m^\star(A)$, $M^\star(A) \in \mathcal{B}(n,1)$ in the following way. For any $i \in N$, we put

$$m^{(i)}(A) := \max_{k>j} a_{jk} \qquad M^{(i)}(A) := \max_{k\geq j} a_{jk}$$
$$m_i^\star(A) := \max_{j\leq i} m^{(j)}(A) \qquad M_i^\star(A) := \min_{j\geq i} M^{(j)}(A)$$

Theorem 5.2. [10] *Let $A \in \mathcal{B}(n,n)$ and let $b \in \mathcal{B}(n,1)$ be a strictly increasing vector. Then $b \in \mathcal{F}(A)$ if and only if $m^\star(A) \leq b \leq M^\star(A)$. In formal notation we can write*

$$\mathcal{F}^{<}(A) = \langle m^\star(A), M^\star(A)\rangle \cap \mathcal{B}^{<}(n,1)$$

Based on this and similar theorems presented in [10], the structure of the eigenspace of a given matrix as a union of 'monotonicity'intervals can be completely described.

6. Simple image sets

The aim of this section is to describe the set consisting of all vectors with a unique pre-image (in short: the simple image set) of a given max-min linear mapping. We present a close connection of the simple image set with the eigenspace of the corresponding matrix (the set of all fixed points of the mapping). The topological aspects of the simple image set problem are described. The questions considered in this section are analogous to those in [3], where matrices and linear mappings in a max-plus algebra are studied.

For a square matrix $A \in \mathcal{B}(n,n)$ and for a permutation $\pi : N \to N$, we denote

$$S_A := \{\, b \in \mathcal{B}(n);\ (\exists! x \in \mathcal{B}(n))\, A \otimes x = b \,\}$$

$$\mathcal{F}_\pi(A) := \{\, b \in \mathcal{B}(n);\ A \otimes b_\pi = b \,\}$$

where b_π is created from b by permutation π. If π is the identity permutation on N, then we write simply $\mathcal{F}(A)$ instead of $\mathcal{F}_\pi(A)$. The set S_A is called the simple image set of the matrix A, in short: the simple image set.

The unique solvability of the equation $A \otimes x = b$ for a given vector $b \in S_A$ can be described as follows. A square matrix in a max-min algebra is strongly regular if the matrix represents a uniquely solvable system of linear equations, for some right-hand side vector. For $A \in \mathcal{B}(n,n)$, $b \in \mathcal{B}(n,1)$ we say that A is b-normal, if $a_{ii} \geq b_i$ and $b_i = \min\{\, b_j;\ a_{ji} > b_j \,\}$. Further, we say that A is generally trapezoidal if there is an increasing vector $b \in \mathcal{B}(n,1)$ such that A is b-normal and the system $A \otimes x = b$ is uniquely solvable.

Theorem 6.1. [12] *Let $A \in \mathcal{B}(n,n)$. The following statements are equivalent*

(i) *A is strongly regular*
(ii) *A can be permuted to a generally trapezoidal matrix, i.e. there are permutations φ on rows, and ψ on columns, such that the permuted matrix $A_{\varphi\psi}$ is generally trapezoidal*

Theorem 6.2. [14] *Let $A \in \mathcal{B}(n,n)$ be generally trapezoidal. Then $S_A \subseteq \mathcal{F}(A)$.*

Theorem 6.3. [14] *Let $A \in \mathcal{B}(n,n)$ be generally trapezoidal. Then A^2 is generally trapezoidal and $S_{A^2} = S_A$.*

The inclusion $S_A \subseteq \mathcal{F}(A)$ in Theorem 6.2 can be extended to the closure of S_A. We shall consider the ordered set $\mathcal{B}$ as a topological space with the interval topology, in which open intervals form a base of open sets. That means that every open set in $\mathcal{B}$ is a union of some set of open intervals. Further, the vector space $\mathcal{B}(n,1)$, for a fixed n, will be considered as a topological space with the product topology derived from the interval topology in $\mathcal{B}$.

Theorem 6.4. [14] *Let $A \in \mathcal{B}(n,n)$ be generally trapezoidal. Then* $\mathrm{cl}(S_A) \subseteq \mathcal{F}(A)$.

We may remark that, in general, the inclusion sign in Theorem 6.4 cannot be substituted by the sign of equality.

References

[1] R. A. Brualdi, H. J. Ryser, *Combinatorial Matrix Theory*, in: Encyclopaedia of Mathematics and its Applications, vol. 39, Cambridge Univ. Press, Cambridge, 1991.

[2] R. E. Burkard, B. Klinz, R. Rudolf, *Perspectives of Monge properties in optimization*, Discr. Appl. Mathem. **70** (1996), 95-161.

[3] P. Butkovič, *Simple image set of (max, +) linear mappings*, Discrete Appl. Math. **105** (2000), 73-86.

[4] K. Cechlárová, *Eigenvectors in bottleneck algebra*, Lin. Algebra Appl. **175** (1992), 63-73.

[5] K. Cechlárová, *Efficient computation of the greatest eigenvector in fuzzy algebra*, Tatra Mt. Math. Publ. **12** (1997), 73-79.

[6] G. Cohen, D. Dubois, J. P. Quadrat, M. Viot, *A linerar-system-theoretic view of discrete event processes and its use for performance evaluation in manufacturing*, IEE Transactions on Automatic Control, AC-30 (1985), 210-220.

[7] R. A. Cuninghame-Green, *Describing industrial processes with interference and approximating their steady-state behavior*, Oper. Res. Quart. **13** (1962), 95-100.

[8] R. A. Cuninghame-Green, *Minimax algebra*, Lecture Notes in Econom. and Math. Systems **166**, Springer-Verlag, Berlin, 1979.

[9] M. Gavalec, *Linear periods of Monge matrices in max-plus algebra*, Proc. of the 17th Conf. Mathem. Methods in Economics, Jindř. Hradec (1999), 85–94.

[10] M. Gavalec, *Monotone eigenspace structure in max-min algebra*, Lin. Algebra Appl. **345** (2002), 149–167.

[11] M. Gavalec, J. Plavka, *An $O(n^2)$ algorithm for maximum cycle mean of Monge matrices in max-algebra*, Discrete Appl. Math. **127** (2003), 651-656.

[12] M. Gavalec, J. Plavka, *Strong regularity of matrices in general max-min algebra*, Lin. Algebra Appl. **371** (2003), 241–254.

[13] M. Gavalec, J. Plavka, *Computing an eigenvector of a Monge matrix in max-plus algebra*, Math. Methods Oper. Res. **63** (2006), 543-551.

[14] M. Gavalec, J. Plavka, *Simple image set of linear mappings in a max-min algebra*, Discrete Appl. Math. **155** (2007), 611–622.

[15] M. Gavalec, J. Plavka, *Structure of the eigenspace of a Monge matrix in max-plus algebra*, Discrete Appl. Math. to appear.

[16] M. Gavalec, J. Plavka, *Eigenspace structure of a concave matrix*, Abstracts of the 15th Inter. Scient. Conf. on Mathematical Methods in Economics and Industry MMEI 2007, Herľany (Slovakia) (2007), p. 12–13.

[17] M. Gondran, *Valeurs propres et vecteurs propres en classification hiérarchique*, R. A. I. R. O., Informatique Théorique **10** (1976), 39-46.

[18] M. Gondran and M. Minoux, *Eigenvalues and eigenvectors in semimodules and their interpretation in graph theory*, Proc. 9th Prog. Symp. (1976), 133-148.

[19] M. Gondran and M. Minoux, *Valeurs propres et vecteurs propres en théorie des graphes*, Colloques Internationaux, C.N.R.S., Paris, 1978, 181-183.

[20] B. Heidergott, G.J. Olsder, J. van der Woude, *Max Plus at Work*, Princeton University Press, Princeton, New Jersey, 2006.

[21] R.M. Karp, *A characterization of the minimum cycle mean in a digraph*, Discrete Math. **23** (1978), 309–311.

[22] E.L. Lawler, *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehart and Winston, New York, 1976.

[23] G. Olsder, *Eigenvalues of dynamic max-min systems*, in Discrete Events Dynamic Systems 1, Kluwer Academic Publishers, 1991, 177-201.

[24] J. Plavka, *Eigenproblem for circulant matrices in max-algebra*, Optimization **50** (2001), 477–483.

[25] J. Plavka, M. Gavalec, *Multiparametric eigenproblem in max-plus algebra*, Abstracts of the 22nd European Conf. on Operational Research EURO 2007, Prague (2007), p. 201.

[26] E. Sanchez, *Resolution of eigen fuzzy sets equations*, Fuzzy Sets and Systems **1** (1978), 69-74.
[27] Yi-Jia Tan, *Eigenvalues and eigenvectors for matrices over distributive lattices*, Lin. Algebra Appl. **283** (1998), 257-272.
[28] U. Zimmermann, *Linear and Combinatorial Optimization in Ordered Algebraic Structures*, North Holland, Amsterdam, 1981.

1. Department of Information Technologies, Faculty of Informatics and Management, University Hradec Králové, Rokitanského 62, 50003 Hradec Králové, Czech Republic
E-mail address: martin.gavalec@uhk.cz

2. Department of Mathematics, Faculty of Electrical Engineering and Informatics, Technical University in Košice, B. Němcovej 32, 04200 Košice, Slovakia
E-mail address: jan.plavka@tuke.sk

Stability of Approximation Algorithms or Parameterization of the Approximation Ratio*

Hans-Joachim Böckenhauer Juraj Hromkovič

Department of Computer Science, ETH Zurich, Switzerland
{hjb,juraj.hromkovic}@inf.ethz.ch

Abstract

The investigation of the possibility to efficiently compute approximate solutions to instances of hard optimization problems is one of the central and most fruitful areas of current algorithm and complexity theory. One tool for investigating the tractability of optimization problems is the concept of stability of approximation algorithms. The key idea behind this concept is to parameterize the set of the instances of an optimization problem and to look for a polynomial-time achievable approximation ratio with respect to this parameterization. Whenever the approximation ratio grows with the parameter, but is independent of the size of the input instances, we speak of *stable* approximation algorithms.

It has been shown that there exist stable approximation algorithms for problems like TSP which are not approximable within any polynomial approximation ratio in polynomial time (assuming P is not equal to NP). Investigating the stability of approximation in this way overcomes the trouble with measuring the approximation ratio in a worst-case manner since it may succeed in partitioning the set of all input instances of a hard problem into infinitely many classes with respect to their approximation hardness.

We believe that approaches like this will become the core of the algorithmics, because they provide a deeper insight into the hardness of specific problems and in many applications we are not interested in the worst-case problem hardness, but in the hardness of actual problem instances.

Keywords: stability of approximation, approximation algorithms, parameterization

1 Introduction

The design of approximation algorithms has proven to be one of the most successful approaches to solving hard optimization problems. Nevertheless, there exist many problems which, under some standard complexity-theoretic assumptions like P $\neq$ NP, do not admit a polynomial-time algorithm computing an arbitrarily good approximation ratio on all input instances. For an overview of approximation algorithms and the theory of inapproximability, see for instance [Hr03, Va03, ACG$^+$99, Go07, MPS98].

Another approach for attacking hard problems is to look for easier subproblems, i.e., for a subset of input instances on which the problem appears to be solvable more

*This work was partially supported by SNF grant 200021-109252/1.

easily. One of the most remarkable examples for this is the traveling salesman problem (TSP), i.e., the problem of finding a minimum-cost Hamiltonian tour in a complete edge-weighted graph. The TSP is not approximable with any polynomial approximation ratio in its general formulation (see e.g. [Hr03]), unless P = NP, but admits for a 3/2-approximation when restricted to metric inputs [Chr76], i.e., to inputs where the cost function c on the edges satisfies the triangle inequality $c(\{u,v\}) \leq c(\{u,w\}) + c(\{w,v\})$ for all vertices u, v, and w.

In other words, we have identified a core of the problem which is much easier to approximate. Our goal is now to partition the set of all input instances into infinitely many subclasses, starting from this core, using a parameter measuring the distance of a particular instance from the core, i.e., in our TSP example, measuring how far away the instance is from being metric, and then to design an approximation algorithm for general instances whose approximation ratio grows with this parameter.

More generally speaking, the concept of *stability* of approximation algorithms deals with the following scenario. We are given an optimization problem P for two sets of inputs L_1 and L_2 where $L_1 \subset L_2$. For P on inputs from L_1 there exists a polynomial-time α-approximation algorithm A, but there exists no constant-factor approximation which works for all inputs from L_2, unless P = NP. Now we pose the question if the usefulness of algorithm A is really restricted to inputs from L_1. We consider a metric distance function *dist* on L_2 measuring the distance between any two input instances from L_2. If we now look at an input instance $x \in L_2 - L_1$ for which there exists some $y \in L_1$ such that $\mathit{dist}(x,y) \leq k$ holds for some positive number k, we can ask how good the algorithm A performs on x. If, for every $k > 0$ and every x with distance at most k to L_1, A computes a $\delta_{\alpha,k}$-approximation of an optimal solution for x (where $\delta_{\alpha,k}$ is considered to be a constant depending on k and α only, but not on the size of the input x), we say that algorithm A is *(approximation) stable* according to *dist*.

This approach is similar to the concept of parameterized complexity which was introduced by Downey and Fellows [DF95, DF99]. Both approaches strive to overcome the usual worst-case analysis of complexity or approximation ratio by partitioning the set of inputs into a hierarchy of infinitely many classes according to some parameterization. While the parameterized complexity tries to measure the time complexity of an exact exponential-time algorithm in terms of the parameter, the stability approach does the same for the approximation ratio of a polynomial-time approximation algorithm.

We believe that approaches like these will be at the heart of future algorithmics since they provide some deeper insight into the hardness of problems and allow us to measure more precisely the hardness of actually forthcoming problem instances, which in many applications is more relevant than the worst-case problem hardness.

2 Definition of Approximation Stability

In this section, we give a short definition of the concept of approximation stability. For a more comprehensive overview, see [BHK+02, BHS07, Hr03].

We start with an extended version of the standard definition of an optimization problem. An *optimization problem* U can be represented as $U = (L, L_I, \mathcal{M}, \mathit{cost}, \mathit{goal})$, where L is the set of all *feasible input instances*, $L_I \subseteq L$ is the set of *actually considered inputs*, $\mathcal{M}$ is a function describing the *set of feasible solutions* for each input, *cost* is a *cost function* assigning a cost to every feasible solution of an input instance, and

$goal \in \{\min, \max\}$ is the *optimization goal*.

For every $x \in L$, let $Opt_U(x)$ denote the cost of an optimal solution for the problem U on input x.

An algorithm A is called *consistent* for U, if it computes, for every $x \in L_I$, a feasible solution $y \in \mathcal{M}(x)$. The *time complexity* of A is defined as $Time_A(n) = \max\{Time_A(x) \mid x \in L_I \text{ and } |x| = n\}$, where $n \in \mathbb{N}$ and $Time_A(x)$ is the length of the computation of A on x.

The *approximation ratio* $R_A(x)$ of A on x is defined as $R_A(x) = \frac{cost(A(x))}{Opt_U(x)}$ if $goal = \min$, and as $R_A(x) = \frac{Opt_U(x)}{cost(A(x))}$ if $goal = \max$. For any input size n, we define $R_A(n) = \max\{R_A(x) \mid x \in L_I \text{ and } |x| = n\}$. If, for some $\delta > 1$, $R_A(x) < \delta$ for all $x \in L_I$, we call A a *δ-approximation algorithm.* Analogously, if, for some function $f : \mathbb{N} \to \mathbb{R}^+$, $R_A(n) < f(n)$ for all $n \in \mathbb{N}$, we call A an *$f(n)$-approximation algorithm.*

We now define how to measure the distance between input instances in L. Let $\overline{U} = (L, L, \mathcal{M}, cost, goal)$ be an optimization problem, let $U = (L, L_I, \mathcal{M}, cost, goal)$, where $L_I \subsetneq L$, be a subproblem of $\overline{U}$. Any function $h_L : L \to \mathbb{R}^+$ satisfying $h_L(x) = 0$ for all $x \in L_I$ is called a *distance function for U according to L_I*. For any $r \in \mathbb{R}^+$, we define $Ball_{r,h}(L_I) = \{w \in L \mid h(w) \leq r\}$ to be the set of instances from L which are at distance at most r from L_I.

This definition now enables us to formally define stable approximation algorithms. We consider an ε-approximation algorithm A for U for some $\varepsilon \in \mathbb{R}^+$ which is consistent for $\overline{U}$ and some $p \in \mathbb{R}^+$. We say that A is *p-stable according to h* if, for every real number $0 \leq r \leq p$, there exists some $\delta_{r,\varepsilon} \in \mathbb{R}^+$ such that A is a $\delta_{r,\varepsilon}$-approximation algorithm on the subproblem of $\overline{U}$ restricted to the instances in $Ball_{r,h}(L_I)$. The algorithm A is called *stable according to h* if it is p-stable according to h for every $p \in \mathbb{R}^+$. If A is not p-stable for any $p > 0$, we call it *unstable*.

3 Stability of TSP Algorithms

To illustrate the usefulness of the concept of approximation stability, we present some results on the stability of TSP algorithms. The best known approximation algorithm for the metric TSP is Christofides' algorithm [Chr76] which achieves an approximation ratio of $\frac{3}{2}$. This algorithm computes a minimum spanning tree T on the input graph G and a minimum-cost perfect matching M on the odd-degree vertices of the tree. The tree and the matching together form an Eulerian graph, and the algorithm computes an Eulerian tour D on $T \cup M$. Then it constructs a Hamiltonian tour H in G by removing all repetitions of vertices in D.

Shortening the Eulerian tour to a Hamiltonian tour in the last step of Christofides' algorithm heavily depends on the triangle inequality which makes it possible to substitute a subpath of D by a direct edge without increasing the cost. This means that, if we try to apply Christofides' algorithm to a broader class of inputs, we should look for inputs "almost" satisfying the triangle inequality.

The following two examples show that the choice of the distance measure is indeed crucial. For some TSP instance $X = (G_X, c_X)$ where $G_X = (V_X, E_X)$ is a complete graph with edge weight function c_X, let

$$distance(x) = \max\left\{0, \max\left\{\frac{c_X(\{u,v\})}{\sum_{i=1}^{m} c_X(\{p_i, p_{i+1}\})} - 1 \,\middle|\, u, v \in V_X, u = p_1, v = p_{m+1}, \right.\right.$$
$$\left.\left. \text{and } p_1, p_2, \ldots, p_{m+1} \text{ is a simple path between } u \text{ and } v \text{ in } G_X\right\}\right\}.$$

Intuitively speaking, in any instance X with distance at most r from the metric TSP, the cost of any edge is bounded by $(1+r)$ times the length of the longest simple path between its endvertices.

Theorem 1 ([BHK⁺02]) *Christofides' algorithm is a stable approximation algorithm for TSP according to distance.*

Theorem 1 shows that there exists a distance measure for which Christofides' algorithm is stable. This means that this algorithm can be useful for a much larger class of input instances. Nevertheless, the measure *distance* imposes a hard requirement on any graph since the cost of each edge has to be compared to the cost of all simple paths. A more desirable distance measure is the following. Let, for any TSP instance X,

$$dist(X) = \max\left\{0, \max\left\{\frac{weight(\{u,v\})}{weight(\{u,w\}) + weight(\{w,v\})} - 1 \,\middle|\, u, v, w \in V_X\right\}\right\}$$

In other words, an input instance X has distance $dist(X) = r$ from the metric TSP if, for all vertices u, v, and w, it satisfies the following *relaxed triangle inequality*

$$c_X(\{u,v\}) \leq (1+r) \cdot (c_X(\{u,w\}) + c_X(w,v\})).$$

Unfortunately, according to this more natural distance measure, Christofides' algorithm does not behave so nicely.

Theorem 2 ([BHK⁺02]) *Christofides' algorithm is unstable for TSP according to dist.*

The next question which naturally arises is whether one can modify Christofides' algorithm as to get a stable algorithm for TSP also according to *dist*. As we will see in the following, the answer to this question is positive.

The main problem in Christofides' algorithm is that for constructing the Hamiltonian tour from the Eulerian tour, paths of unbounded length have to be shortcut by single edges. According to *dist*, shortening a path of m edges to a single edge may increase the cost by a factor of $(1+r)^{\lceil \log_2 m \rceil}$. In [BHK⁺02], it was shown that by constructing a path matching, i.e., a set of paths connecting the odd-degree vertices of the spanning tree, instead of the matching it is possible to arrange the construction of the Hamiltonian tour in such a way that only paths of length 4 will be shortened to single edges. Other stable algorithms for TSP were presented in [AB95, BCh99, And01]. A combination of these algorithms yields the following result.

Theorem 3 *There exists a stable approximation algorithm for the TSP which, for instances X at distance r from the metric TSP according to dist, achieves an approximation ratio of* $\min\{\frac{3}{2}(1+r)^2, (1+r)^2 + (1+r), 4(1+r)\}$.

In [FHP⁺04], the ideas underlying these algorithms were developed further for different path variants of TSP where a minimum-cost Hamiltonian path is searched for instead of a Hamiltonian cycle.

4 Lower Bounds

In the previous section, we have seen that there are stable approximation algorithms for TSP with an approximation ratio which is linear in the distance from the metric. But there are also some lower bounds known on the approximability of TSP on instances satisfying a relaxed triangle inequality. If a TSP instance X has distance at most r from the metric TSP according to the distance measure *dist*, we also say that X satisfies the $(1 + r)$-triangle inequality.

In [BS00], an explicit lower bound on the approximability of TSP restricted to instances satisfying a relaxed triangle inequality was shown.

Theorem 4 ([BS00]) *Unless* P = NP*, there is no polynomial-time α-approximation algorithm for the TSP subproblem restricted to instances satisfying the β-triangle inequality, for some $\beta > 1$, if*

$$\alpha < \frac{3803 + 10\beta}{3804 + 8\beta}.$$

This lower bound tends to $\frac{5}{4}$ for β tending to infinity. A linear lower bound of $1 + \varepsilon\beta$ on the approximation ratio was shown in [BCh99] for some very small $\varepsilon > 0$ not given explicitly in the proof.

An even stronger result was shown in [BHK$^+$07] for a generalization of TSP, where we assume a starting point for the TSP tour and where one of the vertices carries some deadline and has to be visited before this deadline by any feasible tour. This problem was shown not to admit even an $o(|V|)$-approximation in polynomial time (where $|V|$ denotes the number of vertices), unless P = NP, for the TSP on instances satisfying the β-triangle inequality for some $\beta > 1$.

5 Conclusion

In this paper we have presented the concept of approximation stability and we have illustrated it using TSP as an example. To summarize the potential applicability and usefulness of this concept, we observe that it can be used on the one hand to derive positive results like the following: An approximation algorithm (or even a PTAS) can be successfully used for a larger set of inputs than usually considered, or some simple modification of such algorithm allows its application on a significantly larger class of inputs. On the other hand, also proving that an algorithm is unstable for all reasonable distance measures, and thus only applicable for the originally considered set of input instances, can help us to search for a spectrum of the hardness of a problem according to some parameterization of the inputs. Results of this kind can essentially contribute to the study of the nature of hardness of specific problems.

References

[And01] T. Andreae: On the traveling salesman problem restricted to inputs satisfying a relaxed triangle inequality. *Networks* 38 (2001), pp. 59–67.

[AB95] T. Andreae, H.-J. Bandelt: Performance guarantees for approximation algorithms depending on parametrized triangle inequalities. *SIAM Journal on Discrete Mathematics* 8 (1995), pp. 1–16.

[ACG+99] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, M. Protasi: *Complexity and Approximation — Combinatorial Optimization Problems and Their Approximability Properties*. Springer 1999.

[BCh99] M. Bender, C. Chekuri: Performance guarantees for TSP with a parametrized triangle inequality. *Proc. of the Sixth International Workshop on Algorithms and Data Structures (WADS 99)*, LNCS 1663, Springer 1999, pp. 1–16.

[BHK+02] H.-J. Böckenhauer, J. Hromkovič, R. Klasing, S. Seibert, W. Unger: Towards the notion of stability of approximation for hard optimization tasks and the traveling salesman problem. *Theoretical Computer Science* 285 (2002), pp. 3–24.

[BHK+07] H.-J. Böckenhauer, J. Hromkovič, J. Kneis, J. Kupke: The parameterized approximability of TSP with deadlines. *Theory of Computing Systems*, to appear.

[BHS07] H.-J. Böckenhauer, J. Hromkovič, S. Seibert: Stability of approximation. In: T. F. Gonzalez (ed.): *Handbook of Approximation Algorithms and Metaheuristics*, Chapman & Hall/CRC, 2007, Chapter 31.

[BS00] H.-J. Böckenhauer, S. Seibert: Improved lower bounds on the approximability of the traveling salesman problem. *RAIRO Theoretical Informatics and Applications* 34 (2000), pp. 213–255.

[Chr76] N. Christofides: Worst-case analysis of a new heuristic for the travelling salesman problem. Technical Report 388, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, 1976.

[DF95] R. G. Downey, M. R. Fellows: Fixed-parameter tractability and completeness I: Basic Results. *SIAM Journal of Computing* 24 (1995), pp. 873–921.

[DF99] R. G. Downey, M. R. Fellows: *Parametrized Complexity*. Springer 1999.

[FHP+04] L. Forlizzi, J. Hromkovič, G. Proietti, S. Seibert: On the stability of approximation for Hamiltonian path problems. *Proc. SOFSEM 2005*, Springer 2005.

[Go07] T. F. Gonzalez (ed.): *Handbook of Approximation Algorithms and Metaheuristics*. Chapman & Hall/CRC, 2007.

[Hr03] J. Hromkovič: *Algorithmics for Hard Problems. Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics*. Springer 2003.

[MPS98] E. W. Mayr, H. J. Prömel, A. Steger (eds.): *Lecture on Proof Verification and Approximation Algorithms*. Lecture Notes in Computer Science 1967, Springer 1998.

[Va03] V. V. Vazirani: *Approximation Algorithms*. Springer 2003.

INTERIOR POINT METHODS: WHAT HAS BEEN DONE IN LAST 20 YEARS

Janez Povh
University in Maribor, Faculty of logistics
email: janez.povh@uni-mb.si

Abstract
We provide a synopsis of the development in interior point methods for linear programming after Karmarkar [5] in 1984 (re)discovered how important this topic is. We also mention the main extensions of these methods to areas from nonlinear programming.

Keywords: interior point methods, central path, linear programming, semidefinite programming.

1. INTRODUCTION

In 1984, Karmarkar [5] excited the mathematical programming community with the method for linear programming, which seemed promising to outperform the simplex method not only in the theory, but also in practise. Fifteen years later Feund and Mizuno [2] wrote: "Interior-point methods in mathematical programming have been the largest and most dramatic area of research in optimization since the development of the simplex method... Linear programming is no longer synonymous with the celebrated simplex method, and many researchers now tend to view linear programming more as special case of nonlinear programming due to these developments." By the year 1997 there have been published more than 3000 papers on the topic of interior point methods [3].

The purpose of this paper is to give an overview of the major results about interior point methods for linear programming, obtained after the premium result of Karmarkar. At the end we give a short introduction to IPM area for nonlinear programming.

We review the definition of linear programming problem and some important results from duality theory in Section 2. In Section 3 we explain the basic idea of the interior point methods, and describe primal-dual path-following IPM, potential reduction IPM and affine scaling IPM. In Section 4 we give a quick overview of IPM results in nonlinear programming.

Notation: For $x \in \mathbb{R}^n$ we denote by $x \geq (>)0$ that $x_i \geq (>)0, \forall i$. Given an optimization problem P, we denote by OPT_P its optimal value. For a vector x we denote by $x^{-1} = (1/x_1, \ldots, 1/x_n)^T$.

2. Linear programming

2.1 Definitions

A linear programming problem in the standard primal form is an optimization problem of the following type:

$$(PLP) \quad \inf\ c^T x \ \text{ such that } Ax = b, \quad x \geq 0\,,$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Its dual linear program is

$$(DLP) \quad \sup\ b^T y \ \text{ such that } A^T y + s = c, \quad y \ldots \text{free}, s \geq 0\,.$$

Let us denote the feasible set for PLP by $\mathcal{P}$ and for DLP by $\mathcal{D}$. Therefore

$$\mathcal{P} = \{x \in \mathbb{R}^n \colon x \geq 0,\ Ax = b\} \ \text{ and } \ \mathcal{D} = \{(y, s) \in \mathbb{R}^m \times \mathbb{R}^n \colon s \geq 0,\ A^T y + s = c\}.$$

The interiors of $\mathcal{P}$ and $\mathcal{D}$ are the sets of strictly feasible solutions, i.e.

$$\mathcal{P}^+ = \{x \in \mathbb{R}^n : x > 0,\ Ax = b\} \text{ and } \mathcal{D}^+ = \{(y,s) \in \mathbb{R}^m \times \mathbb{R}^n : s > 0,\ A^T y + s = c\}.$$

2.2 Duality theory

The duality theory for linear programming is very rich. We expose only the most important results:

- **Weak duality:** if DLP is not infeasible, then any dual feasible solution gives a lower bound for the optimal value of PLP, i.e. $b^T y \leq c^T x$, for any $(y,s) \in \mathcal{D},\ x \in \mathcal{P}$. The nonnegative quantity $c^T x - b^T y = x^T s$ is called *duality gap.*
- **Strong duality:** If both problems PLP and DLP are feasible, then $OPT_{PLP} = OPT_{DLP}$ and both optimums are attained (we can replace inf by min and sup by max).
- **Attainability:** If one of the problems is infeasible then the other is unbounded or infeasible.
- **Complementary slackness:** If x and (y,s) are optimal solutions of PLP and DLP, resp., then $x_i s_i = 0$, for $1 \leq i \leq n$.

By setting $\inf \emptyset = \infty$ and $\sup \emptyset = -\infty$ we have $OPT_{PLP} = OPT_{DLP}$, provided at least one of the problems is feasible.

2.3 Optimality conditions

Karush-Kuhn-Tucker (KKT) optimality conditions are in general only necessary optimality conditions. For the case of convex optimization, which includes linear programming, these conditions are also sufficient. They can be derived straightforwardly from the strong duality property. A pair $x \in \mathbb{R}^n$ and $(y,s) \in \mathbb{R}^m \times \mathbb{R}^n$ is optimal for PLP and DLP, resp., if and only if it satisfies the following constraints

$$\left.\begin{array}{lrcllcl} \text{(primal feasibility)} & Ax & = & b, & x & \geq & 0 \\ \text{(dual feasibility)} & A^T y + s & = & c, & s & \geq & 0 \\ \text{(zero duality gap)} & x^T s & = & 0. & & & \end{array}\right\} \quad (1)$$

Solving PLP and DLP to optimality therefore amounts to solving this system of (nonlinear) equations.

3. Interior point methods

3.1 Introduction and assumptions

Since the Karmarkar's breakthrough there has been a wide range of interior point methods developed. The main property of interior point methods (IPM) is that these methods generate iterates of solutions which are asymptotically approaching to an optimum solution and stop as soon as they come close enough to an optimum. These iterates may not be (at least at the beginning) feasible for the linear equations. In this case their infeasibility is decreasing by approaching to an optimal solution.

We can classify these methods according to the following properties [3],[4]:

- **Iterate space:** A method is said to be primal, dual or primal-dual when its iterates belong respectively to the primal space, the dual space or the Cartesian product of these spaces.

- **Type of iterate:** A method is said to be feasible when its iterates are feasible. In the case of an infeasible method, the iterates need not satisfy the equality constraints, but are still required to satisfy the nonnegativity(strict positivity) conditions.

- **Type of algorithm:** This is the main difference between the methods. We distinguish path-following algorithms, affine-scaling algorithms and potential reduction algorithms. We will pay most of the attention to the path-following algorithms.

- **Type of step:** In order to preserve their polynomial complexity, some algorithms are obliged to take very small steps at each iteration, leading to a high total number of iterations when applied to practical problems. These methods are called short-step methods and are mainly of theoretical interest. Therefore long-step methods, which are allowed to take much longer steps, have been developed and are the only methods used in practice.

Interior point methods need the following assumptions.

Assumption 1 *The matrix A has linearly independent rows.*

Assumption 2 *Problems (2.1) and (2.1) are strictly feasible, i.e. $\mathcal{P}^+ \neq \emptyset$ and $\mathcal{D}^+ \neq \emptyset$.*

The first assumption implies one-to-one correspondence between the y and the s variables in the $\mathcal{D}$ and is also important for the numerical reasons, see Subsection 3.2.1.

3.2 Path following interior point methods

3.2.1 Central path

Solving PLP and DLP is equivalent to finding a pair x and (y, s) that satisfies the optimality conditions (1). This conditions contain nonlinear equation due to the last constraint. The idea underlaying the path following IPM is to replace optimality condition (1) by the following set of constraints:

$$\left.\begin{array}{llcllcl} \text{(strict primal feasibility)} & Ax & = & b, & x & > & 0 \\ \text{(dual feasibility)} & A^T y + s & = & c, & s & > & 0 \\ \text{(centrality constraint)} & x_i s_i & = & \mu\,,\ \forall i, & & & \end{array}\right\} \qquad (2)$$

where $\mu > 0$. Under Assumptions 1 and 1 has system (2) unique optimal solution. The set of solutions $\{(x_\mu; y_\mu, s_\mu) \colon \mu > 0\}$ is called *central path*, which is an analytic curve and always converges to the analytic center of the optimal face [11].

The main idea of the primal-dual path following methods is to apply the Newton method (see any textbook from nonlinear programming) to the system of non-linear equations (2). This means that if we have a current iterate $(x; y, s)$ then we have to solve the following system of linear equations:

$$A(x + \Delta x) = b \qquad (3)$$

$$A^T(y + \Delta y) + s + \Delta s = c \qquad (4)$$

$$x_i s_i + \Delta x_i s_i + x_i \Delta s_i = \mu, \quad 1 \leq i \leq n\,,\ \forall i, \qquad (5)$$

in variables $(\Delta x, \Delta y, \Delta s)$.

We can eliminate Δs using (4) and Δx using (5). Substituting Δx in (3) we get the system $M\Delta y = m$, where $M = ADA^T$ and D is a diagonal matrix with $d_{ii} = x_i/s_i > 0$. The right hand side m is defined by $m = AD(c - \mu x^{-1} - A^T y) + b - Ax$. The matrix M is positive definite if Assumption 1 holds, therefore system (3)–(5) has a unique solution, obtained by $\mathcal{O}(n^3)$ arithmetic operations.

Typically we perform only few steps of Newton method and before we reach the point on the central path, corresponding to the current μ, we decrease μ and repeat the procedure. Therefore the iterates generated this way do not lie on the central path but in some neighborhood which is getting very tiny as we approach to the limit point.

The conceptual algorithm for the primal dual path-following method is in Figure 1.

INPUT: $A, b, c, \mu_0 > 0, (x_0; y_0, s_0) \in \mathcal{P}^+ \times \mathcal{D}^+$.

1. *Set* $k = 0$.
2. ***Repeat***
   - 2.1 $k := k + 1$. *Choose* $0 < \mu_k < \mu_{k-1}$.
   - 2.2 *Approximately solve system (2) with* $\mu = \mu_k$ *to obtain new iterate* $(x_{k+1}; y_{k+1}, s_{k+1})$.
3. ***Until:*** $\|Ax_{k+1} - b\|$, $\|A^T y_{k+1} + s_{k+1} - c\|$ *and* $x_{k+1}^T s_{k+1}$ *are small enough.*

Figure 1: Primal-dual path following algorithm

3.2.2 Short-step primal-dual path-following algorithm

This variant of IPM performs step 2.2 from Figure 1 such that it executes one single step of Newton method, i.e. it solves system (3)–(5) only once. In equation (5) it employs $\mu = \sigma\mu_k$, for $\mu_k = x_k^T s_k/n$ and some $\sigma \in (0, 1)$.

If we choose $\sigma = 1 - 0.4/n$ and if the starting point $(x_0; y_0, s_0) \in \mathcal{P}^+ \times \mathcal{D}^+$ is close enough to the central path then this method computes a feasible point $(x_k; y_k, s_k) \in \mathcal{P}^+ \times \mathcal{D}^+$ with $x_k^T s_k/n \le \varepsilon$ in $\mathcal{O}(\sqrt{n} \log \frac{\mu_0}{\varepsilon})$ iterations [11].

It has been noticed that this method has poor practical performance due to small reduction of μ (σ is very close to 1). To improve the practical behavior we allow stronger reductions of μ leading to methods, described in the following subsection.

3.2.3 Long-step primal-dual path-following algorithm

This version of the primal-dual path-following IPM tends to overcome the main limitation of the short-step methods: small step size. Here we still decrease μ_k to $\sigma\mu_k$, but σ is much smaller. It may happen that the new iterate $(x_k + \Delta x; y_k + \Delta y, s_k + \Delta s)$ is no longer strictly feasible, causing breaking down the method. Therefore we must perform the so-called damped Newton step: we search for the largest α_k such that $(x_k + \alpha_k\Delta x; y_k + \alpha_k\Delta y, s_k + \alpha_k\Delta s)$ is strictly feasible. It may also happen that the new iterate is too far from the central path. In this case we perform more than one Newton step with the same value of μ_k.

3.2.4 The Mehrotra predictor-corrector algorithm

The description of the methods from the previous section has underlined the fact that the constant σ, defining the new value of μ, has a very important role in determining the algorithm efficiency. Choosing σ nearly equal to 1 allows us to take a full Newton step, but this step is usually very short and does not make much progress towards the solution. However it has the advantage of increasing the proximity to the central path. On the other hand, choosing a smaller σ produces a larger Newton step making more progress towards optimality, but this step is generally infeasible and has to be damped. Moreover this kind of step usually tends to move the iterate away from the central path.

Mehrota [6] proposed the following a strategy how to balance between these two goals. We decompose the update into two parts. The first part is the step towards the global optimum (the predictor step), which we obtain by solving (3)–(5) with $\mu = 0$ and determining the step length by a line search. This way we obtain new (predictor) iterate $(x^p; y^p, s^p)$. We compute new $\mu^p = (x^p)^T s^p/n$ and compute the second part of the update (the corrector part) by solving (3)–(5), where we use the predictor iterate $(x^p; y^p, s^p)$ and $\mu_{new} = \sigma\mu$. Mehrota suggested to take $\sigma = (\mu^p/\mu)^3$. The new iterate is obtained by separate line search on the primal and the dual side. This version of the primal-dual path following IPM has quadratic convergence [11].

3.3 Potential-reduction methods

Instead of targeting a point on the central with smaller duality gap, the method of Karmarkar [5] made use of a potential function to monitor the progress of its iterates. A potential function is a way to measure the quality of an iterate. Its main two properties are the following [4]: it should tend to $-\infty$ if and only if the iterates tend to optimality and it should tend to $+\infty$ when the iterates tend to the boundary of the feasible region without tending to an optimal solution. The main goal of a potential reduction algorithm is simply to reduce the potential function by a fixed amount δ at each step, hence its name. Convergence follows directly from the structure of the potential function. The most famous potential function is so-called Tanabe-Todd-Ye potential function:

$$\Phi_\rho(x, s) = \rho \log(x^T s) - \sum_i \log(x_i s_i),$$

where ρ is a constant required to be greater than n.

This method proceeds similarly to path following methods. It also solves system (2), but the step length is defined such that the potential function is minimized in this direction. We reach ε-optimal solution again in $\mathcal{O}(\sqrt{n} \log \frac{\mu_0}{\varepsilon})$ iterations.

3.4 Affine-scaling methods

Affine scaling methods were motivated my Karmarkar [5] who used projective transformations in his method. Affine-scaling algorithms do not explicitly follow the central path and do not even refer to it. The main idea is not to optimize over the polyhedron but to optimize over the inscribed ellipsoid centered at the current iterate x_k, which should be easier than on a polyhedron, and take this optimum as the next iterate. For more details see [4, 9, 10].

4. Extensions to general cone programming

The main results about IPM for linear programming were obtained in first 10 years after Karmarkar's result. In the first half of 1990s many of these methods were extended to convex quadratic programming and semidefinite programming [11, 1]. Nesterov and Nemirovskii [7] followed a more general approach. They showed that an arbitrary linear problem over a convex set can be solved to a precision ε with an IPM, which needs polynomially many iterations, if we have a self-concordant barrier function for the convex set. This is a smooth convex function with second derivatives which are Lipschitz continuous with respect to a local metric (the metric induced by the Hessian of the function itself) - for more details see also [8, Section 2.2].

In particular, they showed that for the set (cone) of nonnegative vectors $\mathbb{R}^n_+$ and the positive semidefinite cone $\mathcal{S}_n^+$ we have an easy to compute self-concordant barrier function (for $\mathbb{R}^n_+$ this is the function $\sum_i \log x_i$ and for $\mathcal{S}_n^+$ this is $-\log\det X$). This results nicely describe previous results and gave a framework to search for new IPM. We also point out that efficient IPM, designed for semidefinite programming, were the main motivator for the incredible extensive research of this part of the convex optimization.

5. Conclusions

In this paper we give and overview of the main results about interior point methods, obtained after the breakthrough of Karmarkar [5]. We focus on linear programming, but at the end we also consider extensions to nonlinear programming, where IPM were actually initiated and were after a big success in linear programming rediscovered for convex quadratic and semidefinite programming.

References

[1] E. de Klerk. *Aspects of semidefinite programming*, volume 65 of *Applied Optimization*. Kluwer Academic Publishers, Dordrecht, 2002. Interior point algorithms and selected applications.

[2] R. M. Freund and S. Mizuno. Interior point methods: current status and future directions. In *High performance optimization*, volume 33 of *Appl. Optim.*, pages 441–466. Kluwer Acad. Publ., Dordrecht, 2000.

[3] F. Glineur. Interior-point methods for linear programming: a guided tour. *Belg. J. Oper. Res. Statist. Comput. Sci.*, 38(1):3–30 (1999), 1998.

[4] F. Glineur. *Topics in convex optimization: interior-point methods, conic duality and approximations.* PhD thesis, Faculté Polytechnique de Mons,(Mons), 2001.

[5] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[6] S. Mehrotra. On the implementation of a primal-dual interior point method. *SIAM J. Optim.*, 2(4):575–601, 1992.

[7] Y. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994.

[8] J. Renegar. *A mathematical view of interior-point methods in convex optimization.* MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[9] R. Saigal. *Linear programming.* International Series in Operations Research & Management Science, 1. Kluwer Academic Publishers, Boston, MA, 1995. A modern integrated analysis.

[10] T. Tsuchiya. Affine scaling algorithm. In *Interior point methods of mathematical programming*, volume 5 of *Appl. Optim.*, pages 35–82. Kluwer Acad. Publ., Dordrecht, 1996.

[11] S. J. Wright. *Primal-dual interior-point methods.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

Virtual private network design*

Leen Stougie†

Abstract

Virtual Private Network (VPN) design is the problem that emerges if a set of users wishes to select a set of routes on a communication network for internal communication amongst them and rent enough capacity on this network, such as to support any possible internal communication demand scenario within the limits specified by each of the users. The model of the VPN problem was proposed for the first time by Fingerhut et al. [6], and later independently by Duffield et al. [3]. There are several variations of the problem based on wether or not there is symmetry in the demand patterns, in the unit capacity costs, and in the routing. Over the last ten years, the problem has drawn attention of many researchers, because of its simple formulation and its intriguing basic open questions, of which an outstanding one still resisted the many attempts to solve it.

In this lecture I will give a survey of the problem, of its variations and present a selection of the results obtained. Of course, I will describe the open problem in the form of what has become known as the “Tree Conjecture for the VPN problem”.

1 The VPN-problem

In this survey, we consider a problem that is known as the *virtual private network* (VPN) problem, a problem emerging in telecommunication. We will start by describing the so-called symmetric VPN problem. Think of a large communication network represented by an undirected graph $G = (V, E)$, with a vertex for each user and an edge for each link in the network. Within this network, a subgroup $W \subseteq V$ of the users wishes to reserve capacity on the links of the network for communication among themselves: they wish to establish a virtual private network. Vertices in W are also called *terminals.*

On each link, capacity (bandwidth) has a certain price per unit, $c : E \to \mathbb{R}_+$. The problem is to select one or more communication paths between every pair $\{i, j\}$ of users in W and to reserve enough capacity on the edges of the selected paths to accommodate any possible communication pattern amongst the users in W. Possible communication patterns are defined through an upper bound on the amount to be communicated (transmitted and received) for each node in W, specified by $b : W \to \mathbb{R}_+$. More precisely, a *communication scenario* for the symmetric VPN problem can be defined as a symmetric matrix $D = (d_{ij})_{\{i,j\}\subseteq W}$ with zeros on the diagonal, specifying for each *unordered* pair of distinct nodes $\{i, j\} \subseteq W$ the amount of communication $d_{ij} \geq 0$ between i and j. A communication scenario $D = (d_{ij})_{\{i,j\}\subseteq W}$ is said to be *valid* if $\sum_{j\in W\setminus\{i\}} d_{ij} \leq b(i),\ \forall i \in W$. We denote the collection of valid communication scenarios by $\mathcal{D}$.

A network consisting of the selected communication paths with enough capacity reserved on the edges to accommodate every valid communication scenario we call a *feasible VPN.* The (symmetric)

*The work has been supported partially by the Dutch BSIK-BRICKS project and by the FET Unit of EC (IST priority - 6th FP), under contract no. FP6-021235-2 (project ARRIVAL). This extended abstract is based and indeed partially copied from a paper the author wrote together with Cor Hurkens and Judith Keijsper [11]

†Dept. of Mathematics and Computer Science, Eindhoven University of Technology, the Netherlands and CWI Amsterdam, the Netherlands, leen@win.tue.nl.

VPN problem is to find the cheapest feasible VPN. There are several variants of the problem emerging from additional routing requirements.

- SPR *Single path routing*: For each pair $\{i, j\} \subseteq W$, exactly one path $P_{ij} \subseteq E$ is to be selected to accommodate all possible demand between i and j. The problem is to choose the paths P_{ij} so as to minimize $\{\sum_{e \in E} c(e)x_e \mid x_e \geq \sum_{\{i,j\}:e \in P_{ij}} d_{ij}, \forall e \in E \; \forall D = (d_{ij}) \in \mathcal{D}\}$.
- TTR *Terminal tree routing*: This is single path routing with the additional restriction that $\cup_{j \in W} P_{ij}$ should form a tree in G for all $i \in W$.
- TR *Tree routing*: This is SPR with the extra restriction that $\cup_{\{i,j\} \subseteq W} P_{ij}$ is a tree in G.
- MPR *Multi-path routing*: For each pair $\{i, j\} \subseteq W$, and for each possible path between i and j, the fraction of communication between i and j to be routed along that path has to be specified.
- FR *Flexible routing*: No communication paths have to be selected beforehand. Different demand scenarios are allowed to use different sets of paths.

The following lemma summarizes the rather obvious relations between the optimal solution values of these variants. By $OPT(\mathrm{SPR})$ we denote the cost of an optimal solution for the SPR variant of the VPN problem. Similar notation is used for the other optimal values.

Lemma 1

$$OPT(\mathrm{FR}) \leq OPT(\mathrm{MPR}) \leq OPT(\mathrm{SPR}) \leq OPT(\mathrm{TTR}) \leq OPT(\mathrm{TR}).$$

Proof: SPR is the MPR problem with the extra restriction that all fractions must be 0 or 1. The other inequalities are similarly trivial. □

Rumours say that FR is Co-NP-hard [7]. There are instances (even on circuits) where $OPT(\mathrm{FR}) < OPT(\mathrm{MPR})$: if we take for G a triangle, $c \equiv 1$, $b \equiv 1$, then for the optimal solution to FR it suffices to buy all three edges with capacity $1/2$, whereas for MPR it is optimal to buy two edges with capacity 1.

Several groups showed independently that MPR $\in P$, Erlebach and Rüegg [5], Altin et al. [1], Hurkens et al. [10]. In the last two papers this was shown through a linear programming formulation of the MPR-VPN problem.

Kumar et al. [13] have shown that TR $\in P$ (see also [8]). The proof of this result is very elegant and is presented in the following section. Gupta et al. [8] show that $OPT(\mathrm{TR}) = OPT(\mathrm{TTR})$ and that $OPT(\mathrm{TR}) \leq 2OPT(\mathrm{FR})$.

A prominent open question in VPN design is if SPR is polynomially solvable ($\mathrm{SPR} \in P$), cf. e.g. Italiano et al. [12]. This question would be answered affirmatively if one could prove that $OPT(\mathrm{SPR}) = OPT(\mathrm{TR})$, which is indeed what has become known as the "Tree Conjecture" for the symmetric VPN problem.

Tree Conjecture: $OPT(\mathrm{SPR}) = OPT(\mathrm{TR})$.

In fact, in [11] this conjecture is even strengthened:

Strong Tree Conjecture: $OPT(\mathrm{MPR}) = OPT(\mathrm{TR})$.

They prove this conjecture on subclasses of graphs, among which the most significant one is the class of circuits. The proof boils down to showing that the cost of an optimal solution to TR equals the value of an optimal dual solution in a formulation of MPR as a linear program (LP). I will present the LP-formulation and give a sketch of the proof of this result during the lecture.

In the same paper the Strong Tree Conjecture has been proved for any graph G and any cost function c, if the communication bound of some terminal is larger than the sum of the bounds of the other terminals, for any graph on at most 4 vertices, and for any complete graph if the cost function c is identical to 1. They also proved that the property $OPT(\mathrm{MPR}) = OPT(\mathrm{TR})$ is preserved under taking 1-sums of graphs, implying a common generalization of all the aforementioned results. All the positive results will be summarized in a theorem in the last section. Erlebach and Rüegg [5] report that none of a large number of computational experiments has contradicted the Strong Tree Conjecture. The conjecture remains unsettled for general graphs.

The model of the VPN problem presented above was proposed for the first time by Fingerhut et al. [6], and later independently by Duffield et al. [3]. They also formulated the asymmetric version of the problem in which for each node there is a distinction between a bound $b^- : W \to \mathbb{R}_+$ for incoming communication and a bound $b^+ : W \to \mathbb{R}_+$ for outgoing communication. Gupta et al. [8] prove that even the TR problem is NP-hard for the VPN problem with asymmetric communication bounds. However, the TR problem is solvable in polynomial time if $b^-(v) = b^+(v)$ for all $v \in W$. Italiano et al. [12] show that this is true already if $\sum_{v\in W} b^-(v) = \sum_{v\in W} b^+(v)$. Gupta et al. [8] claim that FR is co-NP hard for the asymmetric problem. The polynomial time algorithm for MPR by Erlebach and Rüegg [5] has been derived for the asymmetric problem. Independently, Altin et al. [1] and Hurkens et al. [10] present an LP-formulation of the general asymmetric MPR VPN problem of polynomial size, immediately implying polynomial solvability of this problem. In [10], the LP formulation in that report covers MPR-variants with four types of asymmetry: asymmetric bounds ($b^-(v) \neq b^+(v)$), asymmetric costs ($c_{uv} \neq c_{vu}$), asymmetric routing, and asymmetric communication scenarios ($d_{ij} \neq d_{ji}$). If $b^-(v) = b^+(v)$ for all v, and either cost or routing is symmetric, then attention can be restricted to symmetric communication scenarios, and $OPT(\mathrm{MPR}) = OPT(\mathrm{TR})$ still holds if G is a circuit. For symmetric routing this is easy to see. In §6 of [10] it is argued that allowing asymmetric routing under symmetric arc costs does not yield any advantage; there is always an optimal LP-solution with *symmetric* routing patterns.

As soon as both cost and routing are allowed to be asymmetric, equality of $OPT(\mathrm{MPR})$ and $OPT(\mathrm{TR})$ on any circuit is no longer true, even for symmetric bounds: if we consider a (bidirected) circuit where clockwise arcs have zero cost and counterclockwise arcs have cost 1, then buying all clockwise arcs is cheaper than buying any tree.

Gupta et al. [8] and [9] and Eisenbrand et al. [4] study approximation algorithms for NP-hard versions of the VPN problem. More hardness results appear in Chekuri et al. [2].

2 Polynomial time solvability of the Tree Routing problem

This section is essentially extracted from [8] and summarizes how to compute the cost of a given tree solution. Let (G, b, c) be given, and let W be the set of terminals. We write $b(U)$ for $\sum_{v\in U} b(v)$, $U \subseteq V$. Given a tree $T \subseteq E$ spanning a vertex set $V(T) \supseteq W$ a directed tree can be constructed by directing the edges of T towards the *lighter* side: if L_e and R_e are the components of $T - e$, and if $b(L_e) < b(R_e)$, direct e towards L_e, if $b(L_e) = b(R_e)$, direct e away from some fixed leaf l of the tree (the latter is a correction of what is written in [8]). This directed tree has a unique vertex r of in-degree zero which is what we call a *balance-point* of the tree: every edge in the directed tree is directed away from r. The cost of the tree T is clearly equal to

$$\sum_e \min\{b(L_e), b(R_e)\}c(e). \qquad (1)$$

Another expression for the cost of the tree is given in the following proposition from [8]. Here, we denote by $d^c_G(u,v)$ the distance from u to v in a graph G with respect to the length function c.

Proposition 2 ([8]) *Let $G = (V,E), b : V \to \mathbb{R}_+, c : E \to \mathbb{R}_+$ be given. Then the cost of an optimal tree solution T equals*

$$\sum_{v \in W} b(v) d^c_T(r,v)$$

for some balance-point r. This cost is bounded from below by $\sum_{v\in W} b(v)d^c_G(r,v)$, and bounded from above by $\sum_{v\in W} b(v)d^c_T(u,v)$ for any $u \in V(T)$.

As a consequence, we have that an optimal tree solution can be found by computing a shortest path tree T_u from every vertex $u \in V$ and taking the one with minimal cost $\sum_{v\in W} b(v)d^c_{T_u}(u,v) = \sum_{v\in W} b(v)d^c_G(u,v)$. Hence TR is solvable in polynomial time.

3 Subclasses for which the Strong Tree Conjecture holds

For completeness, I present the most general statement concerning the Strong Tree Conjecture that is obtained so far [11]. We use the following definitions. The *connectivity* of a graph $G = (V,E)$ is the minimum size of a subset U of V for which $G - U$ is not connected. If no such U exists (or equivalently, if G is complete), then the connectivity is ∞. A graph is k-connected if its connectivity is at least k. Now, a k-connected component of a graph $G = (V,E)$ is an inclusion-wise maximal subset U of V for which $G[U]$ (the subgraph of G induced by the vertices in U) is k-connected. A *block* is a 2-connected component U with $|U| \geq 2$. We identify the blocks of a graph with the subgraphs they induce. A connected graph may be obtained from its blocks by taking repeated 1-sums.

Theorem 3 *Suppose $G = (V,E)$ is a connected graph and $c \in \mathbb{R}^{|E|}_+$ is a cost function such that every block $H = (V',E')$ of G endowed with the cost function $c|_{E'}$ is either a circuit, or a graph on at most 4 vertices, or a complete graph with uniform edge costs. Then the cost of an optimal tree solution equals the value of an optimal dual for the instance (G,b,c), for any $b \in \mathbb{R}^{|V|}_+$.*

4 Conclusion

The challenge remains to prove or disprove that SPR is polynomially solvable on any graph.

References

[1] A. Altin, E. Amaldi, P. Belotti, and M.Ç. Pinar, *Virtual private network design under traffic uncertainty*, Proceedings of CTW04, 2004, pp. 24–27, extended version at http://www.elet.polimi.it/upload/belotti/.

[2] C. Chekuri, G. Oriolo, M.G. Scuttellà, and F.B. Shepherd, *Hardness of robust network design*, Proceedings INOC 2005, Lisbon, 2005, pp. 455–461.

[3] N.G. Duffield, P. Goyal, A. Greenberg, P. Mishra, K.K. Ramakrishnan, and J.E. van der Merwe, *A flexible model for resource management in virtual private networks*, ACM SIGCOMM Computer Communication Review **29** (1999), no. 4, 95–108.

[4] F. Eisenbrand, F. Grandoni, G. Oriolo, and M. Skutella, *New approaches for virtual private network design*, Automata, languages and programming, Lecture Notes in Comput. Sci., vol. 3580, Springer, Berlin, 2005, pp. 1151–1162. MR MR2184708 (2006g:68021)

[5] T. Erlebach and M. Rüegg, *Optimal bandwidth reservation in hose-model vpns with multi-path routing*, Proceedings of the 23rd INFOCOM Conference of the IEEE Communications Society, 2004.

[6] J.A. Fingerhut, S. Suri, and J.S. Turner, *Designing least-cost nonblocking broadband networks*, Journal of Algorithms **24** (1997), no. 2, 287–309.

[7] A. Gupta, *Personal communication*.

[8] A. Gupta, J. Kleinberg, A. Kumar, R. Rastogi, and B. Yener, *Provisioning a virtual private network: A network design problem for multicommodity flow*, Proceedings of the 33rd Annual ACM Symposium on Theory of Computing (STOC), 2001, pp. 389–398.

[9] A. Gupta, A. Kumar, and T. Roughgarden, *Simpler and better approximation algorithms for network design*, Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC), 2003, pp. 365–372.

[10] C.A.J. Hurkens, J.C.M. Keijsper, and L. Stougie, *Virtual private network design: a proof of the tree routing conjecture on ring networks*, Tech. report, SPOR 2004-15, Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, 2004, http://www.win.tue.nl/math/bs/spor/2004-15.pdf.

[11] ______, *Virtual private network design: a proof of the tree routing conjecture on ring networks*, SIAM Journal on Discrete Mathematics (to appear).

[12] G. Italiano, S. Leonardi, and G. Oriolo, *Design of networks in the hose model*, Proceedings of the 3rd Workshop on Approximation and Randomization Algorithms in Communication Networks (ARACNE), Carleton Scientific, 2002, pp. 65–76.

[13] A. Kumar, R. Rastogi, A. Silberschatz, and B. Yener, *Algorithms for provisioning virtual private networks in the hose model*, IEEE/ACM Transactions on Networking **10** (2002), no. 4, 565–578.

SIMPLEX ALGORITHM – HOW IT HAPPENED 60 YEARS AGO

Lidija Zadnik Stirn
Biotechnical Faculty, Večna pot 83, 1000 Ljubljana, Slovenia
lidija.zadnik@bf.uni-lj.si

Abstract:
The paper is devoted to the 60th anniversary of the invention of the simplex algorithm. In 1947 George B. Dantzig proposed the simplex algorithm, an efficient method to solve a linear programming problem. In the introduction the linear programming is discussed in a wider context of the decision-making process, i.e. within the field of operations research. Further, the ideas to fundamental linear models are explained. They are followed by simplex algorithm and a short biography of its finder, G. B. Dantzig. Some alternative methods for solving linear programming problems and open problems regarding these techniques and software are reviewed in the concluding part of the paper.

Key words: operations research, optimization, linear programming, simplex algorithm, the 60th anniversary, G. B. Dantzig, simplex extensions

1 INTRODUCTION

The British scientists who were asked during World War II to analyze several military problems, and who applied mathematics and the scientific methods for solving military operations, called these applications operations research (OR). Although their work was concerned primarily with the optimum allocation of the limited war materiel, the OR team included scientists from several disciplines, as sociology, psychology, behavior sciences, in recognition of the importance of their contribution to the decision making process. Thus, today, the term operations research, often also management science, means a scientific approach to decision making, which seeks to determine how best to operate a system, usually under conditions requiring the allocation of scarce resources. OR, which is concerned with the efficient allocation of resources, is both an art and a science. The art lies in the ability to reflect the concept "efficient" in a well-defined mathematical model of a given situation; the science consists of the derivation of computational methods for solving such models (Bronson, 1982, Winston, 1994).

Specifically, decision problems affect OR models. The principal components of an OR model are alternatives, restrictions and an objective criterion. Generally, the alternatives take the form of unknown variables. These variables are then used to construct restrictions and the objective criterion in appropriate mathematical functions. The end result is a mathematical model relating the variables, constraints and objective function. The solution of the model then yields the values of the decision variables that optimize (maximize or minimize) the value of the objective function while satisfying all the constraints. The resulting solution is referred to as the optimum feasible solution (Taha, 1997). In OR mathematical models, the decision variables may be integer or continuous, and the objective and constraints functions may be linear or nonlinear. Thus, the optimization problems posed to such models give rise to a variety of solution methods, each designed to account for special mathematical properties of the model. The most prominent of these techniques is linear programming (LP), where all the objective and constraints functions are linear, and all the variables are continuous. Other techniques that deal with other types of OR models are integer programming, dynamic programming, quadratic programming, network programming, different heuristic and simulation methods, to mention only a few (Willemain, 1994).

Practically all OR methods result in computational algorithms that are iterative in nature. The iterative nature of the algorithms gives rise to voluminous and tedious computations. It is thus imperative that these algorithms be executed by the computer. This fact is also significant for the sensitivity analysis, an important aspect of the model solution phase.

The concern of this paper is only LP, the most widely used optimization technique. In a survey of Fortune 500 firms, 85% of those responding said that they have used linear programming. As a measure of the importance of LP in OR, approximately 40% of the OR textbooks are devoted to LP and related optimization techniques (Winston, 1994).

A two variable LP problem can be solved graphically. The ideas gleaned from the graphical procedure lay the foundation for the general solution technique, called the simplex algorithm (SA). George B. Dantzig created the SA in the year1947. Thus, the main purpose of the present paper is to pay attention to the occasion of the 60^{th} anniversary of the invention of SA. SA is a powerful optimization method that revolutionized planning, scheduling, network design and other complex functions integral to modern-day business, industry, government, telecommunications, advertising, architecture, circuit design and countless other areas. Further, SA influenced the development of many related optimization algorithms.

2 LINEAR PROGRAMMING (LP)

LP is an important field of optimization for several reasons. Many practical problems in OR can be expressed as LP problems. Certain special cases of LP, such as network flow problems and multicommodity flow problems are considered important enough to have generated much research on specialized algorithms for their solution. A number of algorithms for other types of optimization problems work by solving LP problems as sub-problems. Historically, ideas from LP have inspired many of the central concepts of optimization theory, such as duality, decomposition, and the importance of convexity and its generalizations. Likewise, LP is heavily used in microeconomics and business management, either to maximize the income or minimize the costs of a production scheme. Some examples are food blending, inventory management, portfolio and finance management, resource allocation for human and machine resources, planning advertisement campaigns, etc.

LP arose as a mathematical model developed during the World War II to plan expenditures and returns such that it reduces costs to the army and increases losses to the enemy. It was kept secret until 1947. The founders of the subject are George B. Dantzig, who published the SA in 1947, John von Neumann, who developed the theory of the duality in the same year, and Leonid Kantorovich, a Russian mathematician who used similar techniques in economics before Dantzig. The LP problem was first shown to be solvable in polynomial time by Leonid Khachiyan in 1979, but a larger major theoretical and practical breakthrough in the field came in 1984 when Narendra Karmarkar introduced a new interior point method for solving LP problems. Dantzig's example of finding the best assignment of 70 people to 70 jobs still explains the success of LP.

In mathematics LP problems involve the optimization of a linear objective function, subject to linear equality and inequality constraints. More formally, given a polytope, and a real-valued affine function: $f(x_1,x_2,\dots,x_n) = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$ defined on this polytope, the goal is to find a point in the polytope where this function has the smallest (or largest) value. Such points may not exist, but if they do, searching through the polytope vertices is guaranteed to find at least one of them. Linear programs are problems that can be expressed in canonical form: maximize c^Tx; subject to: $Ax \leq b$; where $x \geq 0$.

x represents the vector of variables, while c and b are vectors of coefficients and A is a matrix of coefficients. The expression to be maximized or minimized is called the objective function (c^Tx in this case). The equations $Ax \leq b$ are the constraints which specify a convex polyhedron over which the objective function is to be optimized.

Every LP problem, referred to as a primal problem, can be converted into a dual problem, which provides an upper bound to the optimal value of the primal problem. The corresponding dual problem is: minimize b^Ty; subject to: $A^Ty \geq c$; where $y \geq 0$ and y is used instead of x as variable vector .

There are two ideas fundamental to duality theory. One is the fact that the dual of a dual linear program is the original primal linear program. Additionally, every feasible solution for a linear program gives a bound on the optimal value of the objective function of its dual. The weak duality theorem states that the objective function value of the dual at any feasible solution is always greater than or equal to the objective function value of the primal at any feasible solution. The strong duality theorem states that if the primal has an optimal solution, x*, then the dual also has an optimal solution, y*, such that $c^Tx^*=b^Ty^*$. A linear program can also be unbounded or infeasible. Duality theory tells us that if the primal is unbounded then the dual is infeasible by the weak duality theorem. Likewise, if the dual is unbounded, then the primal must be infeasible. However, it is possible for both the dual and the primal to be infeasible (Hillier and Lieberman, 1995).

Since each inequality can be replaced by an equality and a slack variable, this means each primal variable corresponds to a dual slack variable, and each dual variable corresponds to a primal slack variable. This relation allows us to complementary slackness. It is possible to obtain an optimal solution to the dual when only an optimal solution to the primal is known using the complementary slackness theorem. The theorem states: suppose that $x = (x_1, x_2, \ldots, x_n)$ is primal feasible and that $y = (y_1, y_2, \ldots, y_m)$ is dual feasible. Let $(w_1, w_2, \ldots, w_m)$ denote the corresponding primal slack variables, and let $(z_1, z_2, \ldots, z_n)$ denote the corresponding dual slack variables. Then x and y are optimal for their respective problems if and only if $x_jz_j = 0$, for $j = 1, 2, \ldots, n$, $w_iy_i = 0$, for $i = 1, 2, \ldots, m$. So if the ith slack variable of the primal is not zero, then the ith variable of the dual is equal zero. Likewise, if the jth slack variable of the dual is not zero, then the jth variable of the primal is equal to zero.

Geometrically, the linear constraints define a convex polyhedron, which is called the feasible region. Since the objective function is also linear, hence a convex function, all local optima are automatically global optima. The linearity of the objective function also implies that an optimal solution can only occur at a boundary point of the feasible region, unless the objective function is constant, when any point is a global maximum. There are two situations in which no optimal solution can be found. First, if the constraints contradict each other then the feasible region is empty and there can be no optimal solution, since there are no solutions at all. In this case, the LP is said to be infeasible. Alternatively, the polyhedron can be unbounded in the direction of the objective function. In this case there is no optimal solution since solutions with arbitrarily high values of the objective function can be constructed. Barring these two pathological conditions (which are often ruled out by resource constraints integral to the problem being represented, as above), the optimum is always attained at a vertex of the polyhedron. However, the optimum is not necessarily unique: it is possible to have a set of optimal solutions covering an edge or face of the polyhedron, or even the entire polyhedron (Gaertner and Matousek, 2006, http://en.wikipedia.org/wiki/linear_programming).

Any LP with only two variables can be solved graphically. Unfortunately, most real-life LPs have many variables, so a method was needed to solve LPs with more than two variables. The SA can be used to solve very large LPs.

3 SIMPLEX ALGORITHM – HOW IT HAPPENED 60 YEARS BEFORE

In 1947 G.B. Dantzig proposed the SA as an efficient method to solve a linear programming problem. He was then working in the SCOOP group (Scientific Computation of Optimum Programs), an American research program that resulted from the intensive scientific activity during the World War II, in the USA, aimed at rationalizing the logistics in the war effort. In the Soviet Union, Kantorovitch had already proposed a similar method for the analysis of economic plans, but his contribution remained unknown to the general scientific community. It seems also that 19th century mathematicians, in particular Fourier (1823), had also thought about similar methods. What made the contribution of G.B. Dantzig so important, is its concomitance with two other phenomena:

- the considerable development of the digital computer that permitted the implementation of the algorithm to solve full size real life problems;
- the parallel development of the paradigm of inter industry exchange table, also known as the input/output matrix, proposed by W.A. Leontieff which showed that the whole economy could be represented in a sort of LP structure.

Therefore the method of LP was providing both an efficient instrument to compute solutions of large scale linear optimization problems and a general paradigm of economic equilibrium between different sectors exchanging resources and services.

When World War II started, Dantzig graduate studies at Berkeley were suspended and he became Head of the Combat Analysis Branch of the Army Air Corp's Headquarters Statistical Control, which had to deal with the logistics of supply chains and management of hundreds of thousands of items and people. The job provided the "real world" problems which linear programming would come to solve. Dantzig received his Ph.D. from Berkeley in 1946. He was originally going to accept a teaching post at Berkeley, but was persuaded by his wife and former Pentagon colleagues to go back to the USAF as a mathematical adviser. It was there, in 1947 that he first posed the LP problem, and proposed the SA to solve it (Dantzig, 1951). In 1952, he became a research mathematician at the RAND Corporation, where he began implementing LP on its computers. He was the author of the pioneering book "Linear Programming and Extensions" (1963), updated in 1997 and 2003 (Dantzig, 1963).

In general, a first presentation of the idea behind the SA uses the graphical presentation/diagram. The coordinate axes represent values of decision variables. A point in this system of axes represents a decision. The geometric figure delineated by constraints is called a polytope (or a convex polyhedron). A point located in the polytope is called an admissible decision. Indeed, such a point will satisfy all the constrained imposed on the problem. Further, one looks for the point of the polytope that touches the highest possible iso-profit line (representing the optimal value of the objective function). We see that, necessarily, the optimum lies on an extreme point of the polytope (which is a point where two constraints are simultaneously active). These extreme points correspond to the concept of basic program that is used in the general SA when it is implemented algebraically.

In the simplex method one moves from one feasible extreme point to another one improving the performance criterion. One can easily check that, at each step, one passes from an extreme point to a neighboring one where the objective function increases (if we look for maximum of the objective function). Indeed,the principle of the method is to enumerate only a small fraction of the very large number of extreme points of the polytope before reaching the optimal solution. Here we shortly present the steps to solving a linear program with SA (http://www-fp.mcs.anl.gov/otc/Guide/simplex):

1. before start, we put the linear program into standard form, i.e. changing inequality constraints to equality constraints
2. we find a first basic feasible solution
3. we calculate the reduced costs
4. we test for optimality
5. we choose the entering variable
6. we calculate the search direction
7. we test for unboundedness
8. we choose the leaving variable by the Min Ratio Test
9. we update the solution
10. we change the basis
11. we go back to Step 3.

The CPLEX, C-WHIZ, FortLP, LAMPS, LINDO, MINOS, OSL, and PC-PROG packages are used to solve large-scale problems with SA. Each of these packages accepts input in the industry-standard MPS format. Additionally, some have their own customized input format (for example, CPLEX LP format for CPLEX, direct screen input for PC-PROG). Others can be operated in conjunction with modeling languages (CPLEX, LAMPS, MINOS, and OSL interface with GAMS; LINDO and OSL interface with the LINGO modeling language; CPLEX, MINOS, and OSL interface with AMPL). Recently, interfaces between spreadsheet programs and LP packages have become available. The What's *Best!* package links a wide range of standard spreadsheets (including Lotus 1-2-3 and Quattro-Pro) to LINDO (http://www-fp.mcs.anl.gov/otc/Guide/simplex).

3.1 George B. Dantzig – father of linear programming

George Bernard Dantzig, a mathematician, known as the father of linear programming and as the inventor of the SA, a formula that revolutionized planning, scheduling, network design and other complex functions integral to modern-day business, industry and government, was born in Portland, Oregon, USA in 1914. His father, Tobias Dantzig, was a Russian mathematician who went to Paris to study with Henri Poincare.

Tobias Dantzig married Anja Ourisson, a student at the Sorbonne who also was studying mathematics, and the couple immigrated to the United States. George B. Dantzig received his bachelor's degree in mathematics and physics from the University of Maryland in 1936 and his master's degree in mathematics from the University of Michigan in 1937. He enjoyed statistics and thus moved back to Washington in 1937 and took a job with the Bureau of Labor Statistics. In 1939, he resumed his studies at the University of California at Berkeley, studying statistics under mathematician Jerzy Neyman. From 1941 to 1946, he was the civilian head of the combat analysis branch of the Air Force's Headquarters Statistical Control. His task was to find a way of managing "hundreds of thousands of different kinds of material goods and perhaps fifty thousand specialties of people," seemingly intractable problems that spurred his search for a mathematical model for what would become LP. He received his doctorate from Berkeley in 1946 and returned to Washington, where he became a mathematical adviser at the Defense Department, charged with mechanizing the planning process. In 1947 Dantzig made the contribution to mathematics for which he is most famous, the simplex method of optimization. It grew out of his work with the U.S. Air Force where he became an expert on planning methods solved with desk calculators. In fact this was known as "programming", a military term that, at that time, referred to plans or schedules for training, logistical supply or deployment of men. Dantzig mechanized the planning process by introducing "programming in a linear structure", where "programming" has the military meaning explained above. The term "linear programming" was proposed by T J Koopmans during a visit Dantzig made to the RAND corporation in 1948 to discuss his ideas (Dantzig, 1951). Having discovered his algorithm, Dantzig made an early application to the problem of eating adequately at minimum cost. He has described this in his *book Linear programming and extensions (1963):* »*One of the first applications of the simplex algorithm was to the determination of an adequate diet that was of least cost. In the fall of* 1947, *Jack Laderman of the Mathematical Tables Project of the National Bureau of Standards undertook, as a test of the newly proposed simplex method, the first large-scale computation in this field. It was a system with nine equations in seventy-seven unknowns. Using hand-operated desk calculators, approximately* 120 *man-days were required to obtain a solution. The particular problem solved was one which had been studied earlier by George Stigler* (*who later became a Nobel Laureate*) *who proposed a solution based on the substitution of certain foods by others which gave more nutrition per dollar. He then examined a "handful" of the possible* 510 *ways to combine the selected foods. He did not claim the solution to be the cheapest but gave his reasons for believing that the cost per annum could not be reduced by more than a few dollars. Indeed, it turned out that Stigler's solution* (*expressed in* 1945 *dollars*) *was only* 24 *cents higher than the true minimum per year $*39.69»; and *"LP is viewed as a revolutionary development giving man the ability to state general objectives and to find, by means of the simplex method, optimal policy decisions for a broad class of practical decision problems of great complexity. In the real world, planning tends to be ad hoc because of the many special-interest groups with their multiple objectives".*

The systematic development of practical computing methods for LP began in 1952 at the Rand Corporation in Santa Monica, under the direction of George B Dantzig. The author worked intensively on this project there until late 1956, by which time great progress had been made on first-generation computers. In 1960, he became a professor at Berkeley and chairman of the Operations Research Center, and in 1966, professor of operations research and computer science at Stanford University. He remained at Stanford until his retirement in the mid-1990s. He died on May 13, 2005, in his home in Stanford, California

In addition to his significant work in developing the SA and furthering LP, Dantzig also advanced the fields of decomposition theory, sensitivity analysis, complementary pivot methods, large-scale optimization, nonlinear programming, and programming under uncertainty. He won numerous awards for his groundbreaking work, including the National Medal of Science in 1975. The Mathematical Programming Society honored him by creating the Dantzig Award, bestowed every three years since 1982 on one or two people who have made a significant impact in the field of mathematical programming. The first issue of the SIAM Journal on Optimization in 1991 was dedicated to him. In 1980 Laszlo Lovasz (Lovasz, 1988) stated: »*If one would take statistics about which mathematical problem is using up most of the computer time in the world, then ... the answer would probably be linear programming*«. In 1991 Dantzig noted that: »I*t is interesting to note that the original problem that started my research is still outstanding - namely the problem of planning or scheduling dynamically over time, particularly planning dynamically under uncertainty. If such a problem could be successfully solved it could eventually through better planning contribute to the well-being and stability of the world*«.

4 ALTERNATIVE SOLUTION METHODS FOR LP PROBLEMS

In spite of impressive developments in computational optimization in the last 20 years, including the rapid advance of interior point methods, the SA has stood the test of time quite remarkably. It is still the pre-eminent tool for almost all applications of LP. The The SA, developed by George Dantzig, solves LP problems by constructing an admissible solution at a vertex of the polyhedron and then walking along edges of the polyhedron to vertices with successively higher values of the objective function until the optimum is reached. Although this algorithm is quite efficient in practice and can be guaranteed to find the global optimum if certain precautions against cycling are taken, it has poor worst-case behavior: it is possible to construct a LP problem for which the SA takes a number of steps exponential in the problem size. In fact, for some time it was not known whether the LP problem was solvable in polynomial time (complexity class P).

This long standing issue was resolved by Leonid Khachiyan in 1979 with the introduction of the ellipsoid method, the first worst-case polynomial-time algorithm for LP. It consists of a specialization of the nonlinear optimization technique developed by Naum Shor, generalizing the ellipsoid method for convex optimization proposed by Arkadi Nemirovski, a 2003 John von Neumann Theory Prize winner, and D. Yudin (Khachiyan, 1979). Khachiyan's algorithm was of landmark importance for establishing the polynomial-time solvability of linear programs. The algorithm had little practical impact, as the SA is more efficient for all but specially constructed families of linear programs. However, it inspired new lines of research in LP with the development of interior point methods, which can be implemented as a practical tool. In contrast to the SA, which finds the optimal solution by progresses along points on the boundary of a polyhedral set, interior point methods move through the interior of the feasible region.

In 1984, N. Karmarkar proposed a new interior point projective method for LP which not only improved on Khachiyan's theoretical worst-case polynomial bound, but also promised dramatic practical performance improvements over the SA. Since then, many interior point methods have been proposed and analyzed. Early successful implementations were based on affine scaling variants of the method. For both theoretical and practical properties, barrier function or path-following methods are the most common nowadays. The announcement by Karmarkar in 1984 (Karmarkar, 1984) that he had developed a fast algorithm that generated iterates that lie in the interior of the feasible set (rather than on the boundary, as SA do) opened up exciting new avenues for research in both the computational complexity and mathematical programming communities. Since then, there has been intense research into a variety of methods that maintain strict feasibility of all iterates, at least with respect to the inequality constraints. Although dwarfed in volume by simplex-based packages, interior-point products such as CPLEX/Barrier and OSL have emerged and have proven to be competitive with, and often superior to, the best simplex packages, especially on large problems.

5 FINAL REMARKS AND OPEN PROBLEMS

There are several open problems in the theory of LP, the solution of which would represent fundamental breakthroughs in mathematics and potentially major advances in our ability to solve large-scale linear programs:

- Does LP admit a polynomial algorithm in the real number model of computation?
- Does LP admit a strongly polynomial-time algorithm?
- Does LP admit a strongly polynomial algorithm to find a strictly complementary solution?

This closely related set of problems has been cited by Stephen Smale as among the 18 greatest unsolved problems of the 21st century. In Smale's words, the first version of the problem "is the main unsolved problem of LP theory." While algorithms exist to solve LP in weakly polynomial time, such as the ellipsoid methods and interior-point techniques, no algorithms have yet been found that allow strongly polynomial-time performance in the number of constraints and the number of variables. The development of such algorithms would be of great theoretical interest, and perhaps allow practical gains in solving large LPs as well:

- Are there pivot rules which lead to polynomial-time simplex variants?
- Do all polyhedral graphs have polynomially-bounded diameter?
- Is the Hirsch conjecture true for polyhedral graphs?

These questions relate to the performance analysis and development of simplex-like methods. The immense efficiency of the SA in practice despite its exponential-time theoretical performance hints that there may be variations of SA that run in polynomial or even strongly polynomial time. It would be of great practical and theoretical significance to know whether any such variants exist, particularly as an approach to deciding if LP can be solved in strongly polynomial time (Smale, 1998).

The SA and its variants fall in the family of edge-following algorithms, so named because they solve LP problems by moving from vertex to vertex along edges of a polyhedron. This means that their theoretical performance is limited by the maximum number of edges between any two vertices on the LP polyhedron. As a result, we are interested in knowing the maximum graph-theoretical diameter of polyhedral graphs. It has been proved that all polyhedra have subexponential diameter, and all experimentally observed polyhedra have linear diameter, it is presently unknown whether any polyhedron has superpolynomial or even superlinear diameter. If any such polyhedra exist, then no edge-following variant can run in polynomial or linear time, respectively. Questions about polyhedron diameter are of independent mathematical interest.

Recent developments in LP include work by Vladlen Koltun to show that LP is equivalent to solving problems on arrangement polytopes, which have small diameter, allowing the possibility of strongly polynomial-time algorithms without resolving questions about the diameter of general polyhedra. Jonathan Kelner and Dan Spielman have also proposed a randomized (weakly) polynomial-time SA (Spielman and Teng, 2004).

References

Bronson, R., 1982. Theory and Problems of Operations Research. McGraw Hill, Schaum's Outline Series, Singapore.

Dantzig, G. B., 1951. Maximization of linear function of variables subject to linear inequalities. In T. C. Koopmans (eds.), Activity Analysis of production and allocation, pp. 339-347.

Dantzig, G.B., 1963. Linear Programming and Extensions, Princeton University Press, NJ.

Hillier F.S., Lieberman, G.J., 1995. Introduction to Operations Research. McGraw, New York.

http://en.wikipedia.org/wiki/linear_programming,1.9.2007.

http://www-fp.mcs.anl.gov/otc/Guide/simplex, 1.9.2007

Gaertner, B., Matousek, J., 2006. Understanding and Using Linear Programming. Springer, Berlin.

N. Karmarkar, 1984. A new polynomial-time algorithm for linear programming, Combinatorica, 4 (1984), pp. 373-395.

Khachiyan, L.G., 1979. A polynomial algorithm in linear programming. Doklady Akademia Nauk SSSR, pp. 1093-1096

Lovász, L. 1988. Algorithmic mathematics: an old aspect with a new emphasis. In: Proceedings of the 6th International Congress on Math. Education, Budapest, Math. Soc., pp. 67-78.

Smale, S., 1998. Mathematical Problems for the Next Century, Math. Intelligencer, 20/2, pp. 7-15.

Spielman, D.A., Teng, S.H., 2004. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. JACM, Journal of the ACM, 51, pp. 385-463.

Taha, H.A., 1997. Operations Research: an Introduction. Prentice Hall, New Delhi.

Willemain, T. R., 1994. Insights of the modeling from a dozen experts. Operations Research, 42/2, pp. 213-222.

Winston, W.L., 1994. Operations Research: Applications and Algorithms. Duxbury Press, Belmont, CA.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 1:
Networks

HORN RENAMABILITY TESTING IN THE CONTEXT OF HYPERGRAPHS

Dušan Hvalica
University of Ljubljana, Faculty of Economics
dusan.hvalica@ef.uni-lj.si

Abstract. We discuss satisfiability testing in the context of directed hypergraphs; an algorithm for testing of Horn-renamability with linear time complexity is described.

Keywords: satisfiability, Horn, renamable, hypergraph

1. Introduction

Many problems in OR can be in a natural way formulated as an instance of the *satisfiability problem* — SAT, i.e., as testing the satisfiability of a propositional formula [6, 12]. Of course, as SAT is $\mathcal{NP}$-complete, they are seldom actually dealt with in this way. However, several subclasses of SAT are solvable in polynomial time [8] (among them Horn formulae [5] and formulae that are reducible to Horn [11, 3, 1, 2, 10]) and in such cases solving in this context may be advantageous. Algorithms that, when restricted to some subclass, perform better than the general algorithm, rely on features of that particular subclass and are not admissible outside of that subclass. Thus, it is safe to apply such an algorithm only after one has verified that that particular instance of SAT belongs to that subclass.

SAT can be translated into the context of directed hypergraphs [7] — testing of satisfiability translates into searching for a zero-cardinality cut. Many polynomially solvable subclasses of SAT are also easily described in the context of hypergraphs, in particular this applies to Horn-SAT and to Horn-renamable formulae. The aim of this paper is to demonstrate this by describing an algorithm which in the context of hypergraphs tests for Horn-renamability, with time complexity, comparable to that of the best known algorithms.

2. Definitions and notation

An (oriented) *hypergraph* G is defined as $G = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ and $\mathcal{A}$ are the sets of *nodes* and *hyperarcs*, respectively. A hyperarc $\mathcal{E}$ is defined as $\mathcal{E} = (T(\mathcal{E}), H(\mathcal{E}))$, where $T(\mathcal{E})$, $H(\mathcal{E}) \subset \mathcal{V}$; the sets $T(\mathcal{E})$ and $H(\mathcal{E})$ are called the *tail* and *head* of $\mathcal{E}$, respectively. A hyperarc, whose head has (only) one element, is called a *B-arc* (backward (hyper)arc), while a hypergraph, the hyperarcs of which are all B-arcs, is a *B-graph*.

A *subhypergraph* of a hypergraph $G = (\mathcal{V}, \mathcal{A})$ is a hypergraph $G_1 = (\mathcal{V}_1, \mathcal{A}_1)$ such that $\mathcal{V}_1 \subset \mathcal{V}$ and $\mathcal{A}_1 \subset \mathcal{A}$. When suitable, $\mathcal{V}_1$ will be denoted by $\mathcal{V}(G_1)$ and $\mathcal{A}_1$ by $\mathcal{A}(G_1)$.

For any node u its *backward star* $BS(u)$ is defined by $BS(u) = \{\mathcal{E}\,; u \in H(\mathcal{E})\}$, while its *forward star* $FS(u)$ is $FS(u) = \{\mathcal{E}\,; u \in T(\mathcal{E})\}$. A node u for which $BS(u) = \emptyset$ or $FS(u) = \emptyset$ will be called a *tip node*.

For any subhypergraph $H \subset G$ the set of its nodes v such that $BS(v) \cap \mathcal{A}(H) = \emptyset$ will be denoted by $B(H)$ while the set of its nodes v such that $FS(v) \cap \mathcal{A}(H) = \emptyset$ will be denoted by $F(H)$.

A *path* is a sequence $u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \dots, \mathcal{E}_{q-1}, u_q$, such that $u_i \in H(\mathcal{E}_{i-1})$ for $1 < i \leqslant q$ and $u_i \in T(\mathcal{E}_i)$ for $1 \leqslant i < q$. If $u_q \in T(\mathcal{E}_1)$, such a path is called a *cycle*.

Let S be any set of nodes in a B-graph G. A *B-hyperpath* or, shortly, *B-path*, based on S and ending at t is any hypergraph P, which is a minimal subhypergraph of G such that

- $t \in \mathcal{V}(P)$,
- for every $v \in \mathcal{V}(P)$ there exists in P a simple cycle-free path from some $u \in S$ to v.

When such a B-path exists, we also say that t is *B-connected* to S.

3. Satisfiability of propositional formulae

Let A be a set of propositional variables and B a set of clauses over A, i.e., of formulae of the form:

$$C_1 \wedge \cdots \wedge C_m \Rightarrow D_1 \vee \cdots \vee D_n\,, \tag{1}$$

where $C_1, \dots, C_m, D_1, \dots, D_n$ belong to $A \cup \{true, false\}$. We say that B is *satisfiable* if there exists a truth assignment $A \to \{true, false\}$ such that every clause in B is *true*.

The *satisfiability problem*, SAT, consists of testing the satisfiability of B. As already pointed out, it is $\mathcal{NP}$-complete.

A clause (1) with $n \leqslant 1$ is known as a *Horn* clause. If every clause in B is a Horn clause, we speak of *Horn-SAT*. It is known that it is solvable in linear time [5]. Some other classes of formulae are also known, for which satisfiability can be tested in polynomial time: *Horn renamable* formulae [11], *generalized Horn* [14], the class SLUR (Single Lookahead Unitary Resolution) [13], extended Horn [2], balanced formulae [4], ...

To every instance of SAT a hypergraph can be assigned in the following way: its nodes are the propositional variables, while its hyperarcs correspond to clauses such that the hyperarc corresponding to clause (1) is $(\{C_1, \dots, C_m\}, \{D_1, \dots, D_n\})$. If every clause is a Horn clause, the resulting hypergraph is clearly a B-graph. Moreover, in that case the problem of satisfiability translates into the problem of verifying B-connectedness in this B-graph (for details, see [7]).

4. Horn renamability

Horn-renamable formulae are CNF formulae, for which a *renaming* (replacing, for some variables x_i, $i \in I$, each occurrence of x_j and $\overline{x}_j$ by $\overline{y}_j$ and y_j, respectively) exists that turns the formula into a Horn formula. Clearly, every replacement of x_j and $\overline{x}_j$ by $\overline{y}_j$ and y_j in the formula corresponds to a switch of position of node x_j in every hyperarc of the corresponding hypergraph — if $x_j \in T(\mathcal{E})$ we move it to $H(\mathcal{E})$ and *vice versa*. Thus, a formula is Horn-renamable iff there exists a switching that turns the corresponding hypergraph into a B-graph.

For instance, consider the set of clauses

$$\{\bar{x}_1 \vee \bar{x}_2 \vee x_3 \vee x_4, \bar{x}_4 \vee x_5\} \tag{2}$$

It can be turned into a set of Horn clauses by renaming x_4 and x_5:

$$\{\bar{x}_1 \vee \bar{x}_2 \vee x_3 \vee \bar{y}_4, y_4 \vee \bar{y}_5\}$$

Clearly, to yield a B-graph, switching must always occur along some path which ends in a tip node or a cycle and appears "reversed" in the resulting hypergraph.

5. Tracks

A *track* is a sequence

$$u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \dots, \mathcal{E}_{q-1}, u_q, \tag{3}$$

such that

- for any triple $u_i, \mathcal{E}_i, u_{i+1}$ on the track we have $u_i \neq u_{i+1}$ and $u_i, u_{i+1} \in T(\mathcal{E}_i) \cup H(\mathcal{E}_i)$,
- for any triple $\mathcal{E}_i, u_{i+1}, \mathcal{E}_{i+1}$ on the track we have $\mathcal{E}_i \neq \mathcal{E}_{i+1}$ and one of the following holds:
  - $u_{i+1} \in H(\mathcal{E}_i)$ and $u_{i+1} \in T(\mathcal{E}_{i+1})$,
  - $u_{i+1} \in T(\mathcal{E}_i)$ and $u_{i+1} \in H(\mathcal{E}_{i+1})$.

If $u_{q-1}, \mathcal{E}_{q-1}, u_q, \mathcal{E}_1, u_2$ is a track, then track (3) is called a *hypercycle* (of course, in such case either $u_q \in T(\mathcal{E}_1)$ or $u_q \in H(\mathcal{E}_1)$; often we have $u_q = u_1$).

Thus, a track consists of paths and "reversed" paths glued together at common hyperarcs.

If $T = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i$ and $T' = u_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j$ are tracks, we shall denote

$$TT' = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j$$

provided that TT' is also a track (which happens when $\mathcal{E}_{i-1}$ and $\mathcal{E}'_1$ do not both belong to $BS(u_i)$ or to $FS(u_i)$).

Similarly, for any tracks $T = u_1, \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}, u_i$ and $T' = u'_i, \mathcal{E}'_1, \ldots, \mathcal{E}'_{j-1}, u'_j$ and hyperarc $\mathcal{E}$, we shall denote

$$T\mathcal{E}T' = u_1, \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}, u_i, \mathcal{E}, u'_i, \mathcal{E}'_1, \ldots, \mathcal{E}'_{j-1}, u'_j$$

provided that $T\mathcal{E}T'$ is a track as well.

For any track $T = u_1, \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}, u_i$ its *reverse track* is $T^{-1} = u_i, \mathcal{E}_{i-1}, \ldots, \mathcal{E}_1, u_1$. Clearly we have $(T_1T_2)^{-1} = T_2^{-1}T_1^{-1}$.

Any track $x, \mathcal{E}_1, \ldots, \mathcal{E}_2, x$ such that either $\mathcal{E}_1, \mathcal{E}_2 \in BS(x)$ or $\mathcal{E}_1, \mathcal{E}_2 \in FS(x)$ will be called a *return twist*. Thus, if T is a return twist and $T'T$ is a track then TT'^{-1} is also a track. If T_1 is a return twist and $T = T'T_1$ is a track, we shall say that T *ends with a return twist*. If T_1, T_2 are return twists and $T = T_1T'T_2$ is a track, we shall say that T *has return twists at both ends*. Clearly, in this case, $T_1T'T_2T'^{-1}$ is a hypercycle.

6. Hypergraphs to B-graphs

The test whether it is possible to turn a hypergraph into a B-graph by switching the position of nodes can be based on the following:

Theorem 1. *If (and only if) the hypergraph does not contain any tracks with return twists at both ends, there exists a set S of the nodes such that for each hyperarc $\mathcal{E}$ we have*

- $|T(\mathcal{E}) \cap S| \leqslant 1$,
- *if* $|T(\mathcal{E}) \cap S| = 0$ *then* $|H(\mathcal{E}) \setminus S| \leqslant 1$,
- *if* $|T(\mathcal{E}) \cap S| = 1$ *then* $H(\mathcal{E}) \subset S$.

(Owing to space limitations the proof is omitted.) Clearly, if such S exists, by switching the position of the nodes in S the hypergraph is turned into a B-graph.

Thus, what we need is an algorithm that in the given hypergraph either finds such a set S or determines a track with return twists at both ends. Such algorithm can be based on the following lemmata:

Lemma 2. *If every hyperarc on the track $T = x_1, \mathcal{E}_1, \ldots, \mathcal{E}_2, x_2$ is to be turned into a B-arc, then*

- *if $\mathcal{E}_1 \in FS(x_1), \mathcal{E}_2 \in BS(x_2)$ and the position of x_1 is switched, then x_2 must be switched as well,*
- *if $\mathcal{E}_1 \in FS(x_1), \mathcal{E}_2 \in FS(x_2)$ and the position of x_1 is switched, then x_2 must not be switched,*
- *if $\mathcal{E}_1 \in BS(x_1), \mathcal{E}_2 \in BS(x_2)$ and the position of x_1 is not switched, then x_2 must be switched,*

- *if $\mathcal{E}_1 \in BS(x_1), \mathcal{E}_2 \in FS(x_2)$ and the position of x_1 is not switched, then x_2 must not be switched either.*

Lemma 3. *If there exists a return twist $T = x, \mathcal{E}_1, \dots, \mathcal{E}_2, x$, then — if all hyperarcs on T are to be turned into B-arcs — the position of x*

- *must be switched when $\mathcal{E}_1, \mathcal{E}_2 \in BS(x)$,*
- *must not be switched when $\mathcal{E}_1, \mathcal{E}_2 \in FS(x)$.*

(The proofs are omitted.)

The algorithm can be designed as follows: it starts by tentatively labeling (in a depth-first manner) the nodes of the hypergraph according to Lemma 2 — label '*to be switched*' is propagated in the direction of hyperarcs, while label '*not to be switched*' is propagated in the opposite direction. This can proceed until a return twist is found; when this occurs final labels are set according to Lemma 3 — label '*must be switched*' is propagated along the hyperarcs while label '*must not be switched*' is propagated in the opposite direction. If another return twist is found in this phase the algorithm stops with the output '*a track with return twists at both ends*', otherwise the set of the nodes labelled '*must be switched*' or '*to be switched*' is S from Theorem 1.

As the label of any node can change only twice (when a tentative label is replaced by a fixed one), the propagation of labels through any node can occur only twice. Consequently, the time complexity of the algorithm is $O(|G|)$, where $|G| = \sum_{\mathcal{E}\in\mathcal{A}}(|H(\mathcal{E})|+|T(\mathcal{E})|)$ (which is comparable to the time complexity of the algorithms from [2, 10]).

7. Conclusion

We have demonstrated that the concept of a track gives rise to criteria of which nodes are to be switched if a given hypergraph is to be turned into a B-graph, and that these criteria enable one to design an algorithm for testing for Horn renamability with linear time complexity. Thus, it seems that the hypergraph approach to SAT is worthy of further study.

References

[1] E. Boros, P.L. Hammer and X. Sun, Recognition of q-Horn formulae in linear time, Discrete Applied Mathematics 55 (1994), 1-13.

[2] V. Chandru, C. R. Coullard, P. L. Hammer, M. Montañez and X. Sun, On renamable Horn and generalized Horn functions, Annals of Math. and Art. Intel. 1 (1990), 33-47.

[3] V. Chandru and J.N. Hooker, Extended Horn sets in propositional logic, Journal of the Association for Computing Machinery 38 (1991), 205-221.

[4] M. Conforti, G. Cornuéjols, A class of logical inference problems solvable by linear programming, Foundations of Computer Science 33 (1992), 670-675.

[5] W.F. Dowling, and J.H. Gallier, Linear time algorithms for testing the satisfiability of Horn formulae, J. Logic Programming 1 (1984), 207-284.

[6] M. Ernst, T.D. Millstein, and D.S. Weld, Automatic SAT-compilation of planning problems, in Proceedings of the International Joint Conference on Artificial Intelligence (1997), 1169-1177.

[7] G. Gallo, G. Longo, S. Pallottino, S. Nguyen, Directed Hypergraphs and applications, Discrete Applied Mathematics 42 (1993), 177-201.

[8] G. Gallo and M.G. Scutella, Polynomially solvable satisfiability problems, Information Processing Letters 29 (1988), 221-227.

[9] G. Gallo, C. Gentile, D. Pretolani and G. Rago, Max Horn SAT and the Minimum Cut Problem in Directed Hypergraphs, Mathematical Programming 80 (1998), 213-237.

[10] J.-J. Hebrard, A linear algorithm for renaming a set of clauses as a Horn set, Theoretical Computer Science 124 (1994), 343-350.

[11] H. R. Lewis, Renaming a set of clauses as a Horn set, Journal of the ACM 25 (1978), 134-135.

[12] B. Randerath, E. Speckenmeyer, E. Boros, P. Hammer, A. Kogan, K. Makino, B. Simeone and O. Cepek, A satisfiability formulation of problems on level graphs, in H. Kautz and B. Selman (editors), Electronic Notes in Discrete Math. 9 (2001), 1-9.

[13] J.S. Schlipf, F. Annexstein, J. Franco, and R. Swaminathan, On finding solutions for extended Horn formulas, Information Processing Letters, 54 (1995), 133-137.

[14] S. Yamasaki and S. Doshita, The Satisfiability Problem for a Class Consisting of Horn Sentences and Some Non-Horn Sentences in Propositional Logic, Information and Control, 59 (1983), 1-12.

HORN RENAMABILITY AND B-GRAPHS

Dušan Hvalica
University of Ljubljana, Faculty of Economics
dusan.hvalica@ef.uni-lj.si

Abstract. We discuss satisfiability testing in the context of directed hypergraphs. We give a characterization of Horn-renamable formulae and describe a subclass of SAT that belongs to $\mathcal{P}$.

Keywords: satisfiability, Horn, renamable, hypergraph

1. Introduction

The *satisfiability problem*, SAT, consisting of testing the satisfiability of a propositional formula, is known to be $\mathcal{NP}$-complete. On the other hand, several subclasses of propositional formulae are known, such that the restriction of SAT to such a subclass is solvable in polynomial time [6] — among them Horn formulae [4] and formulae that are reducible to Horn, such as Horn-renamable formulae [9, 3, 1, 2, 8]. Naturally, these classes have attracted much attention.

SAT can be translated in a natural way into the context of directed hypergraphs — testing of satisfiability translates to searching for a zero-cardinality cut [5, 7].

The aim of this paper is to show that the hypergraph approach to SAT gives rise to new concepts, advantageous in designing algorithms and in specifying classes of formulae. We demonstrate this by giving a characterization of Horn-renamable formulae in terms of directed hypergraphs; as a by-product we obtain a subclass of SAT that belongs to $\mathcal{P}$.

2. Definitions and notation

An (oriented) *hypergraph* G is defined as $G = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ and $\mathcal{A}$ are the sets of *nodes* and *hyperarcs*, respectively. A hyperarc $\mathcal{E}$ is defined as $\mathcal{E} = (T(\mathcal{E}), H(\mathcal{E}))$, where $T(\mathcal{E})$, $H(\mathcal{E}) \subset \mathcal{V}$; the sets $T(\mathcal{E})$ and $H(\mathcal{E})$ are called the *tail* and *head* of $\mathcal{E}$, respectively. A hyperarc, whose head has (only) one element, is called a *B-arc* (backward (hyper)arc), a hypergraph, the hyperarcs of which are all B-arcs, is a *B-graph*.

A *subhypergraph* of a hypergraph $G = (\mathcal{V}, \mathcal{A})$ is a hypergraph $G_1 = (\mathcal{V}_1, \mathcal{A}_1)$ such that $\mathcal{V}_1 \subset \mathcal{V}$ and $\mathcal{A}_1 \subset \mathcal{A}$. When suitable, $\mathcal{V}_1$ will be denoted by $\mathcal{V}(G_1)$ and $\mathcal{A}_1$ by $\mathcal{A}(G_1)$.

For any node u its *backward star* $BS(u)$ is defined by $BS(u) = \{\mathcal{E}\,; u \in H(\mathcal{E})\}$, while its *forward star* $FS(u)$ is $FS(u) = \{\mathcal{E}\,; u \in T(\mathcal{E})\}$. A node u for which $BS(u) = \emptyset$ or $FS(u) = \emptyset$ will be called a *tip node*.

A *path* is a sequence $u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \ldots, \mathcal{E}_{q-1}, u_q$, such that $u_i \in H(\mathcal{E}_{i-1})$ for $1 < i \leqslant q$ and $u_i \in T(\mathcal{E}_i)$ for $1 \leqslant i < q$. If $u_q \in T(\mathcal{E}_1)$, such a path is called a *cycle*.

Let S be any set of nodes in a B-graph G. A *B-hyperpath* or, shortly, *B-path*, based on S and ending at t is any hypergraph P, which is a minimal subhypergraph of G such that

- $t \in \mathcal{V}(P)$,
- for every $v \in \mathcal{V}(P)$ there exists in P a simple cycle-free path from some $u \in S$ to v.

When such a B-path exists, we also say that t is *B-connected* to S.

3. Satisfiability of propositional formulae

Let A be a set of propositional variables and B a set of clauses over A, i.e., of formulae of the form:

$$C_1 \wedge \cdots \wedge C_m \Rightarrow D_1 \vee \cdots \vee D_n\,, \tag{1}$$

where $C_1, \ldots, C_m, D_1, \ldots, D_n$ belong to $A \cup \{true, false\}$. We say that B is *satisfiable* if there exists a truth assignment $A \to \{true, false\}$ such that every clause in B is $true$.

The *satisfiability problem*, SAT, consists of testing the satisfiability of B. As already pointed out, it is $\mathcal{NP}$-complete.

A clause (1) with $n \leqslant 1$ is known as a *Horn* clause. If every clause in B is a Horn clause, we speak of *Horn-SAT*. It is known that it is solvable in linear time.

To every instance of SAT a hypergraph can be assigned in the following way: its nodes are the propositional variables, while its hyperarcs correspond to clauses such that the hyperarc corresponding to clause (1) is $(\{C_1, \ldots, C_m\}, \{D_1, \ldots, D_n\})$. If every clause is a Horn clause, the resulting hypergraph is clearly a B-graph. Moreover, in that case the problem of satisfiability translates into the problem of verifying B-connectedness in this B-graph; in general, a formula is satisfiable if and only if there is a 0-cardinality cut in the corresponding hypergraph (for details, see [5]).

One of the polynomially solvable subclasses of SAT is the class of *Horn-renamable formulae*, i.e., CNF formulae, for which a *renaming* (replacing, for some variables x_i, $i \in I$, each occurrence of x_j and $\overline{x}_j$ by $\overline{y}_j$ and y_j, respectively) exists that turns the formula into a Horn formula. Clearly, every replacement of x_j and $\overline{x}_j$ by $\overline{y}_j$ and y_j in the formula corresponds to a switch of position of node x_j in every hyperarc of the corresponding hypergraph — if $x_j \in T(\mathcal{E})$ we move it to $H(\mathcal{E})$ and *vice versa*. Thus, a formula is Horn-renamable iff there exists a switching that turns the corresponding hypergraph into a B-graph.

4. Hypergraphs to B-graphs

Let us introduce the following new concept:

Definition 1. *A* track *is a sequence*

$$u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \dots, \mathcal{E}_{q-1}, u_q, \tag{2}$$

such that

- *for any triple $u_i, \mathcal{E}_i, u_{i+1}$ on the track we have $u_i \neq u_{i+1}$ and $u_i, u_{i+1} \in T(\mathcal{E}_i) \cup H(\mathcal{E}_i)$,*
- *for any triple $\mathcal{E}_i, u_{i+1}, \mathcal{E}_{i+1}$ on the track we have $\mathcal{E}_i \neq \mathcal{E}_{i+1}$ and one of the following holds:*
  - *$u_{i+1} \in H(\mathcal{E}_i)$ and $u_{i+1} \in T(\mathcal{E}_{i+1})$,*
  - *$u_{i+1} \in T(\mathcal{E}_i)$ and $u_{i+1} \in H(\mathcal{E}_{i+1})$.*

If $u_{q-1}, \mathcal{E}_{q-1}, u_q, \mathcal{E}_1, u_2$ is a track, then track (2) is called a hypercycle *(of course, in such case either $u_q \in T(\mathcal{E}_1)$ or $u_q \in H(\mathcal{E}_1)$; often we have $u_q = u_1$).*

Thus, a track consists of paths and "reversed" paths glued together at common hyperarcs.

For instance, $u_1, \mathcal{E}_1, u_4, \mathcal{E}_2, u_5, \mathcal{E}_4, u_6, \mathcal{E}_3, u_4, \mathcal{E}_5, u_3$ is a track in the hypergraph in Fig. 1.

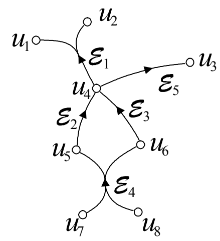


Figure 1

If $T = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i$ and $T' = u_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j$ are tracks, we shall denote

$$TT' = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j.$$

provided that TT' is also a track (which happens when $\mathcal{E}_{i-1}$ and $\mathcal{E}'_1$ do not both belong to $BS(u_i)$ or to $FS(u_i)$.

Similarly, for any tracks $T = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i$ and $T' = u'_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j$ and hyperarc $\mathcal{E}$, we shall denote

$$T\mathcal{E}T' = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i, \mathcal{E}, u'_i, \mathcal{E}'_1, \dots, \mathcal{E}'_{j-1}, u'_j$$

provided that $T\mathcal{E}T'$ is a track as well.

For any track $T = u_1, \mathcal{E}_1, \dots, \mathcal{E}_{i-1}, u_i$ its *reverse track* is $T^{-1} = u_i, \mathcal{E}_{i-1}, \dots, \mathcal{E}_1, u_1$. Clearly we have $(T_1T_2)^{-1} = T_2^{-1}T_1^{-1}$.

Now, consider the hypergraph in Fig. 2:

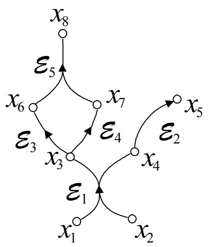


Figure 2

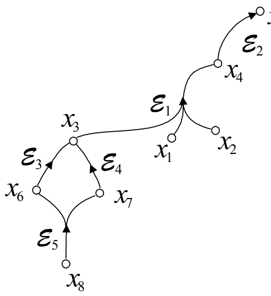


Figure 3

If we start the attempt to turn it into a B-graph by switching the position of x_3 then eventually the position of all nodes on paths $x_3, \mathcal{E}_3, x_6, \mathcal{E}_5, x_8$ and $x_3, \mathcal{E}_4, x_7, \mathcal{E}_5, x_8$ must be switched, which yields the hypergraph in Fig. 3, where $\mathcal{E}_5$ is no more a B-arc.

Similarly, whenever we start the process of turning the hypergraph into a B-graph by switching the position of a node x such that in the hypergraph there exist paths, starting at x and joining at some hyperarc $\mathcal{E}$, eventually two nodes must move from $T(\mathcal{E})$ to $H(\mathcal{E})$, so that after these switches $\mathcal{E}$ cannot be a B-arc. Moreover, any attempt to make $\mathcal{E}$ a B-arc must result in node x moving back to the head of the hyperarc where we started.

Thus, when for some $x \in H(\mathcal{E})$ in the hypergraph there is a track $P_1\mathcal{E}'P_2^{-1}$, where P_1 and P_2 are paths, starting at x, then our problem cannot be resolved by switching the position of the nodes on that track. Of course, we can try to make $\mathcal{E}$ a B-arc by switching along a path, starting at some other $y \in H(\mathcal{E})$ — in the hypergraph in Fig. 2 it can be done by starting at x_4 — but if for some other $x' \in H(\mathcal{E})$ there exists a track $P_1'\mathcal{E}''P_2'^{-1}$, where P_1' and P_2' are paths, starting at x', then neither x nor x' can be switched without making some other hyperarc a non-B-arc. In other words, $\mathcal{E}$ cannot be made a B-arc without spoiling some other hyperarc so that in that case the hypergraph cannot be turned into a B-graph.

In the above situation $P_1\mathcal{E}'P_2^{-1}\mathcal{E}P_1'\mathcal{E}''P_2'^{-1}\mathcal{E}\{x\}$ is clearly a hypercycle through x. As we have seen, the existence of such a hypercycle is sufficient for the hypergraph not to be turnable into a B-graph. It turns out that a little weaker condition is still sufficient, but necessary as well.

Any track $x, \mathcal{E}_1, \ldots, \mathcal{E}_2, x$ such that either $\mathcal{E}_1, \mathcal{E}_2 \in BS(x)$ or $\mathcal{E}_1, \mathcal{E}_2 \in FS(x)$ will be called a *return twist*. Thus, if T is a return twist and $T'T$ is a track then TT'^{-1} is also a track. If T_1 is a return twist and $T = T'T_1$ is a track, we shall say that T *ends with a return twist*. If T_1, T_2 are return twists and $T = T_1T'T_2$ is a track, we shall say that T *has return twists at both ends*. Clearly, in this case, $T_1T'T_2T'^{-1}$ is a hypercycle.

Now, we can state our main result — the following applies:

Theorem 2. *A hypergraph can be turned into a B-graph by switching the position of nodes if and only if it does not contain any tracks with return twists at both ends.*

(Due to space limitations the proof must be omitted.)

For instance, the hypergraph in Fig. 4 cannot be turned into a B-graph, as for $T_1 = x_3, \mathcal{E}_3, x_6, \mathcal{E}_5, x_7, \mathcal{E}_4, x_3,\ T_2 = x_5, \mathcal{E}_7, x_{12}, \mathcal{E}_6, x_{13},\ \mathcal{E}_8, x_{15}, \mathcal{E}_9, x_{14}, \mathcal{E}_7, x_5$ and $T' = x_3, \mathcal{E}_1, x_4, \mathcal{E}_2, x_5$, the track $T = T_1T'T_2$ is a track with return twists at both ends.

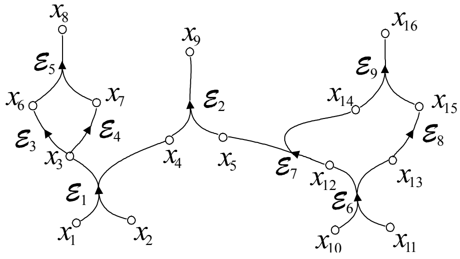


Figure 4

One can design an algorithm that in any hypergraph G finds a set of the nodes, such that by switching the position of these nodes G is turned into a B-graph, or determines a track with return twists at both ends (because of space limitations the details must be omitted — we describe the algorithm in a separate paper); its time complexity is $O(|G|)$, where $|G| = \sum_{\mathcal{E}\in\mathcal{A}}(|H(\mathcal{E})| + |T(\mathcal{E})|)$. Thus, Horn-renamability can be tested in the context of hypergraphs, with the same time complexity as the algorithms from [2, 8].

Clearly, if $T = T_1T'T_2$ is a track with return twists at both ends, then $T_1T'T_2T'^{-1}$ is a hypercycle. Hence, we have the following

Corollary 3. *By switching the position of nodes every hypergraph without hypercycles can be turned into a B-graph.*

Now we can apply our results to the issue of satisfiability.

Theorem 4. *A propositional formula is Horn-renamable if and only if the corresponding hypergraph does not contain any tracks with return twists at both ends.*

Corollary 5. *A propositional formula for which the corresponding hypergraph does not contain any hypercycle is Horn-renamable.*

The class of propositional formulae for which the corresponding hypergraph is without hypercycles is therefore a subclass of the class of Horn-renamable formulae. Thus, we have found a new class such that the restriction of SAT to it is solvable in polynomial time. Of course, this class being a subclass of Horn-renamable formulae our result is not something spectacular, but it turns out that the corresponding hypergraphs have some nice properties.

5. Conclusion

We have introduced the concept of a track and, based on it, found a necessary and sufficient condition that a hypergraph can be turned into a B-graph. This yields a characterization of Horn-renamable formulae. Furthermore, we have found a new class of formulae such that the restriction of SAT to it is solvable in polynomial time.

It turns out that it is possible to generalize these concepts even further, which allows one to describe how resolution is reflected in the corresponding hypergraph and to give a new necessary and sufficient condition for a propositional formula to be satisfiable in terms of directed hypergraphs.

References

[1] E. Boros, P.L. Hammer and X. Sun, Recognition of q-Horn formulae in linear time, Discrete Applied Mathematics 55 (1994), 1-13.

[2] V. Chandru, C. R. Coullard, P. L. Hammer, M. Montañez and X. Sun, On renamable Horn and generalized Horn functions, Annals of Math. and Art. Intel. 1 (1990), 33-47.

[3] V. Chandru and J.N. Hooker, Extended Horn sets in propositional logic, Journal of the Association for Computing Machinery 38 (1991), 205-221.

[4] W.F. Dowling, and J.H. Gallier, Linear time algorithms for testing the satisfiability of Horn formulae, J. Logic Programming 1 (1984), 207-284.

[5] G. Gallo, G. Longo, S. Pallottino, S. Nguyen, Directed Hypergraphs and applications, Discrete Applied Mathematics 42 (1993), 177-201.

[6] G. Gallo and M.G. Scutella, Polynomially solvable satisfiability problems, Information Processing Letters 29 (1988), 221-227.

[7] G. Gallo, C. Gentile, D. Pretolani and G. Rago, Max Horn SAT and the Minimum Cut Problem in Directed Hypergraphs, Mathematical Programming 80 (1998), 213-237.

[8] J.-J. Hebrard, A linear algorithm for renaming a set of clauses as a Horn set, Theoretical Computer Science 124 (1994), 343-350.

[9] H. R. Lewis, Renaming a set of clauses as a Horn set, Journal of the ACM 25 (1978), 134-135.

FREQUENCY ASSIGNMENT – CASE STUDY
PART I – PROBLEM DEFINITION

Igor Pesek[1], Iztok Saje[2] and Janez Žerovnik[1,3]
[1]IMFM, Jadranska 19, Ljubljana
[2]Mobitel d.d., Vilharjeva 23, Ljubljana
[3]FME, University of Maribor, Smetanova 17, Maribor
igor.pesek@imfm.uni-lj.si
iztok.saje@mobitel.si
janez.zerovnik@imfm.uni-lj.si

Abstract: The rapid development of cellular telephone networks in recent years has increased the need for good solution techniques for the frequency assignment problem on cellular networks. The solution methods can be divided in two parts: exact optimization methods on the one hand, and heuristic search techniques on the other hand. As most variants of the problem are NP-hard, therefore use of heuristic is the only choice. In this report we give a formal definition of the optimization problem that appears in a part of design of a practical GSM network.

Keywords: frequency assignment, metaheuristics, local optimization

1. Introduction

Wireless communication is used in many different situations such as mobile telephony, radio and TV broadcasting, satellite communication, and military operations. In each of these situations a frequency assignment problem arises with application specific characteristics. Researchers have developed different modeling ideas for each of the features of the problem, such as the handling of interference among radio signals, the availability of frequencies, and the optimization criterion.

The rapid development of cellular telephone networks in recent years has increased the need for good solution techniques for the frequency assignment problem on cellular networks. A major difference between radio / television broadcasting and cellular phone networks is the need for an individual connection for every customer. In the GSM network the connections are established by assigning different channels to the connections with possible noise due to interference. While the definition of a channel can be more complex (using for example, time division) we will not distinguish between channel and frequency here. Frequencies are limited resource, so it is mandatory to reuse them. As soon as frequencies are reused at different transmitters, we might get interference at radio receiver. With the rapid growth of radio technology (broadcast, cellular systems, etc.) frequencies are heavily reused, thus it was mandatory to find efficient methods to minimize interference. Automatic Frequency Planning tools, mostly based on “graph coloring” methods, are widely used in practice. However, the commercial products usually work as “black box”, which prevent advanced users to explore all the possibilities to optimize their network using methods or settings that are adapted to the instances they need to solve. Therefore, although the problem itself is not new, it is interesting to study and test different algorithms, in particular to tune their parameters, on a real or realistic data. In this study, real GSM 900 MHz network data was used, and results are compared with the frequency plan implemented in the real network at the same time.

This report is structured as follows. In the next section we explain the practical problem and give its formal definition. In Section 3 we briefly outline the local search algorithms that are used in the experimental study that will be reported on in more detail in Part II of this paper.

2. Problem

Although new technologies are emerging, such as UMTS, which do not need explicit assignment of the channels, old technologies are still in use and with growing traffic and new services there is still a need to find the best solutions possible in order to have both content customers and higher profits. Companies have real problems. Our GSM network consists of 1822 900 MHz cells. Each cell needs one BCCH frequency from a given range. BCCH stands for Broadcast Control Channel and used by cell to continuously sending out identifying information about its cell site, the area code for the current location and some other important information.

There are several constraints to be taken into consideration:

- interference between two cells has to be minimized
- each cell has a different set of possible frequencies
- adjacent frequency is to be considered as a possible source of interference
- some cells have BCCH already allocated (given range is one channel only)
- some cells have additional penalty for usage of specific frequencies (border area, intermodulation, fixed and known non-GSM interferes)

This leads us to the formal definition of the problem which we will give below. The problem is called MIFAP (Minimum Interference Frequency Assignment Problem) because it is essentially equivalent to the MI-FAP problem of [4], or, more precisely, to its extension applied to some of the CALMA instances [cf. CALMA website, 1995].

Problem: MIFAP
Instance: set of base stations, each having lower and upper limit of the allowed frequencies, additional penalties for selecting some of the frequencies

Task: Find frequency assignment for each of the base stations that minimizes the cost:

$$f = \sum_{i=1}^{n} \left(c(i,s) + \sum_{j \in N(i)} \frac{p(i,j)}{2} \right)$$

Where:
p(i,j) ... is the cost for interaction between frequencies for cells *i* and *j*
N(i) … is the set of neighbors of cell
c(i,s) … cost for using frequency *s* at cell *i*

We were interested in computing competitive solution on the instance that corresponds to the existing GSM network. The instance is very large, containing more than 1800 cells, thus computing the optimal solution by searching entire solution space would take too long. We therefore decided to find solutions with help of the metaheuristics. Local search heuristics were chosen, and in particular we used two well-known general metaheuristics including the Tabu search and the Simulated Annealing, and the Petford-Welsh algorithm, which is a special heuristics that was designed for the problems that are of "graph-coloring type". We use two different neighborhoods in our search.

3. Local Search Based Algorithms and Their Neighborhoods

The use of a local search algorithm presupposes definitions of a problem and a neighborhood. Roughly speaking, a local search algorithm starts with an initial solution and then continually tries to find better solutions by searching the neighborhoods. The basic version of the local search is iterative improvement (also called hill climbing, steepest descent, local improvement) which, at each step, selects a random neighbor and moves only if the cost of the neighbor is better (or equal) to the cost of current solution. While the moves are always directed towards good (or better) solutions, iterative improvement tends to get trapped in local minima. Multistart of iterative improvement is an obvious solution to the problem. However, there are many other possibilities that are used and studied. Among the most popular are the Tabu search and the Simulated Annealing [10]. In this section we first describe how we build initial solution and in next section explain two neighborhoods used.

3.1 Initial Solution

In order to optimize and find good solutions by local search we have to build initial solution(s). We use a very simple construction: assign to each cell one frequency, which is chosen randomly among all possible frequencies for that respective cell. In this step we obviously did not try to optimize the solution and therefore the initial solution may be unfeasible or very expensive.

3.2 Neighborhoods

Given an instance p of a problem P, we associate a search space S to it. Each element $s \in S$ corresponds to a potential solution of p, and is called a state of p. Local search relies on a function N (depending on the structure of P) which assigns to each $s \in S$ its neighborhood $N(s) \subseteq S$. Each state $s' \in N(s)$ is called a neighbor of s. A local search algorithm starts from an initial state s_0 and enters a loop that navigates the search space, stepping from one state s_i to one of its neighbor's s_{i+1}. The neighborhood is usually composed by the states that are obtained by some local changes (called moves) from the current one.

EASYLOCAL++ is an object-oriented framework [2] that can be used as a general tool for the development and the analysis of local search algorithms in C++.

The two neighborhoods which we use are named Random choice and All Frequencies, respectively.

In the Random choice (or short RC) neighborhood we randomly select one cell between all the unfeasible or very expensive cells. Under unfeasible cells we understand the cells which have too strong interferences with some neighboring cell if using the same frequency and therefore would make solution unusable. Expensive cells on other hand don't make unfeasible solution, but there is very high penalty for using same frequency and therefore using them increases the solution cost significantly.

Next we check for the selected cell all the frequencies and between all the feasible frequencies we select one randomly and pass it to the heuristics.

In All Frequencies (AF) we check all the cells and for each cell select all the frequencies (also unfeasible ones) and pass them to the heuristics.

4. Metaheuristics

We used three of the local search algorithms: Tabu Search, Simulated Annealing and Petford-Welsh algorithm. In this section we will describe each metaheuristic in some detail.

4.1 Tabu Search

Tabu search is a local search strategy based on the idea to add memory to the search [1]. At each state s_i, tabu search explores a subset V of the current neighborhood $N(s_i)$. Among the

elements in V, the one that gives the minimum value of the cost function becomes the new current state s_{i+1}, independently of the fact whether $f(s_{i+1})$ is less or greater than $f(s_i)$.
Such a choice allows the algorithm to escape from local minima, but creates the risk of cycling among a set of states. In order to prevent cycling, the so-called tabu list is used, which determines the forbidden moves. This list stores the most recently accepted moves and the inverses of the moves in the list are forbidden (i.e., the moves that are leading again toward the just visited local minimum).

The simplest way to run the tabu list is as a queue of fixed size k, that is, when a new move is added to the list, the oldest one is discarded. There is a more general mechanism that assigns to each move that enters the list a random number of iterations, ranging from k_{min} to k_{max} (the values k_{min} and k_{max} are parameters of the method), that it should be kept in the tabu list. A move is removed from the list when its tabu period is expired. In this way the size on the list is not fixed, but varies dynamically in the interval k_{min} - k_{max}.

4.2 Simulated Annealing

Simulated Annealing (SA) [3] is a probabilistic algorithm, for which inspiration comes form annealing in metallurgy. By analogy with this physical process, each step of the SA algorithm replaces the current solution by a random "nearby" solution, chosen with a probability that depends on the difference between the corresponding function values and on a global parameter T (called the *temperature*), that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when T is large, but increasingly forces moves "downhill" as T goes to zero. The algorithm stops when the temperature is close enough to zero, which can also be implemented by stopping the algorithm if there is no move (or, alternatively, improvement of the solution) in a prescribed number of steps. We have also implemented the *constant temperature SA*, because in the literature it is reported to achieve good results [11]. In this case we do not decrease temperature, and we add another stopping criterion in terms of total number of steps (or CPU). It is well known that for SA fine tuning of the parameters is very important. Furthermore, it seems that there is no general tuning possible, and hence the tuning has to be done for each instance, or at least for problem domain. (i.e. if a set of problems comes from similar origin it is likely that the same or similar parameters will be suitable.)

4.3 Petford Welsh (PW)

Petford and Welsh proposed a randomized algorithm for 3-coloring which mimics the behavior of a physical process based on multi-particle system of statistical mechanics called the antivoter model [6]. The algorithm starts with a random initial 3-coloring of the input graph and then applies an iterative process. In each iteration a vertex creating a conflict is randomly taken uniformly and recolored according to some probability distribution. This distribution favors colors which are less represented in the neighborhood of the chosen vertex. There is a straightforward generalization of this algorithm to k-coloring, which behaves reasonably well on various types of graphs [5], [8], [9].

We use the same main idea for the MIFAP problem, with some natural generalizations. A set of colors is replaced by a finite subset of natural numbers corresponding to available frequencies. Simple constraints requesting different colors or frequencies at adjacent vertices are generalized to constraints depending on the edge weights and applying to frequencies assigned to adjacent vertices.

The frequency assignment algorithm is

Algorithm PW (Temperature, Time limit)

```
1.   assign available frequencies to nodes of the graph uniformly
random
2.   while  not stopping condition do
2.1        select a bad vertex v (randomly)
2.2        assign a new frequency to v
```

A bad vertex is selected uniformly random among vertices which are endpoints of some edge which violates a constraint. A new frequency is assigned at random from the set *F*. Sampling is done according to the probability distribution defined as follows:
The probability of frequency $i \in F$ to be chosen as a new frequency of the vertex *v* is proportional to

$$\exp(- S_i /T) = \theta^{-S\,i}$$

where S_i is the number of edges with one endpoint at *v* and violating the constraint provided frequency of *v* is set *i*. *T* is a parameter of the algorithm, called temperature for reasons explained later.

The second parameter of the algorithm is the time limit, given as the maximal number of iterations of the while loop. This is at the same time the number of calls to the function which computes a new frequency and also the number of feasible solutions of the problem generated (some of them may be counted more than once).

The stopping condition is: either a proper assignment was found or a time limit was reached. In the later case, the best solution found (with fewest constraints violated) is reported as an approximate solution.

Choosing the optimal temperature and the time limit is in general an open problem. In the rest of the section we give some remarks on the parameters of the algorithm PW.

In the coloring problem, S_i is simply the number of neighbors of vertex *v* colored by *i*. The original algorithm of Petford and Welsh (for coloring) uses probabilities proportional to 4^{Si}, which corresponds to $T\approx 0.72$. Larger values of *T* result in higher probability of accepting a move which increases the number of bad edges. With low values of *T*, the algorithm behaves very much like iterative improvement.

T is a parameter of the algorithm, which may be called temperature because of the analogy to the temperature of the simulated annealing algorithm and to the temperature of the Boltzmann machine neural network. These analogies follow from the following simple observation. Let us denote the old color of the current vertex by *i* and the new color by *j*. The number of bad edges *E'* after the move is

$$E' = E - S_i + S_j$$

where *E* is the number of bad edges before the change. We define $\Delta E = E - E' = S_i - S_j$. At each step, the frequency *i* is fixed and hence S_i and *E* are fixed. Consequently, it is equivalent to define the probability of choosing color *j* to be proportional to either $\exp(-S_j /T)$, $\exp(\Delta E /T)$ or $\exp(-E' /T)$.

Finally, recall that the number of bad edges is a usual definition of energy function in simulated annealing and in Boltzmann machine. Therefore, the algorithm PW is in close relationship to constant temperature operation of the generalized Boltzmann machine (for details, see [7] and the references there). The major difference is in the "firing" rule. While in Boltzmann machine model all neurons are fired with equal probability, in PW algorithm only bad vertices are activated. The algorithm PW differs from constant temperature simulated annealing in the acceptance criteria for the moves improving the cost function.

These are always accepted by simulated annealing and only according to some (high) probability in the PW algorithm.

5. Conclusion

In this report we described a problem that occurs in real world and proposed two new neighborhoods that may be used with three different local search based heuristics. Experimental results are given in the Part II of this paper that appears in the same Proceedings.

References

[1] F. Glover and M. Laguna, Tabu search. Kluwer Academic Publishers, 1997.

[2] L. Di Gaspero and A. Schaerf, EasyLocal++: An object-oriented framework for flexible design of local search algorithms. *Software—Practice and Experience* 33 (2003) 733–765.

[3] S. Kirkpatrick and C. D. Gelatt and M. P. Vecchi, Optimization by Simulated Annealing, *Science* 220 (1983) 671 – 680.

[4] K. I. Aardal, S.P.M. van Hoesel, A.M.C. Koster, C. Mannino, and A. Sassano, Models and Solution Techniques for Frequency Assignment Problems, *4OR* 1 (2003) 261–317. An update with same title and authors appeared in: *Annals of Operations Research* 153 *(2007)* 79 – 129.

[5] B. Chamaret, S. Ubeda and J. Žerovnik, A Randomized Algorithm for Graph Coloring Applied to Channel Allocation in Mobile Telephone Networks, in *Proceedings of the 6th International Conference on Operational Research KOI'96* (T.Hunjak, Lj.Martić, L.Neralić, eds.) 1996, 25-30.

[6] A. Petford and D.Welsh, A Randomised 3-colouring Algorithm, *Discrete Mathematics* 74 (1989) 253-261.

[7] J. Shawe-Taylor and J. Žerovnik, Boltzmann Machines with Finite Alphabet, *Artificial Neural Networks* 1 (1992) 391-394.

[8] J. Shawe-Taylor and J. Žerovnik, Analysis of the Mean Field Annealing Algorithm for Graph Colouring, *Journal of Artificial Neural Networks*, 2 (1995) 329-340.

[9] J. Žerovnik, A Randomized Algorithm for k-colorability, *Discrete Mathematics* 131 (1994) 379-393.

[10] E. Aarts and J.K. Lenstra, (eds.) Local Search in Combinatorial Optimization, Wiley New York 1997.

[11] A. Vesel and J. Žerovnik, Improved lower bound on the Shannon capacity of C_7. *Information Procesing Letters* 81 (2002) 277 - 282.

FREQUENCY ASSIGNMENT – CASE STUDY
PART II – COMPUTATIONAL RESULTS

Igor Pesek[1], Iztok Saje[2] and Janez Žerovnik[1,3]
[1]IMFM, Jadranska 19, Ljubljana
[2]Mobitel d.d., Vilharjeva 23, Ljubljana
[3]FME, University of Maribor, Smetanova 17, Maribor
igor.pesek@imfm.uni-lj.si
iztok.saje@mobitel.si
janez.zerovnik@imfm.uni-lj.si

Abstract: The rapid development of cellular telephone networks in recent years has increased the need for good solution techniques for the frequency assignment problem on cellular networks. The solution methods can be divided in two parts: exact optimization methods on the one hand, and heuristic search techniques on the other hand. As most variants of the problem are NP-hard, therefore use of heuristic is the only choice. In this report we give a formal definition of the optimization problem that appears in a part of design of a practical GSM network. In this part we show and describe use of several heuristics and neighborhoods. Preliminary results of the computational experiments are very promising and can be used by the company.

Keywords: frequency assignment, metaheuristics, local optimization

1. Introduction

Wireless communication is used in many different situations such as mobile telephony, radio and TV broadcasting, satellite communication, and military operations. In each of these situations a frequency assignment problem arises with application specific characteristics. Researchers have developed different modeling ideas for each of the features of the problem, such as the handling of interference among radio signals, the availability of frequencies, and the optimization criterion.

Frequencies are limited resource, so it is mandatory to reuse them. As soon as frequencies are reused at different transmitters, we might get interference at radio receiver. With rapid growth of radio technology (broadcast, cellular systems etc.) frequencies are heavily reused, thus it was mandatory to find efficient methods to minimize interference. Automatic Frequency Planning tools are widely used, based on "graph coloring" methods. While problem itself is not new, it is interesting case to test different algorithms. Real GSM 900 MHz network data was used, and results are compared with frequency plan implemented in the real network at the same time.

This report is structured as follows. Theoretical aspects of the problem are described in the Part I [1], whereas in this part experimental results are described and discussed.

2. The Problem Description

Problem instance, its cost function and goal are described in Part I [1]. In this section we will explain cost function more precisely and describe the dataset formed from three files in details.

Problem: MIFAP
Instance: set of base stations, each having lower and upper limit of the allowed frequencies, additional penalties for selecting some of the frequencies
Task: Find frequency assignment for each of the base stations that minimizes the cost:

$$f = \sum_{i=1}^{n} \left(c(i,s) + \sum_{j \in N(i)} \frac{p(i,j)}{2} \right)$$

Where:
p(i,j) ... cost for interaction between frequencies for cells *i* and *j*
N(i) … the set of neighbors of cell *i*
c(i,s) … cost for using frequency *s* at cell *i*

More precisely:
c(i,s) = g(i,s) + k(i,s)
where
$g(i,s)$ … cost for using frequency *s* at cell *i*
$k(i,s)$ … additional cost for using frequency *s* at cell *i*

The neighbours *N(i)* for each cell were determined experimentally measuring interferences and using statistical methods. The cost *c(i,s)* represents the magnitude of the interferences between two cells. It is important to note that the cost *c(i,s)* is composed from two values *g(i,s)* and *k(i,s)*. First value represent interferences and second value represent cost that is the result of some special interference, such as border area, intermodulation, fixed and known non-GSM interferers.

2.1 Dataset

Dataset contains three different files. First file contains list of all cells, each line containing the name of the cell, the range (or span) from which we can choose the frequencies for the cell and also it’s currently assigned frequency. Second file contains the neighbors for each cell and the penalties for using some frequency *s* at cell *i*. For each neighbor we state penalty for using same frequency and also first and second adjacent frequency. Finally, the last file contains additional penalties for using particular frequency.

3. Algorithms and Their Neighborhoods

In this section we first describe the framework EasyLocal++ that we used for our experiments, followed by explanation of how we build initial solution and in next section explain two neighborhoods used. In last section we describe heuristics we used in searching the best solution.

3.1 EasyLocal++

Easylocal++ is an object-oriented framework that can be used as a general tool for the development and the analysis of local search algorithms in C++. The basic idea behind Easylocal++ is to capture the essential features of most local search techniques and their possible compositions. The framework provides a principled modularization for the design of local search algorithms and exhibits several advantages with respect to directly implementing the algorithm from scratch, not only in terms of code reuse but also in methodology and conceptual clarity.

3.2 Initial Solution

In order to optimize and find good solutions we had to build initial solution. We assigned to each cell one frequency, which was chosen randomly between all possible frequencies for that respective cell. In this step we didn’t optimize solution and therefore initial solution could be unfeasible and very expensive. We decided to leave the optimization part of the solution to heuristics.

3.3 Neighborhoods

Using EasyLocal++ enabled us to define various neighborhoods and then later use them with different heuristics.

First we will define two neighborhoods which we named Random choice and All Frequencies, respectively.

In the Random choice (or short RC) neighborhood we randomly select one cell between all the unfeasible or very expensive cells. Under unfeasible cells we understand the cells which have to strong interferences with some neighboring cell if using the same frequency and therefore would make solution unusable. Expensive cells on other hand don't make unfeasible solution, but there is very high penalty for using same frequency and therefore using them increases the solution cost significantly.

Next we check for the selected cell all the frequencies and between all the feasible frequencies we select one randomly and pass it to the heuristics.

In All Frequencies (AF) we check all the cells and for each cell select all the frequencies (also unfeasible ones) and pass them to the heuristics.

3.4 Metaheuristics

In our experiments we employed three different metaheuristics Tabu Search, Simulated Annealing and Petford Welsh, respectively. For description and variants of the metaheuristics used we refer to Part I [1].

4. Experimental Analysis

This section reports on the preliminary results of the computational studies of our neighborhoods and used heuristics. First we will describe our experimental setting, then we will describe experiments with parameter tuning. Finally, we will report on the best results found.

4.1 Experimental Setting

Experiments were performed on an Intel Pentium 4 (3.0 GHz) processor running Microsoft Windows XP. The algorithms have been coded in C++ exploiting the framework EasyLocal++ , and the executables were obtained using the GNU C/C++ compiler (v. 3.4.4). Stopping criterion for experiments with Tabu Search was the number of idle iterations. By later we mean the number of the iterations with no improvement of the cost function. In other cases we used the total number of iterations as stopping criterion.

A typical number of idle iterations in our experiments were 40000. Because of the size of the problem, running times were from 6000 to 100000 seconds. Note that the running times depend mainly on the parameters used.

We tested how the heuristics perform with random starting solution and its performance when improving some good solution, respectively.

4.2 Experiments for Parameter Tuning

In our experiments we tested several different heuristics combined with our neighborhoods. Since all the heuristics are sensitive to the parameter tuning, we had to tune them accordingly.

In the Tabu Search heuristic one of the parameters is tabu list size, which plays crucial role. We decided to test which of the following tabu lists sizes is the best.

Each of the tables presenting the experimental results has following columns: Column Neighborhood tells us which neighborhood we used, namely Random Choice (RC) or All Frequencies (AF), respectively. Then follow columns for specific parameters of the heuristics, which were already described above. Column Initial solution tells us the cost of the initial solution. Here we used two different initial solutions; first solution was randomly generated as is described in section 3.2. Second initial solution was a good solution, which was a result of a preliminary run of one of the heuristics.

Stopping criterion for Tabu Search heuristics was the number of idle iterations, which is the number of iterations, elapsed from the last strict improvement. Stoping criterion for Simulated Annealing was either the point when temperature reached value 0 or total number of iterations was reached. Stoping criterion for the Petford Welsh heuristic was the total number of iterations.

Table 1: Results for Tabu search heuristic with expensive initial solution

Neighborhood	**Tabu List**	**Idle iterations (stopping criterion)**	**Initial solution**	**Result**
RC	5 10	20000	119476	36123
RC	50 100	20000	119476	35772
AF	5 10	20000	119476	38547
AF	50 100	20000	119476	38655

Table 2: Results for Tabu search heuristic with good initial solution

Neighborhood	**Tabu List**	**Idle iterations**	**Initial solution**	**Result**
RC	5 10	40000	38655	35364
RC	50 100	40000	38655	35910
AF	5 10	40000	38655	38547
AF	50 100	40000	38655	37909

As it is shown in Table 1 and 2, neighborhood Random Choice clearly outperforms All Frequency neighborhood. It is interesting to see that we don't obtain much better results when we start our search with good initial solution.

Simulated annealing uses three parameters: temperature, cooling rate and the number of samples at each temperature. We experimented with both neighborhoods. The stopping criterion for this heuristic was either the temperature near to 0 or the maximum number of the iterations was reached.

Table 3: Results for Simulated Annealing heuristic with randomly builded initial solution

Neighbor-hood	Tempe-rature	Cooling rate	Number of samples	Initial solution	Number of iterations	Result
RC	200	0.98	300	119476	50000	57646
RC	300	0.98	300	119476	50000	57268
AF	200	0.98	300	119476	50000	39568
AF	300	0.98	300	119476	50000	38542

Based on these results we decided that All Frequencies neighborhood performs better with Simulated Annealing. Therefore we ran some additional experiments with this neighborhood and with some new parameters.

Table 4: Results for Simulated Annealing heuristic with different starting temperature

Neighbor-hood	Tempe-rature	Cooling rate	Number of samples	Initial solution	Number of iterations	Result
AF	200	0.975	500	48657	60000	41526
AF	100	0.975	500	48657	60000	41201
AF	50	0.975	500	48657	60000	40963
AF	20	0.975	500	48657	60000	41431
AF	5	0.975	500	48657	60000	41617
AF	3	0.975	500	48657	60000	41687
AF	1	0.975	500	48657	60000	42041

For last experiments we used well known Petford Welsh algorithm. This algorithm needs only one parameter temperature and uses Random Choice neighborhood.

Table 5: Results for Petford Welsh heuristic with constant temperature

Neighborhood	Temperature	Number of iterations (stopping criterion)	Initial solution	Result
AF	200	50000	47884	61004
AF	100	50000	47884	48710
AF	50	50000	47884	47440
AF	20	50000	47884	46931
AF	5	50000	47884	46173
AF	3	50000	47884	46394
AF	1	50000	47884	47320

Experiments show that for Petford Welsh algorithm the best temperature for this problem is 5, although some more experiments with longer runs and different initial solution should be made. As for Simulated Annealing with AF neighborhood experiments show that currently there is not any preferable temperature, but it is important to note, that there should be done more extensive experiments in order to make more precise conclusion.

4.3 Best Solution

In this initial study we continued with more extensive experiments with Simulated Annealing and AF neighborhood. The cost of the best solution we found so far is 30271. The search was performed using Simulated Annealing with Random Choice neighborhood and following parameters. Starting temperature was set to 3, cooling rate was 0.985 and number of samples was equal to 300. Cost of the initial solution was 30452.

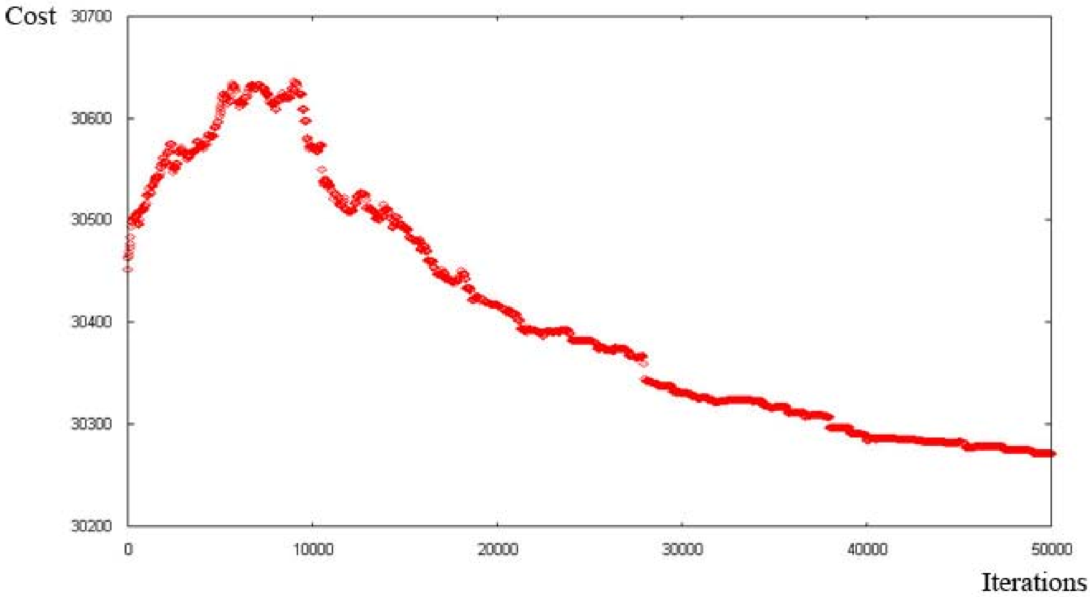


Figure 1: The graph of the best solution search

Experiments were also independently performed in the company that provided us with the real data for this problem. The best result they obtained with their implementation of some optimizing techniques was 30452 and was very close to our best solution.

5. Conclusion

In this report we described a problem that occurs in real company and proposed two new neighborhoods and tested three different heuristics with parameters. Preliminary results show that the best combination of heuristics and neighborhood is Simulated Annealing with All frequencies neighborhood.

For future we plan to optimize search methods and neighborhoods and then do extensive parameter tuning on all described heuristics.

References

[1] I. Pesek, I. Saje and J. Žerovnik, Frequency Assigment – Case study, Part I – Problem Definition, this volume of proceedings.

CIRCULAR CHROMATIC NUMBER OF TRIANGLE-FREE HEXAGONAL GRAPHS

Petra Šparl and Janez Žerovnik
University of Maribor
Smetanova 17
SI-2000 Maribor, Slovenia
and
IMFM,
Jadranska 19,
SI-1000 Ljubljana, Slovenia.

An interesting connection between graph homomorphisms to odd cycles and circular chromatic number is presented. By using this connection, bounds for circular chromatic number of triangle-free hexagonal graphs (i.e. induced subgraphs of triangular lattice) are given.

Keywords: graph homomorphism, circular chromatic number, triangle-free hexagonal graph
2000 Mathematics Subject Classification: 05C15, 68R10

Introduction

Suppose G and H are graphs. A homomorphism from G to H is a mapping f from $V(G)$ to $V(H)$ such that $f(x)f(y) \in E(H)$ whenever $xy \in E(G)$. Homomorphisms of graphs are studied as a generalization of graph colorings. Indeed, a vertex coloring of a graph G with n-colors is equivalent to a homomorphism from G to K_n. Therefore, the term H*-coloring* of G has been employed to describe the existence of a homomorphism of a graph G into the graph H. In such a case graph G is said to be H*-colorable*. Graph homomorphisms are widely studied in different areas, see [2,3] and the references there. One of the approaches is deciding whether an arbitrary graph G has a homomorphism into a fixed graph H. The main result, regarding the complexity of H-coloring problem, was given by Hell and Nešetril in 1990 [4]. They proved that H-coloring problem is NP-complete, if H is non-bipartite graph and polynomial otherwise. Several restrictions of the H-coloring problem have been studied [3]. One of the restricted H-coloring problems was studied in [5], where H is an odd cycle and G an arbitrary, the so-called, *hexagonal graph*, which is an induced subgraph of a triangular lattice. It was shown that any triangle-free hexagonal graph G is C_5-colorable. This result will be used in section 4 to obtain upper bounds for circular chromatic number of triangle-free hexagonal graphs.

Another interesting approach regarding homomorphisms can be found in the literature. In [8] author discuses the connection between graph homomorphisms and so called circular colorings. A partial result of this connection in a slightly different form is given in Section 3.

Circular coloring and circular chromatic number are natural generalizations of ordinary graph coloring and chromatic number of a graph. The circular chromatic number was introduced by Vince in 1988, as "the star-chromatic number" [6]. Here we present an equivalent definition of Zhu [7].

Definition 1 *Let C be a circle of (Euclidean) length r. An r-circular coloring of a graph G is a mapping c which assigns to each vertex x of G an open unit length arc $c(x)$ of C, such that for every edge $xy \in E(G)$, $c(x) \cap c(y) = \varnothing$. We say a graph G is r-circular colorable if there is an r-circular coloring of G. The circular chromatic number of a graph G, denoted by $\chi_c(G)$, is defined as*

$$\chi_c(G) = \inf\{r : G \text{ is } r\text{-circular colorable}\}.$$

For finite graphs G it was proved [1,6,7] that the infimum in the definition of the circular chromatic number is attained, and the circular chromatic numbers $\chi_c(G)$ are always rational.

In this paper we present a connection between homomorphisms to odd cycles and circular chromatic number. Using this connection we prove:

• For an arbitrary graph G the following two statements are equivalent:

(i) k is the biggest positive integer for which there exists a homomorphism $f : G \to C_{2k+1}$,

(ii) $2 + \frac{1}{k+1} < \chi_c(G) \le 2 + \frac{1}{k}$.

• For any triangle-free hexagonal graph G it holds $2 \le \chi_c(G) \le \frac{5}{2}$.

• For any triangle-free hexagonal graph G with odd girth $2K+1$ it holds $\frac{2K+1}{K} \le \chi_c(G) \le \frac{5}{2}$.

The rest of the paper is organized as follows. In Section 2 some definitions and results, which will be needed later on, are given. In Section 3 the connection between graph homomorphisms and circular chromatic number is presented. In Section 4 the proposition presented in Section 3 is improved and bounds for circular chromatic number of triangle-free hexagonal graphs are given. In the last section a conjecture regarding circular chromatic number of triangle-free hexagonal graphs is set up.

Preliminaries

Let G and H be simple graphs. It is well known that the existence of a homomorphism $\varphi : G \to H$ implies the inequality $\chi(G) \le \chi(H)$. Namely, for a homomorphism $\psi : H \to K_n$, the compositum $\psi \circ \varphi : G \to K_n$ is a proper n-coloring of G.

It is not difficult to see that similar holds for circular chromatic numbers of graphs G and H.

Lemma 2 *If there is a homomorphism* $f : G \to H$, *then* $\chi_c(G) \le \chi_c(H)$.

Proof. Let the Euclidean length of the cycle C be equal to r and let $c : V(H) \to C$ be an r-circular coloring of H. Let us show that the compositum $c \circ f : V(G) \to C$ is an r-circular coloring of G. For any edge $xy \in E(G)$ holds $f(x)f(y) \in E(H)$. Since c is an r-circular coloring of H it holds $c(f(x)) \cap c(f(y)) = \varnothing$ for any $xy \in E(G)$ and hence $c \circ f$ is an r-circular coloring of G. Therefore $\chi_c(G) \le \chi_c(H)$. □

Let us present another approach to r-circular coloring, which will be needed in the following section.

The circle C may be cut at an arbitrary point to obtain an interval of length r, which may be identified with the interval $[0, r)$. For each arc $c(x)$ of C, we let $c'(x)$ be the initial point of $c(x)$ (where $c(x)$ is viewed as going around the circle C in the clockwise direction). An r-circular coloring of G can be identified with a mapping $c' : V \to [0, r)$ such that

$1 \le \left| c'(x) - c'(y) \right| \le r - 1$, [7].

For a later reference we introduce the following definition:

Definition 3 *For an arbitrary odd cycle* C_{2k+1} *let* $F : [0, \frac{2k+1}{k}) \to C_{2k+1}$ *be a mapping such that*

$$\text{for } x \in [\frac{i}{k}, \frac{i+1}{k}); i \in \{0,1,...,2k\} : F(x) = \begin{cases} 0 & ; \quad i = 0 \\ 2k - 2i + 1 & ; \quad 1 \le i \le k \\ 4k - 2i + 2 & ; \quad k < i \le 2k \end{cases}.$$

It is not difficult to see that F maps the interval $[0, 2 + \frac{1}{k})$ into vertices $\{0,1,...,2k\}$ of the cycle C_{2k+1} as Figure 1 shows.

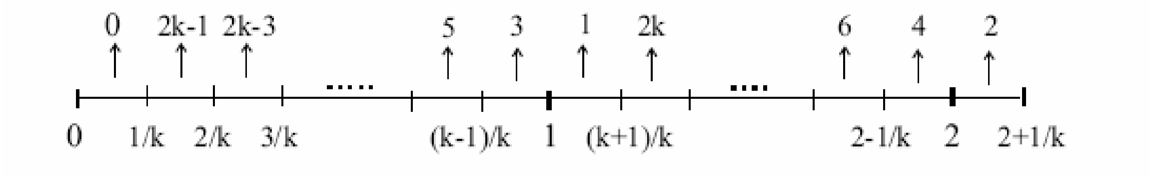


Figure 1. The functional values of subintervals $[\frac{i}{k}, \frac{i+1}{k})$ of the interval $[0, \frac{2k+1}{k})$ of function F defined in Definition 3.

Considering Definition 3 and Figure 1 one can easily find out that the following lemma holds, thus we omit technical details of the proof:

Lemma 4 *Let* $F : [0, 2 + \frac{1}{k}) \to C_{2k+1}$ *be a mapping from Definition 3. For any vertices* $x, y \in [0, 2 + \frac{1}{k})$ *the following statements are equivalent:*

(*i*) $\left| F(x) - F(y) \right| = 1$,

(*ii*) $1 - \frac{1}{k} < \left| x - y \right| < 1 + \frac{2}{k}$.

The connection between graph homomorphisms to odd cycle and circular chromatic number

The Proposition 5 below follows from results given in [8]. For completeness we give an independent proof of the Proposition in the continuation.

Proposition 5 *For any finite graph* G *there exists a homomorphism* $f : G \to C_{2k+1}$ *iff* $\chi_c(G) \le \frac{2k+1}{k} = 2 + \frac{1}{k}$.

Proof. Let $f : G \to C_{2k+1}$ be a homomorphism. Considering Lemma 2 and the well known equality $\chi_c(C_{2k+1}) = \frac{2k+1}{k}$, we have $\chi_c(G) \le \frac{2k+1}{k}$.

Now suppose $\chi_c(G) \le \frac{2k+1}{k}$. Therefore, there exists a $\frac{2k+1}{k}$-circular coloring of G, which can be identified with a mapping $c^{'}: V(G) \to [0, \frac{2k+1}{k})$, such that

$$1 \le \left| c^{'}(x) - c^{'}(y) \right| \le 1 + \frac{1}{k}, \text{ for every edge } xy \in E(G). \tag{1}$$

Let $F: [0, \frac{2k+1}{k}) \to C_{2k+1}$ be a mapping from Definition 3. We will prove that the composition $F \circ c^{'}: V(G) \to C_{2k+1}$ is a homomorphism from G to C_{2k+1}.

Let $xy \in E(G)$. We have to show that $F(c^{'}(x))F(c^{'}(y)) \in E(C_{2k+1})$, which is equivalent to $\left| F(c^{'}(x)) - F(c^{'}(y)) \right| = 1$. Suppose the opposite ($\exists x_0 y_0 \in E(G)$ such that $\left| F(c^{'}(x_0)) - F(c^{'}(y_0)) \right| \ne 1$). From Lemma 4 it follows that the assertion $(1 - \frac{1}{k} < \left| c^{'}(x) - c^{'}(y) \right| < 1 + \frac{2}{k})$ is not true. Hence either $\left| c^{'}(x) - c^{'}(y) \right| \le 1 - \frac{1}{k} < 1$ or $\left| c^{'}(x) - c^{'}(y) \right| \ge 1 + \frac{2}{k} > 1 + \frac{1}{k}$. Both cases are contradictious to the inequalities (1). Therefore, $\left| F(c^{'}(x)) - F(c^{'}(y)) \right| = 1$ for every $xy \in E(G)$ or mapping $F \circ c^{'}: G \to C_{2k+1}$ is a homomorphism.

Corollaries of Proposition 5

The Proposition 5 can be improved further.

Corollary 6 *For an arbitrary graph G the following two statements are equivalent:*
(i) k is the biggest positive integer for which there exists a homomorphism $f: G \to C_{2k+1}$,
(ii) $2 + \frac{1}{k+1} < \chi_c(G) \le 2 + \frac{1}{k}$.

Proof. $(i) \Rightarrow (ii)$: Since $f: G \to C_{2k+1}$ is a homomorphism, by Proposition **5**, we have $\chi_c(G) \le 2 + \frac{1}{k}$. Because there does not exist a homomorphism from G to $C_{2(k+1)+1}$, the Proposition 5 implies $\chi_c(G) > \frac{2(k+1)+1}{k+1} = 2 + \frac{1}{k+1}$.

$(ii) \Rightarrow (i)$: Because of the inequality $\chi_c(G) \le 2 + \frac{1}{k}$, by Proposition 5, there exists a homomorphism $f: G \to C_{2k+1}$. Suppose that there exists a positive integer $n \ge k+1$ such that there is a homomorphism from G to C_{2n+1}. By Proposition 5 we have $\chi_c(G) \le 2 + \frac{1}{n} \le 2 + \frac{1}{k+1}$, which is a contradiction. □

Let G be an arbitrary triangle-free hexagonal graph. It is interesting to ask what is the circular chromatic number of G. Since an even cycle C_{2n} can be a subgraph of a hexagonal graph, the well known equality $\chi_c(G) = 2$ implies the lower bound i.e. $2 \le \chi_c(G)$. To obtain the upper bound we use the result from [5]:

Theorem 7 *Let G be a triangle-free hexagonal graph. Then there exists a homomorphism* $\varphi : G \to C_5$.

Since there exists a homomorphism from G into C_5 Proposition 5 implies the inequality $\chi_c(G) \le \frac{5}{2}$. So we proved the following result:

Proposition 8 *For any triangle-free hexagonal graph G it holds* $2 \le \chi_c(G) \le \frac{5}{2}$.

Odd girth of a graph G is the length of a shortest odd cycle in G. If there is no odd cycle, i.e. the graph is bipartite, then the odd girth is undefined. Note that the smallest odd cycle which can be realized as a triangle free hexagonal graph is C_9. Clearly, for a graph with odd girth $2K+1$, there is no homomorphism $f : G \to C_{2(K+1)+1}$, and hence

Proposition 9 *For any triangle-free hexagonal graph G with odd girth $2K+1$ it holds*
$2 + \frac{1}{K+1} \le \chi_c(G) \le \frac{5}{2}$.

Final remarks

In [5] we conjectured that every triangle-free hexagonal graph is C_7-colorable. If this conjecture is true then it improves the upper bound of Proposition 8. Therefore, we set another conjecture

Conjecture 10 *For any triangle-free hexagonal graph G it holds* $2 \le \chi_c(G) \le \frac{7}{3}$.

References

[1] A. Bondy and P. Hell, *A note on the star chromatic number*, J.Graph Theory 14 (1990), 479-482.

[2] G. Hahn and G. McGillivray, *Graph homomorphisms: computational aspects and infinite graphs*, submitted for publication.

[3] G. Hahn and C. Tardif, *Graph homomorphisms: structure and symmetry, in: Graph symmetry*, ASI ser.C, Kluwer, 1997, pp.107-166.

[4] P.Hell and J. Nešetril, *On the complexity of H-colorings*, J.Combin. Theory B 48 (1990), 92-110.

[5] P. Šparl, J. Žerovnik, *Homomorphisms of hexagonal graphs to odd cycles*, Discrete mathematics 283 (2004), 273-277.

[6] A.Vince, *Star chromatic number*, J. Graph Theory 12 (1988), 551-559.

[7] X. Zhu, *Circular chromatic number*: a survey, Discrete mathematics 229 (2001), 371-410.

[8] X. Zhu, *Circular coloring and graph homomorphism*, Bulletin of the Australian Mathematical Society 59 (1999), 83-97.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 2:

Stochastic and Combinatorial Optimization

A NETWORK FLOW IMPLEMENTATION OF A MODIFIED WORK FUNCTION ALGORITHM FOR SOLVING THE k-SERVER PROBLEM

Alfonzo Baumgartner
Faculty of Electrical Engineering, University of Osijek
Kneza Trpimira 2b, 31000 Osijek, Croatia
E-mail: Alfonzo.Baumgartner@etfos.hr

Robert Manger
Department of Mathematics, University of Zagreb
Bijenička cesta 30, 10000 Zagreb, Croatia
E-mail: Robert.Manger@math.hr

Željko Hocenski
Faculty of Electrical Engineering, University of Osijek
Kneza Trpimira 2b, 31000 Osijek, Croatia
E-mail: Zeljko.Hocenski@etfos.hr

Abstract. We study a modification of the well known work function algorithm (WFA) for solving the on-line k-server problem. Our modified WFA is based on a moving window, i.e. on the approximate work function that takes into account only a fixed number of most recent on-line requests. In this paper we describe in detail a network flow implementation of the modified WFA. We also present theoretical estimates and experimental measurements dealing with the computational complexity of the implemented algorithm.

Keywords: on-line problems, on-line algorithms, the k-server problem, the work function algorithm (WFA), moving windows, implementation, network flows, computational complexity.

1. Introduction

This paper deals with the *k-server problem* [8], which belongs to a broader family of *on-line problems* [5]. In the k-server problem we must decide how k mobile servers should serve a sequence of on-line requests. To solve the k-server problem, we need a suitable *on-line algorithm* [5]. The goal of such an algorithm is not only to serve requests as they arrive, but also to minimize the total cost of serving.

There are various algorithms for solving the k-server problem found in literature. From the theoretical point of view, the most important one is the *work function algorithm* (WFA) [1,7]. In spite of its interesting properties, the WFA is seldomly used in practice due to its prohibitive and ever-increasing computational complexity.

In a previous paper [2] we have proposed a simple modification of the WFA, which is based on a moving window. Our modified WFA is much more suitable for practical purposes since its computational complexity can be controlled by the window size. We have demonstrated in [2] that the performance of the WFA in terms of the incurred total cost is not degraded too much by the introduced modification. More precisely, we have shown that with a reasonably large window the modified WFA achieves the same or almost the same performance as the original WFA.

The aim of this paper is to specify in more detail how the modified WFA from [2] can efficiently be implemented by using network flows. Also, the aim is to give theoretical estimates and experimental measurements of the associated computational complexity. The ultimate goal is to prove that the modified WFA is really an efficient algorithm for solving the k-server problem.

The paper is organized as follows. Section 2 gives preliminaries about the k-server problem and the corresponding on-line algorithms including the modified WFA. Section 3 describes in detail our implementation of the modified WFA, where a single step of the algorithm is reduced to computing minimal-cost maximal flows in a set of suitably constructed networks. The same section also specifies a network flow algorithm that computes the required flows efficiently by employing special properties of the involved networks. Section 4 presents theoretical and experimental results on computational complexity. The theoretical estimates first determine how the running time of our implemented algorithm is related to problem parameters and window size. The experimental measurements then give a clear idea how much our implementation is indeed faster than the original WFA. The final Section 5 gives concluding remarks.

2. Preliminaries

In the *k-server problem* [8] we have k servers each of which occupies a location (point) in a fixed metric space M consisting of altogether m locations. Repeatedly, a request r_i at some location $x \in M$ appears. Each request must be served by a server before the next request arrives. To serve a new request at x, an on-line algorithm must move a server to x unless it already has a server at that location. The decision which server to move may be based only on the already seen requests $r_1, r_2, \ldots, r_{i-1}, r_i$, thus it must be taken without any information about the future requests $r_{i+1}, r_{i+2}, \ldots$. Whenever the algorithm moves a server from location a to location b, it incurs a cost equal to the distance between a and b in M. The goal is not only to serve requests, but also to minimize the total distance moved by all servers.

As a concrete instance of the k-server problem, let us consider the set M of $m = 5$ Croatian cities shown in Figure 1 with distances given. Suppose that $k = 3$ different hail-defending rocket systems are initially located at Osijek, Zagreb and Split. If the next hail alarm appears for instance in Karlovac, then our hail-defending on-line algorithm has to decide which of the tree rocket systems should be moved to Karlovac. Seemingly the cheapest solution would be to move the nearest system from Zagreb. But such a choice could be wrong if, for instance, all forthcoming requests would appear in Zagreb, Karlovac and Osijek and none in Split.

The simplest on-line algorithm for solving the k-server problem is the *greedy algorithm* (GREEDY) [5]. It serves the current request in the cheapest possible way, by ignoring history altogether. Thus GREEDY sends the nearest server to the requested location.

A slightly more sophisticated solution is the *balanced algorithm* (BALANCE) [8], which attempts to keep the total distance moved by various servers roughly equal. Consequently, BALANCE employs the server whose cumulative distance traveled so far plus the distance to the requested location is minimal.

The most celebrated solution to the k-server problem is the *work function algorithm* (WFA) [1,7]. To serve request r_i, the WFA switches from the current server configuration $S^{(i-1)}$ to a new configuration $S^{(i)}$, obtained from $S^{(i-1)}$ by moving one server

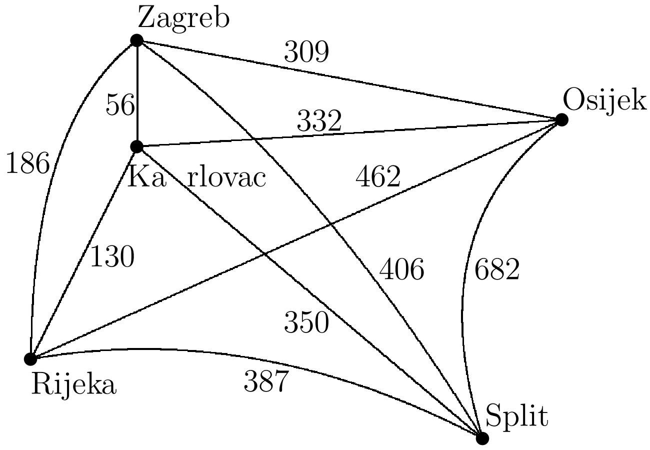


Figure 1: An instance of the k-server problem.

into the requested location (if necessary). Among k possibilities (any of k servers could be moved) $S^{(i)}$ is chosen so that it minimizes the so-called *work function* (WF). More precisely, $S^{(i)}$ is chosen so that $C_{\text{OPT}}(S^{(0)}, r_1, r_2, \ldots, r_i, S^{(i)}) + d(S^{(i-1)}, S^{(i)})$ becomes minimal. The WF is defined here as a sum of two parts. The first part is the optimal cost of starting from $S^{(0)}$, serving in turn $r_1, r_2, \ldots, r_i$, and ending up in $S^{(i)}$. The second part is the distance traveled by a server to switch from $S^{(i-1)}$ to $S^{(i)}$.

Our modification of the WFA, denoted as the w-WFA, is based on the idea that the sequence of previous requests and configurations should be examined through a moving window of size w. More precisely, in its i-th step the w-WFA acts as if r_{i-w+1}, r_{i-w+2}, $\ldots$, r_{i-1}, r_i was the whole sequence of previous requests, and as if $S^{(i-w)}$ was the initial configuration of servers.

Note that an on-line algorithm can only approximate the performance of the corresponding optimal off-line algorithm OPT, which knows the whole input in advance and deals with input data as they arrive at minimum total cost. A desirable property of an on-line algorithm is its *competitiveness* [11]. Vaguely speaking, an algorithm is competitive if its performance is only a bounded number of times worse than that of OPT on each input. It has been proved [1,5,8] that among the considered algorithms only the WFA is competitive. This is the reason why the WFA is so important, and why it is worth trying to mimic its behaviour by the w-WFA.

3. Implementation

In order to implement the w-WFA, we first consider the off-line version of the k-server problem, i.e. the version where the whole sequence of requests is known in advance. We start from the fact that the optimal off-line algorithm OPT can be implemented relatively easily by network flow techniques [3]. Namely, according to [4], finding the optimal strategy to serve a sequence of requests $r_1, r_2, \ldots, r_n$ by k servers can be reduced to computing the minimal-cost maximal flow on a suitably constructed network with $2n + k + 2$ nodes. The details of this construction are shown in Figure 2.

As we can see from Figure 2, the network corresponding to the off-line problem consists of a source node, a sink node, and three additional layers of nodes. The first layer represents the initial server configuration $S^{(0)}$, i.e. node $s_j^{(0)}$ corresponds to the initial location of the j-th server. The remaining two layers represent the whole sequence of requests, i.e. nodes r_p and r'_p both correspond to the location of the p-th request.

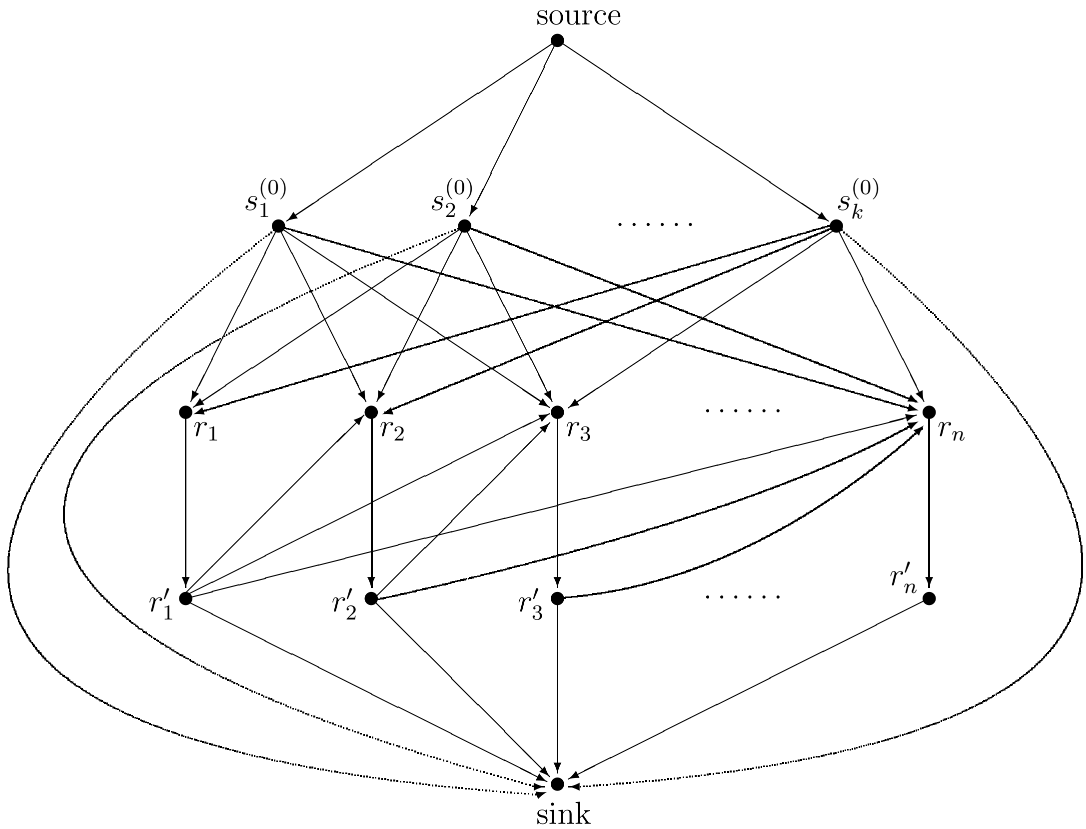


Figure 2: Finding the optimal solution to the off-line k-server problem.

Some pairs of nodes are connected by arcs, as shown in Figure 2. Note that an r_p is connected only to the associated r'_p. Also, a link between an r'_p and an r_q exists only if $q > p$. All arcs are assumed to have unit capacity. The costs of arcs leaving the source or entering the sink are 0. An arc connecting r_p with r'_p has the cost $-L$, where L is a suitably chosen very large positive number. All other arc costs are equal to *distances* between corresponding locations.

It is obvious that the maximal flow through the network shown in Figure 2 must have the value k. Moreover, the maximal flow can be decomposed into k disjunct unit flows from the source to the sink. Each unit flow determines the trajectory of the corresponding server and the requests that are accomplished by that server. If the chosen constant L is large enough, then the minimal-cost maximal flow will be forced to use all arcs between r_p and r'_p, thus assuring that all requests will be served at minimum cost. More details on solving the off-line problem by network flows can be found in [10].

According to the definition from Section 2, the i-th step of the WFA consists of k optimization problem instances, plus some simple arithmetics. So there is a possibility to implement the WFA by using the above mentioned network flow techniques. It is true, however, that the optimization problems within the WFA are not quite equivalent to off-line problems, namely there is an additional constraint regarding the final configuration of servers. Still, the construction from [4,10] can be used after a slight modification. More precisely, the i-th step of the WFA can be reduced to k minimal-cost maximal flow problems, each on a network with $2i + 2k$ nodes. One of the involved k networks is shown in Figure 3. Note that the network size rises with i.

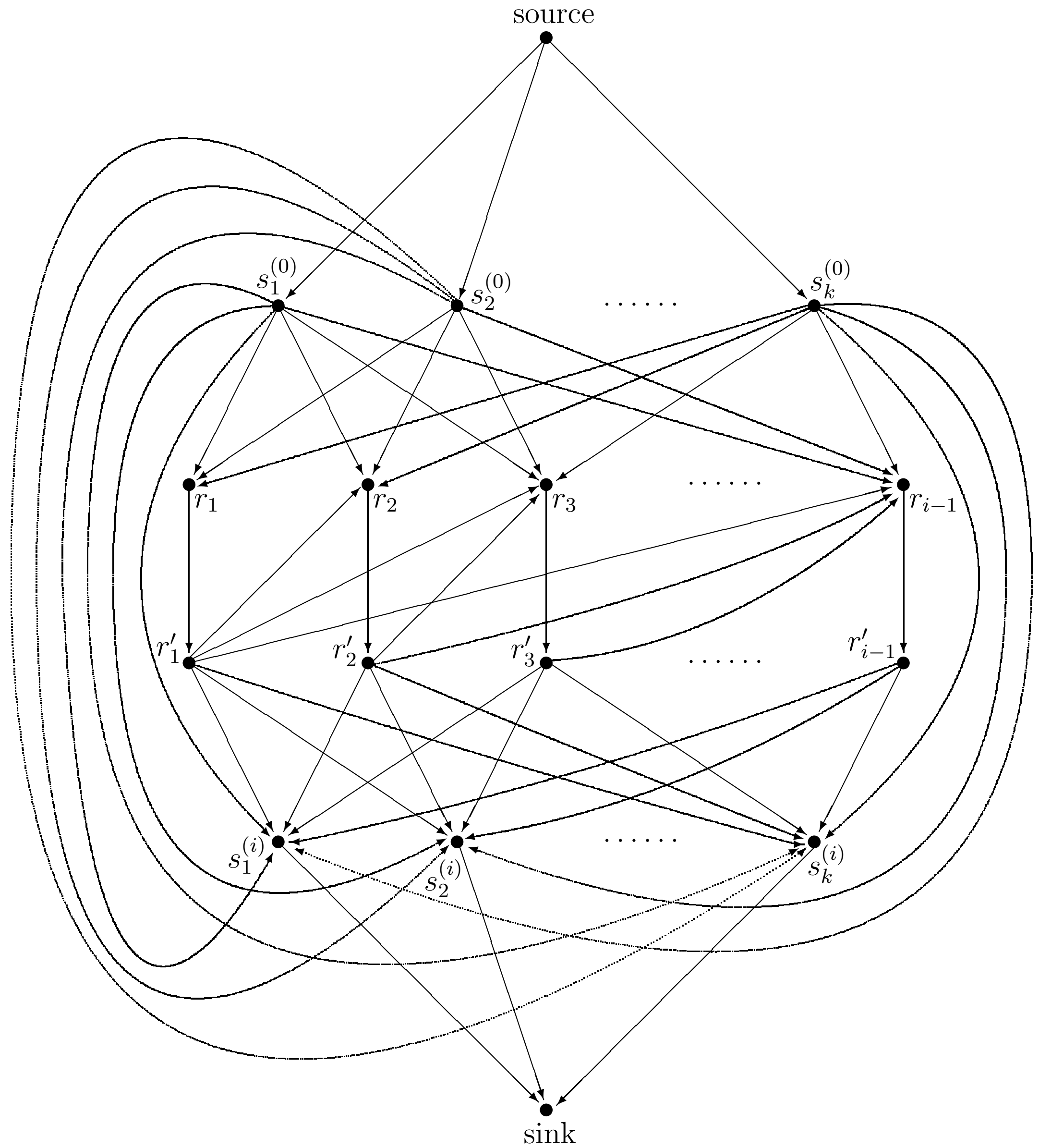


Figure 3: Solving one of k optimization problems within the i-th step of the WFA.

As we can see from Figure 3, one of k networks used to implement the i-th step of the WFA is very similar to the previously described network used to find the optimal solution of the off-line problem. The main difference is that the fourth layer of nodes has been added, which is analogous to the first layer, and which specifies the currently chosen version of the final server configuration $S^{(i)}$. Note that the second and third layer now correspond only to requests $r_1, r_2, \ldots, r_{i-1}$. Still, since the final configuration $S^{(i)}$ always covers the location of the last request r_i, we are sure that r_i will also be served with no additional cost. When we switch from one particular version of $S^{(i)}$ to another, the structure of the whole network remains the same, only the costs of arcs entering the fourth level must be adjusted in order to reflect different final setting of servers.

As it has been explained in Section 2, the w-WFA is only a modified version of the WFA, where the sequence of previous requests and configurations is examined through a moving window of size w. Obviously, the w-WFA can be implemented by network flow techniques in exactly the same way as the original WFA. More precisely, the i-th step of the w-WFA can again be reduced to k minimal-cost maximal flow problems, but now each of those problems is posed on a network built as shown in Figure 3 with $2w + 2k$ nodes. Note that the network size now does not change any more with i.

To complete the proposed implementation of the w-WFA, it is necessary to incorporate a suitable procedure for finding network flows. Our chosen procedure for solving minimal-cost maximal flow problems follows the well known generic *flow augmentation* method [3] with some adjustments. Thus we start with a flow that is not of maximal value but has the minimal cost among those with that value. Then in each iteration we augment the value of the current flow in such a way that it still has the minimal cost among those with the same value. After a sufficient number of iterations we obtain the minimal-cost maximal flow.

In our particular case the procedure can be started with the null flow. Namely, since the involved networks are acyclic, the null flow obviously has the minimal cost among those with value 0. In each iteration, augmentation is achieved by finding a shortest path in the corresponding *displacement network* [3]. Since the maximal flow has value k and each augmentation increments the flow value by one unit, finding the minimal-cost maximal flow reduces to exactly k single-source shortest path problems.

The last unexplained detail within our implementation of the w-WFA is the choice of an appropriate algorithm for finding shortest paths. It is well known that the fastest among such algorithms is the one by Dijkstra [6]. However, Dijkstra's algorithm can be applied only to networks whose arc costs are nonnegative. On the first sight, our networks do not qualify since they contain negative costs $-L$. However, thanks again to acyclicity, it turns out that Dijkstra's algorithm still can be used after a suitable transformation of arc costs.

4. Computational complexity

As we have explained in Section 3, our implementation of the the original WFA for solving the k-server problem is based on reducing the i-th step of the WFA to k minimal-cost maximal flow problems, each on a network with $2i + 2k$ nodes. Any of those minimal-cost maximal flow problems is further reduced to k single-source shortest path problems on networks with the same size. All path problems are finally solved by Dijkstra's algorithm. The w-WFA is implemented in the same way, except that the networks involved in the i-th step have size $2w + 2k$.

It is well known that Dijkstra's algorithm has a quadratic computational complexity. Since the i-th step of the WFA consists of k^2 applications of Dijkstra, and all those applications are on networks of size $2i + 2k$, it follows that the i-th step of the WFA has computational complexity $\mathcal{O}(k^2 \cdot (i+k)^2)$. Similarly, the computational complexity of the w-WFA is $\mathcal{O}(k^2 \cdot (w + k)^2)$ per step.

The above estimates are in accordance with our expectations. Indeed, the complexity of the i-th step of the w-WFA does not rise with i as for the original WFA, but it still exhibits a nonlinear dependency on k and w. Consequently, the w-WFA is faster than the original WFA, but it is still rather complex compared to simple heuristics such as GREEDY or BALANCE, whose steps can easily be implemented in $\mathcal{O}(k)$ operations. Note that we deal here with worst-case estimates, which take into account only input size, while ignoring actual input values such as actual distances among requested locations.

The described implementation of the w-WFA has been realized as a C++ program and tested on a number of k-server problem instances. To allow comparison, we have also realized some other on-line algorithms, such as GREEDY, BALANCE and the original WFA. In addition, we have made a program that implements the corresponding optimal off-line algorithm OPT.

Testing was performed on a Linux cluster whose each node consists of two 2.8 GHz CPU-s with 2GB of memory. Only one node was employed to run one program. Thanks to using the *MPI package* [9], our programs were able to distribute their workload among both CPU-s. Such limited form of parallelism resulted in speeding-up all algorithms approximately by factor of two. Still, relative speeds of different algorithms remained roughly the same as for sequential computing.

The main purpose of our testing has been to measure the performance of the w-WFA in terms of the incurred total cost. The results on total costs have already been presented in the previous paper [2]. During experimenting, we also measured actual computing times of the w-WFA and other algorithms. The results on computing times have been skipped from [2], so they are now summarized here in Table 1.

# of locations m	25	25	25	25	15112	15112	15112	15112
# of servers k	3	3	10	10	3	3	10	10
# of requests n	100	500	100	500	100	500	100	500
OPT	127	19455	392	57505	96	18725	362	57011
BALANCE	3	11	7	36	1	3	1	6
GREEDY	2	11	8	35	1	3	2	6
2-WFA	63	335	762	5651	30	189	780	3510
5-WFA	95	509	966	7512	51	202	1213	4523
10-WFA	208	1125	1555	16943	107	446	1587	10101
20-WFA	567	3502	4263	39654	254	1413	5548	27730
50-WFA	4032	18637	32206	317706	1413	22934	43399	238622
original WFA	9716	5672871	61711	49071292	4490	7182014	103071	48658852

Table 1: Experimental results - total computing time in milliseconds.

Each row of Table 1 corresponds to a particular algorithm, and each column to a particular problem instance. Each entry records the total computing time needed by the corresponding algorithm to serve the whole request sequence from the corresponding problem instance. Any problem instance is characterized by its number of locations m (ranging from 25 to 15112), number of servers k (being 3 or 10) and number of consecutive requests n (100 or 500). All instances with the same m are based on the same metric space M, i.e. they use identical distances among m possible locations. In each instance, the initial server configuration is specified by hand, while the sequence of requests is produced automatically by a random number generator.

The data shown in Table 1 are more or less consistent with the previously presented theoretical estimates of computational complexity. However, small anomalies and discrepancies in measured computing times can be spotted, and they should be attributed to peculiarities of the employed cluster. In addition, it can be observed that computing time in reality also depends on actual distances among requested locations, which is not captured by our worst-case theoretical analysis. We see that even with a fairly large window the w-WFA is indeed dramatically faster than the original WFA. We also notice that even with smaller windows the w-WFA cannot compete in speed with GREEDY or BALANCE.

5. Conclusion

We have studied a modified work function algorithm (WFA) for solving the on-line k-server problem, which is based on a moving window. In the previos paper [2] we have shown that, with a reasonably large window, our modified WFA achieves the same performance in terms of the incurred total cost as the original WFA. In this paper we have demonstrated that the modified WFA can be implemented efficiently by using network flow techniques. Also, we have shown that our implementation runs dramatically faster than the original WFA, thus becoming suitable for practical use.

The computational complexity of the modified WFA is still large compared to simple heuristics, such as the greedy or the balanced algorithm. However, this additional computational effort can be tolerated since it assures better performance, i.e. smaller total cost of responding to requests.

Our future plan is to develop a truly distributed network flow implementation of the modified WFA. By employing a much larger number of processors, it should be possible to further speed up the algorithm in order to meet strict response time requirements that are sometimes imposed by on-line computation.

References

1. Bartala Y., Koutsoupias E.. "On the competitive ratio of the work function algorithm for the k-server problem", *Theoretical Computer Science*, Vol 324 (2004), 337-345.
2. Baumgartner A., Manger R., Hocenski Z., "Work function algorithm with a moving window for solving the on-line k-server problem", in: Lužar-Stiffler V., Hljuz Dobrić V. (editors), *Proceedings of the 29-th Conference on Information Technology Interfaces - ITI 2007, Cavtat, Croatia, June 25-28, 2007*, University Computing Centre, Zagreb, 2007.
3. Bazaraa M.S., Jarvis J.J., Sherali H.D., *Linear Programming and Network Flows*, Third edition, Wiley-Interscience, New York NY, 2004.
4. Chrobak M., Karloff H., Payne T.H., Vishwanathan S., "New results on server problems", *SIAM Journal on Discrete Mathematics*, Vol 4 (1991), 172-181.
5. Irani S., Karlin A.R., "Online computation", in: Hochbaum D. (editor), *Approximation Algorithms for NP-Hard Problems*, PWS Publishing Company, Boston MA, 1997, 521-564.
6. Jungnickel D., *Graphs, Networks and Algorithms*, Second edition, Springer, Berlin, 2005.
7. Koutsoupias E., Papadimitrou C., "On the k-server conjecture", in: Leighton F.T., Goodrich M. (editors), *Proceedings of the 26-th Annual ACM Symposium on Theory of Computing, Montreal, Quebec, Canada, May 23-25, 1994*, ACM Press, New York NY, 1994, 507-511.
8. Manasse M., McGeoch L.A., Sleator D., "Competitive algorithms for server problems", *Journal of Algorithms*, Vol 11 (1990), 208-230.
9. Quinn M.J., *Parallel Programming in C with MPI and OpenMP*, McGraw-Hill, New York NY, 2003.
10. Rudec T., *The k-Server Problem*, MSc Thesis (in Croatian), Department of Mathematics, University of Zagreb, 2001.
11. Sleator D., Tarjan R.E., "Amortized efficiency of list update and paging rules", *Communications of the ACM*, Vol 28 (1985), 202-208.

DECOMPOSITION PROPERTY OF THE M/G/1 RETRIAL QUEUE WITH FEEDBACK AND GENERAL RETRIAL TIMES

Natalia Djellab
Zina Boussaha
Department of Mathematics, Faculty of Sciences
University of Annaba, BP 12, 23000, Algeria
e-mail: djellab@yahoo.fr
boussaha_z@yahoo.fr

Abstract: In this paper, we investigate the stochastic decomposition property of the M/G/1 retrial queue with feedback when the retrial times follow a general distribution, and study the rate of convergence to the ordinary M/G/1 queue with feedback.

Key-words: retrial queue, feedback, embedded Markov chain, decomposition property, rate of convergence.

1. Introduction: model description

Retrial queueing systems are characterized by the requirement that customers finding the service area busy, join the retrial group and reply for service at random intervals. These models arise in the analysis of different communication systems. For surveys on retrial queues see Templeton [5] and also monograph by Falin and Templeton [2].

We consider a single server queueing system with no waiting space at which primary customers arrive according to a Poisson process with rate $\lambda > 0$. An arriving customer receives immediate service if the server is idle, otherwise he leaves the service area to join the retrial group (orbit). Successive inter-retrial times (the time between two consecutive attempts) of any orbiting customer are governed by an arbitrary probability distribution function $F(x)$ having finite mean $1/\theta$. The service times follow a general distribution with distribution function $B(x)$ having finite mean $1/\gamma$ and Laplace-Stieltjes transform $\tilde{B}(s) = \int_0^\infty e^{-sx} dB(x)$, where s is the complex variable with the real part $\mathrm{Re}(s) > 0$ [3]. After the customer is served, he will decide either to join the orbit for another service with probability c or to leave the system forever with probability $\bar{c} = 1 - c$. Finally, we admit the hypothesis of mutual independence between all random variables defined above.

The state of the system at time t can be described by means of the process $\{C(t), N_o(t), \zeta(t), \varepsilon(t), t \geq 0\}$, where $N_0(t)$ is the number of customers in the orbit, $C(t)$ is 0 or 1 depending on whether the server is idle or busy. If $C(t) = 1$, $\zeta(t)$ represents the elapsed service time of the customer being served. When $C(t) = 0$ and $N_o(t) > 0$, the random variable $\varepsilon(t)$ represents the elapsed retrial time.

2. Notations

Let ξ_n be the time when the server enters the idle state for the n-th time; ς_n be the time at which the n-th fresh customer arrives at the server; X_i^n be the time elapsed since the last attempt made the i-th customer in the orbit until instant ξ_n^+; $q_n = N_o(\xi_n^+)$ be the number of customers in the orbit at instant ξ_n^+.

We assume that the system is in steady state, that is $\rho = \frac{\lambda}{\gamma} + c < 1$ [1]. Let $q = \lim_{n\to\infty} q_n$; $X_i = \lim_{n\to\infty} X_i^n$. When $q > 0$, we have a vector $X = (X_1, X_2, ..., X_q)$. We denote by $f_q(x_1, x_2, ..., x_q) = f_q(x)$ the joint density function of q and X, and define

$r_{ij} = \lim_{n\to\infty} P(C(\varsigma_n^-) = i, N_o(\varsigma_n^-) = j) \quad i = 0,1 \quad j = 0,1,2,...;$

$p_{ij} = \lim_{t\to\infty} P(C(t) = i, N_o(t) = j) \quad i = 0,1 \quad j = 0,1,2,...;$

$d_k = \lim_{n\to\infty} P(N_o(\xi_n^+) = k) \quad k = 0,1,2,...$.

From [6], we have that the steady-state probabilty d_k can be also expressed on terms of the joint density function $f_k(x)$: $d_k = \int_0^\infty f_k(x)dx \quad k = 1,2,...$.

We introduce the generating functions, such as

$D(z) = \sum_{k=0}^{\infty} d_k z^k$ and $R_i(z) = \sum_{j=0}^{\infty} r_{ij} z^j \quad i = 0,1$.

3. Stochastic decomposition property

Consider a sequence of random variables $\{q_n, n \geq 1\}$. This is an embedded Markov chain for our model. Its fundamental equation is

$$q_{n+1} = q_n - \delta(q_n; X^n) + v_{n+1} + u, \tag{1}$$

where $\delta(q_n; X^n)$ is 0 or 1 depending on whether the $(n+1)st$ served customer is an orbiting customer or a primary one. When $q_n = 0$, $P(\delta(0; X^n) = 0) = P(\delta(0) = 0) = 1$. Here, the total retrial intensity at idle epochs of the server depends on the number of orbiting customers. Moreover, the times $X_1, X_2, ..., X_k$ of the $k > 0$ orbiting customers depend on each other. The random variable v_{n+1} represents the number of primary customers arriving at the system during the $(n+1)st$ service time interval. Its distribution has the generating function $K(z) = \tilde{B}(\lambda - \lambda z)$ [2]. The random variable u is 0 or 1 depending on whether the served customer leaves the system or goes to orbit. We have also that $P(u = 0) = \bar{c}$ and $P(u = 1) = c$.

Since the random variables v_{n+1}, $\delta(q_n; X^n)$ and u are mutually independent,

$$E[z^{q_{n+1}}] = E[z^{q_n - \delta(q_n; X^n)}]E[z^{v_{n+1}}]E[z^u].$$

Let $n \to \infty$. We find that

$$D(z) = E[z^{q - \delta(q; X)}]\tilde{B}(\lambda - \lambda z)(\bar{c} + cz). \tag{2}$$

Using the rule of conditional expectation, one can obtain

$$E[z^{q-\delta(q;X)}] = \sum_{j=0}^{\infty} \int_0^\infty f_j(x) E[z^{j-\delta(j;x)}]dx =$$

$$\sum_{j=0}^{\infty} \int_0^\infty f_j(x)[z^j P(\delta(j;x) = 0) + z^{j-1}(1 - P(\delta(j;x) = 0))]dx =$$

$$= \frac{1}{z}\sum_{j=0}^{\infty} z^j d_j + (1-\frac{1}{z})\sum_{j=0}^{\infty} z^j \int_0^{\infty} f_j(x)P(\delta(j;x)=0)dx .$$

Consider $\int_0^{\infty} f_j(x)P(\delta(j;x)=0)dx$. This is the probability that an arriving customer finds j customers in the orbit and no customer at the server. This event takes place if and only if the last served customer leaves j customers in the orbit, he still did not decide to join the orbit or to leave the system and the new arrival occurs before any of the j orbiting customers retry for service. Therefore,

$$r_{0j} = \int_0^{\infty} f_j(x)P(\delta(j;x)=0)dx$$

and

$$E[z^{q-\delta(q;X)}] = \frac{1}{z}D(z) + (1-\frac{1}{z})R_0(z) .$$

Finally, the equation (2) becomes

$$D(z) = \frac{(1-\rho)\widetilde{B}(\lambda-\lambda z)(1-z)}{(\overline{c}+cz)\widetilde{B}(\lambda-\lambda z)-z} \times \frac{(\overline{c}+cz)R_0(z)}{1-\rho} . \tag{3}$$

One can see that the first factor on the right hand part of (3) is the generating function for the number of customers in the M/G/1 queueing system with Bernoulli feedback [4], the remaining one is the generating function for the number of customers in the retrial queue with feedback given that the server is idle.

Stochastic decomposition property of the considered system can be expressed in the following manner:

$$\{C_\theta(t), N_{o\theta}(t), \zeta(t), \varepsilon(t)\} = \{C_\infty(t), N_{q_\infty}(t), \zeta(t)\} + \{0, L_\theta(t), \varepsilon(t)\} \tag{4}$$

or

$$\{N_\theta(t) = C_\theta(t) + N_{o\theta}(t), \zeta(t), \varepsilon(t)\} = \{N_\infty(t) = C_\infty(t) + N_{q_\infty}(t), \zeta(t)\} + \{0 + L_\theta(t), \varepsilon(t)\}.$$

The processes $\{C_\theta(t), N_{o\theta}(t), \zeta(t), \varepsilon(t)\}$, $\{0, L_\theta(t), \varepsilon(t)\}$, $\{N_\theta(t), \zeta(t), \varepsilon(t)\}$ and $\{L_\theta(t), \varepsilon(t)\}$ are related to the M/G/1 retrial queue with feedback and retrial rate $\theta > 0$, where $L_\theta(t)$ represents the number of customers in the orbit at time t given that the server is idle. The processes $\{C_\infty(t), N_{q_\infty}(t), \zeta(t)\}$ and $\{N_\infty(t), \zeta(t)\}$ are associated with the ordinary M/G/1 queue with feedback, where $N_{q\infty}(t)$ is the number of customers in the queue at time t.

Let $p_{ij}(\theta)$ be the steady-state distribution of $\{C_\theta(t), N_{o\theta}(t), \zeta(t), \varepsilon(t)\}$, $p_{ij}(\infty)$ be the corresponding one of $\{C_\infty(t), N_{q\infty}(t), \zeta(t)\}$ and $q_j(\theta) = \lim_{t\to\infty} P(L_\theta(t) = j)$.

Theorem *The following inequalities take place*

$$2(1-\rho)(1-q_0(\theta)) < \sum_{i=0}^{1}\sum_{j=0}^{\infty} \left| p_{ij}(\theta) - p_{ij}(\infty) \right| < 2(1-q_0(\theta)) ,$$

where $q_0(\theta)$ *is obtained from* $\frac{(\overline{c}+cz)R_0(z)}{1-\rho}$ *by putting* $z = 0$ *and* $\rho = \frac{\lambda}{\gamma} + c$.

As $\theta \to \infty$, $\sum_{i=0}^{1}\sum_{j=0}^{\infty} \left| p_{ij}(\theta) - p_{ij}(\infty) \right| = 0(\frac{1}{\theta})$.

Proof

From (4), it is easy to see that $p_{ij}(\theta)$ is a convolution of two distributions: $p_{ij}(\infty)$ and $q_j(\theta)$, that is

$$p_{ij}(\theta) = \sum_{k=0}^{j} p_{ik}(\infty) q_{j-k}(\theta) . \qquad (5)$$

Consider

$$p_{ij}(\theta) - p_{ij}(\infty) = p_{ij}(\infty) q_0(\theta) - p_{ij}(\infty) + (1-\delta_{j0}) \sum_{k=0}^{j-1} p_{ik}(\infty) q_{j-k}(\theta) . \qquad (6)$$

With the help of (5)-(6), we obtain that

$$\sum_{i=0}^{1}\sum_{j=0}^{\infty} \left| p_{ij}(\theta) - p_{ij}(\infty) \right| < (1-2q_0(\theta)) \sum_{i=0}^{1}\sum_{j=0}^{\infty} p_{ij}(\infty) + \sum_{i=0}^{1}\sum_{j=0}^{\infty} p_{ij}(\theta) .$$

Thus, the upper inequality follows.

Now, by using the inequality $|x-y| \geq x-y$, we obtain that

$$\sum_{i=0}^{1}\sum_{j=0}^{\infty} \left| p_{ij}(\theta) - p_{ij}(\infty) \right| \geq \sum_{i=0}^{1} \left| p_{i0}(\theta) - p_{i0}(\infty) \right| + \sum_{i=0}^{1}\sum_{j=1}^{\infty} (p_{ij}(\theta) - p_{ij}(\infty)) =$$

$$= (1-q_0(\theta)) \sum_{i=0}^{1} p_{i0}(\infty) + 1 - \sum_{i=0}^{1} p_{i0}(\theta) - 1 + \sum_{i=0}^{1} p_{i0}(\infty) =$$

$$= 2(1-q_0(\theta)) \sum_{i=0}^{1} p_{i0}(\infty) = 2(1-q_0(\theta)) \frac{1-\rho}{\bar{c}\tilde{B}(\lambda)} > 2(1-q_0(\theta))(1-\rho) .$$

Here, $\frac{1-\rho}{\bar{c}\tilde{B}(\lambda)}$ is obtained from the generating function $\frac{(1-\rho)(1-z)}{(\bar{c}+cz)\tilde{B}(\lambda-\lambda z)-z}$ of the random variable $N_{q\infty} = \lim_{t\to\infty} N_{q\infty}(t)$ by putting $z=0$.

End of proof

4. Conclusion

We have established that the number of customers in the M/G/1 retrial queue with feedback (at idle epochs of the server) is equal to the sum of two independent random variables: the number of customers in the ordinary M/G/1 queue with feedback and the number of customers in retrial queue with feedback given that the server is idle. Consequently, the obtained result states that we need only study how retrial time affects the number of customers in the system given that the server is idle. It allows also to estimate the rate of convergence of the considered model to the ordinary M/G/1 queue with feedback.

References

[1] N.V. Djellab. On the M/G/1 retrial queue with feedback. *Proceedings of the International Conference "Mathematical Methods of Optimisation of Telecommunication Networks"*, pp. 32-35, 22-24 February 2005, Minsk, Byelorussia.

[2] G.I. Falin and J.G.C. Templeton. *Retrial Queues*. Chapman and Hall, 1997.

[3] L. Kleinrock. *Queueing Systems. Volume 1: Theory.* John Wiley and Sons, 1975.

[4] I. Takacs. A single server queue with feedback. *Bell System Technical Journal*, 42, 505-519, 1963.

[5] J.G.C. Templeton. Retrial queues. *TOP*, 7, 351-353, 1999.

[6] T. Yang, M.J.M. Posner, J.G.C. Templeton and H. Li. An approximation method for the M/G/1 retrial queue with general retrial times. *European Journal of Operational Research*,76, 552-562, 1994.

EIGENVALUE AND SEMIDEFINITE APPROXIMATIONS FOR GRAPH PARTITIONING PROBLEM[1]

Janez Povh
University in Maribor, Faculty of logistics
email: janez.povh@uni-mb.si

Abstract
Partitioning the nodes of a graph into sets with prescribed cardinalities is an NP-hard problem. Solving it to optimality often relies on good lower bounds. Donath and Hoffman and later also Rendl and Wolkowicz presented lower bounds which are based on graph eigenvalues.

We show how to rewrite the graph partitioning problem as a linear program over the cone of completely positive matrices and then analyze the semidefinite lower bounds, obtained by relaxing the copositive program. We show that these new lower bounds are significantly tighter than existing spectral and semidefinite lower bounds.

Keywords: semidefinite programming, graph partitioning problem, spectral lower bound, semidefinite lower bound.

1. INTRODUCTION

In this paper we consider the graph partitioning problem, which is defined as follows. Given a graph $G = (V, E)$ with $|V| = n$, a number $k > 1$ and a vector $m = (m_1, m_2, \ldots, m_k) \in \mathbb{N}^k$ with $1 \le m_1 \le m_2 \le \ldots \le m_k$, $\sum_i m_i = n$, we are interested in a partition $(S_1, S_2, \ldots, S_k)$ of vertex set V such that $|S_i| = m_i$, which minimizes the total sum of edges between different sets S_i. If $m_1 = \lfloor \frac{n}{2} \rfloor$ and $m_2 = \lceil \frac{n}{2} \rceil$, we get the NP-complete bisection problem as a special case (see [4]).

The graph partitioning problem appears in a wide range of applications from numerical linear algebra to floor planning and analysis of bottlenecks in communication networks. In parallel computing, partitioning the set of tasks among processors in order to minimize the communication between processors is another instance of graph partitioning problem. A comprehensive survey with results in this area up to 1995 is contained in [1]. The special case when we consider only 3-partitioning and try to minimize the total number of edges between two sets, is called the min-cut problem and has been studied in [8].

We represent any partition of graph vertices into k blocks with prescribed sizes by a matrix $X \in \{0, 1\}^{n\times k}$, where $x_{ij} = 1$ if and only if the ith vertex belongs to jth set. With this notation the total sum of edges between different sets S_i, defined by X, is exactly $0.5\langle X, AXB\rangle$, where A is the adjacency matrix of the graph (i.e. $a_{ij} = 1$ if (ij) is an edge and $a_{ij} = 0$ otherwise) and B is defined by $b_{ij} = 1$ if $i \ne j$ and $b_{ij} = 0$ otherwise. Using this notation and observation we can write the graph partitioning problem as

$$\text{(GP)}\qquad \begin{array}{rrcl} \min & \multicolumn{3}{l}{\frac{1}{2}\langle X, AXB\rangle} \\ \text{s. t.} & Xu_k & = & u_n, \\ & X^T u_n & = & m, \\ & X & \in & \{0, 1\}^{n\times k}. \end{array}$$

It is the purpose of this paper to present how to use semidefinite programming (SDP) to obtain tight lower bounds for OPT_{GP}. In Chapter 3 we review Donath-Hoffman [3] and Rendl-Wolkowicz [9] spectral lower bounds. In Chapter 4 we present the strongest known

[1]Project was partially supported by Slovene Ministry of higher education, science and technology under contract BI-HU/06-07-04.

semidefinite lower bound from Wolkowicz and Zhao [10]. We also present how to reformulate the graph partitioning problem as a linear program over the cone of completely positive matrices. Relaxing this problem yields two semidefinite lower bounds for which we show in Chapter 5 (proof is omitted) that they are tighter than Wolkowicz-Zhao lower bounds. Preliminary numerical results on random test graphs confirm this relations and also show that the new semidefinite lower bounds are the strongest available. By this approach we can obtain even stronger SDP bounds, but the underlying SDP models would have large time complexity.

1.1 Notation

We denote the ith standard unit vector by e_i. The vector of all ones is $u_n \in \mathbb{R}^n$ (or u if the dimension n is obvious) and the vector of all zeros is 0 or 0_n. The square matrix of all ones is denoted by J_n or J, the identity matrix by I and $E_{ij} = e_i e_j^T$. We also use the following square matrices: $B_{ij} = \frac{1}{2}(E_{ij} + E_{ji})$ and $E_i = e_i u_k^T \in \mathbb{R}^{k \times k}$.

In this paper we refer to various sets of matrices. The vector space of real nonnegative $n \times n$ matrices we denote by $\mathcal{N}_n = \{X \in \mathbb{R}^{n \times n} : x_{ij} \geq 0\}$, the vector space of real symmetric matrices of order n is $\mathcal{S}_n = \{X \in \mathbb{R}^{n \times n} : X = X^T\}$, the cone of positive semidefinite matrices of order n we denote by $\mathcal{S}_n^+ = \{X \in \mathcal{S}_n \colon y^T X y \geq 0, \forall y \in \mathbb{R}^n\}$ and the cone of completely positive matrices of order n is $\mathcal{C}_n^* = \{X \in \mathcal{S}_n \colon X = \sum_{i=1}^k y_i y_i^T,\ k \geq 1,\ y_i \in \mathbb{R}_+^n,\ \forall i = 1, \ldots, k\}$.

We also use $X \succeq 0$ for $X \in \mathcal{S}_n^+$. A linear program over $\mathbb{R}_+^n$ is called a linear program, a linear program over $\mathcal{S}_n^+$ is called a semidefinite program while a linear program over $\mathcal{C}_n^*$ is called a copositive program.

The sign $\otimes$ stands for the Kronecker product. When we consider the matrix $X \in \mathbb{R}^{m \times n}$ as a vector from $\mathbb{R}^{mn}$, we write this vector as $\mathrm{vec}(X)$ or x. For $u, v \in \mathbb{R}^n$ we define $\langle u,\, v\rangle = u^T v$ and for $X, Y \in \mathbb{R}^{m \times n}$ we set $\langle X,\, Y\rangle = \mathrm{trace}(X^T Y)$, where trace of a square matrix is the sum of its diagonal. If $a \in \mathbb{R}^n$, then $\mathrm{Diag}(a)$ is an $n \times n$ diagonal matrix with a on the main diagonal and $\mathrm{diag}(X)$ is the main diagonal of a square matrix X.

For a matrix $Z \in \mathcal{S}_{kn}$ we often use the following block notation:

$$Z = \begin{bmatrix} Z^{11} & \cdots & Z^{1k} \\ \vdots & \ddots & \vdots \\ Z^{k1} & \cdots & Z^{kk} \end{bmatrix}, \tag{1}$$

where $Z^{ij} \in \mathbb{R}^{n \times n}$.

Given an optimization problem P, we denote its optimal value by OPT_P.

2. SPECTRAL LOWER BOUNDS

2.1 Donath-Hoffmam lower bound

Donath and Hoffmam [3] used the fact that for any partition matrix $X \in \mathbb{R}^{n \times k}$ has the matrix $Y = XM^{-1/2}$ orthonormal columns, where $M = \mathrm{Diag}(m)$. Let $L = \mathrm{Diag}(Au) - A$ be the Laplacian matrix of the graph. For any partition matrix X we have on one hand $\langle X,\, LX\rangle = \langle X,\, AXB\rangle$ and on the other hand $\langle X,\, LX\rangle = \langle Y,\, LYM\rangle$, hence

$$\begin{aligned} OPT_{GP} &= \min\{\frac{1}{2}\langle Y,\, LYM\rangle \colon YM^{1/2} \in \{0,\, 1\}^{n \times k},\ Y\bar{m} = u_n,\ Y^T u_n = \bar{m}\} \\ &\geq \min\{\frac{1}{2}\langle Y,\, LYM\rangle \colon Y^T Y = I\} \ = \ \frac{1}{2}\sum_{i=1}^{k} m_{k-i+1}\lambda_i(L) \end{aligned}$$

where $\lambda_1(L) \leq \lambda_2(L) \leq \ldots \leq \lambda_n(L)$ are eigenvalues of matrix L and $\bar{m} = (\sqrt{m_1}, \sqrt{m_1}, \ldots, \sqrt{m_k})^T$. Since for any diagonal matrix D with $\mathrm{trace}(D) = 0$ and any partition matrix X we have $\langle X, LX\rangle = \langle X, (L+D)X\rangle$, we get the following lower bound [3]

$$OPT_{GP} \geq \max\Big\{\frac{1}{2}\sum_{i=1}^{k} m_{k-i+1}\lambda_i(L+D)\colon D = \mathrm{Diag}(d),\ u^T d = 0\Big\} =: OPT_{DH} \qquad (2)$$

Anstreicher and Wolkowicz [2] showed that OPT_{DH} is the optimal solution of a semidefinite program, obtained by Lagrangian relaxation of an appropriate quadratic relaxation of GP. In particular they showed

$$\begin{aligned} OPT_{DH} \ = \ & \max\ \mathrm{trace}(S) + \mathrm{trace}(T) \\ & \text{s. t.}\ \ \bar{M} \otimes (L + \mathrm{Diag}(v)) - I \otimes S - T \otimes I \succeq 0 \\ & \qquad u_n^T v = 0,\ v \in \mathbb{R}^n,\ S, T \in \mathcal{S}_n, \end{aligned}$$

where

$$\bar{M} = \frac{1}{2}\begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}.$$

2.2 Rendl-Wolkowicz lower bound

Rendl and Wolkowicz [9] used the projection technique to establish a lower bound for OPT_{GP}, which is comparable with Donath-Hoffman lower bound. In particular, they used that for each partition matrix X it holds

$$\mathrm{trace}(X^T A(d) X) = \mathrm{trace}(X^T A X) + s(d),$$

where $A(d) = A + \mathrm{Diag}(d)$ and $s(d) = u^T d$, and proved (see [9, Corollary 4.2])

$$OPT_{GP} \geq \frac{s(A)}{2} - \min_{d\in\mathcal{D}} \max\{\frac{1}{2}\mathrm{trace}(X^T A(d) X)\colon X u_k = u_n,\ X^T u_n = m,\ X^T X = M\},$$

where $\mathcal{D} = \mathcal{D}(A) = \{d \in \mathbb{R}^n\colon d^T u = 0,\ A(d)u = \alpha_d u\}$. Moreover, in [9] is also presented the optimum of the inner maximization problem:

$$\begin{aligned} & \max\{\frac{1}{2}\mathrm{trace}(X^T A(d) X\colon X u_k = u_n,\ X^T u_n = m,\ X^T X = M\} \\ & = \frac{1}{2}\sum_{j=1}^{k-1} \lambda_j(\hat{A}(d))\lambda_j(\hat{M}) + \frac{\alpha_d}{2n} s(M^2). \end{aligned}$$

Here λ_i is the ith largest eigenvalue (i.e. $\lambda_1 \geq \lambda_2 \geq \cdots$), α_d is the eigenvalue of $A(d)$, which corresponds to eigenvector u, $\hat{A}(d) = V_n^T A(d) V_n$ and $\hat{M} = W_k^T M W_k$. We may take for V_n and W_k arbitrary basis of $u_n^\perp$ and $\sqrt{m}^\perp$, respectively.

This lower bound is stronger than Donath-Hoffman lower bound, but is hard to compute. One possibility is to use a special $d \in \mathcal{D}$ instead of minimizing over the whole set $\mathcal{D}$. Using $\bar{d} = s(A)/n \cdot u - Au$ we get the Rendl-Wolkowicz lower bound, [9, Theorem 5.1]:

$$OPT_{GP} \geq \frac{s(A)}{2} - \frac{1}{2}\sum_{j=1}^{k-1} \lambda_j(\hat{A}(\bar{d}))\lambda_j(\hat{M}) - \frac{s(A)s(M^2)}{2n^2} =: OPT_{RW} \qquad (3)$$

3. SEMIDEFINITE LOWER BOUNDS

3.1 Wolkowicz-Zhao lower bounds

Several SDP approaches to the graph partitioning problem have been studied in the last decade. Wolkowicz and Zhao [10] have extended approach, designed for the Quadratic assignment problem, to GP. They have proposed two semidefinite models, one in the cone $\mathcal{S}^+_{1+kn}$ and the other in $\mathcal{S}^+_{1+(k-1)(n-1)}$. The second is obtained from the first by projecting the feasible set to the minimal face with certain properties. This among others also made the Slater condition to hold and therefore enabled efficient employment of the interior-point methods.

Wolkowicz and Zhao [10] proposed the following semidefinite models for the GP. The first model is in the cone $\mathcal{S}^+_{1+kn}$ and is denoted by GP_{WZ}:

$$(GP_{WZ})\qquad \begin{array}{rrcl} \min & \langle L_A, Y\rangle & & \\ \text{s. t.} & \operatorname{Arrow}(Y) & = & 0,\ \ Y_{00} \ = \ 1 \\ & \langle D_1, Y\rangle & = & 0,\ \ \langle D_2, Y\rangle \ = \ 0 \\ & \mathcal{G}_J(Y) & = & 0,\ \ Y \ \in \ \mathcal{S}^+_{kn} \end{array}$$

where

$$L_A = \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{2}I\otimes L\end{bmatrix},\ D_1 = \begin{bmatrix} n & -u_k^T\otimes u_n^T \\ -u_k\otimes u_n & J_k\otimes I_n\end{bmatrix},\ D_2 = \begin{bmatrix} m^Tm & -m^T\otimes u_n^T \\ -m\otimes u_n & I_k\otimes J_n\end{bmatrix}$$

and $\mathcal{G}_J(\cdot)$ is the operator that forces the zero pattern, i.e. all non-diagonal square blocks must have zeros on the main diagonal. The operator Arrow guarantees that the main diagonal of Y is equal to the first row $Y^{0,:}$ and therefore represents the 0-1 constraint in the original problem.

The second model from [10] is in the projected cone $\mathcal{S}^+_{1+(k-1)(n-1)}$ and is denoted by GP_{PWZ}:

$$(GP_{PWZ})\qquad \begin{array}{rrcl} \min & \langle L_A,\, \hat{V}Z\hat{V}^T\rangle & & \\ \text{s. t.} & \mathcal{G}_J(\hat{V}Z\hat{V}^T) & = & 0,\ \ (\hat{V}Z\hat{V}^T)_{00} \ = \ 1, \\ & \operatorname{Arrow}(\hat{V}Z\hat{V}^T) & = & 0,\ \ Z \ \in \ \mathcal{S}^+_{1+(k-1)(n-1)}. \end{array}$$

Matrix $\hat{V}$ is defined as follows:

$$\hat{V} = \begin{bmatrix} e_0^T \\ W\end{bmatrix},\quad W = \begin{bmatrix}\frac{1}{n}m\otimes u_n \mid V_k\otimes V_n\end{bmatrix},\quad V_p = \begin{bmatrix} I_{p-1} \\ -u_{p-1}^T\end{bmatrix}.$$

3.2 New semidefinite lower bounds

Povh [7, Chapter 8] formulated GP as a linear program over the cone of completely positive matrices:

$$(GP_{CP})\qquad \begin{array}{rrcll} \min & \frac{1}{2}\langle B^T\otimes A, Y\rangle & & & \\ \text{s. t.} & \langle J_k\otimes E_{ii},\, Y\rangle & = & 1,\ \ 1\le i\le n, & (4) \\ & \langle E_{ii}\otimes J_n,\, Y\rangle & = & m_i^2,\ \ 1\le i\le k, & (5) \\ & \langle E_j\otimes E_i^T,\, Y\rangle & = & m_j,\ \ 1\le i\le n,\ 1\le j\le k, & (6) \\ & \langle B_{ij}\otimes I,\, Y\rangle & = & m_i\delta_{ij},\ \ 1\le i\le j\le k & (7) \\ & Y & \in & \mathcal{C}^*_{kn}. & \end{array}$$

Solving this problem to optimality is still an NP-hard problem since the separation problem for the cone of completely positive matrices is an NP-hard problem [5].

We can relax the hard constraint $Y \in \mathcal{C}_{kn}^*$ to obtain approximation models. If we relax it to $Y \in \mathcal{S}_{kn}^+ \cap \mathcal{N}_{kn}$, we get a strong SDP model, which we name GP_{SDP-1}. Therefore we can write

$$OPT_{SDP-1} = \min \left\{ \frac{1}{2} \langle B^T \otimes A, Y \rangle \ : \ Y \in \mathcal{S}_{kn}^+ \cap \mathcal{N}_{kn}, \text{ Y feasible for (4)–(7)} \right\}$$

This model is very time consuming since it contains $\binom{kn}{2}$ sign constraints. We can get a simpler (and weaker) SDP model if we demand that Y is positive semidefinite and impose sign constraints only on few positions in the matrix Y. If we force sign constraints on the diagonals of off-diagonal blocks (i.e. we demand $\mathrm{diag}(Y^{ij}) \geq 0,\ i \neq j$), we get the second SDP model, called GP_{SDP-2}:

$$OPT_{SDP-2} = \min \left\{ \frac{1}{2} \langle B^T \otimes A, Y \rangle \ : \ Y \in \mathcal{S}_{kn}^+, \text{ Y feasible for (4)–(7)}, \ \mathrm{diag}(Y^{ij}) = 0_n, i \neq j \right\}.$$

We point out that by stronger approximation of the completely positive cone we get stronger SDP lower bound. Using hierarchy of cones, based on the sum of squares concept from [6], would give rise to a sequence of increasingly tight lower bounds, which would rely on SDPs, whose complexity would increase exponentially.

4. Comparison of spectral and semidefinite lower bounds

Povh [7, Theorem 8.5] proved that

$$OP_{SDP-1} \geq OPT_{SDP-2} = OPT_{PWZ} \geq OPT_{WZ}.$$

Theoretical comparison of OPT_{SDP-1} and OPT_{SDP-2} with spectral lower bounds is not easy. We can find instances where OPT_{SDP-2} is weaker than OPT_{DH} and OPT_{RW} and also many instances where the situation is reversed (see [7, Example 8 and Remark 17]). However, on all test instances we noticed that $OPT_{SDP-1} \geq \max\{OPT_{DH}, OPT_{RW}\}$ and we conjecture that this holds for all graphs and all partition vectors.

n	seed	$\|E\|$	OPT_{SDP-1}	OPT_{SDP-2}	OPT_{DH}	OPT_{RW}
35	50302	137	50.463	45.522	37.945	32.963
35	50303	192	85.568	80.669	66.987	57.705
35	50304	258	126.355	119.731	101.954	95.235
35	50305	287	144.795	139.698	119.901	115.596
35	50306	347	188.034	181.802	154.770	142.247
35	50307	447	263.824	255.613	221.129	222.677
35	50308	476	285.353	277.763	241.623	256.962

Table 1: Numerical results on random graphs, $n = 35$, $m = (10, 10, 15)$

We demonstrate proven and conjectured relations in Table 1. We computed spectral and SDP lower bounds on random graphs with 35 nodes, where edge density is varying from 0.2 to 0.8. We used the random graph generator from Kim Toh, which is available also on the website `http://www2.arnes.si/~jpovh/research`. We partition graphs with $m = (10, 10, 15)$. The reason why we didn't do computations with larger graphs is memory limitation. Computing OPT_{SDP-1} is very time consuming, since it has $\mathcal{O}(n^2k^2)$ linear constraints (in our case this is about 5000, which is already on the boundary for the computer with AMD Athlon XP 2100+ processor and 512 MB of RAM, which we used). In column 1 we have the number of graph vertices, in column 2 we provide the seed for the random graph generator, so we can reproduce the instances. Column 3 contains the number of edges of the graph while the last 4 columns contain the lower bounds OPT_{SDP-1}, OPT_{SDP-2}, OPT_{DH} and OPT_{RW}. We can see that OPT_{SDP-1} is the strongest lower bound as we proved and conjectured. SDP lower bounds are reasonable stronger than spectral lower bounds.

5. Conclusions

In this paper we present a way how to improve spectral lower bounds for the graph partitioning problem (GP) using semidefinite programming. We reformulate GP as a linear program over the cone of completely positive matrices. Relaxing this problem yields two semidefinite lower bounds, which are at least as strong as the SDP lower bounds from Wolkowicz and Zhao [10] and are significant improvements comparing to Donath-Hoffman [3] and Rendl-Wolkowicz [9] spectral lower bounds. Our approach could give even stronger bounds if we considered stronger relaxations of the cone of completely positive matrices. However, all this models are very time consuming and for practical purposes OPT_{SDP-1} is already out of reach for standard SDP solvers.

References

[1] C. J. Alpert and A. B. Kahng. Recent directions in netlist partition: A survey. *Integr., VLSI J.*, 19:1–81, 1995.

[2] K. Anstreicher and H. Wolkowicz. On lagrangian relaxation of quadratic matrix constraints. *SIAM J. Matrix Anal. Appl.*, 22:41–55, 2000.

[3] W. E. Donath and A. J. Hoffman. Lower bounds for the partitioning of graphs. *IBM J. Res. Develop.*, 17:420–425, 1973.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability: a guide to the Theory of NP-Completeness.* Freeman, 1979.

[5] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Programming*, 39:117–129, 1987.

[6] P. A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization.* PhD thesis, California Institute of Technology (Pasadena), 2000.

[7] J. Povh. *Application of semidefinite and copositive programming in combinatorial optimization.* PhD thesis, University of Ljubljana, Faculty of mathematics and physics (Slovenia), November 14, 2006.

[8] J. Povh and F. Rendl. A copositive programming approach to graph partitioning. *SIAM J. Optim.*, 18:223–241, 2007.

[9] F. Rendl and H. Wolkowicz. A projection technique for partitioning the nodes of a graph. *Ann. Oper. Res.*, 58:155–179, 1995.

[10] H. Wolkowicz and Q. Zhao. Semidefinite programming relaxations for the graph partitioning problem. *Discr. Appl. Math.*, 96-97:461–479, 1999.

THE APPLICATION OF THE EXTENDED METHOD FOR RISK ASSESSMENT IN THE PROCESSING CENTRE WITH DEXi SOFTWARE

Marko Potokar[1], Mirjana Rakamarić Šegić[2]
and Gregor Miklavčič[3]
[1]Bankart d.o.o., Celovška 150, 1000 Ljubljana, Slovenia,
email: marko.potokar@bankart.si
[2]Politechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia
email: mrakams@veleri.hr
[3]Bank of Slovenia, Slovenska 35, 1000 Ljubljana, Slovenia
email: gregor.miklavcic@bsi.si

Abstract: In this paper we show a model that we developed for analysing risks using a software for multicriteria decision making DEXi. We extended the standard risk assessment method with addition criteria and developed a qualitative model for risk assessment in DEXi. The model was applied in the process of operational risk assessment in the processing centre.
In the article we present the application of the extended qualitative method for operational risks assessment with DEXi on a process ON-LINE PROCESSING.

Keywords: Risk management, risk, control, risk assessment, risk reporting matrix, DEXi.

1. INTRODUCTION

Any potential impact on the goals of the organisation caused by an unplanned event should be identified, analysed and assessed. Risk mitigation strategies should be adopted to minimise residual risk to an accepted level.

Uncertainty is the central issue of risk. There are two fundamentally different metric schemes applied to the measurement of risk elements: qualitative and quantitative. Both approaches have advantages and disadvantages. In quantitative models the assessment and results are based substantially on independently objective processes and metrics. Thus meaningful statistical analysis is supported. But calculations are complex. If they are not understood or effectively explained, management may mistrust the results of "black box" calculations. There is also a problem with gathering a substantial amount of data to calculate the parameters of the model. Although in qualitative models the risk assessment and results are essentially subjective, calculations, if any, are simple and readily understood and executed. Furthermore general indication of significant areas of risk that should be addressed is provided. In our model we use a qualitative approach.

Because the results of the risk assessment should be understandable to the management and expressed in clear term, to enable management to align risk to an acceptable level of tolerance, one has to use the appropriate tools in the process of risk management.

Risk management

Risk management is a continuous process that is accomplished throughout the life cycle of a system. It is an organized methodology for continuously identifying and measuring the unknowns; developing mitigation options; selecting, planning, and implementing appropriate risk mitigations; and tracking the implementation to ensure successful risk reduction.

Risk management is the process of reducing risks to an acceptable level.

The risk management process model includes the following key activities, performed on a continuous basis:

- Risk Identification,
- Risk Assessment,

- Risk Mitigation Planning,
- Risk Mitigation Plan Implementation,
- Risk Tracking.

Risk

Risk is a measure of future uncertainties in achieving goals and objectives within defined cost, schedule and performance constraints. Risk addresses a potential variation in the planned approach and its expected outcome.

Risks have three components:

- A future root cause (yet to happen), which, if eliminated or corrected, would prevent a potential consequence from occurring,
- A probability (or likelihood) assessed at the present time of that future root cause occurring, and
- The consequence (or effect) of that future occurrence.

A future root cause is the most basic reason for the presence of a risk. Accordingly, risks should be tied to future root causes and their effects.

Risk identification

In the process of risk identification we identify any event (threat and vulnerability) with a potential impact on the goals or processes of the organisation, including business, regulatory, legal, technology, trading partner, human resources and operational aspects.

Control

Control is defined as the policies, procedures, practices and organisational structures designed to provide reasonable assurance that business objectives will be achieved and undesired events will be prevented or detected and corrected.

2. STANDARD RISK ASSESSMENT METHOD

The intent of risk assessment is to answer the question "How big is the risk?" by:

- Considering the likelihood of the root cause occurrence;
- Identifying the possible consequences in terms of performance, schedule, and cost;
- Identifying the risk level using the Risk Reporting Matrix shown in Figure 1.

Risk Reporting Matrix

Each undesirable event that might affect the success of the operations of the organisation should be identified and assessed as to the likelihood and consequence of occurrence. A standard format for evaluation and reporting of risk assessment findings facilitates common understanding of operational risks at all levels of management. The Risk Reporting Matrix below is typically used to determine the level of risks identified within a risk identification. The level of risk for each root cause is reported as low (green), moderate (yellow), or high (red).

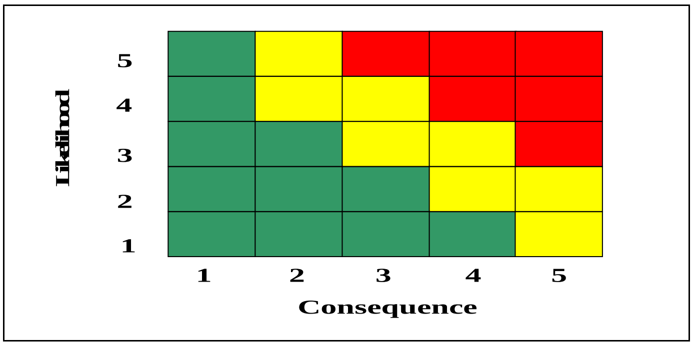


Figure 1: Risk Reporting Matrix

Levels of likelihood criteria

The level of likelihood of each root cause is established utilizing specified criteria (*Figure 2*). For example, if the root cause has an estimated 50% probability of occurring, the corresponding likelihood is Level 3.

Likelihood

Level	Likelihood	Probability of Occurrence
1	Not Likely	~10%
2	Low Likelihood	~30%
3	Likely	~50%
4	Highly Likely	~70%
5	Near Certainty	~90%

Figure 2: Levels of Likelihood Criteria

Levels of consequence criteria

The levels of consequences of each risk are established utilizing criteria such as those described in *Figure 3*.

Level	Cosequence
1	Minimal or no consequence on the process and on the goals of the organisation
2	Minor consequence on a process, can be tolerated with little or no impact on the goals of the organization
3	Moderate consequence on a process with limited impact on the goals of the organisation
4	Significant consequence on a process; may jeopardize the goals of the organisation
5	Severe consequence on a process; will jeopardize the goals of the organisation

Figure 3: Levels of consequence criteria

3. THE APPLICATION OF THE EXTENDED METHOD FOR RISK ASSESMENT IN THE PROCESSING CENTRE

In the process of risk assessment we took a subprocess ON-LINE PROCESSING from the process model, which is one of the most important subprocess of the organisation (*Figure 4*).

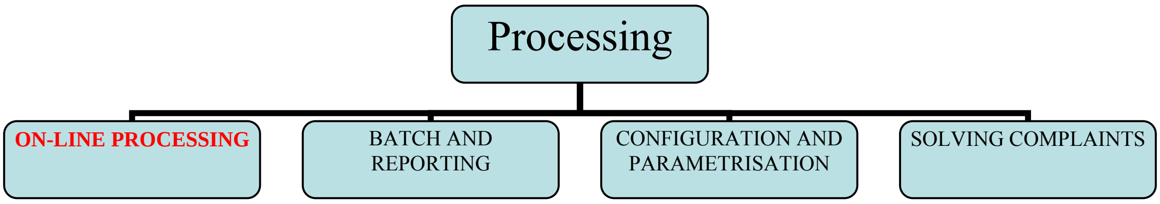


Figure 4: The subprocess ON-LINE PROCESSING

For chosen subprocess we identified events with a potential impact on the subprocess. In the next step we extended the standard method with additional criteria and evaluated the values of parameters for each event. In *Figure 5* the structure of parameters is shown. We implemented the method by developing a qualitative model for risk assessment with DEXi as shown in *chapter 4*.

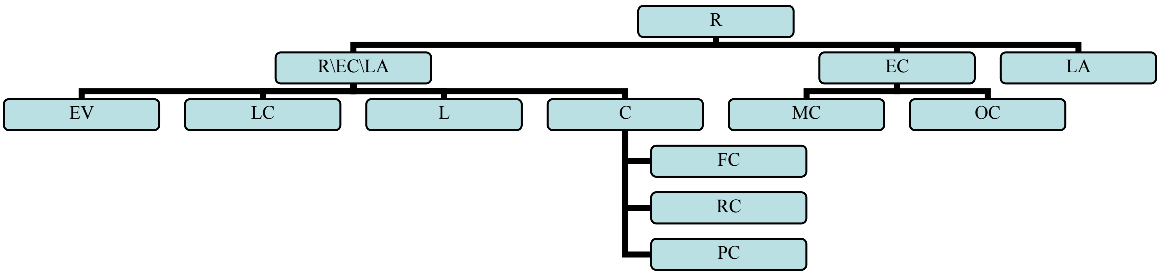


Figure 5: events and their parameters for subprocess On-line processing

R – evaluated Risk
- LA – Last Audit (last time the audit for the process was provided)
- EC – Existing Controls
  - MC – Management Controls
  - OC – Operational Controls
- R\EC\LA – evaluated Risk without considering Existing Controls and Last Audit
  - EV – Event Velocity (how fast the event is happening)
  - LC – Level of Control (how much the event can be controlled)
  - L – Likelihood of the event
  - C – Consequence of the event
    - FC – Financial Consequence of the event
    - RC – Reputation Consequence of the event
    - PC – Performance Consequence

4. THE MODEL FOR RISK ASSESSMENT IN DEXi

DEXi is a software for multicriteria decision making. It is a very useful tool for qualitative decision models not just because of validation and verification of the results but also because of making the decision transparent i.e. the understanding why did one take a certain decision.

4.1 Multicriteria decision making model

In the process of multicriteria decision making we divide a primary problem Y (a problem we want to resolve, evaluate) into subproblems. Subproblems are presented as parameters (attributes, criteria) $X_1, X_2, \ldots, X_m$. Utility function $F(X_1, X_2, \ldots, X_m)$ merges values of parameters in final evaluation (utility) of the primary problem (*Figure 6*).

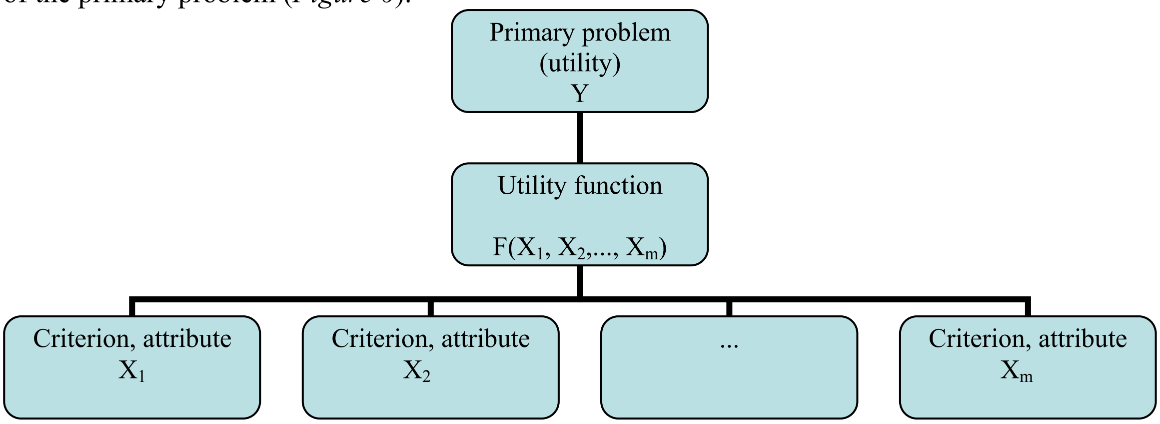
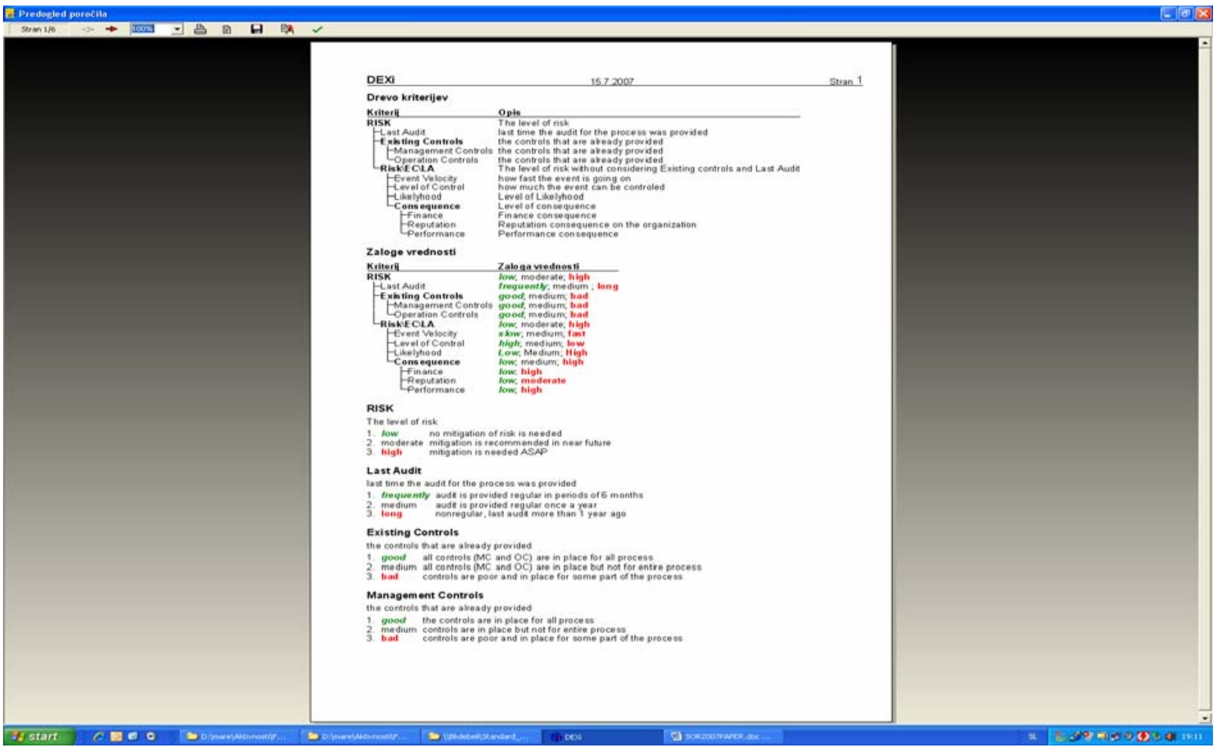


Figure 6: multicriteria decision making model

4.2 The risk assessment model for processing centre

In our risks assessment model we define each event as a primary problem which is then divided into subproblems and described with parameters (criteria) from *Figure 5*. For each parameter we defined possible values. In the next step utility functions for composed parameters were defined. *Figure 7* shows the risk assessment model in DEXi.



DEXi 15.7.2007

Drevo kriterijev

Kriterij	Opis
RISK	The level of risk
Last Audit	last time the audit for the process was provided
Existing Controls	the controls that are already provided
Management Controls	the controls that are already provided
Operation Controls	the controls that are already provided
Risk\EC\LA	The level of risk without considering Existing controls and Last Audit
Event Velocity	how fast the event is going on
Level of Control	how much the event can be controled
Likelyhood	Level of Likelyhood
Consequence	Level of consequence
Finance	Finance consequence
Reputation	Reputation consequence on the organization
Performance	Performance consequence

Zaloge vrednosti

Kriterij	Zaloga vrednosti
RISK	low; moderate; high
Last Audit	frequently; medium ; long
Existing Controls	good; medium; bad
Management Controls	good; medium; bad
Operation Controls	good; medium; bad
Risk\EC\LA	low; moderate; high
Event Velocity	slow; medium; fast
Level of Control	high; medium; low
Likelyhood	Low; Medium; High
Consequence	low; medium; high
Finance	low; high
Reputation	low; moderate
Performance	low; high

RISK

The level of risk

1. low no mitigation of risk is needed
2. moderate mitigation is recommended in near future
3. high mitigation is needed ASAP

Last Audit

last time the audit for the process was provided

1. frequently audit is provided regular in periods of 6 months
2. medium audit is provided regular once a year
3. long nonregular, last audit more than 1 year ago

Existing Controls

the controls that are already provided

1. good all controls (MC and OC) are in place for all process
2. medium all controls (MC and OC) are in place but not for entire process
3. bad controls are poor and in place for some part of the process

Management Controls

the controls that are already provided

1. good the controls are in place for all process
2. medium controls are in place but not for entire process
3. bad controls are poor and in place for some part of the process

Figure 7: the risk assessment model in DEXi

Example 1

Event: *Malicious attack – manipulation with data or SW*

For each parameter of the event that is not composed we evaluate its value as shown in *Table 1*.

Last Audit	Management Control	Operation Control	Event Velocity	Level of Control	Likelyhood	Finance	Reputation	Performance
Frequent	Good	Good	Fast	Medium	Medium	Low	Moderate	high

Table 1: non-composed parameters and their values

Values for composed parameters are calculated with their utility functions. The example of the utility function for the composed parameter *Consequence* is shown in *Table 2*.

Finance	Low	Low	Low	Low	High	High	high	High
Reputation	Low	Low	Moderate	Moderate	Low	Low	Moderate	Moderate
Performance	Low	High	Low	High	Low	High	Low	High
Consequence	**Low**	**Medium**	**Low**	**High**	**Medium**	**High**	**High**	**High**

Table 2: the utility function for the composed parameter Consequence

In our example the value of composed parameter *Consequence* equale *high*.

4.3 The results

In *Figure 8* the assessed risks for subprocess *ON-LINE* PROCESSING are shown. On the graph is clearly shown the estimated level for specific risk.

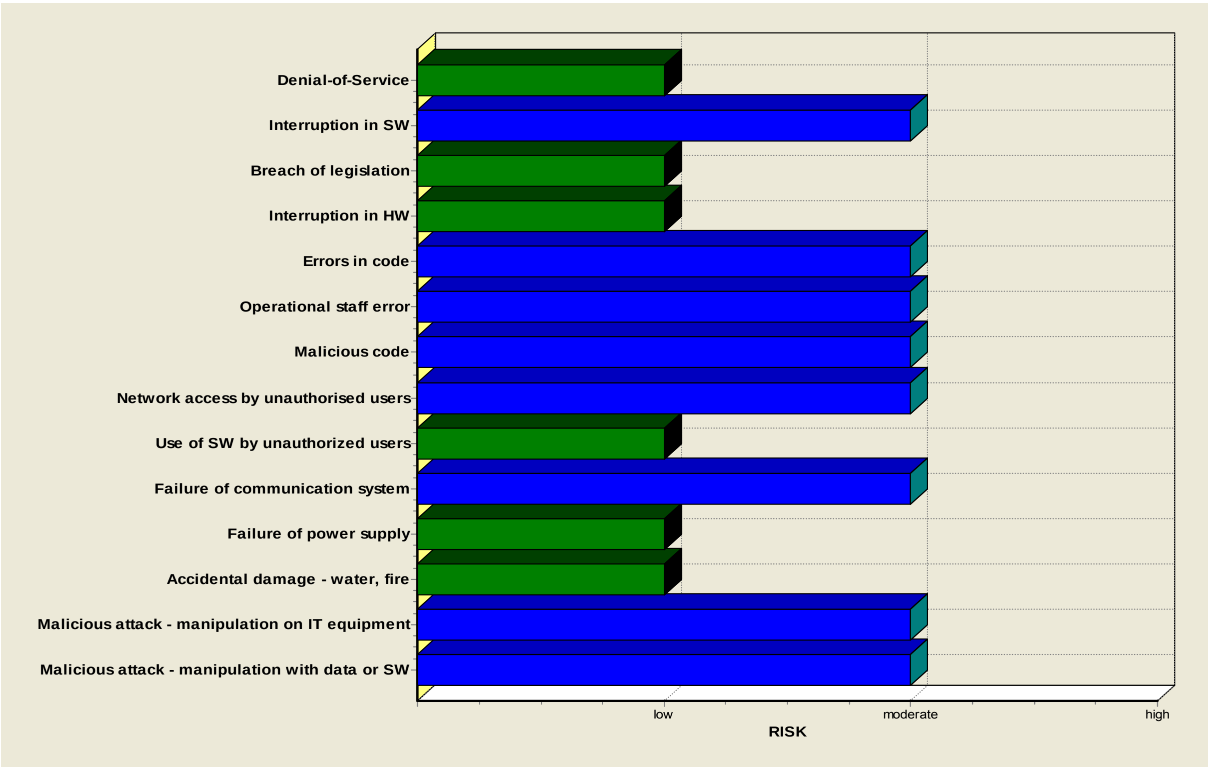


Figure 8: estimated risks for subprocess ON-LINE PROCESSING

Because we are more interested in risks with higher estimated level we can examine this kind of risks more thoroughly (*Figure 9*). We can also perform what-if analyse e.g. changing values of some parameters and observing what happened with the level of specific risk. In this way we can chose more appropriate controls to mitigate the risk.

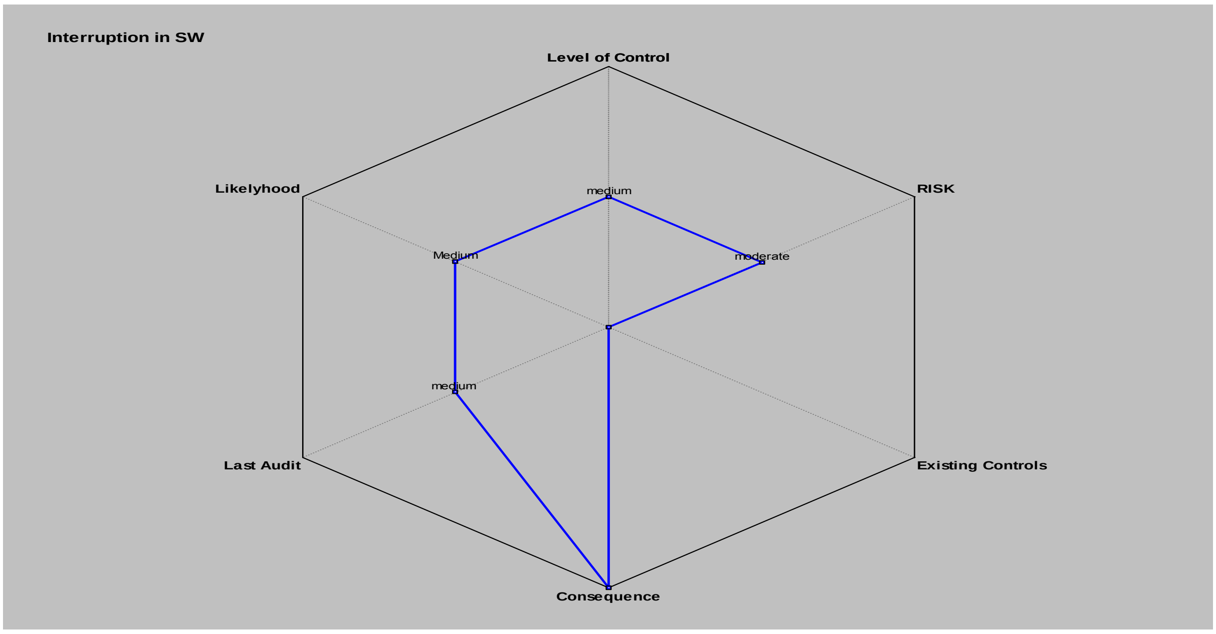


Figure 9: deeper look at specific risk

5. CONCLUSION

In this paper the qualitative model for risk assessment developed with DEXi software and its application on concrete business process was shown.

Althought the model is qualitative thus the risk assessment and results are subjective, general indication of significant areas of risk that should be addressed is provided.

The model developed with DEXi provides the result of the risk assessment that is transparent and understandable to the management. Results are shown on graphs and as such easily interpretible. If one desires one can easily see inside of specific risk. The model is appropriate for what-if analysis because we can easily change values of different parameters of the model.

Once the structure of the model is developed it can easily be fixed for other types of risks or similar events.

REFERENCES

1. Auerbach Publications: *Information Security Management Handbook.* CRC Press LLC, 2004.
2. Bankart d.o.o.: *Metodologija upravljanja tveganj.* Ljubljana, 2007.
3. Bankart d.o.o.: *Procesna shema.* Ljubljana, 2007.
4. Bertoncelj B.: *Operativna tveganja*. Banka Slovenije, 2007.
5. Bradeško L., Kušar J., Starbek M.: *Obvladovanje tveganj pri projektih naročil izdelkov/storitev.* Podčetrtek: Projektni forum, 2007.
6. BS ISO/IEC 17799:2005: *Code of practice for information security management.* BSI, 2005.

7. BS ISO/IEC 27001:2005: *Information Security Management Systems - Requirements.* BSI, 2005.
8. Department of Defence: *Risk Management Guide for DoD Acquisition; sixth edition*. USA, august 2006.
9. Information Systems Audit and Control Association: *CISM Review Manual.* IL USA, 2006.
10. IT Governance Institute: *COBIT*. USA, 2005.
11. Jereb E., Bohanec M., Rajkovič V.: *DEXi - Računalniški program za večparametrsko odločanje*. Kranj: Založba Moderna organizacija, 2003.
12. NLB d.d.: *Analiza tveganosti.* Ljubljana, 1999.
13. http://lopesl.fov.uni-mb.si/dexi

THE MULTI-CRITERIA MODEL FOR FINANCIAL STRENGTH RATING OF INSURANCE COMPANIES

Danijel Vukovič, Vesna Čančer
Faculty of Economics and Business Maribor, Razlagova 14, 2000 Maribor, Slovenia
E-mail: danijel.vukovic@uni-mb.si, vesna.cancer@uni-mb.si

Abstract: This paper presents the multi-criteria model for the creditworthiness evaluation of insurance companies. Besides to structuring the model, where the most important quantitative factors that influence financial strength ratings were taken into consideration, special attention is paid to the creation of value functions to measure the local alternative values with respect to each attribute. Because international rating agencies hide the importance of some of the factors, sensitivity analysis was used to determine the weights so that appropriate financial strength ratings were obtained.

Keywords: financial strength rating, multi-criteria decision analysis, sensitivity analysis, synthesis, value function, weight

1. INTRODUCTION

Insurance financial strength ratings provide the information about the creditworthiness of insurance companies, needed for selecting an insurance partner. Ratings are presented by world-renowned rating agencies (see [1, 5, 8, 9]), which use comparable models based on the analysis of quantitative and qualitative factors. Since rating agencies do not disclose their models in detail, the weights of the considered factors as well as the methods for measuring the local alternatives' (insurance companies') values with respect to each criterion on the lowest level are not known.

Slovene insurance companies have not yet entered the process of acquiring the financial strength rating. Therefore we built the multi-criteria model for financial strength rating of insurance companies. When verifying its applicability for the evaluation of creditworthiness, and therefore for the selection of insurance companies, we paid special attention to the following steps of multi-criteria decision-making process (see [3]):

- Problem structuring/building a model,
- Measuring the local alternative values,
- Expressing judgements on the factors' importance/weights' determination,
- Synthesis to obtain the final alternative values,
- Verification by sensitivity analysis.

The model was verified on the selected sample of German insurance companies which had already got the financial strength rating from Standard & Poor's [9]. This rating agency appoints most of the financial strength ratings in European countries. The final model was applied to the creditworthiness evaluation of Slovene insurance companies.

2. PROBLEM STRUCTURING/BUILDING A MODEL

On the basis of the methodologies of international rating agencies and surveys that have already closely examined the theme of ratings [1, 5, 8, 9], we structured the decision tree presented in Figure 1. It includes only quantitative factors because they can be examined on the basis of public available insurance companies' reports. The hierarchy in Figure 1 is composed of eight criteria: 'Profitability', 'Liquidity', 'Capital Adequacy', 'Asset Risk', 'Insurance Company Profile', 'Reserve Adequacy', 'Financial Flexibility' and 'Reinsurance Program'. We set out twenty accounting ratios – sub-criteria (see Table 2) that were studied in a four-year period, as presented for 'Profitability' in Figure 1; attributes (criteria on the lowest level in the hierarchy) of other criteria can be presented similarly.

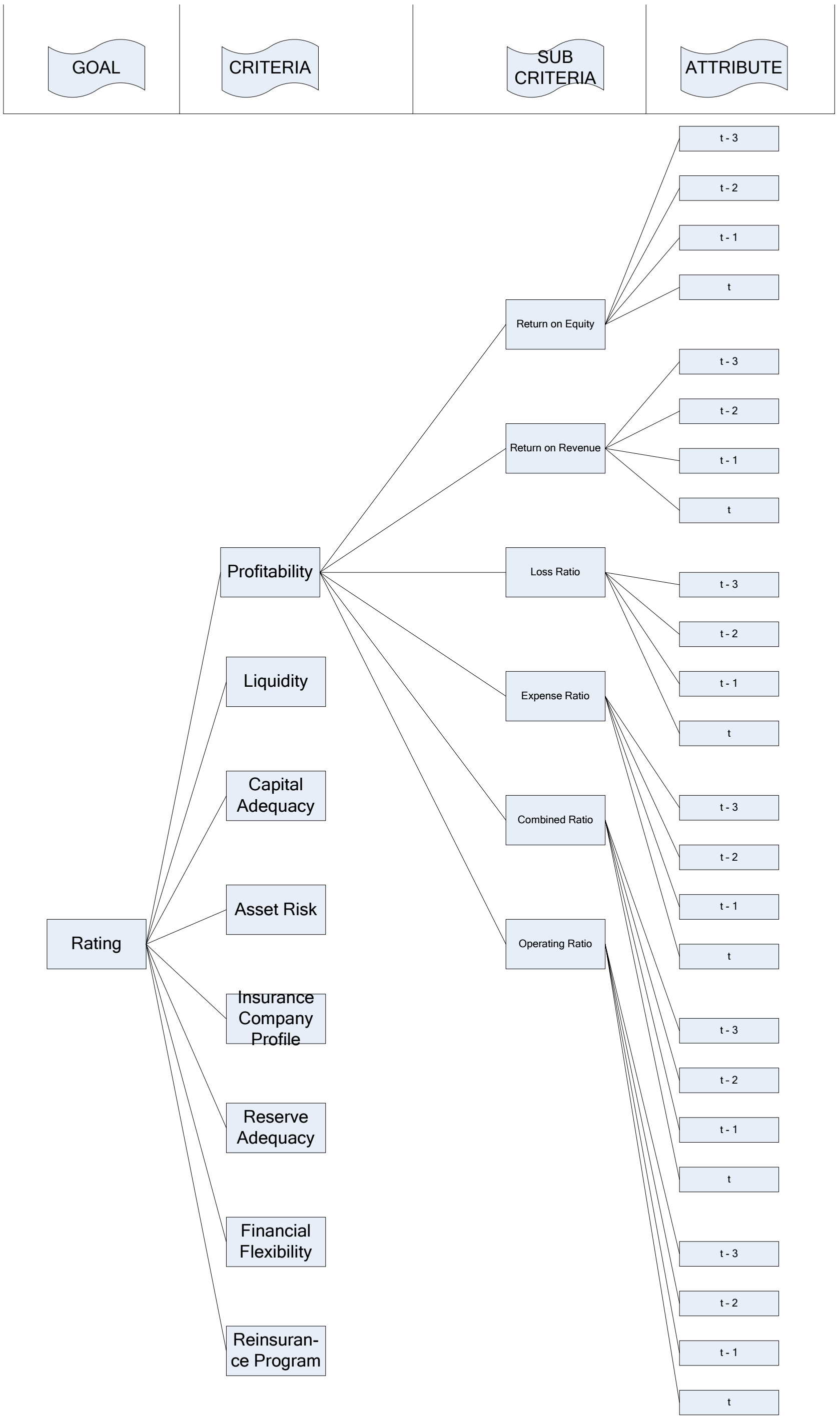


Figure 1: The criteria structure.

3. MEASURING THE LOCAL ALTERNATIVE VALUES BY VALUE FUNCTIONS

The purpose of the value function elicitation is to model and describe the desirability of achieving different performance levels of the given attribute [7]. According to [7], the main value measurement techniques can be divided in two main classes: numerical estimation (direct rating, category estimation, ratio estimation, assessing the form of value function) and indifference methods (difference standard sequence, bisection).

A value function can be defined as a mathematical representation of human judgements, because it translates the performances of the alternatives into a value score, which represents the degree to which a decision objective is matched [2]. Therefore, a value function maps the data of alternatives with respect to each attribute to the local value of alternatives. According to own experience of Čančer [4], Web-HIPRE is especially applicable for the measurement of alternatives' values with respect to each attribute by value functions. Using Web-HIPRE [7], we can create linear, piece-wise linear or exponential value functions.

In our research, the local alternative values were measured mainly by assessing the form of value function and by creating the value function by the bisection method. In solving our problem, these techniques required expertise and prior accounting knowledge. Table 1 shows the forms of value functions used for the evaluation of the considered insurance companies with respect to the attributes of sub-criteria. We formed them on the basis of sub-criteria's influence on the final rating value, distribution of ratios of German insurance companies and own experience of Vukovič [10].

Table 1: The forms of value functions in financial strength ratings.

Form	Sub-criteria (ratios)
Increasing linear function	Return on Revenue, Ln (Assets)
Decreasing linear function	Financial Leverage
Increasing piece-wise linear function	Return on Equity, Ratio of Net Operating Cash Flow and Net Written Premiums, Ratio of Capital and Total Assets, Market Share Ratio, Ratio of Loss Reserves and Net Premiums Earned, Earnings Coverage, Reinsurance Leverage, Ratio of Net Reserves and Gross Reserves
Decreasing piece-wise linear function	Loss Ratio, Expense Ratio, Combined Ratio, Operating Ratio, Gross Underwriting Leverage, Kenney Ratio, Ratio of Common Stock Investments and Invested Assets, Ratio of Reinsurance Recoverables and Capital
Decreasing convex exponential function	Ratio of Investments in Affiliates and Capital & Surplus

Let us explain how to create increasing piece-wise linear value functions by using the bisection method. In this method, two objects are presented to a decision maker; he/she is asked to define the attribute level that is halfway between the objects in respect of the relative strengths of the preferences (see [3, 7]). First, the two extreme points, the least preferred evaluation object x_{min} and the most preferred evaluation object x_{max} are identified and associated with values

$v(x_{min}) = 0,$

$v(x_{max}) = 1.$

Then, a decision maker is asked to define a midpoint x_1, for which

$(x_{min}, x_1) \sim (x_1, x_{max}),$

where ~ indicates the decision maker's indifference between the changes in value levels. While x_1 is in the middle of the value scale, we must have

$v(x_1) = 0.5v(x_{min}) + 0.5\, v(x_{max}) = 0.5.$

For the midpoint x_2 between x_{min} and x_1, and the midpoint x_3 between x_1 and x_{max} we obtain

$v(x_2) = 0.5v(x_{min}) + 0.5\, v(x_1) = 0.25,$

$v(x_3) = 0.5v(x_1) + 0.5\, v(x_{max}) = 0.75.$

For 'Return on Equity', $x_{min} = –70$ % and $x_{max} = 120$ %. To experts, the increase of 'Return on Equity' from -70 % to 9 % is equally favorable as its increase from 9 % to 120 %.

Therefore, the local value of 9 % is 0.5. Further, the increase of ‘Return on Equity’ from –70 % to 5 % is equally preferred as its increase from 5 % to 9 %; the local value of 5 % is 0.25. Finally, the increase of ‘Return on Equity’ from 9 % to 15 % is equally favorable as its increase from 15 % to 120 %; the local value of 15 % is therefore 0.75.

4. THE DETERMINATION OF WEIGHTS AND THE MODEL’S VERIFICATION

Table 2 shows that the hierarchical weighting was used in our model: weights were defined for each hierarchical level separately. Although several weight elicitation methods which base on an ordinal (SMARTER), interval (SWING, SMART) and a ratio scale (AHP) have already been successfully used (see [4]), we decided for the direct determination of weights. Namely, in the first model we used the weights, published by international rating agencies [1, 5, 8, 9]. Since they do not specifically disclose the importance of single criteria, the authors’ own comprehension of the criteria’s and sub-criteria’s importance was taken into consideration, as well.

Table 2: The criteria structure and weights.

Influence factors (criteria)	Influence factor’s weight		Ratios (sub-criteria)	Ratio’s weight
	Initial model	Final model		
Profitability	*0.25*	*0.23*		
			Return on Equity	0.10
			Return on Revenue	0.20
			Loss Ratio	0.10
			Expense Ratio	0.10
			Combined Ratio	0.30
			Operating Ratio	0.20
Liquidity	*0.10*	*0.095*		
			Investments in Affiliates/Capital & Surplus	0.50
			Net Operating Cash Flow/Net Written Premiums	0.50
Capital Adequacy	0.15	0.15		
			Gross Underwriting Leverage	0.30
			Kenney Ratio	0.30
			Capital/Total Assets	0.40
Asset Risk	0.05	0.05		
			Common Stock Investments/Invested Assets	0.40
			Reinsurance Recoverables/Capital	0.60
Insurance Company Profile	*0.20*	*0.255*		
			Market Share Ratio	0.75
			Ln (Assets)	0.25
Reserve Adequacy	*0.10*	*0.088*		
			Loss Reserves/Net Premiums Earned	1
Financial Flexibility	*0.05*	*0.044*		
			Financial Leverage	0.70
			Earnings Coverage	0.30
Reinsurance Program	*0.10*	*0.088*		
			Reinsurance Leverage	0.50
			Net Reserves/Gross Reserves	0.50

When processing the insurance financial strength ratings, it is necessary to consider the information of several years. In our model, the four-year data (2002 - 2005) are included. The importance (weights) of yearly information (time sub-criteria or attributes) is presented in Table 3.

Table 3: The weights of time sub-criteria.

Influence factors (criteria)	t – 3	t - 2	t – 1	t
Profitability	0.15	0.20	0.30	0.35
Liquidity	0.04	0.06	0.20	0.70
Capital Adequacy	0.04	0.06	0.20	0.70
Asset Risk	0.04	0.06	0.20	0.70
Insurance Company Profile	0.15	0.20	0.30	0.35
Reserve Adequacy	0.04	0.06	0.20	0.70
Financial Flexibility	0.04	0.06	0.20	0.70
Reinsurance Program	0.04	0.06	0.20	0.70

For verifying our model, we selected the sample of 28 German insurance companies (they represent 57.79% of gross premium, written in German property-casualty insurance companies in 2005) because their annual financial statements are very similar to Slovene insurance companies' statements. According to Standard & Poor's financial strength ratings, insurance companies can be classified in four groups: α (the highest rating), β, γ and δ. There is no representative insurance company in the first group; there are 5 insurance companies in β group, 14 insurance companies in γ group and 9 insurance companies in δ group.

We used the computer program Web-HIPRE to calculate the aggregated values of alternatives. Table 4 shows that in the initial model, we can not reject the hypothesis that the means of the aggregated values in γ and β groups are equal; there are several insurance companies in γ group with higher aggregated value, which could lead to their classification in higher group. Similar conclusions can be drawn for several insurance companies in δ group, although the results of independent samples *t*-test show that we can confirm the claim that the mean in γ group is significantly different from the mean in δ group ($p < 0.01$).

To improve the initial model, we used sensitivity (value) analysis - a tool for gaining information about tendency and force of the aggregated values' changes caused by the modifications of weights. We made several simulations. Based on the outcome of value analysis in several steps we decided to consider the decreased weights of 'Profitability', 'Liquidity', 'Reserve Adequacy', 'Financial Flexibility' and 'Reinsurance Program', and the increased weight of the 'Insurance Company Profile' criterion. They are presented in Table 2. Table 4 shows that in the final model, the means of the aggregated values in different groups are significantly different (even when comparing β and γ, $p < 0.05$ – 1-tailed).

Table 4: Comparisons of the groups' means.

Group	Initial model		Final model	
	Mean	*t*-Statistic	Mean	*t*-Statistic
β	0.64		0.634	
γ	0.599	(β, γ): 1.525, $p = 0.146$ (2-tailed)	0.583	(β, γ): 1.933, $p = 0.07$ (2-tailed)
δ	0.512	(γ, δ): 3.649, $p = 0.001$ (2-tailed)	0.492	(γ, δ): 4.182, $p = 0$ (2-tailed)

5. CONCLUSIONS

By using the final multi-criteria model for financial strength rating of insurance companies, we correctly classified 23 sample companies (82.14 % success of the final model). Some sample insurance companies are still not classified correctly; namely, the presented model includes only the quantitative factors. It should be completed by appropriate qualitative factors, as well.

It can be concluded that from the synthesis point of view, a main advantage of using Web-HIPRE is as follows: when changing the sample's size (number of alternatives), the aggregated alternative values remain unchanged. This enabled us the application of the presented model for financial strength rating of Slovene insurance companies (see Table 5) and their comparison to their potential German and Austrian competitors.

Table 5: Aggregated values of Slovene insurance companies and foreign competitors.

Insurance company	Aggregated value	Group
Zavarovalnica Triglav	0.683	β
Merkur zavarovalnica	0.652	β
Zavarovalnica Maribor	0.586	γ
Grawe zavarovalnica	0.549	γ
Zavarovalnica Tilia	0.527	δ
Generali zavarovalnica	0.451	δ
Wiener Städtische Versicherung AG	0.705	β
Allianz Versicherung AG	0.689	β
Uniqua Versicherung AG	0.644	γ

References

1. A. M. Best (2006): *Best's Ratings Methodology: Non – US Domiciled Companies.* http://www.ambest.com/ratings/intlpreface.pdf , consulted September 2006.
2. Beinat, E. (1997): *Value Functions for Environmental Management.* Dordrecht, Boston, London: Kluwer Academic Publishers.
3. Čančer, V. (2003): *Analiza odločanja (Decision-making analysis. In Slovenian).* Maribor: University of Maribor, Faculty of Economics and Business.
4. Čančer, V. (2005): *Comparison of the Applicability of Computer Supported Multi-Criteria Decision-Making Methods.* In: L. Zadnik Stirn and S. Drobne: SOR '05 Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research.
5. Fitch (2006): *The Rating Process.* http://www.fitchratings.com/corporate/reports/, consulted September 2006.
6. Forman, E. H., Saaty, T. L., Shvartsman, A., Forman, M. R., Korpics, M., Zottola, J., Selly, M. A. (2000): *Expert Choice 2000.* Pittsburgh: Expert Choice; Inc.
7. Helsinki University of Technology: *Value Tree Analysis.* http://www.mcda.hut.fi/value_tree/theory, consulted June 2007.
8. Moody's (2006): *Moody's Global Rating Methodology for Property and Casualty Insurers.* http://www.moodys.com/moodys/cust/research/MDCdocs, consulted September 2006.
9. S & P – Standard & Poor's (2004): *Insurance Ratings Criteria – Property/Casualty Edition.* http://www.2standardandpoors.com/spf/pdf/fixedincome, consulted July 2006.
10. Vukovič, D. (2007): *Presoja bonitete zavarovalnic s sistemom ratingov (The evaluation of creditworthiness of insurance companies with financial strength ratings. In Slovenian).* Master Diss. Maribor: University of Maribor, Faculty of Economics and Business.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 3:

Algorithms

ALGORITHM FOR PERTURBED MATRIX INVERSION PROBLEM

Hossein Arsham, Janez Grad, Gašper Jaklič
University of Baltimore, MIS Division, Baltimore, MD21201-5779, USA
University of Ljubljana, Faculty of Administration, 1000 Ljubljana, Slovenia
University of Ljubljana, Institute of Mathematics, Physics and Mechanics, 1000 Ljubljana, Slovenia
harsham@ubalt.edu, janez.grad@fu.uni-lj.si, gasper.jaklic@fmf.uni-lj.si

Abstract: In linear programming solving procedures the problem of computing the inverse of a perturbed matrix $A+D$, where $A \in \Re^{n \times n}$ is nonsingular basis matrix and D is sparse, frequently appears. In this paper an algorithm for computing the inverse matrix $(A+D)^{-1}$ is analysed, if the matrix A^{-1} is given in advance. The non-singularity requirement for D is removed.

Keywords: Linear programming, Sparse simplex, Perturbed matrix inversion.

1. Introduction

In the paper [3], the authors discuss the use of dense matrix techniques within a sparse simplex. Referring to this they provide a short description and analyse a number of known methods for solving the linear programming (LP) problem. Computation of the inverse of the perturbed LP basis matrix A by D, where D is sparse, appears at each iteration step within the discussed methods. Different authors use different strategies in defining the iterative solution procedures of the LP problem (see [3] and references therein).

In [3] the authors developed and analysed a method based on the updating of the dense Schur complement matrix S of a given matrix A. Their main attention was given to a dense orthogonal QR factorisation technique. In addition, numerical experiments have been carried out with the inverse matrix S^{-1}, stored as a dense matrix.

In the continuation we show how the updating of S^{-1} into $(S+D)^{-1}$ with the help of Sherman-Morrison-Woodbury formula, where D is sparse, can be performed and eventually utilized within the sparse simplex process. The use of the explicit inverse is perhaps the most natural, but this approach does not necessarily assure the numerical stability. Numerical experiments, computed in [3], show that numerical stability problems of Sherman-Morrison-Woodbury formula rarely appear in practical LP problems, and could be always resolved a posteriori by computing a fresh basis factorisation.

The paper is organized as follows. In Section 2 computation of $(A+D)^{-1}$ is analysed and in Section 3 an example is given.

2. Computation of $(A+D)^{-1}$

It is well known that the inverse A^{-1} of a matrix $A \in \Re^{n \times n}$ is rarely computed in practice, since it is usually not really needed.

But some LP methods require the inverse of the basis matrix. The inverse of a dense matrix A is usually computed by solving the linear system

$$AX = I \tag{1}$$

using LU decomposition with partial pivoting in $2n^3$ floating point operations. If the matrix A is sparse, the system (1) can be efficiently solved by using some iterative method (Gauss-Seidel, SOR, ...). The inverse is not necessarily sparse. In [1] a symbolic approach has been applied for the inverse computation, showing both CPU time and memory requirement superiority over the standard Gaussian row operations method.

It has been noted in [3] that for a sparse matrix A that contains a dense submatrix in some cases, where A is being updated during the algorithm, dense matrix techniques could be applied for the computation of the inverse A^{-1}.

Let us consider the Sherman-Morrison-Woodbury formula ([4]), a generalization of the well known Sherman-Morrison formula for computing the inverse of a matrix $A+uv^T$, a rank 1 perturbation of the matrix A, if A^{-1} is already known. Now let $A \in \Re^{n\times n}$ be a nonsingular sparse matrix and let its inverse A^{-1} be given. Further, let $D \in \Re^{n\times n}$. Our goal is to efficiently compute the inverse of the perturbed matrix $A+D$.

Using the Sherman-Morrison-Woodbury formula, the inverse may be written as

$$(A+D)^{-1} = A^{-1} - A^{-1}(A^{-1}+D^{-1})^{-1}A^{-1}.$$

This is not a feasible formula to find $(A+D)^{-1}$ and it requires D and $A^{-1}+D^{-1}$ to be nonsingular.

Let us simplify the notation by denoting $B := A^{-1}$. Then

$$\begin{aligned}(A+D)^{-1} &= B - B(DB+I)^{-1}DB \\ &= B - BD(BD+I)^{-1}B,\end{aligned}$$

and the nonsingularity requirement for D is thus removed. Now let us consider the structure of the matrix D. First assume

$$D = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, \tag{2}$$

and

$$B_c := \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}, \quad B_r := \begin{bmatrix} B_{11} & B_{12} \end{bmatrix},$$

where $D_{11} \in \Re^{m_1\times m_2}$, $B_{11} \in \Re^{m_2\times m_1}$, $B_c \in \Re^{n\times m_1}$, $B_r \in \Re^{m_2\times n}$. The inverse of the matrix $A+D$ can now be written as

$$(A+D)^{-1} = B+H,$$

where

$$\begin{aligned} H &= -B_c(I+D_{11}B_{11})^{-1}D_{11}B_r \\ &= -B_cD_{11}(I+B_{11}D_{11})^{-1}B_r. \end{aligned} \tag{3}$$

Note that in (3) the matrix $I+D_{11}B_{11}$ is of order m_1 and similarly the matrix $I+B_{11}D_{11}$ is of order m_2. If $m_1 << n$ or $m_2 << n$ the formula (3) can be applied to efficiently compute the inverse. Especially if $m_1 = 1$ or $m_2 = 1$ this can be easily evaluated without going through the usual matrix inversion processing.

Note that with suitable permutations of rows and columns an arbitrary matrix D can be written in the form (2). The matrix B has to be permuted accordingly.

Let us study an efficient implementation of the algorithm for evaluating (3). Consider **Table 1**. In the second column are the steps of the algorithm and in the last column the number of floating point operations needed are given.

	Algorithm	Number of operations
1	$C_1 = D_{11}B_{11}$	$2m_1^2 m_2$
2	$C_2 = I + C_1$	m_1
3	$C_3 = D_{11}B_r$	$2m_1 m_2 (n - m_1)$
4	$C_4 = C_2^{-1}C_3$	$2/3\, m_1^3 + 2m_1^2 n$
5	$H = -B_c C_4$	$2m_1 n^2$
6	$(A+D)^{-1} = B + H$	n^2

Table 1. **An algorithm for evaluating (3).**

Note that in step 4 the matrix C_4 should be calculated by solving the system $C_2 C_4 = C_3$. If m_1 and m_2 are small in comparison to n, only the term $n^2(1+2m_1)$ matters for large n. An implementation of an algorithm for evaluating the second expression in (3) is similar, and its analysis will be omitted.

Note that the presented algorithm is a slight improvement of the algorithm, given in [2]. An implementation of the improved algorithm in Matlab ([5]) is available at

http://www.fmf.uni-lj.si/~jaklicg/pertinv.html.

In order to compare our algorithm with the implemented method **inv** for calculating the inverse of the sparse matrix in Matlab, we ran some numerical tests. In **Table 2** are the times needed for the computation of $(A+D)^{-1}$ if $A \in \Re^{n\times n}$ is a random sparse matrix with approximately δn^2 uniformly distributed nonzero entries and $D \in \Re^{n\times n}$ has just one nonzero column with random approximately δn nonzero entries with $\delta = 0.5$, the case usually encountered in sparse LP problems. Computations were done on Pentium M 1.5 processor.

n	inv(A+D)	our algorithm
200	0.22	0.05
400	2.13	0.14
600	7.55	0.29
800	17.27	0.50
1000	34.76	0.77
2000	306.26	3.21

Table 2. **Time (in seconds) needed for computing $(A+D)^{-1}$ by Matlab's function inv and the presented algorithm.**

Note that for matrices D with widely distributed nonzero elements our algorithm is not so efficient, and usually Matlab's method **inv(A+D)** is faster.

3. Example

Let us conclude the paper with an example. Let

$$A = \begin{bmatrix} 1 & 0 & 2 & -3 & 5 \\ 4 & -1 & -1 & 0 & 1 \\ 7 & 2 & -1 & 0 & -1 \\ -2 & 1 & 3 & 4 & 0 \\ 0 & 1 & 2 & 3 & -2 \end{bmatrix},$$

and a perturbation

$$D = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Further, assume that the inverse

$$A^{-1} = \begin{bmatrix} 0.0500 & 0.1391 & 0.0283 & -0.0978 & 0.1804 \\ -0.0500 & -0.4435 & 0.3630 & 0.3587 & -0.5283 \\ 0.2500 & 0.0435 & -0.1630 & -0.3587 & 0.7283 \\ -0.1500 & 0.1478 & 0.0457 & 0.3804 & -0.3239 \\ 0.0000 & 0.0435 & 0.0870 & 0.3913 & -0.5217 \end{bmatrix}$$

has already been computed. Then

$$D_{11} = \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}, \quad B_{11} = \begin{bmatrix} -0.0500 & 0.3630 & -0.5283 \end{bmatrix},$$

and

$$B_c = \begin{bmatrix} -0.0500 & 0.3630 & -0.5283 \\ 0.0500 & 0.0283 & 0.1804 \\ 0.2500 & -0.1630 & 0.7283 \\ -0.1500 & 0.0457 & -0.3239 \\ 0.0000 & 0.0870 & -0.5217 \end{bmatrix},$$

$$B_r = \begin{bmatrix} -0.0500 & 0.3630 & -0.5283 & -0.4435 & 0.3587 \end{bmatrix}.$$

Hence

$$(I + B_{11}D_{11})^{-1} = -1.4984.$$

The resulting matrix H is

$$H = \begin{bmatrix} 0.1249 & -0.9070 & 1.3198 & 1.1080 & -0.8962 \\ -0.0109 & 0.0792 & -0.1153 & -0.0968 & 0.0783 \\ -0.1099 & 0.7982 & -1.1615 & -0.9751 & 0.7887 \\ 0.0458 & -0.3323 & 0.4835 & 0.4059 & -0.3283 \\ 0.0586 & -0.4257 & 0.6195 & 0.5200 & -0.4206 \end{bmatrix},$$

thus the inverse of the perturbed matrix reads

$$(A + D)^{-1} = \begin{bmatrix} 0.0391 & 0.0423 & 0.1075 & -0.0195 & 0.0651 \\ 0.0749 & 0.6645 & -0.5440 & -0.5375 & 0.7915 \\ 0.1401 & -0.9316 & 0.6352 & 0.4300 & -0.4332 \\ -0.1042 & 0.5537 & -0.2866 & 0.0521 & 0.1596 \\ 0.0586 & 0.5635 & -0.3388 & -0.0293 & 0.0977 \end{bmatrix}.$$

Acknowledgement

This work is supported by the National Science Foundation grant CCR-9505732 and Ministry of Higher Education, Science and Technology of Slovenia grant Z1-7330-0101.

References

[1] H. Arsham, A Linear Symbolic-Based Approach to Matrix Inversion, *Journal of Mathematics and Computers in Simulation*, **35** (1993), 493-500.

[2] H. Arsham, J. Grad, G. Jaklič, Perturbed Matrix Inversion with Application to LP Simplex Method, *Applied Mathematics and Computation*, to appear.

[3] J. Barle, J. Grad, On the use of dense matrix techniques within sparse simplex, *Annals of Operations Research* **43** (1993), 3-14.

[4] G. H. Golub, C. F. van Loan, *Matrix computations*. 3rd ed., Baltimore, London: The Johns Hopkins University Press, 1996.

[5] D. Hanselman, B. Littlefield, *Mastering MATLAB 7*, Pearson/Prentice Hall, 2005.

AN EP THEOREM FOR DLCP AND INTERIOR POINT METHODS

Tibor Illés*, Marianna Nagy*, Tamás Terlaky+

* Eötvös Loránd University of Science, Departement of Operations Research
Budapest, Hungary

+ McMaster University, Department of Computing and Software
Hamilton, Ontartio, Canada

illes@math.elte.hu, nmariann@cs.elte.hu, terlaky@mcmaster.ca

Abstract: The linear complementarity problem (LCP) belongs to the class of $\mathbb{NP}$-complete problems. Therefore we can not expect a polynomial time solution method for LCPs without requiring some special property of the matrix of the problem. We show that the dual LCP can be solved in polynomial time if the matrix is row sufficient, moreover in this case all feasible solutions are complementary. Furthermore we present an existentially polytime (EP) theorem for the dual LCP with arbitrary matrix.

Keywords: Linear complementarity problem, row sufficient matrix, $\mathcal{P}_*$-matrix, EP theorem

1 Introduction

Consider the *linear complementarity problem* (LCP): find vectors $\mathbf{x}$, $\mathbf{s} \in \mathbb{R}^n$, that satisfy the constraints

$$-M\mathbf{x} + \mathbf{s} = \mathbf{q}, \quad \mathbf{x}\mathbf{s} = \mathbf{0}, \quad \mathbf{x},\ \mathbf{s} \geq \mathbf{0}, \tag{1}$$

where $M \in \mathbb{R}^{n\times n}$ and $\mathbf{q} \in \mathbb{R}^n$, and the notation $\mathbf{xs}$ is used for the coordinatewise (Hadamard) product of the vectors $\mathbf{x}$ and $\mathbf{s}$.

Problem LCP belongs to the class of $\mathbb{NP}$-complete problems, since the feasibility problem of linear equations with binary variables can be described as an LCP [8]. Therefore we can not expect an efficient (polynomial time) solution method for LCPs without requiring some special property of the matrix M. The matrix classes that are important for our goals are discussed in Section 2, along with the LCP duality theorem and an EP form of the duality theorem.

Consider the *dual linear complementarity problem* ($DLCP$) [5, 6]: find vectors $\mathbf{u}$, $\mathbf{z} \in \mathbb{R}^n$, that satisfy the constraints

$$\mathbf{u} + M^T\mathbf{z} = \mathbf{0}, \quad \mathbf{q}^T\mathbf{z} = -1, \quad \mathbf{u}\,\mathbf{z} = \mathbf{0}, \quad \mathbf{u},\, \mathbf{z} \geq \mathbf{0}. \tag{2}$$

We show that the dual LCP can be solved in polynomial time if the matrix is row sufficient, as for this case all feasible solutions are complementary (see Lemma 6). This result yields an improvement compared to earlier known polynomial time complexity results, namely an LCP is solvable in polynomial time for $\mathcal{P}_*(\kappa)$-matrices with known $\kappa \geq 0$. Due to the special structure of $DLCP$, the polynomial time complexity of interior point methods depends on the row sufficient property of the coefficient matrix M. Furthermore, we present an EP theorem for the dual LCP with arbitrary matrix M, and apply the results for homogeneous LCP's.

Throughout the paper the following notations are used. Scalars and indices are denoted by lowercase Latin letters, vectors by lowercase boldface Latin letters, matrices by capital Latin letters, and finally sets by capital calligraphic letters. Further, $\mathbb{R}^n_\oplus$ ($\mathbb{R}^n_+$) denotes the nonnegative (positive) orthant of $\mathbb{R}^n$, and X denotes the diagonal matrix whose diagonal elements are the coordinates of vector $\mathbf{x}$, i.e., $X = \mathrm{diag}(\mathbf{x})$ and I denotes the identity matrix of appropriate dimension. The vector $\mathbf{x}\,\mathbf{s} = X\mathbf{s}$ is the componentwise product (Hadamard product) of the vectors $\mathbf{x}$ and $\mathbf{s}$, and for $\alpha \in \mathbb{R}$ the vector $\mathbf{x}^\alpha$ denotes the vector whose ith component is x_i^α. We denote the vector of ones by $\mathbf{e}$. Furthermore

$$\mathcal{F}_P = \{(\mathbf{x}, \mathbf{s}) \geq \mathbf{0} \;:\; -M\mathbf{x} + \mathbf{s} = \mathbf{q}\}$$

is the set of the feasible solutions of problem LCP and

$$\mathcal{F}_D = \{(\mathbf{u}, \mathbf{z}) \geq \mathbf{0} \;:\; \mathbf{u} + M^T\mathbf{z} = \mathbf{0},\ \mathbf{q}^T\mathbf{z} = -1\}$$

is the set of the feasible solutions of problem $DLCP$.

The rest of the paper is organized as follows. The following section reviews the necessary definitions and basic properties of the matrix classes used in this paper. In Section 3 we present our main results about polynomial time solvability of dual LCP's.

2 Matrix classes and LCP's

The class of $P_*(\kappa)$-matrices, that can be considered as a generalization of the class of positive semidefinite matrices, were introduced by Kojima et al. [8].

Definition 1 *Let $\kappa \geq 0$ be a nonnegative number. A matrix $M \in \mathbb{R}^{n\times n}$ is a $\mathcal{P}_*(\kappa)$-*matrix *if*

$$(1+4\kappa)\sum_{i\in\mathcal{I}_+(\mathbf{x})} x_i(Mx)_i + \sum_{i\in\mathcal{I}_-(\mathbf{x})} x_i(Mx)_i \geq 0, \quad \textit{for all } \mathbf{x} \in \mathbb{R}^n,$$

where $\mathcal{I}_+(\mathbf{x}) = \{1 \leq i \leq n : x_i(Mx)_i > 0\}$ *and* $\mathcal{I}_-(\mathbf{x}) = \{1 \leq i \leq n : x_i(Mx)_i < 0\}$.

The nonnegative number κ denotes the weight that need to be used at the positive terms so that the weighted 'scalar product' is nonnegative for each vector $\mathbf{x} \in \mathbb{R}^n$. Therefore, $\mathcal{P}_*(0)$ is the class of positive semidefinite matrices (if we set aside the symmetry of the matrix M).

Definition 2 *A matrix $M \in \mathbb{R}^{n\times n}$ is a $\mathcal{P}_*$-*matrix *if it is a $\mathcal{P}_*(\kappa)$-matrix for some $\kappa \geq 0$, i.e.*

$$\mathcal{P}_* = \bigcup_{\kappa\geq 0} \mathcal{P}_*(\kappa).$$

The class of sufficient matrices was introduced by Cottle, Pang and Venkateswaran [2].

Definition 3 *A matrix $M \in \mathbb{R}^{n\times n}$ is a* column sufficient matrix *if for all $\mathbf{x} \in \mathbb{R}^n$*

$$X(M\mathbf{x}) \leq 0 \quad \textit{implies} \quad X(M\mathbf{x}) = 0,$$

and it is row sufficient *if M^T is column sufficient. The matrix M is* sufficient *if it is both row and column sufficient.*

Kojima et al. [8] proved that any $\mathcal{P}_*$ matrix is column sufficient and Guu and Cottle [7] proved that it is row sufficient too. Therefore, each $\mathcal{P}_*$ matrix is sufficient. Väliaho proved the other direction of inclusion [10], thus the class of $\mathcal{P}_*$-matrices coincides with the class of sufficient matrices.

Fukuda and Terlaky [6] proved a fundamental theorem for quadratic programming in oriented matroids. As they stated in their paper, the LCP duality theorem follows from that theorem for sufficient matrix LCPs.

Theorem 4 *Let a sufficient matrix $M \in \mathbb{Q}^{n\times n}$ and a vector $\mathbf{q} \in \mathbb{Q}^n$ be given. Then exactly one of the following statements hold:*
(1) problem LCP has a solution $(\mathbf{x}, \mathbf{s})$ whose encoding size is polynomially bounded.
(2) problem $DLCP$ has a solution $(\mathbf{u}, \mathbf{v})$ whose encoding size is polynomially bounded.

A direct and constructive proof of the LCP duality theorem can be found in [4].

The concept of EP (existentially polynomial-time) theorems was introduced by Cameron and Edmonds [1]. It is a theorem of the form:

$$[\forall x : F_1(x), F_2(x), \dots, F_k(x)],$$

where $F_i(x)$ is a predicate formula which has the form

$$F_i(x) = [\exists y_i \text{ such that } \|y_i\| \le \|x\|^{n_i} \text{ and } f_i(x, y_i)].$$

Here $n_i \in \mathbb{Z}^+$, $\|z\|$ denotes the encoding length of z and $f_i(x, y_i)$ is a predicate formula for which there is a polynomial size certificate.

The LCP duality theorem in EP form was given by Fukuda, Namiki and Tamura [5]:

Theorem 5 *Let a matrix $M \in \mathbb{Q}^{n\times n}$ and a vector $\mathbf{q} \in \mathbb{Q}^n$ be given. Then at least one of the following statements hold:*

(1) problem LCP has a complementary feasible solution $(\mathbf{x}, \mathbf{s})$, whose encoding size is polynomially bounded.

(2) problem $DLCP$ has a complementary feasible solution $(\mathbf{u}, \mathbf{z})$, whose encoding size is polynomially bounded.

(3) matrix M is not sufficient and there is a certificate whose encoding size is polynomially bounded.

3 Main results

In this section we show that if the matrix is row sufficient then all feasible solutions of $DLCP$ are not only nonnegative, but they are complementary as well. Based on this result we get an EP theorem for problem $DLCP$.

Lemma 6 *Let matrix M be row sufficient. If $(\mathbf{u}, \mathbf{z}) \in \mathcal{F}_D$, then $(\mathbf{u}, \mathbf{z})$ is a solution of problem $DLCP$.*

Corollary 7 *Let matrix M be row sufficient. Then problem $DLCP$ can be solved in polynomial time.*

We have to note that there is no known polynomial time algorithm for checking whether a matrix is row sufficient or not. The following theorem presents what can be proved about a LCP problem with arbitrary matrix using a polynomial time algorithm.

Theorem 8 *Let matrix $M \in \mathbb{Q}^{n\times n}$ and vector $\mathbf{q} \in \mathbb{Q}^n$ be given. Then it can be shown in polynomial time that at least one of the following statements hold:*

(1) problem $DLCP$ has a feasible complementary solution $(\mathbf{u}, \mathbf{z})$, whose encoding size is polynomially bounded.

(2) problem LCP has a feasible solution, whose encoding size is polynomially bounded.

(3) matrix M is not row sufficient and there is a certificate whose encoding size is polynomially bounded.

Observe that Theorem 8 is in EP form. Both Theorems 5 and 8 deal with problem LCP, but Theorem 5 approaches the problem from the primal, while Theorem 8 from the dual side. The advantages of Theorem 8 is to determine certificates in polynomial time. The proof of Theorem 5 is constructive too, it is based on the criss-cross algorithm (for details see [4, 5]). The LCP duality theorem gives in the first two cases not only a feasible, but also complementary solutions.

We deal with the second case of Theorem 8 using a modified interior point method which either solves problem LCP with the given arbitrary matrix, or provides a polynomial size certificate in polynomial time, that the matrix of the problem is not sufficient. We can state our main result:

Theorem 9 *Let an arbitrary matrix $M \in \mathbb{Q}^{n\times n}$ and a vector $\mathbf{q} \in \mathbb{Q}^n$ be given. Then one can verify in polynomial time that at least one of the following statements hold*

(1) the LCP problem (1) has a feasible complementary solution $(\mathbf{x}, \mathbf{s})$ whose encoding size is polynomially bounded.

(2) problem $DLCP$ has a feasible complementary solution $(\mathbf{u}, \mathbf{z})$ whose encoding size is polynomially bounded.

(3) matrix M is not in the class $\mathcal{P}_*(\tilde{\kappa})$ with a given $\tilde{\kappa}$.

Let us note that Theorem 9 and Theorem 5 (a result of Fukuda et al. [5]) are different in two aspects: first, our statement (3) is weaker in some cases then theirs (there is no direct certificate in one case), but on the other hand our constructive proof is based on polynomial time algorithms and a polynomial size certificate is provided in all other cases in polynomial time.

References

[1] K. Cameron and J. Edmonds, Existentially polytime theorems, in: *Polyhedral Combinatorics* (Morristown, NJ, 1990), 83–100, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 1, American Mathematical Society, Providence, RI, 1990.

[2] R.W. Cottle, J.-S. Pang, and V. Venkateswaran, Sufficient matrices and the linear complementarity problem, *Linear Algebra and Its Applications* 114/115 (1989), 231-249.

[3] R.W. Cottle, J.-S. Pang, and R.E. Stone, *The Linear Complementarity Problem.* Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1992.

[4] Zs. Csizmadia and T. Illés, New criss-cross type algorithms for linear complementarity problems with sufficient matrices, *Optimization Methods and Software* 21 (2006), 247-266.

[5] K. Fukuda, M. Namiki, and A. Tamura, EP theorems and linear complementarity problems, *Discrete Applied Mathematics* 84 (1998), 107-119.

[6] K. Fukuda and T. Terlaky, Linear complementary and orientated matroids, *Journal of the Operations Research Society of Japan* 35 (1992), 45-61.

[7] S.-M. Guu and R.W. Cottle, On a subclass of P_0, *Linear Algebra and Its Applications* 223/224 (1995), 325-335.

[8] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise, *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, volume 538 of Lecture Notes in Computer Science. Springer Verlag, Berlin, Germany, 1991.

[9] C. Roos, T. Terlaky, and J.-Ph. Vial, *Theory and Algorithms for Linear Optimization, An Interior Point Approach*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, New York, USA, 1997. (Second edition: *Interior Point Methods for Linear Optimization*, Springer, New York, 2006.)

[10] H. Väliaho, P_*-matrices are just sufficient, *Linear Algebra and Its Applications* 239 (1996), 103-108.

SOLUTION CONCEPTS FOR INTERVAL EQUATIONS - A GENERAL APPROACH WITH APPLICATIONS TO OR.

Karel Zimmermann
Faculty of Mathematics and Physics
Charles University Prague
Karel.Zimmermann@MFF.CUNI.CZ

Abstract

General functional equations with interval inputs are considered. Appropriate solution concepts for such functional interval equations are proposed. Necessary and sufficient conditions satisfied by solutions of functional interval equations are proved. The paper provides a general framework for investigating certain class of operations research problems with interval input parameters. The approach is demonstrated on a synchronization problem with interval input parameters.

Keywords: Interval Equations, Solution Concepts to Interval Equations, Properties of Solution Concepts.

1 Introduction.

In [2], systems of interval $(max, +)$-linear equations with variables only on one side of the equations were studied. The results from [2] are based on results concerning non-interval equation systems of this type published in earlier publications (e.g. [4], [9]). Properties of non-interval systems of this type, in which variables occur on both sides of the equations (so called "two-sided systems") and methods for their solution were studied in [1]. The aim of the present paper is to propose a general framework for investigating properties of interval equations, which encompass also results in [2], [4], [9] as well as some other results from the literature concerning interval equations and inequalities . The extension is carried out using a generalized framework with functions (mappings) having a partially ordered range. Systems of $(max, +)$-linear equations can be applied to solving various types of synchronization problems, in which the coefficients of the equations represent traveling times between two places (e.g. cities) or processing times of activities. Since such traveling or processing times in real world situations may change within some bounds, one way how to manage synchronization under such conditions is to use an interval approach.

2 A Motivating Example.

Let two groups of transport means (e.g. trains and buses) be considered. Let x_j, y_k denote departure times of train j or bus k for $j \in J \equiv \{1, \ldots, n\}$, $k \in K \equiv \{1, \ldots, s\}$ respectively. Let us assume that we have m places (villages, towns, stations) $i \in I \equiv \{1, \ldots, m\}$. Let a_{ij}, b_{ik} denote the traveling time of train j or bus k to place i respectively for all $i \in I,\ j \in J,\ k \in K$. Therefore , $x_j + a_{ij}$, $y_k + b_{ik}$ are arrival times of train j or bus k to i respectively. The last train arrives to place i at time $a_i(x) \equiv max_{j\in J}(a_{ij} + x_j)$ and the last bus arrives to i at time $b_i(y) \equiv max_{k\in K}(b_{ik} + y_k)$, where we set $x = (x_1, \ldots, x_n)^T$ and $y = (y_1, \ldots, y_r)^T$. A synchronization will mean to determine departure times $x_j,\ y_k$ for all $j \in J,\ k \in K$ (i.e. finding values $x_j,\ y_k$) in such a way that $a_i(x),\ b_i(y)$ satisfy some relation $R_i \in \{=, \leq, \geq\}$. Since we cannot choose $x_j,\ y_k$ quite arbitrarily, it is additionally required that $x_j \in [\underline{x}_j, \overline{x}_j]$, $y_k \in [\underline{y}_k, \overline{y}_k]$, where $\underline{x}_j,\ \overline{x}_j,\ \underline{y}_k,\ \overline{y}_k$ are given real numbers. Until now we assumed that traveling times a_{ij} , b_{ik} are given positive numbers. However, in real world situations the traveling times cannot be usually exactly prescribed, they may vary within some intervals. To encompass such situations, we shall assume that $a_{ij} \in [\underline{a}_{ij}, \overline{a}_{ij}]$ and $b_{ik} \in [\underline{b}_{ik}, \overline{b}_{ik}]$, where $\underline{a}_{ij},\ \overline{a}_{ij},\ \underline{b}_{ik},\ \overline{b}_{ik},$ are given real numbers, i.e. $a_{ij},\ b_{ik}$ are given as interval input parameters in the sense of interval mathematics. In the further part of this contribution, we shall propose how to proceed in this situation, i.e. how to define an appropriate concept of "appropriately synchronized" departure times $x_j,\ y_k$.

3 Notations, Basic Concepts.

Let Z be a partial ordered set with order relation $\leq_Z$. We will introduce the following notation

$$[v^{(1)}, v^{(2)}] \equiv \{v \in Z; v^{(1)} \ \leq_Z \ v \ \leq_Z \ v^{(2)}\},$$

where $v^{(1)}, v^{(2)}$ are given elements from Z. Let $X,\ Y$ be given sets, F be a partially ordered set of functions $f : X \longrightarrow Z$ with partial order $\leq_F$, G be a partially ordered set of functions $g : Y \longrightarrow Z$ with partial order $\leq_G$. We will assume that the following implications hold:

$$f^{(1)},\ f^{(2)} \in F, f^{(1)} \ \leq_F \ f^{(2)} \Longrightarrow f^{(1)}(x) \ \leq_Z \ f^{(2)}(x)\ \forall x \in X$$

and similarly

$$g^{(1)},\ g^{(2)} \in G, g^{(1)} \ \leq_G \ g^{(2)} \Longrightarrow g^{(1)}(y) \ \leq_Z \ g^{(2)}(y)\ \forall y \in Y$$

Let us define the interval sets of functions $\mathbf{F},\ \mathbf{G}$ as follows:

$$\mathbf{F} \equiv \{f \in F; \underline{f} \leq_F f \leq_F \overline{f}\},$$

$$\mathbf{G} \equiv \{g \in G; \underline{g} \leq_G g \leq_G \overline{g}\},$$

where $\underline{f} \in F,\ \overline{f} \in F,\ \underline{g} \in G,\ \overline{g} \in G$ are given functions. We will assume that for all $f \in \mathbf{F},\ g \in \mathbf{G}$ sets $f(X) \equiv \{f(x); x \in X\},\ g(Y) \equiv \{g(y); y \in Y\}$ satisfy the equality $f(X) = g(Y) = Z$. Further to simplify the notations, we will omit the subscripts at the inequalities $\leq_F,\ \leq_G,\ \leq_Z$ if the meaning of the inequality follows from the context. Inequality $\geq$ is defined as usual by $h^{(1)} \geq h^{(2)} \Longleftrightarrow h^{(2)} \leq h^{(1)}$, where $h^{(1)},\ h^{(2)}$ are elements of any of the partially ordered sets $Z,\ F,\ G$.
We will consider the set of equations

$$f(x) = g(y),\ f \in \mathbf{F},\ g \in \mathbf{G}.$$

Such set of equations will be denoted further by

$$\mathbf{F}(x) \ = \ \mathbf{G}(y) \tag{3.1}$$

We will call (3.1) interval equation. It arises naturally a question how a solution of interval equation (3.1) should be defined. In the sequel, we propose several solution concepts for (3.1) using some ideas from [2], [3], [8].

Definition 3.1
A pair $(x, y) \in X \times Y$ is called a weak solution of (3.1) if there exist $f \in \mathbf{F},\ g \in \mathbf{G}$ such that $f(x) = g(y)$.

Definition 3.2
A pair $(x, y) \in X \times Y$ is called a strong solution of (3.1) if for all $f \in \mathbf{F},\ g \in \mathbf{G}$ equality $f(x) = g(y)$ holds.

Definition 3.3
(a) A pair $(x, y) \in X \times Y$ is called a left tolerance solution of (3.1) if for any $f \in \mathbf{F}$ we have $f(x) \in [\underline{g}(y), \overline{g}(y)]$.
(b) A pair $(x, y) \in X \times Y$ is called a right tolerance solution of (3.1) if for any $g \in \mathbf{G}$ we have $g(y) \in [\underline{f}(x), \overline{f}(x)]$.
(c) A pair $(x, y) \in X \times Y$ is called a tolerance solution of (3.1) if it is both left- and right tolerance solution of (3.1).

Further we will formulate an assumption under which necessary and sufficient conditions, which the solutions introduced in the preceding definitions satisfy, will be proved. These conditions make possible to describe the set of all weak, strong or tolerance solutions and in some cases find such a solution by a strongly polynomial algorithm from [1]. In other cases, in which no such algorithm exists the corresponding problems can be a challenge for further research.
Let us introduce functions (mappings) $P_x(f)$ defined on $\mathbf{F}$ for any fixed $x \in X$ as follows:

$$P_x(f) \equiv f(x)\ \forall f \in \mathbf{F}$$

Similarly we will define $Q_y(g)$ for any fixed $y \in Y$ as follows:

$$Q_y(g) \equiv g(y)\ \forall g \in \mathbf{G}$$

We will make the following assumptions:

Assumption I.
Let $x \in X,\ y \in Y$ be fixed, and $c \in [\underline{g}(y), \overline{g}(y)]$. Then there exists a function $g^{(c)} \in \mathbf{G}$ such that $c = g^{(c)}(y)$.

Assumption II.
Let $x \in X,\ y \in Y$ be fixed, and $d \in [\underline{f}(x), \overline{f}(x)]$. Then there exists a function $f^{(d)} \in \mathbf{F}$ such that $d = f^{(d)}(x)$.

Remark 3.1
Assumptions I., II. can be formulated also with the aid of mappings $P_x(f),\ Q_y(g)$ introduced above as follows. For any $c \in [\underline{g}(y), \overline{g}(y)]$ there exists $g^{(c)} \in \mathbf{G}$ such that $Q_y(g^{(c)}) = c$ (i.e. mapping $Q_y(g)$ is for any fixed $y \in Y$ surjective). Similarly, for any $d \in [\underline{f}(x), \overline{f}(x)]$ there exists $f^{(d)} \in \mathbf{F}$ such that $P_x(f^{(d)}) = d$ (i.e. mapping $P_x(f)$ is for any fixed $x \in X$ surjective).

Theorem 3.1
A pair $(x, y) \in X \times Y$ is a weak solution of (3.1) if and only if relations

$$\underline{f}(x) \le \overline{g}(y) \ , \ \overline{f}(x) \ge \underline{g}(y) \tag{3.2}$$

hold.

Proof:
If (x, y) is a weak solution of (3.1), it is $f(x) = g(y)$ for some $f \in \mathbf{F}, \ g \in \mathbf{G}$. Therefore we have:

$$\underline{f}(x) \le f(x) = g(y) \le \overline{g}(y),$$

$$\underline{g}(y) \le g(y) = f(x) \le \overline{f}(x)$$

and thus (3.2) holds.
Let us assume further that $x \in X, \ y \in Y$ satisfy inequalities (3.2). It follows that

$$J \equiv [\underline{f}(x), \overline{f}(x)] \cap [\underline{g}(y), \overline{g}(y)] \cap [\underline{g}(y), \overline{f}(x)] \cap [\underline{f}(x), \overline{g}(y)] \ne \emptyset$$

Let c be any element of J. According to Assumption I. there exist elements $f^{(c)} \in \mathbf{F}$, $g^{(c)} \in \mathbf{G}$ such that $P_x(f^{(c)}) = f^{(c)}(x) = c = Q_y(g^{(c)}) = g^{(c)}(y)$. Therefore (x, y) is a weak solution of (3.1). □

Theorem 3.2
A pair $(x, y) \in X \times Y$ is a strong solution of (3.1) if and only if relations

$$\overline{f}(x) = \underline{g}(y) \ , \ \underline{f}(x) = \overline{g}(y) \tag{3.3}$$

hold.

Proof:
If $x \in X$, $y \in Y$ is a strong solution of (3.1), the equalities (3.3) follow directly from Definition 3.1. To prove the opposite direction, let us assume that equalities (3.3) are satisfied and $f \in \mathbf{F}, \ g \in \mathbf{G}$ are arbitrary. We have to prove that $f(x) = g(y)$. We have:

$$\overline{g}(y) = \underline{f}(x) \le f(x) \le \overline{f}(x) = \underline{g}(y) \le g(y) \le \overline{g}(y)$$

Therefore we obtain $f(x) = g(y)$. □

Example 3.1
In the motivating example considered above we have:

$$f(x) = f^A(x) = A \otimes x \ , \ g(y) = g^B(y) = B \otimes y,$$

where $f^A(x) = (f_1^A(x), \ldots, f_m^A(x))^T, \ g^B(y) = (g_1^B(y), \ldots, g_m^B(y))^T$, $X \equiv \{x \in R^n \ ; \ \underline{x} \le x \le \overline{x}\}$, $Y \equiv \{y \in R^s \ ; \ \underline{y} \le y \le \overline{y}\}$, symbols $A, \ B$ denote matrices of appropriate size with elements $a_{ij}, \ b_{ik}$ respectively, $f_i^A(x) = (A \otimes x)_i \equiv \max_{j \in J}(a_{ij} + x_j)$, $g_i^B(y) = (B \otimes y)_i \equiv \max_{k \in K}(b_{ik} + y_k) \ \forall i \in I$, so that $\mathbf{F} = \{f^A \ ; \ \underline{A} \le A \le \overline{A}\}$, $\mathbf{G} = \{g^B \ ; \ \underline{B} \le B \le \overline{B}\}$ with given matrices $\underline{A}, \ \overline{A}, \ \underline{B}, \ \overline{B}$. Assumption I. follows from the continuity of functions $A \otimes x, \ B \otimes y$ with respect to $A, \ B$ for any fixed $x, \ y$.

Remark 3.2
The general framework can be applied in a similar way as in the example above also to other types of functions f, g, e.g. $f^A(x) \equiv A \otimes' x$ with $(A \otimes' x)_i \equiv \min_{j\in J}(a_{ij} + x_j)$, $g^B(y) \equiv B \otimes'' y$ with $(B \otimes'' y)_i = \max_{k\in K}(\min(b_{ik}, y_k))$, or $f^A(x) \equiv A\tilde{\otimes}x$ with $(A\tilde{\otimes}x)_i \equiv min_{j\in J}(a_{ij}x_j)$, which were considered in the literature (e.g. [4], [7], [9]).

Remark 3.3
Similar results as in Theorems 3.1, 3.2 can be proved for the tolerance solutions defined above, as well as for other solution concepts proposed in [6] (after a corresponding extension to two sided equations).

References

[1] Butkovič, P., Zimmermann, K.: Strongly Polynomial Algorithm for Solving Two-sided Systems of (max,plus)-linear Equations, DAA, 2006.

[2] Cechlárová, K.: Solutions of Interval Linear Systems in $(max, +)$-algebra, in Proceedings of the 6th International Symposium on Operational Research Preddvor (Slovenia), September 2001, pp. 321-326.

[3] Cechlárová, K., Cuninghame-Green, R., A.: Interval Systems of Max-separable Linear Equations, LAA, 340/1-3(2002), pp. 215-224.

[4] Cuninghame-Green, R., A.: Minimax Algebra, Lecture Notes in Economics and Mathematical Systems, 166, Springer Verlag, Berlin, 1979.

[5] Cuninghame-Green, R., A., Zimmermann, K.: Equation with Residual Functions, Comment. Math. Univ. Carolinae 42(2001), 4, pp. 729-740.

[6] Myšková, H.: Solvability of Interval Systems of Fuzzy Linear Equations, Proceedings of the international conference Mathematical Methods in Economics and Industry, June 2007, Herlany (Slovak Republic).

[7] Fiedler, M. , Nedoma, J., Ramík, J., Rohn, J., Zimmermann, K.: Linear Optimization Problems with Inexact Data, Springer Verlag, 2006.

[8] Rohn, J. : Systems of Linear Interval Equations,, Linear Algebra and Applications, 126 (1989), pp. 39-78.

[9] Vorobjov, N., N.: Extremal Algebra of Positive Matrices, Elektronische Datenverarbeitung und Kybernetik, 3, 1967, pp. 39-71 (in Russian).

Supported by GA ČR 202-03/2060982, GA ČR # 402/06/1071.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 4

Multicriteria Decision Making

THE ROLE OF INCONSISTENCY IN AUTOMATICALLY GENERATED AHP PAIRWISE COMPARISON MATRICES

Andrej Bregar, Jozséf Györkös, Matjaž B. Jurič
University of Maribor, Faculty of Electrical Engineering and Comupter Science
andrej.bregar@uni-mb.si, jozsef.gyorkos@uni-mb.si, matjaz.juric@uni-mb.si

Abstract: Slight inconsistency of the Analytic Hierarchy Process pairwise comparison matrices is treated with regard to the validity and reliability of derived priority vectors. Three transformation functions for the automatic construction of matrices are experimentally evaluated with a simulation model. They are based on the linear or multiplicative scale, respectively, and exhibit various levels of consistency. It is shown that slightly inconsistent matrices produce more accurate weights than consistent ones.

Keywords: Multi-criteria decision analysis, Analytic Hierarchy Process, Inconsistency rate, Criteria weights, Simulation experiments

1. INTRODUCTION

One of the fundamental and most widely used multi-criteria decision analysis methods is the Analytic Hierarchy Process (AHP) [3, 5]. It is based on pairwise comparisons of criteria and alternatives. It follows the presumption that the human mind is uncapable of processing large sets of items at a time because of its short-term memory capacity [6]. The number of available decision-maker's responses must thus be restricted to 7 ± 2. If the upper scale limit would exceed 9, a higher heterogeneity of judgements would reduce the overall consistency. Analogously, the number of directly compared elements in a hierarchic group should also be accordant with the Miller's limitation of 7 ± 2. Otherwise, the resulting inconsistency rate is too low to give the decision-maker the opportunity to identify the most contradictive element and to modify its relations with other elements, in order to improve the validity of weights [8]. Hence, when pairwise comparisons are (near) consistent, forced adjustments may impair the correctness of the derived priority vector. Moreover, the decision-maker usually avoids alterations of such matrices, which hinders his ability to restructure thought patterns and to deepen the understanding of the problem situation/domain. The implication is that modest inconsistency rates should be strived for. The threshold of acceptable incosistency has been set to 0.1, which is an order of magnitude smaller than the decision-maker's evaluations [3].

Because Saaty suggests that slightly inconsistent matrices are more efficient than totally consistent ones, it is the goal of the presented research to show if the weight vectors derived from them are also more reliable and correct. A case of automatically generated matrices is considered. Such matrices are constructed from existing weights for the purpose of providing the decision-maker with a means of adjusting computationally inferred objective information based on the set of feasible alternatives with his subjective assessment of importance that is bound to his personal experience, knowledge, wishes and the way of thinking. These weights are obtained from the values of other input parameters, such as the veto thresholds [2]. There do not exist any studies that would evaluate the effect of inconsistency in the same context, although some research has been performed to deal with AHP inconsistency rates in general or in different decision analysis settings [1].

2. EXPERIMENTAL MODEL

2.1 Evaluated matrix construction functions and measurement scales

The original AHP [5] derives priorities from a deterministic pairwise comparison matrix with the principal right eigenvector based decomposition. Although some evidence indicates a sufficient efficiency of this approach [7], many researchers believe that it is inappropriate in various decision-making situations. For this reason, several types of measurement scales and several versions of AHP have been defined. Probably the most popular among them is the geometric scale based multiplicative AHP [9]. It is one of the decision-maker's prinicipal tasks to choose the most suitable method/scale since it has been proven that none performs ideally in all cases [4].

To simulate different scales and different levels of inconsistency, three matrix generation functions are introduced. They all take an original automatically inferred weight vector and transform it into a pairwise comparison matrix for the purpose of offering the user an insight into preferences and the possibility to perform interactive adjustments. However, this study is not intended to determine the decision-maker's responses, but to assess the correctness of generated weights with regard to the incosistency of pairwise comparison judgements. Thus, a new weight vector is derived from the unchanged, automatically constructed matrix, and it is observed to what extent the two vectors deviate.

The first function is linear. It results in a classic 1 to 9 scale AHP matrix by transforming differences of original criteria weights into ratios:

$$r_{ij} = \frac{8}{\Delta_{\max}} \cdot \Delta_{ij} + 1 .$$

Only non-negative differences are considered; if $\Delta_{ij} < 0$, a reciprocal value $r_{ij} = 1 / r_{ji}$ is taken. The constant $b = 1$ ensures that $\Delta_{ij} = 0$ is transformed to the $r_{ij} = 1$ ratio, which indicates total equality of criteria. The linear function does not guarantee matrix consistency, since for any ratios r_{ij}, r_{jk} and r_{ik}, which are computed in accordance with Δ_{ij}, Δ_{jk} and $\Delta_{ik} = \Delta_{ij} + \Delta_{jk}$, the following relation holds:

$$r_{ik} = a \cdot \Delta_{ik} + b = a \cdot (\Delta_{ij} + \Delta_{jk}) + b = (r_{ij} - b) + (r_{jk} - b) + b,$$
$$r_{ik} - b = (r_{ij} - b) + (r_{jk} - b).$$

To ensure total consistency of the pairwise comparison matrix, the exponential function is defined:

$$r_{ij} = (r_{\max})^{E} , \ E = \frac{\Delta_{ij}}{\Delta_{\max}} .$$

Now $r_{ij} = 1$ if $\Delta_{ij} = 0$, $r_{ij} = r_{\max} = 9$ if $\Delta_{ij} = \Delta_{\max}$, and $r_{ik} = r_{ij} \cdot r_{jk}$ for $\Delta_{ik} = \Delta_{ij} + \Delta_{jk}$. Transitivity is easily proven:

$$r_{ij} \cdot r_{jk} = (r_{\max})^{\Delta_{ij}/\Delta_{\max}} \cdot (r_{\max})^{\Delta_{jk}/\Delta_{\max}} = (r_{\max})^{(\Delta_{ij}+\Delta_{jk})/\Delta_{\max}} = (r_{\max})^{\Delta_{ik}/\Delta_{\max}} = r_{ik} .$$

In addition to the eigenvalue approach, with which priority vectors are computed from positive reciprocal matrices, the geometric mean method is statistically evaluated. Using the multiplicative AHP, the preference ratios $r_{ij} = (8 / \Delta_{\max}) \cdot \Delta_{ij} + 1$ are converted to values on the geometric scale with the equation $\exp(\gamma \cdot r_{ij})$, where $\gamma = \ln 2$. The multiplicative AHP matrices of human judgments are skew-symmetric, which means that $r_{\max} = 8$, $r_{ii} = 0$ and $r_{ji} = -r_{ij}$.

2.2 Random sampling of experimental data

The approaches were evaluated with simulation consisting of 1000000 test cases. In each trial, a fuzzy relation was randomly generated. It was represented by a matrix of discordance indices $d_j(a_i)$, where $i = 1, \ldots, m$, $j = 1, \ldots, n$. The number of criteria n was held at 8, while the number of alternatives m varied. It was selected from a uniform distribution over the discrete interval [4, 20]. The $d_j(a_i)$ indices were obtained by transformations of random numbers R_N

from a uniform distribution over the interval [0, 1]. For this reason, the P_S and P_W thresholds were introduced to divide the set of indices into 3 subsets:

$$d_j(a_i) = \begin{cases} 1 & , R_N < P_S \, , \\ (R_N - P_S)/(P_W - P_S) & , P_S \leq R_N < P_W \, , \\ 0 & , R_N \geq P_W \, . \end{cases}$$

Here, $P_S \leq P_W$. In each trial of the simulation, the P_S and P_W thresholds were randomly selected from uniform distributions over the intervals [0.1, 0.3] and [0.3, 0.5], respectively. Original weights were inferred from the $d_j(a_i)$ indices by using the selective strengths based algorithm [2]. These weights were then transformed into pairwise comparison matrices by applying the defined functions – linear, exponential and multiplicative. Finally, new weights were derived from generated AHP matrices.

To measure the inconsistency rates of AHP pairwise comparison matrices obtained with the linear transformation operator, the sampling procedure was slightly modified. Instead of being randomly chosen, the number of alternatives was fixed, so that four combinations of m and n were observed: $m=n=6$, $m=8$ and $n=6$, $m=4$ and $n=10$, or $m=30$ and $n=10$. Only 10000 trials were performed for each combination. The inconsistency rate was approximated with the power method.

2.3 Hypotheses

H_1: AHP matrices generated with the linear function have an acceptable inconsistency rate because they preserve the transitivity of ordinal criteria rankings.

Saaty believes that because of the short-term memory capacity the decision-maker is able to accurately and validly process only a few elements at once [6]. Otherwise, the consistency may be too low, which may in turn result in an unreasonable decision. On the other hand, too high, or even total, consistency is not recommended as well since it distracts the decision-maker from identifying poor judgements, and consequently disables knowledge generation. It has been proven that an order of magnitude smaller numerical valuations are relevant for assessing the rate of inconsistency than for expressing preferences [8]. Thus, the upper limit of acceptable inconsistency is set to 0.1 [3].

As the linear transformation function was defined in Section 2.1, it was mathematically proven that it does not assure total consistency. For this reason, it is necessary to confirm its practical usefulness by experimentally determining its average and maximal possible rates of inconsistency. The transformation is:

- unacceptable, if the average inconsistency rate exceeds 0.1;
- conditionally acceptable, if the average and maximal inconsistency rates equal to 0, or are insignificantly higher;
- acceptable, if the maximal inconsistency rate does not exceed 0.1, while the average approaches the centre of the [0, 0.1] interval.

Although the linear function does not result in total consistency, it generally satisfies the characteristic of rank order transitivity. A low rate of inconsistency is hence expected. Slight deviations in cardinal judgements should be its only source.

H_2: AHP matrices generated with the linear transformation function preserve information of original weights irrespectively of the chosen measurement scale, and increase the contribution of weak criteria.

There exist several reasons for the transformation of the original weights into a pairwise comparison matrix:

- a potential occurence of »zero« weights is prevented;

- automatically inferred preferential information on importance of criteria is presented to the decision-maker in a clear and comprehensible way;
- the adjustment of inferred preferences is enabled with regard to the decision-maker's personal expectations and points of view.

As a consequence, the original weights should change, yet under no circumstances can the reachness of criteria discriminating information decrease. In this case, the transformation function would cause the loss of existing relevant preferences that must be incorporated in the specified weights reflecting the problem situation. It is thus necessary to determine the extent to which the discriminating information is preserved, or possibly, enriched. Several metrics are applied for this purpose, including the distances between corresponding elements of compared vectors, as well as the range, extremes and asymmetry of weight intervals.

The AHP derived weights should not deviate considerably from the original weights with respect to the absolute distance measure. Especially, the range of the weight interval should be preserved. It may change only slightly in order to solve the problem of »zero« weights. Based on the theoretical definitions of both linear transformations – to the [1, 9] and [0, 8] scales – it can be assumed that the specified requirements are satisfied. These functions take into consideration the entire $[0, \Delta_{max}]$ interval, and treat the Δ_{ij} differences in an unbiased and uniform manner.

H_3: Information is lost if AHP matrices are generated with the exponential transformation.

The exponential function determines weight ratios by accentuating large differences Δ_{ij} between original weights. Because such differences are uncommon, many pairs of criteria are given the same, or similar, priority ratios. Consequently, criteria tend to become equally important, without exceptionally influential ones. It is hence expected that the discriminating information vanishes.

H_4: The weights that are derived from the AHP matrix are less extreme than the original weights irrespectively of the applied transformation operator.

The ratio of any two weights should not be too high or too low, which means that it must not exceed 75 [10]. One of the crucial tasks of pairwise comparison based priority derivation techniques is thus to ensure adequate weights of uninfluential criteria. Thereby, the highest ratio would fall under 75, which would prevent potential unacceptable extremeness.

3. EXPERIMENTAL RESULTS

Experiments show that the inconsistency rate of a pairwise matrix, which is constructed with the linear function, is very low. The results are summarized in Table 1. All values are considerably better than required. They do not rise above the allowed CR = 0.1 threshold in any test case. Because the measured rates range in the lower part of the [0, 0.1] interval, the identification of those judgements that would improve the validity of weights is not an easy task. However, it is still possible.

Table 1: Inconsistency rates of pairwise comparison matrices generated with the linear function

	$m=6, n=6$	$m=8, n=6$	$m=4, n=10$	$m=30, n=10$
average	0.0106	0.0104	0.0112	0.0106
standard deviation	0.0050	0.0050	0.0032	0.0030
maximum	0.0309	0.0268	0.0236	0.0146

It is experimentally confirmed that the linear transformation function preserves ordinal transitivity of preferences, and also improves cardinal transitivity when m and n increase. As it does not fully prevent the generation of problem domain knowledge, it can be concluded that it reflects judgements properly. Thereby, H_1 is proven.

Table 2 gives L_1-metric distances between original vectors and weight vectors obtained by all three types of transformations. It presents the following measures: the average distance between elements of two vectors, the average difference of corresponding elements deviating the most, and the extreme difference of two elements with the identic index.

Table 2: Distances between weight vectors

	average	average maximum	maximum
linear function, classical 1 to 9 scale	0.0149	0.0466	0.3143
linear function, multiplicative AHP	0.0214	0.0828	0.4075
exponential function	0.0392	0.1120	0.4698

In most cases, a matrix constructed with the linear function results in criteria weights that do not noticeably deviate from the original weights. This is especially evident for the weight derivation method which is based on the principal eigenvalue problem. On the contrary, the exponential function generally causes moderate changes of criteria importance coefficients. The interpretation of data in Table 3 confirms this assumption.

Table 3: Ranges of criteria weights

		original weights	linear	multiplicative	exponential
range	average	0.1851	0.1717	0.2142	0.0603
	deviation	0.0528	0.0458	0.0818	0.0452
minimum	average	0.0439	0.0634	0.0557	0.1010
	deviation	0.0219	0.0121	0.0155	0.0186
maximum	average	0.2291	0.2351	0.2699	0.1614
	deviation	0.0392	0.0405	0.0731	0.0368
skewness	average	0.2125	0.7099	0.8746	0.3422
	deviation	0.5852	0.6053	0.6917	1.3343
deviation from average		0.0580	0.0549	0.0683	0.0232

The ranges of weight vectors produced with the linear transformation function are similar to the ranges of original weights. This means that the richness of discriminating information is preserved, which is additionally confirmed by the deviations indicating that there are no radical discrepancies even in the worst cases. The essential distinction is that minimum and maximum values are shifted to the right. The lowest new criterion weight is thus higher than the lowest original weight. This represents a significant benefit, because no criterion should be assigned a (near) 0 importance coefficient. The weighting of weak criteria is especially efficient in the context of linear transformation, which results in the highest average minimal weight and the narrowest distribution of measured values around it. The multiplicative AHP derives less stable priorities, but it still outperformes the original weights.

Based on the obtained and interpreted results from Tables 2 and 3, the H_2 hypothesis can be confirmed. The same measurements also prove H_3. The intervals of weights, which are inferred in accordance with the exponential function, are observably narrow. Criteria tend to become equally important because the highest weights decrease and the lowest increase. In this way, the discriminating information vanishes.

Interesting results are given in Table 3 by the skewness indices of distributions of criteria importance coefficients. Sets of original weights are more symmetrical than new vectors. In the case of the linear transformation function, weights are clustered to the right of the mean, with a few high values on the left. This should not be regarded as a drawback, since it is unnatural to be confronted with perfectly symmetrical distributions of weights. Rather, it is reasonable that several principal criteria are chosen by the decision-maker. Moreover, the measured average levels of skewness are adequately moderate. In the case of the exponential function, high deviation from the average skewness should be noticed. Because weights are uniform, small perturbations are enough to obstruct symmetry.

Hence, it can be argued that the linear function at least preserves, or even improves, the richness of provided information. On the contrary, information vanishes if the exponential function is used. Slight inconsistency is therefore clearly proven to enable more correct and better preferential judgements than forced consistency. But yet, the exponential function can be sensibly applied in situations when the decision-maker aims at accentuating the influence of a few really strong criteria, while making the remaining criteria uniform.

The H_4 hypothesis is confirmed with the weight ratios of the most important criterion and the least important criterion, which are presented in Table 4. All three transformation types reduce the ratio obtained for the original weights. This is especially true for the exponential function, however the linear transformation results in a substantial improvement as well. The multiplicative AHP is the only approach that provides no significant benefits. The ratios are not overly extreme because they do not reach the proposed limit of 75.

Table 4: Ratios of average weights of the most important criterion and the least important criterion

original weights	linear function	multiplicative AHP	exponential function
5.2187	3.7082	4.8456	1.5980

4. CONCLUSION

The influence of inconsistency in the automatically generated AHP pairwise comparison matrices was discussed and experimentally treated. By proving the H_1 to H_4 hypotheses, the following thesis was confirmed: slightly inconsistent AHP matrices allow for the efficient and reliable presentation of preferential information on criteria weights which is reacher than when it is derived from totally consistent matrices. Within the scope of further research, the decision-makers' reactions to the proposed weights will be empirically tested.

5. REFERENCES

[1] Aull-Hyde, R., Erdogan, S., Duke, J. An experiment on the consistency of aggregated comparison matrices in AHP. European Journal of Operational Research, 171 (1), 290–295, 2006.

[2] Bregar, A., Györkös, J. Semiautomatic determination of criteria weights according to veto thresholds in the case of the localized alternative sorting analysis. Proceedings of the 7th International Symposium on Operational Research in Slovenia, 267–274, 2003.

[3] Forman, E. H., Selly, M. A. Decision by Objectives. World Scientific, 2001.

[4] Ishizaka, A., Balkenborg, D., Kaplan, T. AHP does not like compromises: The role of measurement scales. Proceedings of the EURO Working Group on Decision Support Systems Workshop, 46–54, 2005.

[5] Saaty, T. L. The Analytic Hierarchy Process. McGraw-Hill, New York, 1980.

[6] Saaty, T. L. The seven pillars of the Analytic Hierarchy Process. Proceedings of the 5th International Symposium on the Analytic Hierarchy Process, 1999.

[7] Saaty, T. L. Decision-making with AHP: Why is the principal eigenvector necessary. European Journal of Operational Research, 145 (1), 85–91, 2003.

[8] Saaty, T. L., Ozdemir, M. Why the magic number seven plus minus two. Mathematical and Computer Modelling, 38 (3–4), 233–244, 2003.

[9] Van den Honert, R. C. Stochastic group preference modelling in the multiplicative AHP. European Journal of Operational Research, 110 (1), 99–111, 1998.

[10] Zanakis, S., Solomon, A., Wishart, N., Dublish, S. Multi-attribute decision making: A simulation comparison of select methods. European Journal of Operational Research, 107 (3), 507–529, 1998.

MULTI-CRITERIA ASSESSMENT OF CONFLICTING ALTERNATIVES: EMPIRICAL EVIDENCE ON SUPERIORITY OF RELATIVE MEASUREMENTS

Andrej Bregar, Jozsef Györkös, Matjaž B. Jurič
University of Maribor, Faculty of Electrical Engineering and Comupter Science
andrej.bregar@uni-mb.si, jozsef.gyorkos@uni-mb.si, matjaz.juric@uni-mb.si

Abstract: Application of absolute and relative measurements in multi-criteria preference aggregation is treated. Four methods/operators which exhibit various levels of relativeness are defined and evaluated with a simulation based experimental model. Several variables are considered – probability of strict/weak preference, number of alternatives, type of random distribution, richness of information, sensitivity to inputs, ability to discriminate conflicting alternatives, and rank reversal. Contrarily to the established beliefs, superior efficiency of relative over absolute assessment is empirically proven.

Keywords: Multi-criteria decision analysis, Preference aggregation, Simulation experiments

1. INTRODUCTION

In multi-criteria decision analysis, two types of judgements exist – absolute and relative [8]. The former are characteristic of the utility theory [4] and ideal solution based methods, such as TOPSIS [11]. Each alternative is either evaluated with regard to predefined quality levels or compared with a single ideal/antiideal solution. This is a normative process, which assumes mutual preferential independence of available alternatives. The latter are applied by the Analytic Hierarchy Process [7] and several outranking methods [1, 6]. Such descriptive approaches rely on pairwise comparisons. Each alternative is compared to all, or a subset of, other alternatives. Its numerical evaluation and rank are hence dependent on the number and quality of competing options. It has been argued that nonconformation to the independency axiom is the primary drawback of methods based on relative measurements. If an alternative is eliminated or if its identic copies are added, rank order of existing alternatives can change, which may potentially result in a different decision.

Saaty has studied rank preservation and reversal [8]. He has concluded that robustness of rank orders is implied by the chosen preference aggregation technique. Absolute assessment must unconditionally preserve ranks, so that the presence/absence of any alternative does not influence evaluations of other alternatives. However, when relative pairwise comparisons are made, rank preservation is not required, since as a consequence of structural dependency, the increase in the number of copies generally reduces their value, except in the case of synergy that originates from functional dependency.

The goal of the presented research is to experimentally prove that absolute preference aggregation methods do not evaluate alternatives more reliably than those which are based on relative pairwise comparisons, although they satisfy the independence axiom and hence do not cause rank order reversals. To fulfil the goal by maximally reducing the complexity of experiments, decision making models with conflicting alternatives are considered. In real life situations, it is often the case that certain options perform well on some and poor on the other criteria, while exactly the contrary characteristics may be observed for disjunctive subsets of alternatives and criteria. This typically occurs when cost is considered. Alternatives with low (i.e. good) cost are usually unacceptable with regard to the majority of other criteria, and the opposite. In mathematical terms the concept can be defined as: alternatives from the A' (A'') subset are good according to criteria from the X' (X'') subset and bad according to criteria from X'' (X'), where A' and A'' respectively X' and X'' are disjunctive, such that $A' \cap A'' = \varnothing$,

$X' \cap X'' = \emptyset$, $A = A' \cup A''$ and $X = X' \cup X''$. When the A' subset is small in comparison with A'', so that $|A'| << |A''|$, and cost/benefit criteria are balanced, so that $|X'| \approx |X''|$, alternatives $a_i \in A'$ should be preffered over alternatives $a_j \in A''$ because of the exclusion effect. This means that if some criteria give good evaluations to many alternatives, their influence must be reduced since they do not have enough discriminating power to differentiate between choices [12].

It is reasonable to presume that absolute assessment methods are not able to distinguish between conflicting alternatives, to identify weakly discriminating criteria, and to decrease the influence of these criteria. To gain such possibilities, criterion-wise characteristics of all feasible alternatives must be compared, since the obtained information is relative to a given problem situation. Therefore, four aggregation methods/operators are defined, which exhibit various levels of absoluteness/relativeness. They have already been applied to infer criteria weights from the selective intensity of veto thresholds [2], but can also be easily generalized with the purpose of evaluating and rank ordering alternatives. The aim of the research is to measure the effectiveness of these methods/operators with regard to several variables, as are sensitivity to input parameters of the decision model, sensitivity to adding new and copies of existing alternatives, and ability to discriminate conflicting alternatives and criteria.

2. TESTED PREFERENCE AGGREGATION METHODS/OPERATORS

Criterion-wise evaluations of alternatives are expressed in the form of a fuzzy preference relation $P = \{p_j(a_i)\}$, where $0 \le p_j(a_i) \le 1$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, $a_i \in A$ is the i-th alternative, $x_j \in X$ is the j-th criterion and $p_j(a_i)$ is the fuzzy preference degree of a_i with regard to x_j. The most simple and straightforward way to compute the overall preference for a_i is to average partial fuzzy degrees by applying the following operator:

$$p(a_i) = \frac{1}{n} \cdot \sum_{j=1..n} p_j(a_i) \, .$$

This operator is absolute, as it does not compare alternatives. It is a member of the most widely used family of weighted averaging operators [5, 12]. However, to identify conflicting alternatives and to prioritise criteria with sufficient discriminating power, relative pairwise comparisons are required. A possible approach is to infer the so called preferential strengths of alternatives. For this purpose, all alpha-cuts of the fuzzy relation P are taken. The partial strength of each alternative a_i is calculated for each criterion and crisp relation. It indicates to which degree a single alternative outperforms the weakest one:

$$\varphi_{ji}^{k} = \begin{cases} \text{card}(a_l \in A \setminus \{a_i\} : p_j(a_l) < \alpha_k) & , p_j(a_i) \ge \alpha_k \\ 0 & , p_j(a_i) < \alpha_k . \end{cases}$$

Partial preferential strengths are aggregated in two ways. The first approach is based on the weighted sum function in which cut levels α_k are considered as weights:

$$\Phi_i = \sum_{k=1..l} \sum_{j=1..n} \alpha_k \cdot \varphi_{ji}^{k} \, .$$

Since this linear transformation is absolute, only the previous cardinality based equation represents a source of relative measurements. Yet, the level of relativeness may be increased by obeying several principles:

- The alternative a_i gains the highest strength at the first cut for which the $p_j(a_i)$ degree exceeds the α_k threshold.
- If $\varphi_{ji}^{k_1} = \ldots = \varphi_{ji}^{k_h}$ for adjacent $\alpha_{k_1} > \ldots > \alpha_{k_h}$, only the highest level cut is considered.

- If the difference $\delta = \varphi_{ji}^{k} - \varphi_{ji}^{k'}$ exceeds 0 for $\alpha_{k'} < \alpha_k$, the strength of a_i falls by $\alpha_{k'} \cdot \delta$.

The total preferential strength Φ_i of the alternative a_i is not based on average values of partial results. Instead, it relies on similarities between alpha-cuts. It should therefore be able to ensure a more consistent result than would be achieved with the weighted sum function. A simple algorithm may be defined based on the former principles [2]. However, it can also be expressed by means of an aggregation operator:

$$\Phi_i = \sum_{k=1..l} \sum_{j=1..n} \alpha_k \cdot \left(\varphi_{ji}^{k} - \varphi_{ji}^{k-1}\right), \varphi_{ji}^{0} = 0 .$$

The third relative approach constructs a fuzzy binary relation on the set of alternatives $B = \{((a_i, a_j), \mu_B(a_i, a_j)) \mid (a_i, a_j) \in A \times A\}$, which is interpreted with the assertion "a_i is at least as preferred as a_j". It is constructed by applying the triangle superproduct composition:

$$\mu_B(a_i, a_j) = \cap_{k=1..m} (p_k(a_i) \leftarrow p_k(a_j)) .$$

The Lukasiewitcz's implication is used for the inner, while Werners' fuzzy "and" serves as the outer aggregation operator:

$$\mu_B^k(a_i, a_j) = \min(1 - p_k(a_j) + p_k(a_i), 1) ,$$

$$\mu_B(a_i, a_j) = \gamma \cdot \min_{k=1..m} \mu_B^k(a_i, a_j) + \frac{1-\gamma}{m} \cdot \sum_{k=1..m} \mu_B^k(a_i, a_j) , 0 \le \gamma << 1 .$$

For the purpose of being analysed, the obtained binary relation has to be at least a fuzzy quasiorder relation. Therefore, its transitive closure is found. With respect to each cut-level α_k, a different partial order of alternatives is derived. These orders are combined into a single weak order with a procedure based on a distance measure between preference, indifference and incomparability relations. The procedure computes and compares dominance indices. It relies upon the presumption that a_i is the more preffered the more are relations in which it is with the other alternatives $a_j \in X \setminus \{a_i\}$ distant from the antiideal considering all cut-levels:

$$\Theta(a_i) = \sum_k \sum_{j \ne i} \alpha_k \cdot \pi(\prec, R_{ij}^k) \text{, where } R_{ij}^k \in \{\succ, \prec, \approx, ?\} \text{ and } \pi(\cdot,\cdot) \in \{a, 4a/3, 5a/3, 2a\} .$$

3. EXPERIMENTAL MODEL

3.1 Independent variables

In order to compare the efficiency of absolute and relative measurements in the context of multi-criteria decision-making, several independent variables are defined:

- *Method/operator* can compute average preference degrees, weighted sums of partial preference strengths, total preference strengths or dominance indices. The first of the four approaches is absolute, while the last three exhibit various levels of relativeness.
- *Number of criteria* may be $n \in \{4, 12, 20\}$. *Number of observed alternatives* is fixed to $m = 8$ because only the $m : n$ ratio is significant.
- *Probability of strict preference* $p_j(a_i) = 1$ may be $P_S \in \{0.1, 0.3, 0.5\}$, and *probability of weak preference* $p_j(a_i) > 0$ may be $P_W \in \{0.3, 0.6, 0.9\}$. In accordance with the P_S and P_W probabilities, preference matrices are randomly generated.
- *Random distribution* is always uniform, but it can be unbiased or biased. In this way, it is determined whether the characteristics of the observed method/operator change when it is applied to strictly conflicting instead of arbitrarily generated alternatives.

3.2 Random sampling of experimental data

The approaches were evaluated by conducting statistical experiments consisting of 10000 test cases for each type of random distribution and for each of 21 parameter combinations that were determined by the values of n, P_S and P_W variables. In each simulation trial, the fuzzy preference relation P was randomly generated. It was represented by a matrix of $p_j(a_i)$ indices, where $i = 1, \ldots, m$, $j = 1, \ldots, n$. These indices were obtained by transforming random numbers R_N from a uniform distribution over the interval [0, 1] in the following way:

$$p_j(a_i) = \begin{cases} 1 & , R_N < P_S , \\ (R_N - P_S)/(P_W - P_S) & , P_S \le R_N < P_W , \\ 0 & , R_N \ge P_W . \end{cases}$$

In the case of biased distribution, random number matrices were modified. The upper left quadrant represented l alternatives from the A' subset which were prioretized with regard to criteria $x_1, \ldots, x_{n/2}$ from the X' subset, and the lower right quadrant included $m-l$ disjunctive alternatives from A'' which were preffered according to criteria $x_{n/2+1}, \ldots, x_n$ from X''. In the upper right and lower left quadrants, all $p_j(a_i)$ indices were set to 0 to simulate conflictness. Two different situations were considered – for $l = |A'| = 1$ and $l = |A'| = 2$. It is not sensible to choose higher values of l, because the unbiased uniform distribution is obtained for $l = 4$.

3.3 Dependent variables

Sensitivity to input parameters of the decision model shows how numerical evaluations of alternatives deviate when the n, P_S and P_W levels change and thereby result in a different experimental combination. It is not reasonable to expect that diverse problem settings imply consistent or even identical numerical assessments. Several metrics are used to quantify this and the next two variables. They are all based on averages, deviations, ratios and distances of valuations obtained for different simulation trials.

Richness of discriminating information determines to what extent numerical assessments of alternatives vary. It is undesirable that all options have similar values, since in this case it is difficult to distinguish between them in terms of preference, and consequently to select the most appropriate one. However, evaluations should not be too extreme. It has been proven that the highest acceptable ratio of the best and worst ranked alternatives' values is 75 [10].

Ability to discriminate conflicting alternatives and criteria is determined by comparing numerical evaluations of alternatives resulting from different distribution types – biased and unbiased. A method which has the observed ability should not produce the same results for these two distributions. Particularly, l best ranked alternatives should stand out.

Sensitivity to adding new and copies of existing alternatives is one of the most important issues of multi-criteria decision analysis. According to the belief of many researchers, each useful and efficient method should satisfy the independency axiom, which states that if an alternative is removed or its identical copies are added, rank ordering of original alternatives should stay the same. However, Saaty disagrees with this opinion [8]. Because his arguments are sound, the following directions are established within the scope of presented research:

- If distinct new alternatives are added to the A set, rank order preservation is generally neither required nor expected. In this case, new columns of randomly generated $p_j(a_i)$ indices expand the fuzzy preference matrix P. Thereby, the decision-making context is substantially and nondeterministically changed.
- Methods/operators based on absolute assessments are obliged to preserve rankings of original alternatives if their copies are made. This means that the absence/presence of a certain alternative does not influence other alternatives.

- Because of structural dependency, relative pairwise comparisons do not require rank preservation. Nonetheless, rank reversals must be moderate and reasonable.

In this experimental study, rank reversals are measured to determine the »relativeness« degrees of particular aggregation methods/operators, and to potentially expose those of them which cause unmodest perturbations. Only situations of adding alternatives are considered, as it is presumed that removal has a similar effect. To identify possible convergence, $l \in \{1, 2, 3, 4\}$ alternatives are added. The procedure for measuring rank reversals is as follows:

1. the P matrix is randomly generated for m original alternatives;
2. m alternatives are rank ordered according to their computed numerical values;
3. the P matrix is extended with l new/copied alternatives;
4. numerical values of $m+l$ alternatives are calculated;
5. l added alternatives are discarded, so that m original alternatives are ordered again;
6. the discrepancies between both rank orders of m original alternatives are measured.

Steps 3 to 6 are repeated 100 times. Three metrics are used to measure the discrepancies between rank orders. The first is the percentage of simulation trials in which the best ranked alternative changes. The second is the Kemeny-Snell distance between rank orders [3]. Since this metric does not take into acount at which ranks reversals occur, the weighted distance is introduced. It assumes that higher ranked alternatives are more relevant for decision-makers than lower ranked alternatives. It satisfies all three axioms of a distance metric – symmetry, nonnegativity and triangle inequality – and is a slight improvement of an existing metric [9]:

$$d(A,B) = \frac{1}{2} \cdot \sum_{i=1..m} \left(\left(m+1-R_i^A \right) + \left(m+1-R_i^B \right) \right) \cdot \left| R_i^A - R_i^B \right|,$$

$$d_{\max}(A,B) = \frac{m+1}{2} \cdot \sum_{i=1..m} \left| i - (m+1-i) \right|.$$

4. EXPERIMENTAL RESULTS

4.1 Rank preservation

As can be seen from Figure 1, all four methods/operators perform similarly when a new randomly generated distinct alternative is added to the A set. According to expectations, rank reversals occur because a nondeterministic expansion of the P matrix changes the decision-making situation. The smaller the matrix is, the higher impact the addition of a new column has. Hence, robustness increases proportionally with the number of criteria. On Figure 1, and in the subsequent text, PS stands for total preference strengths, WS denotes weighted sums, FD are average fuzzy preference degrees and DI are dominance indices.

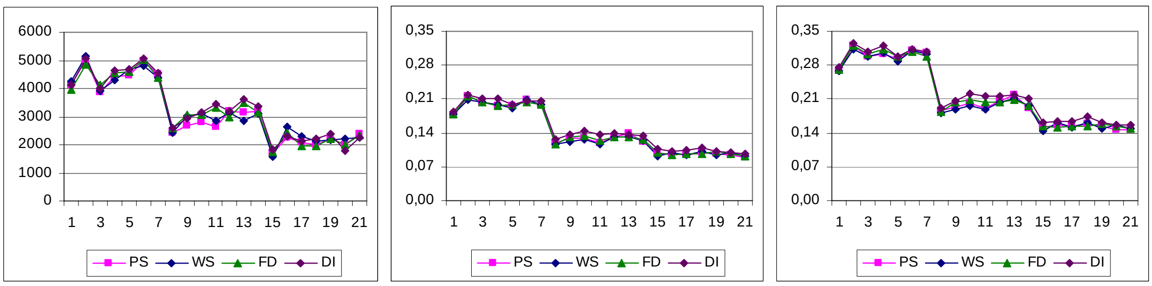


Figure 1: Best rank reversals, Kemeny-Snell distances and weighted distances for $l = 1$ new alternative

A high correlation between rank reversal metrics may be observed. Paerson correlation coefficients range from 0.96 to 1.00 for various combinations of metrics and l values. Also, when l rises to 2, 3 or 4, almost identic results are obtained. The weakest correlation is 0.91.

Differences between the applied methods/operators become clearer as copies of existing alternatives are added. The FD operator is absolute, and hence totally preserves rank orders. The DI method causes minor rank reversals. New elements $\mu_B(a_{m+1}, a_j) = \mu_B(a_j, a_{m+1}) = 1$, and $2 \cdot m - 2$ copies of existing elements, which result from the Lukasiewitcz implication, do not alter original m^2 relations in the fuzzy binary matrix B because the set of criteria according to which alternatives are compared remains the same for all alpha-cuts. However, new relations are introduced. They affect partial rank orderings that are aggregated with the antiideal based algorithm. Dominance indices thus change in the aggregation phase.

In the case of PS and WS methods, moderate rank reversals occur. These methods have almost identic characteristics. They infer partial preference strengths by determining to what extent a certain alternative outperformes the weakest one. Hence, they introduce structural dependency of fuzzy preference degrees. If a copy of a_i is made, its strength decreases since another alternative a_{m+1} appears that is preffered with regard to the same criteria. So, a_i and a_{m+1} conjointly weaken the discrimination effect and consequently give other alternatives the opportunity to improve. However, Figure 2 indicates that rank reversals are acceptable and considerably less frequent as by adding distinct new alternatives.

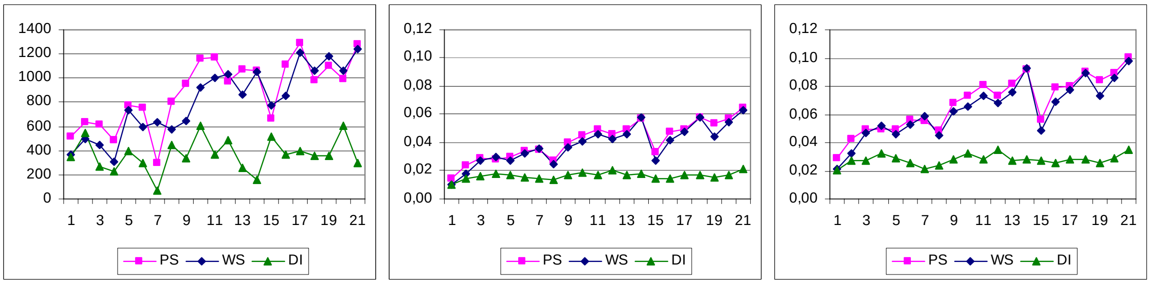


Figure 2: Best rank reversals, Kemeny-Snell distances and weighted distances for $l = 1$ copied alternative

Rank reversals become more frequent when l is increased. It is evident from Figure 3 that an upper limit is approached. Finding the convergence limit is a subject of further research.

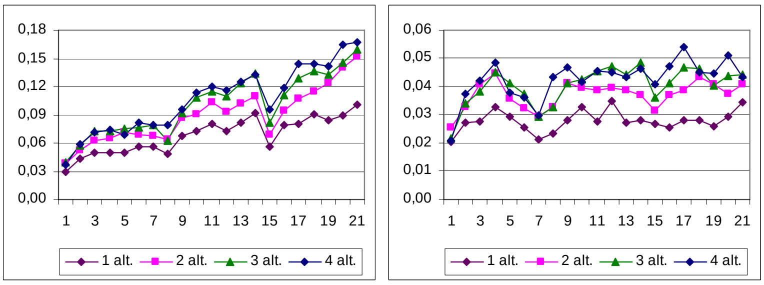


Figure 3: Increase of weighted distances for preference strengths (left) and dominance indices (right)

4.2 Unbiased distribution

On Figure 4, where the number of criteria and the probabilities of weak/strict preference increase from right to left, average unbiased preference strengths, fuzzy preference degrees and dominance indices are presented. PS and FD are sensitive to input parameters. This is a consequence of the unbiased uniform random distribution. If the number of criteria is small and the P_S respectively P_W probabilities are low, only a few alternatives are assigned high or at least nonzero $p_j(a_i)$ levels. The implication is a considerable discrimination of alternatives. However, such situations are very unlikely to occur in practice.

On the other hand, DI remain constant when independent variables change. The reason for their robustness is that partial rank orders are derived for various alpha-cuts of the binary matrix *B*. The original numerical information is thereby lost. Based on the distance metric, it is replaced with more robust, yet less rich, cardinal information in the aggregation phase.

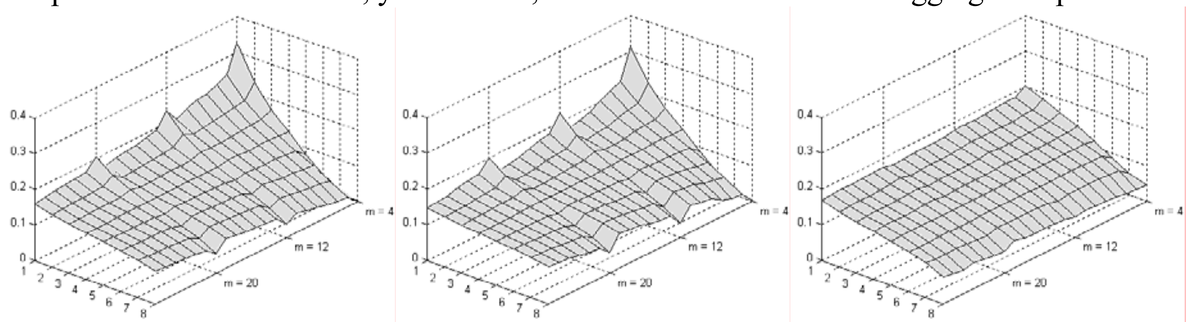


Figure 4: Average unbiased preference strengths, fuzzy preference degrees and dominance indices

It is crucial to notice that the richness of discriminating information does not depend on the level of relativeness inherent in the decision-making method. Totally absolute average fuzzy preference degrees provide reacher information than dominance indices which exhibit some relativeness, and at the same time poorer information than preference strengths which are based on pairwise comparisons. Figure 5 confirms this assumption. It shows that WS are the most discriminating approach. Further statistical experiments have revealed that they are overly extreme. The highest obtained ratio of the best and worst evaluations of alternatives is 63.94 for PS, 423.77 for WS, 32.94 for FD and 4.16 for DI. Ratios exceeding 75 have been proven to be unacceptable [10].

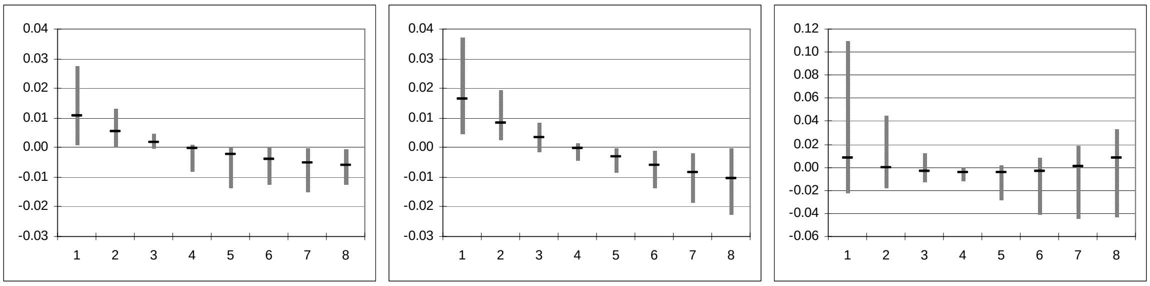


Figure 5: Distances between unbiased WS and PS (left), PS and FD (middle), FD and DI (right)

4.3 Biased distribution

Figure 6 shows that the two true relative methods – PS and WS – are the only ones which are able to cope with one conflicting alternative. They assign this alternative a priority that is considerably higher than priorities of other presumably indistinguishable alternatives. FD and DI which exhibit total absoluteness or a low level of relativeness, respectively, cannot identify conflictness. They perform the same as in the case of unbiased distribution.

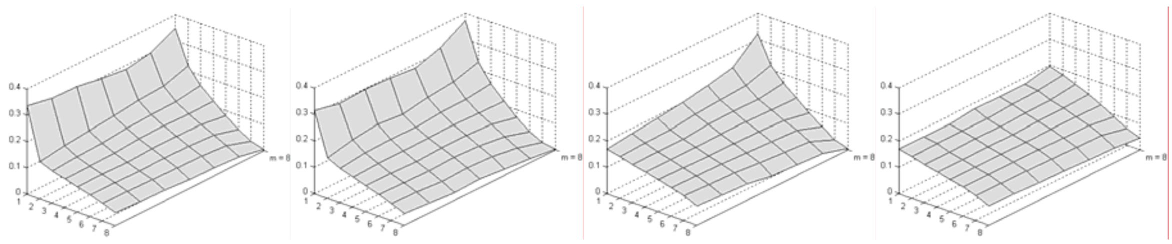


Figure 6: Average biased preference strengths, weighted sums, fuzzy degrees and dominance indices for $l = 1$

Superiority of relative measurements based PS is even more evident if $l = 2$ conflicting alternatives are introduced, as is depicted on Figure 7, which gives a comparison of results for both distribution types. PS improve the preferences of outstanding alternatives correctly, while WS produce only slight, yet positive, increases. Absolute FD perform exactly the same as in the case of unbiased distribution, which means that they are uncapable of dealing with various problem situations. For DI, an undesirable negative effect may be observed. The two alternatives which should improve actually deteriorate. Because priorities are normalized, they must decrease (increase) for the alternatives a_3 to a_8 if they increase (decrease) for the alternatives a_1 and a_2.

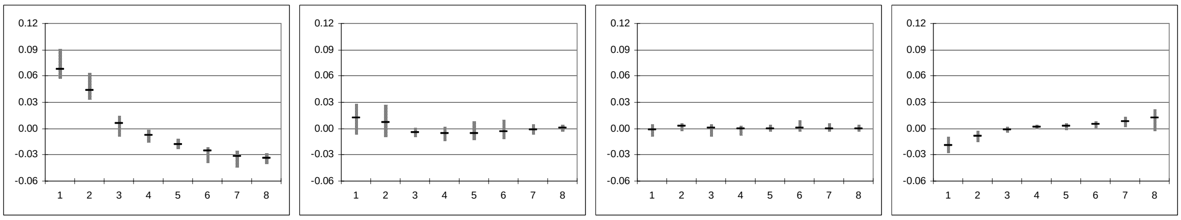


Figure 7: Distances between biased and unbiased PS, WS, FD and DI for $l = 2$

5. CONCLUSION

Many researchers believe that absolute preference aggregation methods or operators are more efficient than their relative counterparts since they preserve rank orders of alternatives. In this experimental study, it was proven that such prejudiced positions are untrue. Several methods/operators that exhibit different levels of absoluteness/relativeness were evaluated with a simulation model. It was shown that relative measurements based methods perform efficiently in the case of outstanding conflicting alternatives, while keeping rank reversals at an acceptable level. Thereby, Saaty's statements on structural dependency were empirically confirmed and generalized from AHP to other types of decision models.

6. REFERENCES

[1] Brans, J. P., Vincke, P. A preference ranking organisation method: The PROMETHEE method. Management Science, 31 (6), 647–656, 1985.

[2] Bregar, A., Györkös, J. Semiautomatic determination of criteria weights according to veto thresholds in the case of the localized alternative sorting analysis. Proceedings of the 7th International Symposium on Operational Research in Slovenia, 267–274, 2003.

[3] Emond, E. J., Mason, D. W. A new rank correlation coefficient with application to the consensus ranking problem. Journal of Multi-Criteria Decision Analysis, 11 (1), 17–28, 2002.

[4] Keeney, R. L., Raiffa, H. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. John Willey & Sons, New York, 1976.

[5] Ribeiro, R. A., Marques Pereira, R. A. Generalized mixture operators using weighting functions: A comparative study with WA and OWA. European Journal of Operational Research, 145 (2), 329–342, 2003.

[6] Roy, B. The outranking approach and the foundations of ELECTRE methods. Theory and Decision, 31 (1), 49–73, 1991.

[7] Saaty, T. L. The Analytic Hierarchy Process. McGraw-Hill, New York, 1980.

[8] Saaty, T. L. Rank from comparisons and from ratings in the analytic hierarchy/network process. European Journal of Operational Research, 168 (2), 557–570, 2006.

[9] Triantaphyllou, E., Baig, K. The impact of aggregating benefit and cost criteria in four MCDA methods. IEEE Trans. on Engineering Management, 52 (2), 213–226, 2005.

[10] Zanakis, S., Solomon, A., Wishart, N., Dublish, S. Multi-attribute decision making: A simulation comparison of select methods. European Journal of Operational Research, 107 (3), 507–529, 1998.

[11] Zavadskas, E. K., Zakarevicius, A., Antucheviciene, J. Evaluation of ranking accuracy in multi-criteria decisions. Informatica, 17 (4), 601–618, 2006.

[12] Zeleny, M. Multiple Criteria Decision Making. McGraw-Hill, New York, 1982.

OPTIMISATION AND MODELLING WITH SPREADSHEETS

Josef Jablonsky
Department of Econometrics, University of Economics
Praha, 130 67 Czech Republic
jablon@vse.cz, URL: http://nb.vse.cz/~jablon/

Abstract: The main aim of the paper is to discuss wide possibilities of spreadsheets in solving optimisation problems and mathematical modelling of economic processes in various fields of operations research. A special attention will be given to modelling languages as *LINGO* and others and their linking to spreadsheets in the process of building of end-user applications. Efficiency of using spreadsheets will be demonstrated on several applications – add-ins for solving multicriteria decision making problems (*Sanna*) and DEA models (*DEA Excel solver*) written in VBA.

Keywords: spreadsheets, optimisation, modelling languages, MS Excel, multiple criteria decision making, data envelopment analysis

1. Introduction

Spreadsheets belong among software products with very wide possible applications. Unfortunately their properties are used in everyday practice just for working with tables, simple recalculations by means of standard mathematical operators and functions, etc. Spreadsheets have much wider usage – they contain tools for financial decisions, statistical analyses, working with databases, graphical representation of data and last but not least for optimisation and mathematical modelling. In this paper we will discuss how it is possible to use spreadsheets for mathematical modelling and optimisation. In the next sections of the paper we will work with the typical spreadsheet product which is MS Excel.

Among the most often used operational research fields belong linear programming, project management, supply chain management, waiting lines analyses, simulation, etc. Each of the mentioned fields needs its own tools for solving various problems. It is not possible to discuss more than one or two fields within this brief paper. That is why we will show how it is possible solve data envelopment analysis (DEA) models and problems of multiple criteria decision making. DEA models are used as a tool for evaluation of efficiency, productivity and performance of decision making units. They are based on solving linear programming optimisation problems. Multiple criteria decision making (MCDM) problems (evaluation of alternatives) are very simple to understand even for non-expert persons in modelling and that is why they find very frequent usage. MS Excel can be used for solving analytical problems in several ways. We can see the following three ones:

1. Standard way characterised by using built-in tools in MS Excel (mathematical operators, functions, add-in applications coming with common MS Excel installation, etc). It is the easiest way that may suppose some advanced experience in using add-ins and other tools.
2. Linking spreadsheets to modelling languages as *LINGO*, *MPL for Windows*, *GAMS* and others. The advantage of this approach consists in possibility to use modelling and solving features of such products that are much more powerful than the same ones included directly in MS Excel or other spreadsheets.
3. Building end-user applications by means of VBA (Visual Basic for Applications) or by using other programming tools.

In the next three sections of the paper we will show the possibility of solving DEA and/or MCDM problems by using the three above presented different approaches. Before we start we will shortly formulate the standard DEA model and MCDM problem of evaluation of alternatives.

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 $\mathbf{X} = \{x_{ij}, i = 1, 2, \ldots, m, j = 1, 2, ..., n\}$ the matrix of inputs and $\mathbf{Y} = \{y_{ij}, i = 1, 2, \ldots, r, j = 1, 2, ..., n\}$ the matrix of outputs. For evaluation of efficiency of the unit U_q one of the most well known DEA models - CCR (Charnes, Cooper and Rhodes) model - can be used. Below is the input oriented formulation of this model:

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

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

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

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

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, $\lambda \geq 0$, is the vector of weights assigned to decision making 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 unit U_q is efficient if the following two conditions hold:

- the optimum value of the variable θ^* equals to 1,
- the optimum values of all slacks $\mathbf{s}^+$ and $\mathbf{s}^-$ equal to zero.

The problem (1) is standard LP problem with $(n+m+r+1)$ variables and $(m+r)$ constraints. For evaluation of efficiency of all units of the set it is necessary to solve the slightly modified problem (1) *n*-times. More about DEA models and their solving can be found in Cooper (2000).

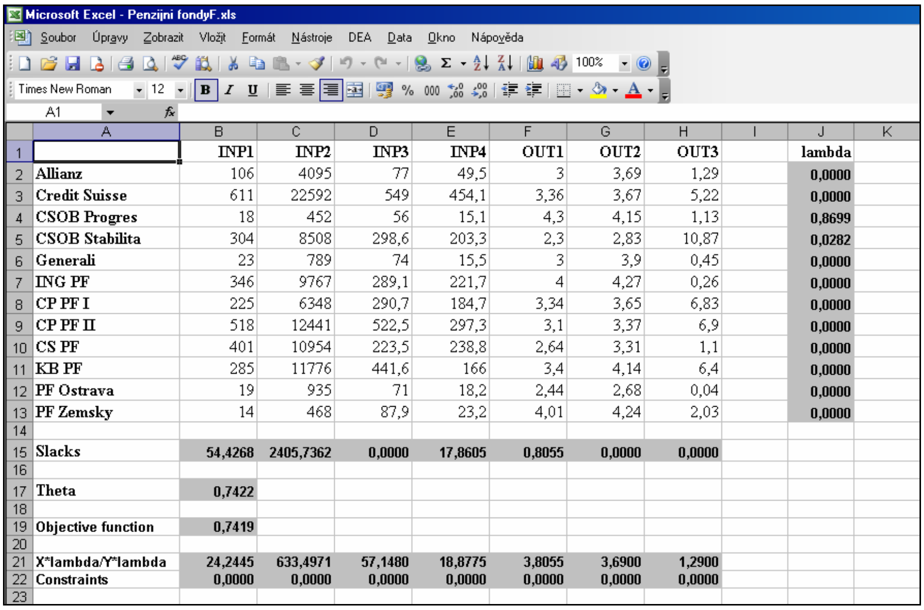
The MCDM problem of evaluation of alternatives is formulated very simply by the criterion matrix:

$$\begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array} \begin{array}{c} \begin{array}{cccc} Y_1 & Y_2 & \cdots & Y_k \end{array} \\ \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \end{array}$$

where $X_1, X_2, \ldots, X_n$ are alternatives, $Y_1, Y_2, \ldots, Y_k$ are criteria and y_{ij}, $i=1,2,\ldots n$, $j=1,2,\ldots k$, are criterion values. The aim of the analysis is to find the "best" (compromise) alternative or rank the alternatives.

2. Using standard MS Excel features

It is not difficult to solve the LP optimisation problem (1) using standard MS Excel features and MS Excel optimisation solver. Figure 1 shows how can be arranged data for evaluation of efficiency of 12 decision making units (pension funds in the Czech Republic) described by 4 inputs and 3 outputs by means of model (1). The shaded cells are variables or formulas necessary for expression of constraints. Evaluation of pension funds in the Czech Republic is discussed in detail in Jablonsky (2004) and (2007)



	A	B	C	D	E	F	G	H	I	J	K
1		INP1	INP2	INP3	INP4	OUT1	OUT2	OUT3		lambda	
2	Allianz	106	4095	77	49,5	3	3,69	1,29		0,0000	
3	Credit Suisse	611	22592	549	454,1	3,36	3,67	5,22		0,0000	
4	CSOB Progres	18	452	56	15,1	4,3	4,15	1,13		0,8699	
5	CSOB Stabilita	304	8508	298,6	203,3	2,3	2,83	10,87		0,0282	
6	Generali	23	789	74	15,5	3	3,9	0,45		0,0000	
7	ING PF	346	9767	289,1	221,7	4	4,27	0,26		0,0000	
8	CP PF I	225	6348	290,7	184,7	3,34	3,65	6,83		0,0000	
9	CP PF II	518	12441	522,5	297,3	3,1	3,37	6,9		0,0000	
10	CS PF	401	10954	223,5	238,8	2,64	3,31	1,1		0,0000	
11	KB PF	285	11776	441,6	166	3,4	4,14	6,4		0,0000	
12	PF Ostrava	19	935	71	18,2	2,44	2,68	0,04		0,0000	
13	PF Zemsky	14	468	87,9	23,2	4,01	4,24	2,03		0,0000	
14											
15	Slacks	54,4268	2405,7362	0,0000	17,8605	0,8055	0,0000	0,0000			
16											
17	Theta	0,7422									
18											
19	Objective function	0,7419									
20											
21	X*lambda/Y*lambda	24,2445	633,4971	57,1480	18,8775	3,8055	3,6900	1,2900			
22	Constraints	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000			
23											

Figure 1: DEA model using built-in Excel solver

The example in Figure 1 calculates the efficiency of the first decision making unit (Allianz). The variables are placed in ranges J2:J13 (λ_1, λ_2,…, λ_{12}), B15:E15 ($\mathbf{s}^+$), F15:H15 ($\mathbf{s}^-$) and B17 (θ). The formula for objective function is put into the cell B19, scalar products on the left side of the constraints of the model (1) are in cells B21:H21 and finally the left hand sides of the constraints are in row 22. The results of the optimisation are clearly given in Figure 1, i.e. the fund Allianz is not efficient and its efficiency score is 0,742. The problem is that the aim is to evaluate all the units and the formulas in row 22 are created for evaluation of the first unit only. In case we want to evaluate the remaining ones the formulas have to be modified and the optimisation run must be repeated. It is not convenient.

In case of MCDM problems is the situation even worse. There exist several methods for multicriteria evaluation of alternatives based on various principles. Only few of them can be simply realised by means of basic MS Excel features. One of them is WSA (weighted sum approach) method. Due to the comparability of the criterion values the following normalisation is applied:

$$y'_{\mathrm{ij}} = \frac{H_{\mathrm{j}} - y_{\mathrm{ij}}}{H_{\mathrm{j}} - D_{\mathrm{j}}}, \qquad i = 1,2,...,n,\ j = 1,2,...,k. \qquad (2)$$

The final utility of the alternative X_i is calculated as the weighted sum of normalised values:

$$u(\mathrm{X}_{\mathrm{i}}) = \sum_{\mathrm{j=1}}^{\mathrm{k}} v_{\mathrm{j}} y'_{\mathrm{ij}}, \qquad i = 1,2,\ldots n,\ j = 1,2,\ldots,k. \qquad (3)$$

where v_j, j=1,2,…n, are the weights of the criteria expressing their importance for the decision maker. It is clear that the WSA method can be very simply prepared in spreadsheet using simple formulas. Nevertheless the more sophisticated methods as AHP, PROMETHEE or ELECTRE class methods can be hardly realised in spreadsheets without using more advanced techniques.

3. Modelling systems and their linking to spreadsheets

Modelling systems were designed in order to improve the process of building mathematical models of (not only) optimisation problems, handling their data sets, their effective solving and presentation of results. Their general structure is given in Figure 2 – see Jablonsky (1998). Among the most often used languages belongs *LINGO*, *AMPL*, *MPL for Windows*, *GAMS*, *XPRESS-MP*, etc. All the modelling systems have many common features:

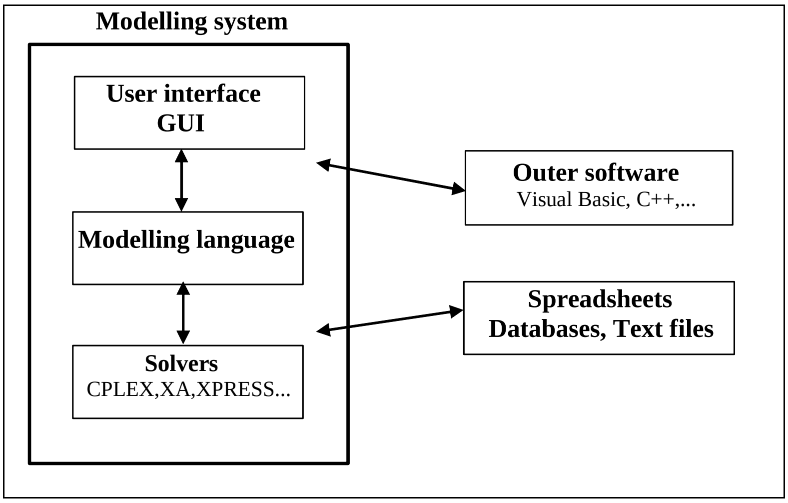


Figure 2: General structure of modelling systems

1. The modelling systems provide high-level languages for the compact representation of the model to users. These languages make it possible to represent the model in the general form similar to its standard mathematical formulation.
2. The general representation of the model by means of the modelling language makes it possible to separate the model and its input data set. The first step in process of solving of the problem consists in linking of these two basic parts. The separation of the model and the data set allows changing the size of the problem without any modifications in the model expression.
3. Any modifications of the model can be done very simply - e.g. adding a new constraint to the model is often possible without any changes to the representation of the data set.
4. The modelling systems support usually several optimisation solvers for different classes of optimisation problems (linear, nonlinear, integer programming, etc.) That is why optimisation problems can be solved by using several optimisation solvers available to users without any changes in the logic of the model or in the data set.
5. The expression of a model in modelling systems is concise and it is easy to understand it for readers. This expression can be used as a specific documentation of the model.
6. The modelling systems are usually available with their own library of sample models. The user can work with any model from the library and if necessary he can modify the model or can easily link the general model from the library with the data set and receive a solution.
7. All the systems have the features that enable linking to spreadsheets, databases, text files or other common software products.

Linking spreadsheets to modelling systems can be organised in several ways. From simple reading of data sets from spreadsheet files and/or returning the results of the model into the spreadsheet file onto building complex applications that uses modelling systems as modelling environment with all their advantages, solvers, etc.

The following example shows how it is possible to create the DEA model (1) within MS Excel environment. The *LINGO* model written directly in MS Excel sheet is presented in Figure 3. The SETS section of the model defines the variables and parameters of the model. They have the same or similar name as in the mathematical formulation (1). The rows 10 to 12 contain the objective function (*eff* is the given name of the function) and the constraints of the model (*inp* and *out* are their names). Variable INDEX corresponds to the index of the evaluated unit (index *q* in the mathematical formulation above). The DATA section contains links to named ranges in the spreadsheet file. E.g. the matrices of inputs/outputs X/Y and variable INDEX are read from the ranges X/Y and INDEX of the file FUNDS.XLS. The results of the optimisation are sent to the given ranges in the spreadsheet file. The model in Figure 3 is completed by *LINGO* commands that allow to start and quit the optimisation process. The optimisation can be launched directly from the Excel sheet by using a simple VBA launching procedure.

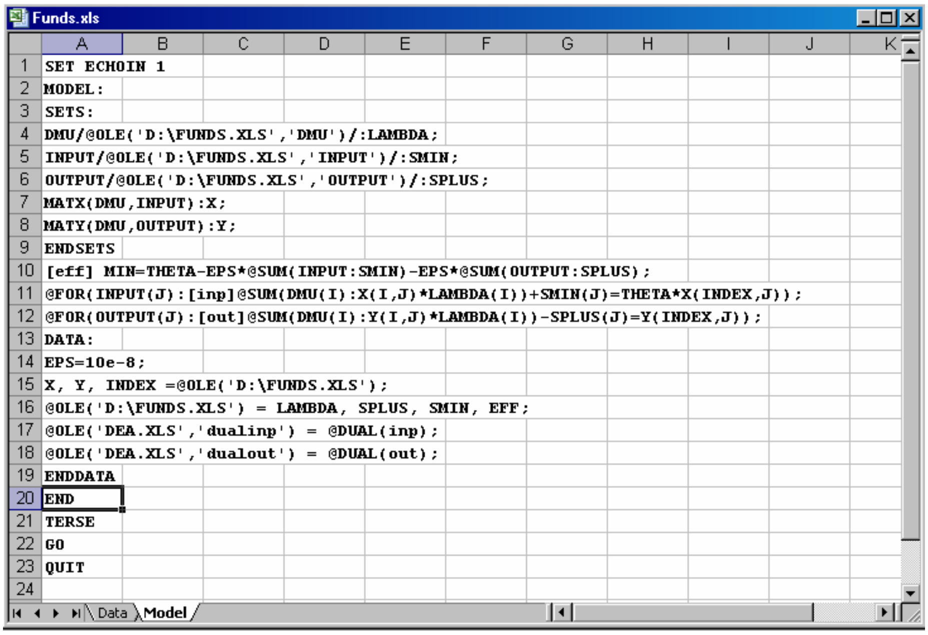


Figure 3: DEA model written in *LINGO*

The problems of multiple criteria evaluation of alternatives are not optimisation problems in the sense of mathematical programming. The problem is to rank the alternatives according to rules defined by different ranking methods. A formal model for WSA method – formulas (2) and (3) in the introductory part of this paper – can be written in *LINGO* language as follows:

```
MODEL:
! it is supposed that all the criteria are to be maximised;
SETS:
ALTER/@OLE('MCDM.XLS','ALTER')/:UTILITY;
CRIT/@OLE('MCDM.XLS','CRIT')/:W, UP, LO;
MATICE(ALTER,CRIT):X, XN;
ENDSETS
@FOR(CRIT(J): UP(J) = @MAX(ALTER(I):X(I,J)); LO(J) =
@MIN(ALTER(I):X(I,J)));
@FOR(ALTER(I): @FOR(CRIT(J): XN(I,J) = (X(I,J)-LO(J))/(UP(J)-LO(J))));
@FOR(ALTER(I): UTILITY(I) = @SUM(CRIT(J): XN(I,J)*W(J)));
```

```
DATA:
X, W = @OLE('d:\mcdm.xls');
@ole('d:\mcdm.xls') = UTILITY;
ENDDATA
END
```

We think that this formal notation need not be further explained. Please note only that this notation is fully general in all parameters, i.e. it need not be changed at all if the problem in MS Excel sheet is modified either in data or size.

4. Building end-user applications in spreadsheets

In this session of the paper we would like to inform very briefly about two applications for solving DEA models (*DEA Excel solver*) and for multiple criteria evaluation of alternatives (*Sanna*). Both the systems are created as add-in MS Excel applications. They are written in VBA and they need not any additional features for their full functionality.

The first version of the DEA Excel solver was presented in Jablonsky (2005). The current one contains several new eatures. The DEA Excel solver is an add-in application that covers several basic DEA models including super-efficiency models. The applications uses built-in MS Excel solver as the tool for solving LP problems. This solver is limited for problems with approx. 250 variables. This limit allows to solve DEA models (1) with n = 200 units and $m = r$ = 20 inputs/outputs. What is the problem is the necessity to repeat the optimisation run n-times in order to receive the appropriate results for all the units of the given set. That is why we decided to build an add-in application in MS Excel environment. In this way the system can be used on any computers with MS Excel spreadsheet, i.e. on almost all computers. The DEA Excel solver appears in the main Excel menu after its activation. As it is clear from Figure 4 the DEA Excel solver includes the following list of models:

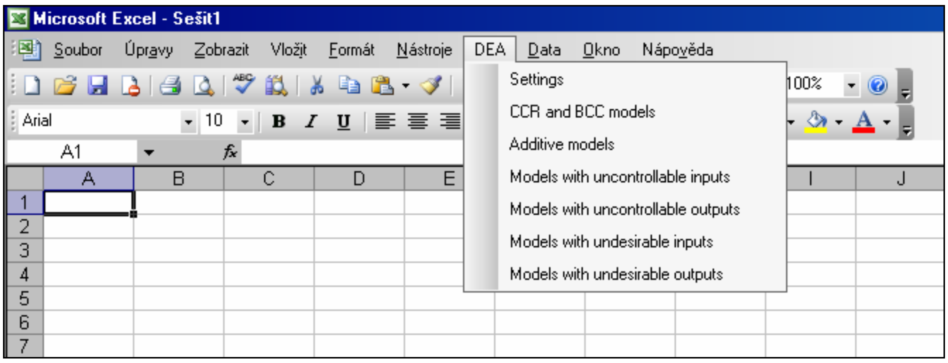


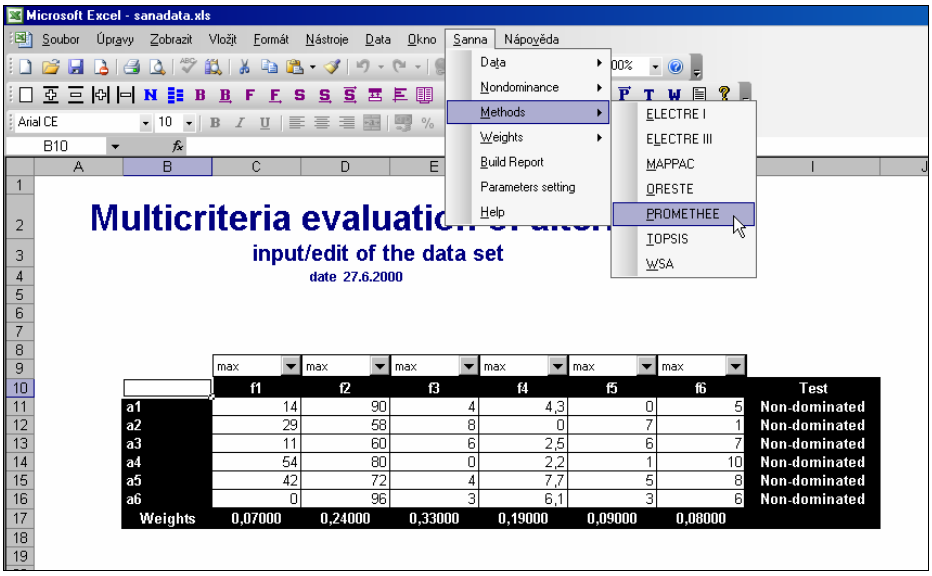
Figure 4: *DEA Excel solver* - available models.

- Standard radial models with constant, variable, non-decreasing or non-increasing returns to scale with input or output orientation.
- Additive models often denoted as SBM models. This group of models measures the efficiency by means of slack variables only.
- Models with uncontrollable inputs or outputs. In many applications some of the inputs or outputs cannot be directly controlled by the decision maker. In this case the uncontrollable characteristics have to be fixed.
- Models with undesirable inputs or outputs. In typical cases inputs are to be minimised and outputs are to be maximised in DEA models, i.e. the lower value of inputs and

higher value of outputs lead to higher efficiency score. It is not difficult to formulate problem where some of the inputs and outputs will be of reverse nature. Such characteristics are denoted as undesirable inputs or outputs. The models with undesirable characteristics are included in the DEA Excel Solver too.

Most of the mentioned models can be extended by super-efficiency option. After the selection of the appropriate models the decision maker specifies the necessary data in a dialog window that appears. The results in two possible forms are then displayed in separate MS Excel sheets. The presented DEA solver is not only attempt to solve DEA models in spreadsheets. Another DEA Excel solver is included e.g. in book Zhu (2003).

Real applications of mathematical models depend often on the availability of appropriate software tools. The same holds for the problems of multicriteria evaluation of alternatives. We have developed the *Sanna* system that covers several the most often used methods for multiple criteria evaluation of alternatives. The current version the system supports the following methods: WSA, ELECTRE I and III, PROMETHHEE, ORESTE, TOPSIS and MAPPAC. The typical form of the worksheet with working with *Sanna* is shown in Figure 5. Except the mentioned methods the system offers some other functions - support for estimating weights of criteria by means of pairwise comparison methods like AHP and Fuller's triangle, testing and filtering of nondominated alternatives, etc. The basic version of *Sanna* can solve multicriteria problems up to 100 alternatives and 30 criteria.



	f1	f2	f3	f4	f5	f6	Test
	max	max	max	max	max	max	
a1	14	90	4	4,3	0	5	Non-dominated
a2	29	58	8	0	7	1	Non-dominated
a3	11	60	6	2,5	6	7	Non-dominated
a4	54	80	0	2,2	1	10	Non-dominated
a5	42	72	4	7,7	5	8	Non-dominated
a6	0	96	3	6,1	3	6	Non-dominated
Weights	0,07000	0,24000	0,33000	0,19000	0,09000	0,08000	

Figure 5: *Sanna* worksheet.

5. Conclusions

Spreadsheets are powerful and popular software products that can be used for solving problems of mathematical modelling and optimisation. We have discussed several possible ways how to use spreadsheets and presented them on solving one of the DEA models and problem of multiple criteria evaluation of alternatives. In the last section of the paper the original MS Excel add-in applications were briefly introduced. It is *DEA Excel solver* that allows solving problems of evaluation of efficiency by means of standard DEA models of the size up to 200 decision making units and 20 inputs and outputs. The second application is *Sanna* that analyses problems of multiple criteria evaluation of alternatives (100 alternatives is maximum). Both the applications are written in VBA, are user-friendly controlled by pull down menus and dialog windows and do not suppose any other software tools installed (except MS Excel including MS Excel solver). They can be downloaded from the download section of the web page http://nb.vse.cz/~jablon/ and used by any interested professional.

Acknowledgements

The research is partially supported by the Grant Agency of the Czech Republic - project no. 402/06/0150.

References

[1] Cooper,W.W., Seiford,L.M. and Tone,K. (2000), Data Envelopment Analysis. *Kluwer Publ.*
[2] Jablonský,J.(1998), Mathematical programming modelling and optimisation systems. *CEJORE* 3-4, pp.279-288.
[3] Jablonský,J. (2004), Models for efficiency evaluation of production units, *Politická ekonomie*, 52, pp. 206-220.
[4] Jablonský,J. (2005), A MS Excel based support system for data envelopment analysis models, In. Skalská,H. (ed.): *Proceedings of the 23rd Conference Mathematical Methods in Economics,* Hradec Králové, pp. 175-181.
[5] Jablonský,J. (2007), Measuring efficiency of production units by AHP models, *Mathematical and Computer Modeling*, 46, (in print).
[6] Zhu,J. (2003), Quantitative Models for Performance Evaluation and Benchmarking. *Kluwer Publ.*

UNDERBAD AND OVERGOOD ALTERNATIVES IN BIPOLAR METHOD

Tadeusz Trzaskalik, Department of Operations Research, The Karol Adamiecki University of Economics in Katowice, ul. Bogicicka 14, 40-226 Katowice, Poland, e-mail: ttrzaska@ae.katowice.pl
Sebastian Sitarz, Institute of Mathematics, University of Silesia in Katowice, ul. Bankowa 14, 40-007 Katowice, Poland, e-mail: ssitarz@ux2.math.us.edu.pl

Abstract: Bipolar is one of the Multiple Criteria Decision Analysis (MCDA) methods, proposed by Ewa Konarzewska-Gubała. The essence of the analysis in the Bipolar method consists in a fact that alternatives are not compared directly to each other, but they are confronted to the two sets of reference objects: desirable and non-acceptable. Some alternatives can be evaluated as overgood, i.e. better than at least one of desirable reference object or underbad, i.e. worse than at least of one non-acceptable object. The aim of the paper is to describe relations between these alternatives.

Keywords: Multiple Criteria Decision Analysis (MCDA), Bipolar method, underbad alternatives, overgood alternatives.

1. Introduction

Bipolar is one of MCDA methods, developed by Ewa Konarzewska-Gubała (1987, 1989). Finite number of decision alternatives are confronted to the two set of really existing or imaged reference objects, divided on desirable and non-acceptable. Final evaluation of alternatives is based on its independent position with regard to both segments of the reference system. The decision maker wishes to select the best alternative or to select a set of satisfying alternatives for further study or to rank all the alternatives from the best to the worst. In the Bipolar method elements of Electre methodology (Roy (1985)) and ideas of Merighi (1980) algorithms of confrontation can be found. The Bipolar method has already been used in applications (for instance Jakubowicz (1987), Dominiak (1997, 2006), Konarzewska-Gubała (2002)). The method has also been applied to model multi-stage multi-criteria decision processes (Trzaskalik (1987)).

Applying Bipolar method some alternatives can be evaluated as overgood, i.e. better than at least one of desirable reference object or underbad, i.e. worse than at least of one non-acceptable object. The question arises is it possible to evaluate an alternative as overgood and underbad simultaneously? Konarzewska-Gubała claims that if none non-acceptable object dominates any desirable object, such situation cannot occur. Practical applications of Bipolar method showed that the condition described above is not sufficient for eliminating such a possibility. It causes some theoretical problems. In the present paper we are looking for a precise mathematical condition to eliminate such a situation.

The paper consists of six parts. In chapters 1-4 new, mathematically-oriented description of the Bipolar method is presented. Example, given in Chapter 5 illustrates the situation, where an alternative is simultaneously underbad and overgood. In Chapter 6 main theorem is proved. The concluding remarks end the paper.

2. Assumptions of Bipolar method

It is assumed, that there are given: the set of decision alternatives $\mathcal{A} = \{\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^m\}$ and the set of criteria functions $\mathcal{F} = \{f_1, \ldots, f_n\}$, where $f_k\colon \mathcal{A} \to \mathcal{K}_k$ for $k=1, \ldots, n$, and $\mathcal{K}_k$ is a cardinal, ordinal or binary scale. The criteria evaluations are to be maximized or minimized or close as much as possible to some desirable values. In the paper we assume that criteria are defined in such a way that higher values are preferred to lower values. It is possible to transform them to the form, considered in the present work. For each criterion the decision maker establishes weight w_k of relative importance (it is assumed, that $\sum_{k=1}^{n} w_k = 1$ and $w_k \geq 0$ for each $k=1, \ldots, n$), equivalence threshold q_k and veto threshold v_k. The decision maker also establishes minimal criteria values concordance level s as the outranking threshold. It is assumed, that condition $0.5 \leq s \leq 1$ holds.

The decision maker establishes a bipolar reference system $\mathcal{R} = \mathcal{D} \cup \mathcal{Z}$, which consists of the set of desirable objects $\mathcal{D} = \{\mathbf{d}^1, \ldots, \mathbf{d}^d\}$ and the set of non-acceptable objects $\mathcal{Z} = \{\mathbf{z}^1, \ldots, \mathbf{z}^z\}$, where d and z denote the number of desirable and non-acceptable objects, respectively. It is assumed, that $\mathcal{D} \cap \mathcal{Z} = \varnothing$. The number of elements of the set $\mathcal{R}$ is equal to d+z. Elements of the set $\mathcal{R}$ are denoted as $\mathbf{r}^h$, $h=1,\ldots,d+z$. Values $f_k(\mathbf{r}^h)$ for k=1,...,n and h=1,...,r are known. Let D be classical domination relation:

$$\boldsymbol{f}(\mathbf{z}) \text{ D } \boldsymbol{f}(\mathbf{d}) \Leftrightarrow \forall_{k=1,..n} \ f_k(\mathbf{z}) \leq f_k(\mathbf{d}) \wedge \exists_{l=1,..n} \ f_l(\mathbf{z}) < f_l(\mathbf{d}).$$

Following Konarzewska-Gubała (1989) we assume, that condition

$$\sim \exists_{\mathbf{d} \in \mathcal{D}} \sim \exists_{\mathbf{z} \in \mathcal{Z}} \quad \boldsymbol{f}(\mathbf{z}) \text{ D } \boldsymbol{f}(\mathbf{d}) \tag{1}$$

is fulfilled.

3. Phase 1: Comparison of alternatives to reference objects

3.1. Outranking indicators

For the pair $(\mathbf{a}^i, \mathbf{r}^j)$, where $\mathbf{a}^i \in \mathcal{A}$, $\mathbf{r}^j \in \mathcal{R}$, the following values:

$$c^+\left(\mathbf{a}^i, \mathbf{r}^j\right) = \sum_{k=1}^{n} w_k \varphi_k^+\left(\mathbf{a}^i, \mathbf{r}^j\right) \text{ where } \varphi_k^+\left(\mathbf{a}^i, \mathbf{r}^j\right) = \begin{cases} 1, & \text{if } f_k\left(\mathbf{a}^i\right) - f_k\left(\mathbf{r}^j\right) > q_k \\ 0, & \text{otherwise} \end{cases}$$

$$c^-\left(\mathbf{a}^i, \mathbf{r}^j\right) = \sum_{k=1}^{n} w_k \varphi_k^-\left(\mathbf{a}^i, \mathbf{r}^j\right) \text{ where } \varphi_k^-\left(\mathbf{a}^i, \mathbf{r}^j\right) = \begin{cases} 1, & \text{if } f_k\left(\mathbf{r}^j\right) - f_k\left(\mathbf{a}^i\right) > q_k \\ 0, & \text{otherwise} \end{cases}$$

$$c^=\left(\mathbf{a}^i, \mathbf{r}^j\right) = \sum_{k=1}^{n} w_k \varphi_k^=\left(\mathbf{a}^i, \mathbf{r}^j\right) \text{ where } \varphi_k^=\left(\mathbf{a}^i, \mathbf{r}^j\right) = \begin{cases} 1, & \text{if } \left| f_k\left(\mathbf{r}^j\right) - f_k\left(\mathbf{a}^i\right) \right| \leq q_k \\ 0, & \text{otherwise} \end{cases}$$

are calculated. Sets of indices:

$$I^+\left(\mathbf{a}^i, \mathbf{r}^j\right) = \left\{k : \varphi_k^+\left(\mathbf{a}^i, \mathbf{r}^j\right) = 1\right\} \qquad I^-\left(\mathbf{a}^i, \mathbf{r}^j\right) = \left\{k : \varphi_k^-\left(\mathbf{a}^i, \mathbf{r}^j\right) = 1\right\}$$

are determined. Let v_k be threshold values given for k=1,...,n by the decision maker.

Condition

$$\underset{k \in I^-}{\forall} f_k\left(\mathbf{a}^i\right) > v_k$$

is called veto test. Conditions

$$\underset{k \in I^-}{\forall} \quad f_k\left(\mathbf{a}^i\right) > v_k \qquad \underset{k \in I^+}{\forall} \quad f_k\left(\mathbf{a}^i\right) > v_k$$

are called non-discordance tests.

Case 1: $c^+(\mathbf{a}^i, \mathbf{r}^j) > c^-(\mathbf{a}^i, \mathbf{r}^j)$

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is positively verified, then outranking indicators are defined as follows:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = c^+(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j), \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is not positively verified, then:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

Case 2: $c^+(\mathbf{a}^i, \mathbf{r}^j) < c^-(\mathbf{a}^i, \mathbf{r}^j)$

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ veto test is positively verified, then:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j) + c^=(\mathbf{a}^i, \mathbf{r}^j)$$

- If for the pair $(\mathbf{a}^i, \mathrm{r}^j)$ veto test is not positively verified, then:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0 \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0$$

<u>Case 3</u>: $c^+(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j)$.

- If for the pair $(\mathbf{a}^i, \mathbf{r}^j)$ two non-discordance tests are positively verified, then:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = c^+(\mathbf{a}^i, \mathbf{r}^j) + c^-(\mathbf{a}^i, \mathbf{r}^j) \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = c^-(\mathbf{a}^i, \mathbf{r}^j) + c^+(\mathbf{a}^i, \mathbf{r}^j)$$

- For the pair $(\mathbf{a}^i, \mathbf{r}^j)$ at least one of non-discordance tests is not positively verified, then:

$$d^+(\mathbf{a}^i, \mathbf{r}^j) = 0, \qquad d^-(\mathbf{a}^i, \mathbf{r}^j) = 0.$$

3.2. Preference structure

By means of outranking indicators three relationships: large preference $\mathrm{L_s}$, indifference $\mathrm{I_s}$ and incomparability $\mathrm{R_s}$ are defined as follows:

$$\mathbf{a}^i \ \mathrm{L_s}\ \mathbf{r}^h \ \text{ iff } \ d^+(\mathbf{a}^i, \mathbf{r}^h) > s \ \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) = 0$$

$$\mathbf{r}^j \ \mathrm{L_s}\ \mathbf{a}^i \ \text{ iff } \ d^+(\mathbf{a}^i, \mathbf{r}^h) = 0 \ \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) > s$$

$$\mathbf{a}^i \ \mathrm{I_s}\ \mathbf{r}^h \text{ iff } d^+(\mathbf{a}^i, \mathbf{r}^h) > s \ \wedge d^-(\mathbf{a}^i, \mathbf{r}^h) > s$$

$$\mathbf{a}^i \ \mathrm{R_s}\ \mathbf{r}^j \text{ otherwise}$$

4. Phase 2: Position of an alternative in relation to the bipolar reference system

4.1. Success achievement degree

For a given $\mathbf{a}^i \in \mathcal{A}$ auxiliary sets of indices are defined as follows:

$$\mathcal{L}_s(\mathbf{a}^i, \mathcal{D}) = \{h\text{: } \mathbf{a}^i \ \mathrm{L_s}\ \mathbf{d}^h, \ \ \mathbf{d}^h \in \mathcal{D}\}$$

$$\mathcal{I}_s(\mathbf{a}^i, \mathcal{D}) = \{h\text{: } \mathbf{a}^i \ \mathrm{I_s}\ \mathbf{d}^h, \ \ \mathbf{d}^h \in \mathcal{D}\}$$

$$\mathcal{L}_s(\mathcal{D}, \mathbf{a}^i) = \{h\text{: } \mathbf{d}^h \ \mathrm{L_s}\ \mathbf{a}^i, \ \ \mathbf{d}^h \in \mathcal{D}\}$$

In the set $\mathcal{L}_s(\mathbf{a}^i, \mathcal{D})$ there are included these indices of desirable objects, for whom the statement $\mathbf{a}^i \mathrm{L_s} \mathbf{d}^h$ is true. The two remaining sets are defined similarly.

Defining the position of an alternative $\mathbf{a}^i$ in relation to the set $\mathcal{D}$ we consider three possibilities:

<u>Case S1.</u> $\quad \mathcal{L}_s(\mathbf{a}^i, \mathcal{D}) \cup \mathcal{I}_s(\mathbf{a}^i, \mathcal{D}) \neq \varnothing$.

The value

$$d_{\mathcal{D}}^+(\mathbf{a}^i) = \max \{d^+(\mathbf{a}^i, \mathbf{d}^h)\text{: } h \in \mathcal{L}_s(\mathbf{a}^i, \mathcal{D}) \cup \mathcal{I}_s(\mathbf{a}^i, \mathcal{D})\}$$

is calculated. The success achievement degree $d_{\mathrm{S}}(\mathbf{a}^i)$ is defined to be equal to $d_{\mathcal{D}}^+(\mathbf{a}^i)$.

<u>Case S2.</u> $\quad \mathcal{L}_s(\mathbf{a}^i, \mathcal{D}) \cup \mathcal{I}_{\mathrm{S}}(\mathbf{a}^i, \mathcal{D}) = \varnothing \wedge \mathcal{L}_s(\mathcal{D}, \mathbf{a}^i) \neq \varnothing$.

The value

$$d_{\mathcal{D}}^-(\mathbf{a}^i) = \min \{d^-(\mathbf{a}^i, \mathbf{d}^h)\text{: } h \in \mathcal{L}_s(\mathcal{D}, \mathbf{a}^i)\}$$

is calculated. The success achievement degree $d_{\mathrm{S}}(\mathbf{a}^i)$ is defined to be equal to $d_{\mathcal{D}}^-(\mathbf{a}^i)$.

<u>Case S3</u>. If conditions described in Cases S1 and S2 are not fulfilled, then the success achievement degree $d_{\mathrm{S}}(\mathbf{a}^i)$ is defined to be equal to 0.

4.2. Failure avoidance degree

For a given $\mathbf{a}^i \in \mathcal{A}$ auxiliary sets of indices are defined as follows:

$$\mathcal{L}_s(\mathcal{Z}, \mathbf{a}^i) = \{h\text{: } \mathbf{z}^h \ \mathrm{L_s}\ \mathbf{a}^i, \mathbf{z}^h \in \mathcal{Z}\}$$

$$\mathcal{I}_s(\mathcal{Z}, \mathbf{a}^i) = \{h\text{: } \mathbf{z}^i \ \mathrm{I_s}\ \mathbf{a}^h, \mathbf{z}^h \in \mathcal{Z}\}$$

$$\mathcal{L}_s(\mathbf{a}^i , \mathcal{Z}) = \{h: \mathbf{a}^i \ \mathrm{L_S}\ \mathbf{z}^h, \mathbf{z}^h \in \mathcal{Z}\}$$

In the set $\mathcal{L}_s(\mathcal{Z},\mathbf{a}^i)$ there are included these numbers of „bad" objects, for whom the statement $\mathbf{z}^h \mathrm{L_S} \mathbf{a}^i$ is true. The two remaining sets are interpreted similarly.

Defining the position of an alternative $\mathbf{a}^i$ in relation to the set $\mathcal{Z}$ we consider three possibilities:

Case N1. $\mathcal{L}_s(\mathcal{Z} , \mathbf{a}^i) \cup \mathcal{I}_s(\mathcal{Z} , \mathbf{a}^i) = \varnothing \wedge \mathcal{L}_s(\mathbf{a}^i , \mathcal{Z}) \neq \varnothing$.

The value

$$d_{\mathcal{Z}}^+(\mathbf{a}^i) = \min \{d^+(\mathbf{a}^i, \mathbf{z}^h): h\in \mathcal{L}_s(\mathbf{a}^i ,\mathcal{Z})\}$$

is calculated. The failure avoidance degree $d_{\mathrm{N}}(\mathbf{a}^i)$ is defined to be equal to $d_{\mathcal{Z}}^+(\mathbf{a}^i)$.

Case N2. $\mathcal{L}_s(\mathcal{Z} , \mathbf{a}^i) \cup \mathcal{I}_s(\mathcal{Z} , \mathbf{a}^i) \neq \varnothing$.

The value

$$d_{\mathcal{Z}}^-(\mathbf{a}^i) = \max \{d^-(\mathbf{a}^i, \mathbf{z}^h): h\in \mathcal{L}_s(\mathcal{Z} , \mathbf{a}^i) \cup \mathcal{I}_s (\mathbf{a}^i, \mathcal{D}) \}$$

is calculated. The failure avoidance degree $\mathrm{d_N}(\mathbf{a}^i)$ is defined to be equal to $\mathrm{d_Z}^-(\mathbf{a}^i)$.

Case N3. If conditions described in Cases S1 and S2 are not fulfilled, then the failure avoidance degree $d_{\mathrm{N}}(\mathbf{a}^i)$ is defined to be equal to 0.

5. Relationships in the set of alternatives

5.1. Mono-sortings

According to the success achievement degree the alternatives from the set $\mathcal{A}$ are sorted to the three subsets:

S_1 consists of the „overgood" alternatives (condition, formulated in Case S1 is fulfilled).

S_2 consists of the alternatives, for which condition, formulated in Case S2 is fulfilled.

S_3 consists of the alternatives, for which condition, formulated in Case S3 is fulfilled (decision variants non-comparable with $\mathcal{D}$).

A way of building above categories implies that each alternative from S_1 should be preferred to any alternative from S_2.

According to the failure avoidance degree the alternatives from the set $\mathcal{A}$ are sorted to the three subsets:

N_1 consists of the alternatives, for which condition, formulated in Case N1 is fulfilled.

N_2 consists of „underbad" alternatives (condition, formulated in Case N2 is fulfilled).

N_3 consists of the alternatives, for which condition, formulated in Case N3 is fulfilled (alternatives non-comparable with $\mathcal{Z}$).

A way of building above subsets implies that each alternative from N_1 should be preferred to any alternative from N_2.

5.2. Bipolar-sorting and Bipolar-ranking

Considering jointly evaluation of success achievement degree and failure avoidance degree, three subsets of alternatives are defined:

B_1 consists of such alternatives $\mathbf{a}^i$, that $d_{\mathcal{D}}^+(\mathbf{a}^i) >0 \wedge d_{\mathcal{Z}}^+(\mathbf{a}^i) >0$

B_2 consists of such alternatives $\mathbf{a}^i$, that $d_{\mathcal{D}}^-(\mathbf{a}^i) >0 \wedge d_{\mathcal{Z}}^+(\mathbf{a}^i) >0$

B_3 consists of such alternatives $\mathbf{a}^i$, that $d_{\mathcal{D}}^-(\mathbf{a}^i) >0 \wedge d_{\mathcal{Z}}^-(\mathbf{a}^i) >0$

Assuming, that each alternative from B_1 is preferred to any alternative from B_2 and each alternative from B_2 is preferred to any alternative from B_3, linear order is given:

for $\mathbf{a}^i, \mathbf{a}^j \in \mathrm{B}_1$ — $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $d_{\mathrm{S}}(\mathbf{a}^i)+ d_{\mathrm{N}}(\mathbf{a}^i)> d_{\mathrm{S}}(\mathbf{a}^j)+d_{\mathrm{N}}(\mathbf{a}^j)$

$\mathbf{a}^i$ equivalent to $\mathbf{a}^j$, iff $d_{\mathrm{S}}(\mathbf{a}^i)+d_{\mathrm{N}}(\mathbf{a}^i) = d_{\mathrm{S}}(\mathbf{a}^j)+d_{\mathrm{N}}(\mathbf{a}^j)$

for $\mathbf{a}^i, \mathbf{a}^j \in \mathrm{B}_2$ — $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $1-d_{\mathrm{S}}(\mathbf{a}^i)+d_{\mathrm{N}}(\mathbf{a}^i) > 1-d_{\mathrm{S}}(\mathbf{a}^j)+d_{\mathrm{N}}(\mathbf{a}^j)$

$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff $1-d_S(\mathbf{a}^i)+d_N(\mathbf{a}^i) = 1-d_S(\mathbf{a}^j)+d_N(\mathbf{a}^j)$

for $\mathbf{a}^i, \mathbf{a}^j \in B_3$ — $\mathbf{a}^i$ is preferred to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i) + d_N(\mathbf{a}^i) < d_S(\mathbf{a}^j) + d_N(\mathbf{a}^j)$

$\mathbf{a}^i$ is equivalent to $\mathbf{a}^j$, iff $d_S(\mathbf{a}^i) + d_N(\mathbf{a}^i) = d_S(\mathbf{a}^j) + d_N(\mathbf{a}^j)$.

6. Example

We consider the set of alternatives $\mathcal{A} = \{\mathbf{a}^1, \mathbf{a}^2, \mathbf{a}^3, \mathbf{a}^4\}$ and the reference system, which consists of the sets: $\mathcal{D} = \{\mathbf{d}^1, \mathbf{d}^2, \mathbf{d}^3\}$ and $\mathcal{Z} = \{\mathbf{z}^1, \mathbf{z}^2\}$. Values of two criteria for alternatives and reference objects are shown in Fig. 1.

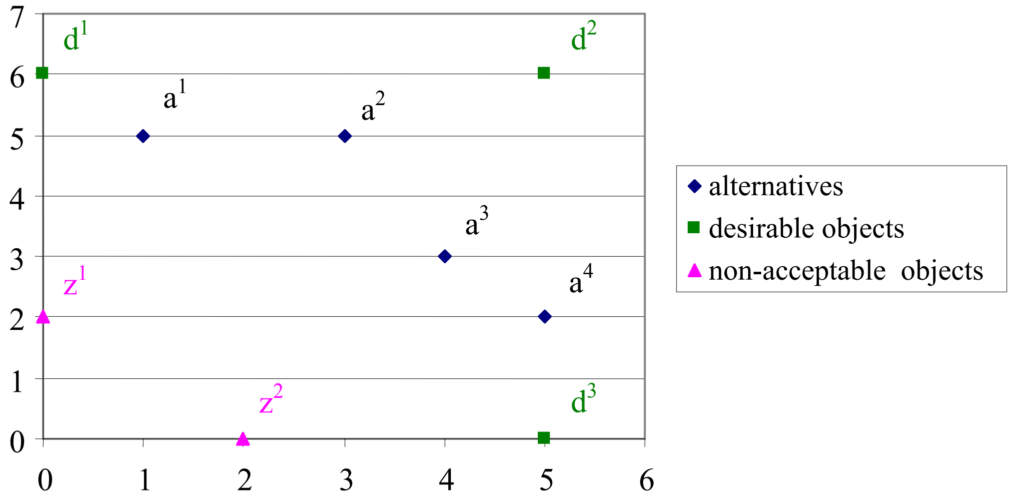


Fig. 1. Values of criteria for alternatives and reference objects

Assuming, that veto thresholds are equal to $v_1 = 0$, $v_2 = 0$, weights are equal to $w_1 = 0{,}6$, $w_2 = 0{,}4$, and the concordance threshold and the equivalence threshold are equal to $s=0{,}5$ and $q_1=0{,}25$, $q_2=0{,}25$, we apply classical Bipolar procedure, described in Chapters 2-4 and obtain the following bipolar ranking:

1: $\mathbf{a}^4$.

2: $\mathbf{a}^2$, $\mathbf{a}^3$ (equivalent decision variants).

We have $d^+_{\mathcal{D}}\left(\mathbf{a}^1\right) > 0$ and $d^-_{\mathcal{Z}}\left(\mathbf{a}^1\right) > 0$. It means that the alternative $\mathbf{a}^1$ is both "overgood" and "underbad" and cannot be classified to any previously defined bipolar categories.

7. Main Theorem

If the condition

$$\forall_{k=1,\ldots,n} \forall_{\mathbf{d}\in\mathcal{D}} \forall_{\mathbf{z}\in\mathcal{Z}} \; f_k(\mathbf{d}) \geq f_k(\mathbf{z}) \tag{2}$$

holds and $s \geq 0{,}5$ then

$$\neg \underset{\mathbf{a}\in\mathcal{A}}{\exists} \; \mathbf{a} \in N_2 \cap S_1.$$

Proof

1.Let us notice that (2) implies

$$\{k : f_k(\mathbf{a}) - f_k(\mathbf{d}) > q_k\} \subset \{k : f_k(\mathbf{a}) - f_k(\mathbf{z}) > q_k\}$$

It means that

$$\underset{\mathbf{a}\in\mathcal{A}}{\forall} \; \underset{\mathbf{d}\in\mathcal{D}}{\forall} \; \underset{\mathbf{z}\in\mathcal{Z}}{\forall} \; c^+(\mathbf{a}, \mathbf{d}) \leq c^+(\mathbf{a}, \mathbf{z}) \tag{3}$$

2. Analogically, from (2) we have

$$\{k : f_k(\mathbf{z}) - f_k(\mathbf{a}) > q_k\} \subset \{k : f_k(\mathbf{d}) - f_k(\mathbf{a}) > q_k\}$$

It means that

$$\underset{\mathbf{a}\in\mathcal{A}}{\forall}\ \underset{\mathbf{d}\in\mathcal{D}}{\forall}\ \underset{\mathbf{z}\in\mathcal{Z}}{\forall}\ c^{-}(\mathbf{a},\mathbf{z}) \le c^{-}(\mathbf{a},\mathbf{d}) \tag{4}$$

3. Suppose that $\underset{\mathbf{a}^*\in\mathcal{A}}{\exists}\ \mathbf{a}^* \in N_2 \cap S_1$.

4. If $\mathbf{a}^* \in N2$ then

$$\mathcal{L}_s(\mathcal{Z}, \mathbf{a}^*) \cap \mathcal{I}_s(\mathcal{Z}, \mathbf{a}^*) \ne \varnothing$$

Thus

$$\underset{\mathbf{z}^*\in\mathcal{Z}}{\exists}\ [\, d^{-}(\mathbf{a}^*, \mathbf{z}^*) > s \wedge d^{+}(\mathbf{a}^*, \mathbf{z}^*) = 0 \,] \vee [\, d^{-}(\mathbf{a}^*, \mathbf{z}^*) > s \wedge d^{+}(\mathbf{a}^*, \mathbf{z}^*) > s \,]$$

From assumption $s \ge 0{,}5$ follows that it is impossible that

$$d^{-}(\mathbf{a}^*, \mathbf{z}^*) > s \wedge d^{+}(\mathbf{a}^*, \mathbf{z}^*) > s$$

Hence the following condition holds

$$d^{-}(\mathbf{a}^*, \mathbf{z}^*) > s \wedge d^{+}(\mathbf{a}^*, \mathbf{z}^*) = 0$$

It means that

$$c^{+}(\mathbf{a}^*, \mathbf{z}^*) < c^{-}(\mathbf{a}^*, \mathbf{z}^*) \tag{5}$$

5. Analogically as in point 4. we obtain that if $\mathbf{a}^* \in S_1$ then

$$\underset{\mathbf{d}^*\in\mathcal{D}}{\exists}\ c^{-}(\mathbf{a}^*, \mathbf{d}^*) < c^{+}(\mathbf{a}^*, \mathbf{d}^*) \tag{6}$$

6. From (4), (5) and (6) we obtain

$$c^{+}(\mathbf{a}^*, \mathbf{z}^*) < c^{-}(\mathbf{a}^*, \mathbf{z}^*) \le c^{-}(\mathbf{a}^*, \mathbf{d}^*) < c^{+}(\mathbf{a}^*, \mathbf{d}^*)$$

thus

$$c^{+}(\mathbf{a}^*, \mathbf{z}^*) < c^{+}(\mathbf{a}^*, \mathbf{d}^*) \tag{7}$$

7. The condition (7) is contradictory to the condition (3). It means that the hypothesis in point 3. is false and the theorem is true.

8. Conclusions

Konarzewska-Gubała (1989) claims, that if reference sets are separate and condition (1), formulated as follows: *there does not exist desirable reference object and non-acceptable reference object such that desirable reference object is dominated by the non-acceptable reference object* holds, there does not exist an alternative which is simultaneously overgood and underbad. Example, described in section 5 shows that condition (1) is not sufficient to eliminate such a situation. We proved in section 6, that the necessary condition (2) should be formulated as follows: *each desirable reference object dominates each non-acceptable reference object.* Anyway, assumption (2) is over-restrictive and in many cases it makes impossible for decision makers to apply the approach in real decision problems. Thus new ideas should be included to the method. In the further research, assuming condition (1), we will try to modify reference system. The second possibility is to extend the number of subsets in Bipolar classification.

9. References

Dominiak C. (2006): „Application of modified Bipolar method". In: T.Trzaskalik (ed.) *Multicriteria Methods on Polish Financial Market*, p.105-113, PWE (in Polish).

Dominiak C. (1997): "Portfolio Selection Using the Idea of Reference Solution". In: G.Fandel,T.Gal (eds.) *Multiple Criteria Decision Making*. Springer-Verlag, p.593-602.

Jakubowicz S.(1987): "Work Characteristics of a „Good" Physics Teacher on the Basis of His Lessons".RPBP.III.30.VI.4.6. The University of Wrocław (in Polish).

Konarzewska-Gubała E. (2002): "Multiple Criteria Company Benchmarking Using the BIPOLAR Method". In T.Trzaskalik, J.Michnik (eds.) *Multiple Objective and Goal Programming. Recent Developments.* Physica-Verlag. Springer-Verlag, p.338-350.

Konarzewska-Gubała E. (1989) *BIPOLAR: Multiple Criteria Decision Aid Using Bipolar Reference System*, LAMSADE, Cahier et Documents no 56, Paris.
Konarzewska-Gubała E. (1987): "Multicriteria Decision Analysis with Bipolar Reference System: Theoretical Model and Computer Implementation". *Archiwum Automatyk i Telemechaniki* vol. 32, no 4, p.289-300.
Merighi D.(1980): *Un modello di valutazione rispetto insiemi di riferimento assegnati*. *Ricerca Operativa* no 13, p.31-52.
Roy B.(1985): *Methodologie Multicritere d'Aide a la Decision*. Economica, Paris.
Trzaskalik T.(1987): "Model of multistage multicriteria decision processes applying reference sets". In: *Decision Models with Incomplete Information,* Scientific Works of the University of Economics in Wrocław, no 413, p.73-93 (in Polish).

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 5
Scheduling and Control

VISCOSITY SOLUTION IN MRP THEORY AND SUPPLY NETWORKS FOR NON ZERO LEAD TIMES

Ludvik Bogataj and Marija Bogataj
University of Ljubljana, Faculty of Economics
ludvik.bogataj@ef.uni-lj.si; marija.bogataj@ef.uni-lj.si

Abstract: The paper describes viscosity solution of MRP model extended to global supply chain. It shows that the theory of Lions of viscosity solution developed in time domain gives the same results as the theory of Grubbström developed in frequency domain and at the same time gives us also the answer what is the optimal logistics policy.

Keywords: MRP theory, Logistics, Viscosity solution.

1 INTRODUCTION TO VISCOSITY SOLUTION AS DEVELOPED BY LIONS AND CRANDALL

The viscosity solution approach was introduced in 1980's by Pierre – Louis Lions and Michael Crandall (see the details in Crandall, Ishii, and Lions, 1992) as a generalization of the classical concept of a solution to a partial differential equation (PDE). It has been found that the viscosity solution is the natural solution concept to use in many applications of PDE's, In 1983 their main contribution was viscosity solutions for Hamilton-Jacobi equations, equations that had been the subject of Pierre's doctoral dissertation, where he had found solutions using techniques from partial differential equations and probability. The method is particularly interesting for OR society for solving the problems in differential games and especially for second-order equations such as the ones arising in stochastic optimal control or stochastic differential games. The classical concept was that a PDE: $H(x,u,Du) = 0$ over a domain $x \in \Omega$ has a solution if we can find a function $u(x)$ continuous and differentiable over the entire domain such that x, u and Du (the differential of u) satisfy the above equation at every point. Under the viscosity solution concept, u need not be everywhere differentiable. There may be points where Du does not exist, i.e. there could be a kink in u and yet u satisfies the equation in an appropriate sense. Although Du may not exist at some point, the *superdifferential* D^+u and the *subdifferential* D^-u, to be defined below, are used in its place where

$$D^+u(x_0) = \left\{p : \limsup_{x_1 \to x_0} \frac{u(x_1) - u(x_0) - p(x_1 - x_0)}{|x_1 - x_0|} \leq 0\right\} \tag{1}$$

$$D^-u(x_0) = \left\{p : \liminf_{x_1 \to x_0} \frac{u(x_1) - u(x_0) - p(x_1 - x_0)}{|x_1 - x_0|} \geq 0\right\} \tag{2}$$

By definition, a continuous function u is a *viscosity supersolution* of the above PDE if

$$H(x, u(x), p) \leq 0, \forall x \in \Omega, \forall p \in D^+u(x) \tag{3}$$

A continuous function u is a *viscosity subsolution* of the above PDE if

$$H(x, u(x), p) \geq 0, \forall x \in \Omega, \forall p \in D^-u(x). \tag{4}$$

A continuous function u is a viscosity solution of the PDE if it is both a viscosity supersolution and a viscosity subsolution.

2 MRP THEORY AS DEVELOPED BY GRUBBSTRÖM

Optimal decisions (i) where to produce, (ii) how to distribute the product and (iii) at what time to order or deliver the items in integrated supply chain can be successfully discussed and evaluated in time or frequency environment, where lead times and other time delays can be considered in linear form. The site and capacity selection, as for instance the problems where it is best to locate a facility and what capacity is needed to achieve the most rapid response, are discussed more easily in transformed environment. Lead times in the entire supply chain can be analysed in compact form using MRP and I-O analysis in Laplace transformed space.

An integrated supply chain includes the purchasing of raw materials, manufacturing with assembly and the distribution of produced goods to the final clients. A third part is the reverse logistics, having the same formal properties in the networks. In a supply chain the key variables that have to be considered at each activity cell are activity level and timing, inventory level and lead times between the order time and the moment of the arrival of items in the required activity cell. The managers of a supply chain have two main goals: (a) to keep the level of inventory in the supply chain as low as possible, to reduce inventory costs; (b) to move the inventory in its continually changing form or location from raw material to final product and its delivery through the physical distribution part of the supply chain to the final consumer at different locations and back in remanufacturing or recycling as fast as possible. The final goal is mostly to achieve the maximal net present value NPV of all activities in the supply chain and not only to reduce the costs of operations.

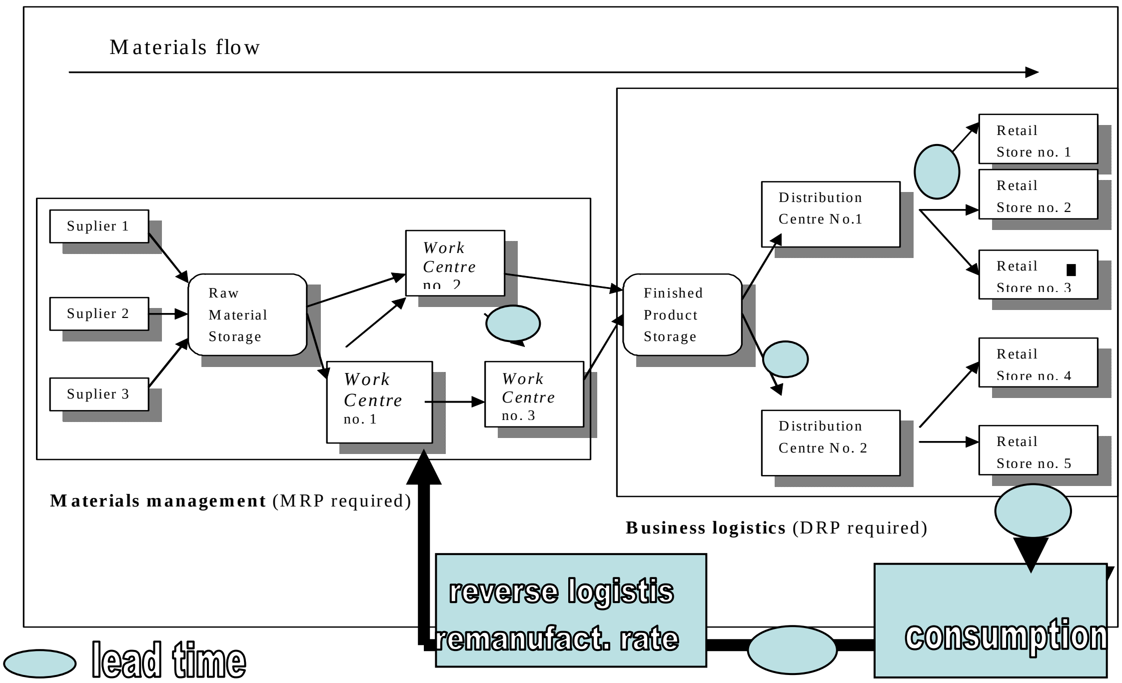


Figure 1: The representation of the inner structure in the logistic chain, having sub-systems of production, distribution, consumption and reverse logistics.

The line of research, now designated *MRP theory (in frequency domain)*, has attempted at developing a theoretical background for multi-level production-inventory systems, Material

Requirements Planning (MRP) in a wide sense. Basic in MRP theory are the rectangular *input* and *output matrices* **H** and **G**, respectively, having the same dimension. Different rows correspond to different items (products) appearing in the system and different columns to different activities (processes). We let *m* denote the number of processes (columns) and *n* the number of item types (rows). If the *j*th process is run on activity level P_j, the volume of required inputs of item *i* is $h_{ij}P_j$ and the volume of produced (transformed) outputs of item *k* is $g_{kj}P_j$. The total of all inputs may then be collected into the column vector **HP**, and the total of all outputs into the column vector **GP**, from which the net production is determined as (**G** - **H**)**P**. In general **P** (and thereby net production) will be a time-varying vector-valued function. If each process produces one type of item only, and this item is produced by this process alone, the output matrix may be written as the identity matrix **G** = **I**, assuming the processes and products to be numbered alike. Such systems are *elementary systems*. We may distinguish between the two clear-cut cases of an *assembly system* and an *arborescent system*. For the *assembly* system, items (on lower levels, upstream) are assembled (processed) into new items (on higher levels, downstream), the material flow being convergent. Adopting a principle of enumeration of the items in such a way that downstream items have lower indices compared to the upstream items they will become part of, the input matrix **H** may be written in a triangular form with positive elements only appearing below its main diagonal. For arborescent systems (having a divergent flow), an item is split into two or several downstream items. For such a system, the items and processes may be enumerated so that the output matrix **G** has the property that its positive elements only appear above its main diagonal. Distribution, extraction and reverse logistics processes typically have this property.

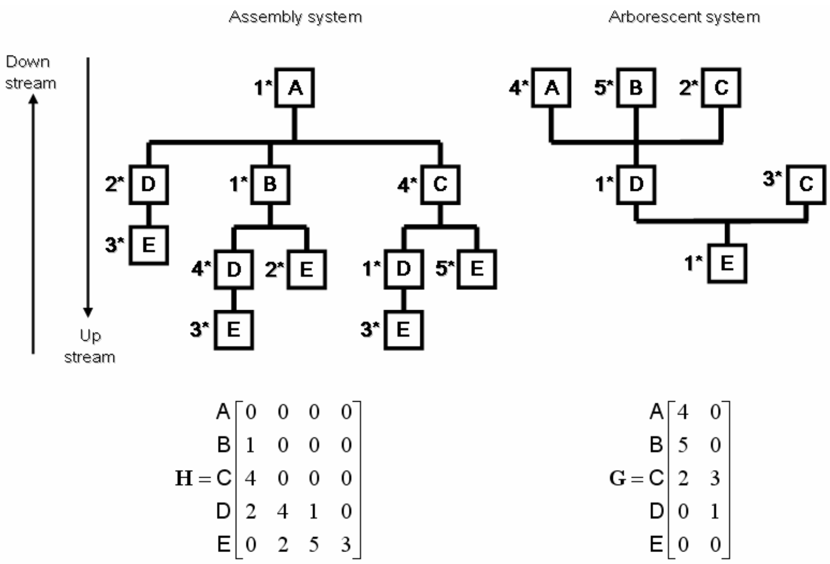


Figure 2. Examples of a pure assembly system and a pure arborescent system, in the form of product structures and their input and output matrices **H** *and* **G**, *respectively. (See the details in Grubbström, Bogataj, Bogataj, 2007)*

Systems may also be combinations of assembly and arborescent systems, in which case the input and output matrix may have arbitrarily complex structures. These are *mixed* systems.

Following Grubbström's MRP theory let $\tilde{\mathbf{F}}(s)$ is the vector of deliveries from the system. These are normally exports satisfying external demand, but could also be surplus items requiring disposal. Then, given a plan $\tilde{\mathbf{P}}(s)$, available inventory $\tilde{\mathbf{R}}(s)$ will develop according to

$$\tilde{\mathbf{R}}(s) = \frac{\mathbf{R}(0) + \left(\tilde{\boldsymbol{\Delta}}(s)\mathbf{G} - \mathbf{H}\tilde{\boldsymbol{\tau}}(s)\right)\tilde{\mathbf{P}}(s) - \tilde{\mathbf{F}}(s)}{s}, \tag{5}$$

where $\mathbf{R}(0)$ collect initial available inventory levels. The division by s represents a time integration of the flows represented by the other terms. The term $\tilde{\boldsymbol{\Delta}}(s)\mathbf{G}\tilde{\mathbf{P}}(s)$ is the inflow of purchasing, production, extraction, distribution etc. into available inventory, the term $\mathbf{H}\tilde{\boldsymbol{\tau}}(s)\tilde{\mathbf{P}}(s)$ is the required outflow representing needs generated by all processes (internal demand, dependent demand) and the term $\tilde{\mathbf{F}}(s)$ represents deliveries (exports) from the system which try to follow demand $\tilde{\mathbf{D}}(s)$. Sometimes it has to cover backlogs $\mathbf{B}(s)$.

For such a production system Grubbström developed also other fundamental equation and NPV expression in frequency domain and at the end he transformed it back to time domain. In order for the plan $\tilde{\mathbf{P}}(s)$ to be feasible, we must always have $\pounds^{-1}\left\{\tilde{\mathbf{R}}(s)\right\} \geq \mathbf{0}$. This is the *available inventory constraint*. As like the expected NPV of the production part was derived by Grubbström (1998), it has been derived on the same way also for total supply by Bogataj and Bogataj (2004) and it has got the extended form on the time horizon $\hat{T}$ when there is final backlog $\mathbf{B}^{tot}(\hat{T}) = 0$ as

$$E(NPV) = \mathbf{r}(E(\mathbf{D}^{tot}(\rho)) - E(\rho\dot{\mathbf{B}}^{tot}(\rho))) - \mathbf{c}\mathbf{P}^{tot}(\rho) - \mathbf{K}^{tot}\boldsymbol{\tau}^{tot}(\rho)\mathbf{v}^{tot}(\rho) \tag{6}$$

where $E(\cdot)$ denotes expected value, *tot* denotes the extension to total supply chain, r_j in $\mathbf{r}$ are sales prices of the products j at different production or distribution or reveres logistics levels and different locations which appear at different production and distribution cells in the supply chain, K_j in $\mathbf{K}$ are set-up costs at $j-th$ production cells or distribution cells for each activity (for example, in transportation there are always some fixed costs, which do not depend on the transportation volume). Here $\mathbf{c}$ is the row vector of unit incremental production costs, $\mathbf{v}^{tot}(\rho)$ is the vector of all completion times of production or distribution activities in the case when complex frequency equals to ρ, $\boldsymbol{\tau}^{tot}(\rho)$ is the lead time matrix, which depends on the time distance between two activities in production, reverse logistics or distribution cells and on input or output capacities which define setup, production or distribution times. In (6) we have unlimited capacities.

3 VISCOSITY SOLUTION OF MRP OR GLOBAL SUPPLY CHAIN MODEL

The same results we can get as viscosity solution without transforming this system in frequency domain. We consider the following production – inventory system with time delays in the state of the system (like the case of short time conservation effects for perishable goods) described by available inventory: $\mathbf{R}(t,\mathbf{a})$, and in production rate $\mathbf{P}(t,\boldsymbol{\tau})$ as the control:

$$\dot{\mathbf{R}}(t) = \mathbf{A}_0\mathbf{R}(t) + \sum_{i=1}^{N}\mathbf{A}_i\mathbf{R}(t-a_i) + \mathbf{P}(t) - \mathbf{H}'\,\mathbf{P}(t,\boldsymbol{\tau}) - \mathbf{F}(t) \tag{7}$$

with initial conditions (8)-(10):

$$\mathbf{R}(0) = \mathbf{R}^0 \tag{8}$$

Because we have delays in the state of the system and in the control, the state is well presented not only in the initial points. Therefore the state should be given by trajectories rather than just by points in time. The initial trajectories are the following:

$$\mathbf{R}\,(\theta) = \mathbf{R}^1(\theta) \qquad \textit{for} \qquad -a_{\max} \le \theta < 0 \tag{9}$$

$$\mathbf{P}\,(\eta) = \mathbf{P}^2(\eta) \qquad \textit{for} \qquad -\tau_{\max} \le \eta < 0 \tag{10}$$

In (7) we have: $\mathbf{A}_0 \in L(\mathbb{R}^N)$, $\mathbf{A}_i \in L(\mathbb{R}^N)$ are matrices describing deterioration and/or conservation effects if the items are perishable (deteriorating) or living stock and $\mathbf{H}' \in L(\mathbb{R}^N)$. Here $\mathbf{H}'$ is the Input Matrix where the elements $h'_{i,j}$ represent the number of units of item i required to produce one unit of its immediate successor j and could be extended by capacity constrains. For the case of simplicity without losing the main direction, here **G** is supposed to be $\mathbf{I}$. $\mathbf{R} \in \mathbb{R}^N$

$$\mathbf{P} \in L_2([-\tau_{\max}, T], \mathbb{R}^N), \qquad \mathbf{P}(t,\boldsymbol{\tau}) = \sum_{i=1}^{N} P_i(t-\tau_i)\cdot \mathbf{e}^i \tag{11}$$

where $\mathbf{e}^i$ is the i-th unit vector. $\mathbf{R}(t,a) = \sum_{i=1}^{N} R_i(t-a_i)\cdot \mathbf{e}^i$ (12)

and: $$\hat{\mathbf{R}} = (\mathbf{R}^0, \mathbf{R}^1, \mathbf{P}^2) \in M^2([-a_{\max}, 0]\;,\; \mathbb{R}^N) \times L_2([-\tau_{\max}, 0]\;, \mathbb{R}^N) \equiv$$
$$\equiv \mathbb{R}^N \times L_2([-a_{\max}, 0], \mathbb{R}^N) \times L_2([-\tau_{\max}, 0], \mathbb{R}^N), \tag{13}$$

In this formulas $a_i, \tau_i \in \mathbb{R}^+ \cup \{0\}$.

The system (7) can be written as the following equivalent abstract evolution equation presented in the paper of Ichikawa (1982):

$$\dot{\hat{\mathbf{R}}}(t) = \tilde{\mathbf{A}}\hat{\mathbf{R}}(t) - \hat{\mathbf{F}}(t) \tag{14}$$

with the initial condition: $\hat{\mathbf{R}}(0) = \hat{\mathbf{R}}_0$ (15)

$$\hat{\mathbf{R}}(t) = \begin{bmatrix} \mathbf{R}(t) \\ \mathbf{R}\,(t+\theta) \\ \mathbf{P}\,(t+\eta) \end{bmatrix}, \tag{16}$$

$$-a_{\max} \le \theta < 0$$

$$-\tau_{\max} \le \eta < 0 \tag{17}$$

$$\hat{\mathbf{R}}(t) \in \mathbf{Z}, \mathbf{Z} = \mathbb{R}^N \times L_2([t - a_{\max}, t], \mathbb{R}^N) \times L_2([t - \tau_{\max}, t], \mathbb{R}^N). \tag{18}$$

The infinitesimal generator $\widetilde{\mathbf{A}}$ of strongly continuous semigroup can be expressed as

$$\widetilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \frac{\partial}{\partial \theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{\partial}{\partial \eta} \end{bmatrix} \tag{19}$$

where $$\mathbf{AR} = \mathbf{A_0 R}(t) + \sum_{i=1}^{N} \mathbf{A_i R}(t - a_i) \tag{20}$$

$$\mathbf{BP} = \mathbf{P}(t) - \mathbf{H}'\mathbf{P}(t, \boldsymbol{\tau}) \tag{21}$$

$$\hat{\mathbf{F}}(t) = \begin{bmatrix} \mathbf{F}(t) \\ 0 \\ 0 \end{bmatrix} \tag{22}$$

$\tilde{\mathbf{A}}(\mathbf{P})$ is in general possibly unbounded operator on domain $D\left(\tilde{\mathbf{A}}(\mathbf{P})\right) \subset E$, where E is reflexive Banach state space. Under a certain assumptions (see Soner, 2002) there is unique mild solution to (14) and (15) such that the value function $NPV(\hat{\mathbf{R}}_0, t_0)$ is given by:

$$NPV(\hat{\mathbf{R}}_0, t_0) = \inf_{\mathbf{P} \in \mathbf{B}_{ad}}$$

$$(-\int_{t_o}^{T} \exp(\int_{t_o}^{t} -\lambda(\mathbf{R}(\xi; \mathbf{R}_0, t_0, \mathbf{P}(\cdot), \xi, \mathbf{P}(\xi)) d\xi) \cdot l(\mathbf{R}(\psi); \mathbf{R}(0), t_0, \mathbf{P}(\cdot)), \psi, \mathbf{P}(\psi)) d\psi +$$

$$+ \exp(\int_{t_o}^{T} -\lambda(\mathbf{R}(\xi; \mathbf{R}_0, t_0, \mathbf{P}(\cdot), \xi, \mathbf{P}(\xi)) d\xi) \cdot g(\mathbf{R}(T; \mathbf{R}_0, t_0, \mathbf{P}(\cdot)))), \tag{23}$$

where

$l(\mathbf{R}(0), t_0, \mathbf{P})$, $\lambda(\mathbf{R}(0), t_0, \mathbf{P})$ and $g(\mathbf{P})$ in general could be any real – valued functions. In (23) B_{ad} is a set of all admissible production vectors.

4 CONCLUSION

In Grubbström's and Bogataj's expression of NPV the general expressions from equation (23) has the following values:

$$l(\mathsf{R}(0),t_0,\mathsf{P},\lambda(\mathsf{R}(0),t_0,\mathsf{P}))\equiv$$

$$(r(t)(\mathsf{D}(t)-\dot{B}(t))-c(t)P(t)-w(t)H''(\overline{P}(t,\tau)-\overline{P}(t))+K(t)v(t))_{t=\psi} \qquad (24)$$

and general expression of λ is simplified to be constant ρ which is not always the case in real world problems:

$$\lambda(\mathsf{R}(\xi;\mathsf{R}_0,t_0,\mathsf{P}(\cdot)),\xi,\mathsf{P}(\xi))\equiv s\equiv\rho \qquad (25)$$

Here we start from the fact that in control theory problems the optimal solution of NPV is achieved as the solution of Hamilton – Jacobi – Bellman equation:

$$-(\frac{\partial}{\partial t_0})NPV(R_0,t_0)+H\,((\mathsf{R}_0,t_0,\mathrm{NPV}(\mathsf{R}_0,t_0)),\,Đ_{\mathbf{R}(0)}\mathrm{NPV}(\mathsf{R}_0,t_0))=0, \qquad (26)$$

$$NPV(\mathbf{R}_0,T)=g(\mathbf{R}_0) \qquad (27)$$

and $\mathbf{R}_0$ belongs to the set of initial available inventory levels.
where $Đ_{\mathbf{R}(0)}$ denotes the Frechet derivative of NPV in the direction of R_0.

References

[1] Bogataj, M., Bogataj, L. On the compact presentation of the lead time perturbations in distribution networks, *International Journal of Production Economics*, Vol. 88(2), 2003, 145–155.

[2] Crandall, M.G., Ishii, H. and Lions, P.L.: User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc., 27 (1992) , p.1-67.

[3] Grubbström, R.W. "On the Application of the Laplace Transform to Certain Economic Problems", *Management Science*, Vol.13, No.7, 1967, 558-567.

[4] Grubbström, R.W., "A Net Present Value Approach to Safety Stocks in Planned Production ", *International Journal of Production Economics*, Vol. 56-57, 1998, 213-229.

[5] Grubbström, R. W., Bogataj, M. and Bogataj, L., A compact representation and optimization of distribution and reverse logistics in the value chain, »Ten years after Storlien«, (Mathematical economics, operational reseach and logistics, serial no. 5). Ljubljana: Faculty of Economics: = Ekonomska fakulteta, KMOR, 2007, p. 1-48.

[6] Ichikawa, A., Stability of semilinear stochastic evolution equations , *Journal of Mathematical Analysis and Applications*, 90 (1), 1982, p.12-44.

[7] Soner, M., Touzi, N., Stochastic target problems, dynamic programming, and viscosity solution. SIAM J. Control. Optim, 41/2, 2002, p. 404-424 .

GLOBAL OPTIMIZATION OF THE SUPPLY CHAIN COSTS

Liljana Ferbar
Faculty of Economics, University of Ljubljana, Kardeljeva pl. 17, 1000 Ljubljana, Slovenia
liljana.ferbar@ef.uni-lj.si

Abstract: Forecasting models are often based on methods using various smoothing techniques. All parameters used in these techniques are determined by minimizing the mean absolute error (MAE), the mean square error (MSE) or some other error measurements. In this paper we show that optimization of the forecasting model should not be treated apart from the inventory model in which we use calculated forecasts. Using global optimization and the Solver optimization tool, we show that initial and smoothing parameters in the forecasting model can be determined to minimize costs – a fact applicable also to other models and other fields.

Keywords: Forecasting, Inventory, Supply chain, Cost model, Optimization

1. Introduction

Forecasting using time series analysis is a quantitative technique frequently used when numbers concerning the future are required. It is a common practice to use computer spreadsheets and Solver to choose the smoothing parameters for exponential smoothing technique, but management science textbooks [1,6,8,9] do not always disclose how the smoothing constants (and initial parameters, if at all) are computed (this is also discussed in [7]) – parameters can be determined by minimizing the mean absolute error (MAE), the mean square error (MSE) or some other error measurements.

In this paper we show how to use spreadsheets and Solver optimization tool for the determination of smoothing and initial parameters in forecasting methods and, what is more important, that the optimization of the forecasting model should not be treated apart from the inventory model in which we use calculated forecasts. We present an example of a supply chain and demonstrate that calculated forecasts of external demand, which are determined by minimizing MSE, are not optimal values for minimizing the supply chain costs. By letting Solver optimize more parameters in our supply chain model, we show that initial and smoothing parameters can be determined to minimize costs.

The paper is organized as follows. We begin by describing different forecasting methods. We then present our model of the supply chain with centralized demand and calculate average costs for all different forecasts to show how different forecasting methods influence the costs of the supply chain. Finally, we present the proposed global optimization and confirm that the initial and smoothing parameters in the forecasting methods can be chosen to minimize costs.

2. Forecasting methods

In this section we describe different exponential smoothing procedures. Smoothing and initial parameters in these methods are estimated by minimizing the mean square error

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - F_t)^2 ,$$

where Y_t is the observed data at time point t, F_t is the forecast made at the previous time point $t-1$, and n is the number of time periods. Estimation of parameters could also be done with respect to some other error measurements, but this paper deals only with MSE.

2.1 SES method

The single exponential smoothing (SES) is defined as

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t ,$$

where α is a given weight value to be selected subject to $0 \le \alpha \le 1$. Thus F_{t+1} is the weighted average of the current observation, Y_t, with the forecast, F_t, made at the previous time point $t-1$.

Since the value for F_1 is not known, we can use the first observed value (Y_1) as the first forecast ($F_1 = Y_1$). This is one possible method of initialization which is used very often. Another possibility would be to average the first four or five values in the data set and use this as the initial forecast. Even though more complicated formulas for the first forecast can be applied, the initial value should be part of the optimization model.

Solver can be used to minimize MSE for $t = 2-24$, but for a comparison with other models, which use first year for initialization, the period $t = 5-24$ is minimized (see Fig. 1, where we use notation Et for $|Y_t - F_t|$).

2.2 Holt's method

Holt's method is the extension of the exponential smoothing that takes into account a possible linear trend. There are two smoothing constants α and β; $0 \le \alpha, \beta \le 1$.

Forecasts at time t for period $t+k$ are made by $F_{t+k} = L_t + kb_t$.

The level L_t is updated as $L_t = \alpha Y_t + (1-\alpha)[L_{t-1} + b_{t-1}]$.

The trend b_t is updated as $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$.

Initial estimates are needed for L_1 and b_1. Some simple (and very often used) choices are $L_1 = Y_1$ and $b_1 = 0$. Solver can be used to minimize MSE regarding the smoothing constants α and β as well as the starting values for level and trend (see Fig. 1).

2.3 Holt-Winter's method

This is an extension of Holt's method to take seasonality into account. There are two versions, additive and multiplicative. Since the multiplicative one is more widely used, we will illustrate only this method.

Forecasts are adjusted for seasonal effects according to $F_{t+k} = (L_t + kb_t)S_{t+k-p}$, where p means the number of seasons per cycle (as 4 quarters per year).

The level L_t is updated as $L_t = \alpha(Y_t / S_{t-p}) + (1-\alpha)[L_{t-1} + b_{t-1}]$.

The trend is updated as $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1}$.

The seasonal parameters are updated according to $S_t = \gamma[Y_t / L_t] + (1-\gamma)S_{t-p}$.

To initialize the level we need one complete cycle of data, i.e. p values (in our case $p = 4$). Then we set $L_p = \frac{1}{p}(Y_1 + Y_2 + \cdots + Y_p)$.

To initialize the trend we use $p+m$ time periods

$$b_p = \frac{1}{m}\left(\frac{Y_{p+1}-Y_1}{p}+\frac{Y_{p+2}-Y_2}{p}+\cdots+\frac{Y_{p+m}-Y_m}{p}\right).$$

If the series is long enough, then $m = p$ is a good choice: however we can use $m = 1$. Initial seasonal indices can be taken as $S_m = Y_m / L_p; m = 1,2,\ldots, p.$

The smoothing parameters α, β and γ lie in the interval $[0,1]$, and can again be selected, as well as initial parameters, to minimize MSE (see Fig. 1).

	A	B	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1			SES method			Holt's method					Holt-Winter's method					
2			Alpha	0,4514		Alpha	0				Alpha	0,3124				
3						Beta	0,8568				Beta	0				
4		Data		MSE	7437,71				MSE	4028,93	Gamma	0			MSE	372,78
5	Month t	Yt	Ft	Et	Et*Et	Lt	bt	Ft	Et	Et*Et	Lt	bt	St	Ft	Et	Et*Et
6	1	362	839,30			333,20	18,50						0,45			
7	2	385	623,86	238,86	57054,82	351,70	18,50	351,70	33,30	1108,97			0,48			
8	3	432	516,05	-84,05	7064,02	370,20	18,50	370,20	61,80	3819,62			0,54			
9	4	341	478,11	137,11	18799,56	388,69	18,50	388,69	-47,69	2274,81	803,53	40,20	0,41			
10	5	382	416,22	-34,22	1171,29	407,19	18,50	407,19	-25,19	634,69	847,07	40,20	0,45	377,23	4,77	22,78
11	6	409	400,78	8,22	67,63	425,69	18,50	425,69	-16,69	278,59	875,27	40,20	0,48	427,50	18,50	342,10
12	7	498	404,49	93,51	8744,43	444,19	18,50	444,19	53,81	2895,60	918,93	40,20	0,54	492,05	5,95	35,46
13	8	387	446,70	-59,70	3563,64	462,69	18,50	462,69	-75,69	5728,56	954,05	40,20	0,41	393,68	-6,68	44,64
14	9	473	419,75	53,25	2835,41	481,19	18,50	481,19	-8,19	67,00	1014,15	40,20	0,45	444,52	28,48	811,12
15	10	513	443,79	69,21	4790,58	499,68	18,50	499,68	13,32	177,33	1057,59	40,20	0,48	508,00	5,00	25,03
16	11	582	475,03	106,97	11443,27	518,18	18,50	518,18	63,82	4072,81	1093,12	40,20	0,54	590,04	-8,04	64,60
17	12	474	523,31	-49,31	2431,55	536,68	18,50	536,68	-62,68	3928,72	1140,03	40,20	0,41	465,18	8,82	77,86
18	13	544	501,05	42,95	1844,39	555,18	18,50	555,18	-11,18	124,94	1191,64	40,20	0,45	527,67	16,33	266,56
19	14	582	520,44	61,56	3789,86	573,68	18,50	573,68	8,32	69,30	1224,37	40,20	0,48	593,52	11,52	132,61
20	15	681	548,23	132,77	17629,20	592,17	18,50	592,17	88,83	7890,12	1265,34	40,20	0,54	679,68	1,32	1,75
21	16	557	608,16	-51,16	2616,84	610,67	18,50	610,67	-53,67	2880,66	1321,63	40,20	0,41	535,87	21,13	446,65
22	17	628	585,07	42,93	1843,38	629,17	18,50	629,17	-1,17	1,37	1375,20	40,20	0,45	608,86	19,14	366,25
23	18	707	604,44	102,56	10517,61	647,67	18,50	647,67	59,33	3520,30	1431,64	40,20	0,48	681,96	25,04	627,21
24	19	773	650,73	122,27	14948,84	666,17	18,50	666,17	106,83	11413,52	1461,33	40,20	0,54	791,08	18,08	326,87
25	20	592	705,92	113,92	12977,98	684,66	18,50	684,66	-92,66	8586,62	1483,02	40,20	0,41	616,31	24,31	590,99
26	21	627	654,50	-27,50	756,30	703,16	18,50	703,16	-76,16	5800,66	1485,47	40,20	0,45	681,02	54,02	2918,25
27	22	725	642,09	82,91	6874,40	721,66	18,50	721,66	3,34	11,15	1519,13	40,20	0,48	735,08	10,08	101,69
28	23	854	679,51	174,49	30446,20	740,16	18,50	740,16	113,84	12959,96	1568,57	40,20	0,54	838,10	15,90	252,73
29	24	661	758,27	-97,27	9461,38	758,66	18,50	758,66	-97,66	9536,74	1609,28	40,20	0,41	660,33	0,67	0,45
30	Solver settings															
31	Minimize		F4			K4					Q4					
32	By changing		E2			H2:H3					M2:M4					
33	and		D6			G6:H6					L9:M9; N6:N9					
34	Subject to		0≤E2≤1			0≤H2,H3≤1					0≤M2,M3,M4≤1					

Fig. 1. Forecasts made with SES, Holt's and Holt-Winter's method where the smoothing and initial parameters are estimated by minimizing MSE.

3. Supply chain model

Demand amplification is a problem in real world supply chains. Many investigations [2,4,5] have shown that providing the supplier upstream with centralized data (all links in the supply chain are instantly aware of the demand change in the market) can significantly reduce the costs in the supply chain.

In this paper we deal with a supply chain model with centralized demand information in order to illustrate that even with a "good inventory model" and "good forecasts" the supply chain costs are still too high. Later we show that these costs can be minimized if we use "global optimization".

In our model we use the following notation:

Y_t ………observed data at time point t,

F_t ………forecast made at the previous time point $t-1$,

D_t^l ....demand of link *l* at time point *t*,

P_t^l ....production of link *l* at time point *t*,

IS_t^l ....initial stock of link *l* at time point *t*,

FS_t^l ....final stock of link *l* at time point *t*,

C_t^l ....inventory or penalty costs of link *l* at time point *t*,

C_t ....inventory or penalty costs of all links in chain at time point *t*

and assumptions:

1. At the time of initial observation the production, P_0^l, and the stock, IS_0^l, of all links in the supply chain are balanced.
2. The information in the supply chain is centralized – all links in the supply chain are instantly aware of the demand change in the market.
3. The production and the stock are nonnegative ($P_t^l \geq 0$, $IS_t^l \geq 0$).
4. The production and the change in production are not bounded (except by nonnegativity in the previous item).
5. Batches ordered at the time period *t* are available at the beginning of the time period *t* + 1 (lead times are assumed to be 1 period).
6. Insufficient stock level provokes the missing amount of products to be supplied from the marketplace (assuming that a perfect substitute for our product exists), which causes penalty costs.
7. Production (P_t^l) + Initial Stock (IS_t^l) = (Internal) Demand (D_t^l) + Final Stock (FS_t^l).

The costs of the supply chain are defined as the sum of the inventory costs and the penalty costs for all links. We assume the penalty costs to be higher than the inventory costs, which is expressed by introducing a weight, *penalty*, that is larger than 1. In other words, using the common notation $X^+ = \max(X, 0)$, the supply chain costs at the time point *t* are expressed as (*n* – total number of links in the supply chain):

$$C(t) = \sum_{l=1}^{n} C_t^l = \sum_{l=1}^{n} \left(\left(IS_t^l - D_t^l\right)^+ + penalty \cdot \left(D_t^l - IS_t^l\right)^+ \right)$$

A typical approach to the manufacturing process is an orientation towards production. In this case production levels for the time point *t* are determined through a postulation: production levels must equal forecast, i.e. $P_t^l = F_t$. The other approach is the inventory-oriented approach, which will be used in the simulation that follows in this paper (see also [3]). In this case our aim is to optimize inventory levels in order to reduce inventory costs, and production levels are adapted accordingly: $FS_t^l = IS_{t+1}^l = F_t$ and

$$P_t^l = \begin{cases} 0, & F_t < IS_t^l - D_t^l \\ F_t, & F_t \geq IS_t^l - D_t^l \text{ and } D_t^l \geq IS_t^l \\ F_t + D_t^l - IS_t^l, & F_t \geq IS_t^l - D_t^l \text{ and } D_t^l < IS_t^l \end{cases}$$

Without a loss of generality and for the sake of simplicity, we now consider a centralized supply chain with a manufacturer and one supplier ($n = 2$). In this case D_t^1 means an external demand for a manufacturer ($D_t^1 = Y_t$) and D_t^2 is an internal demand for a supplier ($D_t^2 = P_t^1$).

Now we can calculate average costs (*penalty* = 2, $t = 5 - 24$) for the forecast obtained with the SES method, where the smoothing and initial parameters were estimated by minimizing MSE (column D in Fig. 1). In Fig. 2 we illustrate these calculations using Excel spreadsheets.

	A	B	C	D	E	F	G	H	I	J	K	L
1			SES method									
2			Alpha	0,4514				Penalty	2			AC(5-24)
3		Data	Forecast	Manufacturer				Supplier				211,80
4	Month t	Yt	Ft	P_m(t)	IS_m(t)	FS_m(t)	C_m(t)	P_s(t)	IS_s(t)	FS_s(t)	C_s(t)	C(t)
5	1	362	839,30	623,86	362,00	623,86	0,00	623,86	362,00	623,86	523,72	523,72
6	2	385	623,86	277,19	623,86	516,05	238,86	169,37	623,86	516,05	346,68	585,54
7	3	432	516,05	394,06	516,05	478,11	84,05	356,13	516,05	478,11	121,98	206,03
8	4	341	478,11	279,11	478,11	416,22	137,11	217,23	478,11	416,22	199,00	336,11
9	5	382	416,22	366,55	416,22	400,78	34,22	351,10	416,22	400,78	49,67	83,90
10	6	409	400,78	404,49	400,78	404,49	16,45	404,49	400,78	404,49	7,42	23,87
11	7	498	404,49	446,70	404,49	446,70	187,02	446,70	404,49	446,70	84,42	271,44
12	8	387	446,70	360,06	446,70	419,75	59,70	333,11	446,70	419,75	86,64	146,34
13	9	473	419,75	443,79	419,75	443,79	106,50	443,79	419,75	443,79	48,07	154,57
14	10	513	443,79	475,03	443,79	475,03	138,43	475,03	443,79	475,03	62,48	200,91
15	11	582	475,03	523,31	475,03	523,31	213,95	523,31	475,03	523,31	96,57	310,51
16	12	474	523,31	451,74	523,31	501,05	49,31	429,49	523,31	501,05	71,57	120,88
17	13	544	501,05	520,44	501,05	520,44	85,89	520,44	501,05	520,44	38,77	124,66
18	14	582	520,44	548,23	520,44	548,23	123,12	548,23	520,44	548,23	55,57	178,70
19	15	681	548,23	608,16	548,23	608,16	265,55	608,16	548,23	608,16	119,86	385,41
20	16	557	608,16	533,91	608,16	585,07	51,16	510,82	608,16	585,07	74,24	125,40
21	17	628	585,07	604,44	585,07	604,44	85,87	604,44	585,07	604,44	38,76	124,63
22	18	707	604,44	650,73	604,44	650,73	205,11	650,73	604,44	650,73	92,58	297,69
23	19	773	650,73	705,92	650,73	705,92	244,53	705,92	650,73	705,92	110,37	354,90
24	20	592	705,92	540,58	705,92	654,50	113,92	489,16	705,92	654,50	165,34	279,26
25	21	627	654,50	614,59	654,50	642,09	27,50	602,17	654,50	642,09	39,91	67,41
26	22	725	642,09	679,51	642,09	679,51	165,82	679,51	642,09	679,51	74,85	240,67
27	23	854	679,51	758,27	679,51	758,27	348,98	758,27	679,51	758,27	157,52	506,49
28	24	661	758,27	617,10	758,27	714,37	97,27	573,19	758,27	714,37	141,17	238,44
29			714,37									

Fig. 2. Average costs (penalty = 2) for a forecast calculated with the SES method where the smoothing and initial parameters were estimated by minimizing MSE.

We calculate average costs (for the period $t = 5 - 24$) in similar way with different penalties for different forecasting methods (Table 1).

		SES method	HOLT's method	HOLT-WINTER's method
Penalty	2	211,80	115,23	155,57
	3	291,17	150,96	210,94
	4	370,53	186,70	266,30
	5	449,89	222,43	321,67

Table 1. Average costs with different penalties for different forecasting methods.

4. Global optimization

In this section we show that optimization of the forecasting model should not be treated apart from the production–inventory model in which we use calculated forecasts. Even though we believe that we get the best possible fit for the future demand, the fact is that the smoothing and initial parameters calculated by optimization of the forecasting model are not optimal values for minimizing the supply chain costs.

Using the Solver optimization tool we show that initial and smoothing parameters in the forecasting model can be determined to minimize costs. In Fig. 3 we illustrate how the smoothing and initial parameters in the SES forecasting method are estimated by minimizing average costs. In this case we get $\alpha = 0.2304$ and $F_1 = 488.47$ and the average costs can be reduced by almost 12% (see Fig. 2 and Fig. 3).

	A	B	C	D	E	F	G	H	I	J	K	L
1			SES method									
2			Alpha	0,2304				Penalty	2			AC(5-24)
3		Data	Forecast	Manufacturer				Supplier				186,42
4	Month t	Yt	Ft	P_m(t)	IS_m(t)	FS_m(t)	C_m(t)	P_s(t)	IS_s(t)	FS_s(t)	C_s(t)	C(t)
5	1	362	488,47	459,33	362,00	459,33	0,00	459,33	362,00	459,33	194,66	194,66
6	2	385	459,33	367,88	459,33	442,21	74,33	350,75	459,33	442,21	91,46	165,79
7	3	432	442,21	429,65	442,21	439,86	10,21	427,30	442,21	439,86	12,56	22,77
8	4	341	439,86	318,23	439,86	417,08	98,86	295,45	439,86	417,08	121,63	220,49
9	5	382	417,08	373,92	417,08	409,00	35,08	365,84	417,08	409,00	43,16	78,25
10	6	409	409,00	409,00	409,00	409,00	0,00	409,00	409,00	409,00	0,00	0,00
11	7	498	409,00	429,50	409,00	429,50	178,00	429,50	409,00	429,50	41,01	219,01
12	8	387	429,50	377,21	429,50	419,71	42,50	367,42	429,50	419,71	52,30	94,80
13	9	473	419,71	431,99	419,71	431,99	106,58	431,99	419,71	431,99	24,55	131,13
14	10	513	431,99	450,65	431,99	450,65	162,02	450,65	431,99	450,65	37,33	199,35
15	11	582	450,65	480,91	450,65	480,91	262,70	480,91	450,65	480,91	60,52	323,22
16	12	474	480,91	472,41	480,91	479,32	6,91	470,82	480,91	479,32	8,50	15,41
17	13	544	479,32	494,22	479,32	494,22	129,36	494,22	479,32	494,22	29,80	159,16
18	14	582	494,22	514,44	494,22	514,44	175,56	514,44	494,22	514,44	40,44	216,01
19	15	681	514,44	552,81	514,44	552,81	333,12	552,81	514,44	552,81	76,74	409,86
20	16	557	552,81	553,78	552,81	553,78	8,37	553,78	552,81	553,78	1,93	10,30
21	17	628	553,78	570,88	553,78	570,88	148,44	570,88	553,78	570,88	34,20	182,64
22	18	707	570,88	602,24	570,88	602,24	272,25	602,24	570,88	602,24	62,72	334,97
23	19	773	602,24	641,58	602,24	641,58	341,53	641,58	602,24	641,58	78,68	420,21
24	20	592	641,58	580,58	641,58	630,16	49,58	569,16	641,58	630,16	61,00	110,57
25	21	627	630,16	626,27	630,16	629,43	3,16	625,55	630,16	629,43	3,88	7,04
26	22	725	629,43	651,45	629,43	651,45	191,14	651,45	629,43	651,45	44,03	235,18
27	23	854	651,45	698,11	651,45	698,11	405,11	698,11	651,45	698,11	93,33	498,44
28	24	661	698,11	652,45	698,11	689,56	37,11	643,90	698,11	689,56	45,66	82,77
29			689,56									
30	Solver settings											
31	Minimize			L3								
32	By changing			D2								
33	and			C5								
34	Subject to			0≤D2≤1								

Fig. 3. Minimized average costs (penalty = 2) obtained with the SES forecasting method where the smoothing parameter α and initial parameter F_1 are estimated by minimizing average costs.

When we carry out the global optimization of our supply chain (costs) model for all variations of forecasting methods (see Fig. 4, where we use notation *minAC* for minimized average costs obtained with regard to smoothing and initial parameters), we observe the following:

a) For the SES method: Average costs AC(SES) (obtained by minimization of MSE with regard to the smoothing and initial parameters) can be reduced (on average with regard to the penalty) by 17% in comparison with minAC(SES)

b) For Holt's method: Average costs AC(HOLT) can be reduced by 11% in comparison with minAC(HOLT)

c) For Holt-Winter's method: Average costs AC(H-W) can be reduced by 62% in comparison with minAC(H-W)

Finally, we can conclude that, with the global optimization of our costs model, we can always reduce average costs, as was actually expected, and that the maximal reduction can be achieved in "the best" forecasting method, i.e., Holt-Winter's method, what is perhaps more surprising.

	SES method	
Penalty	AC(SES)	minAC(SES)
2	211,80	186,42
3	291,17	256,35
4	370,53	321,99
5	449,89	309,36
	HOLT's method	
Penalty	AC(HOLT)	minAC(HOLT)
2	115,23	112,96
3	150,96	139,75
4	186,70	162,01
5	222,43	174,26
	HOLT-WINTER's method	
Penalty	AC(H-W)	minAC(H-W)
2	155,57	78,49
3	210,94	83,98
4	266,30	86,49
5	321,67	88,02

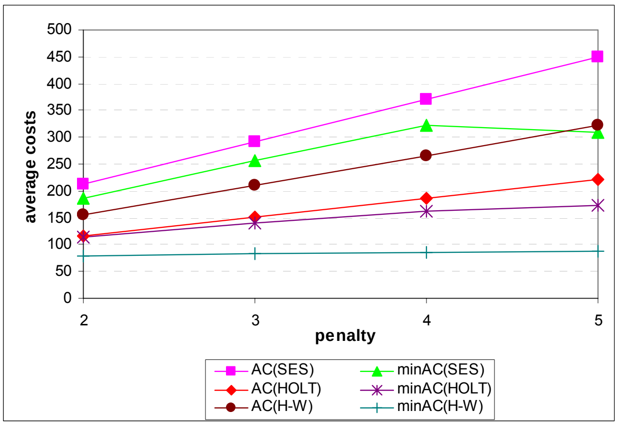


Fig. 4. Minimized average costs with different penalties for different forecasting methods.

5. Conclusion

This paper exposes the problem of the local optimization of the forecasting methods (i.e. selecting appropriate initial and smoothing parameters to get better fit to the time series data), when the calculated forecasts are used in the other model. We propose global optimization of an inventory oriented supply chain model with centralized demand and confirm that the initial and smoothing parameters in the forecasting methods can be chosen to minimize costs. The advantage of this paper is not only the reduction of the supply chain costs but also the usage of spreadsheets and optimization tools to other models and other fields.

References

[1] Camm JD, Evans JR. Management science and decision technology, first ed. Cincinnati, OH: Southwestern College Publishing; 2000. p. 390.

[2] Chen at al., Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times and Information. *Management Science*, Linthicum (MD), 2000; 46(3): 436-443.

[3] Ferbar L, Čreslovnik D, Mojškerc B and Rajgelj M. The Influence of Smoothing Coefficient on Costs of Centralized Supply Chain, Proceedings: KOI 2006 / 11th International conference on operational research (in press).

[4] Lee LH, Padmanabhan V, Whang S. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, Linthicum (MD), 1997; 43(4): 546-558.

[5] Lee LH, Padmanabhan V, Whang S. The Bullwhip Effect in Supply Chains. *Sloan Management Review*, Cambridge (MA), 1997; 38(3): 93-102.

[6] Ragsdale CT. Spreadsheet modeling and decision analysis, third ed. Cincinnati, OH: Southwestern College Publishing; 2001. p. 794.

[7] Rasmussen R. On time series data and optimal parameters. *The International Journal of Management science*, Omega (32); 2004. pp. 111-120.

[8] Render B, Stair RM. Quantitative analysis for management, seventh ed. Englewood Cliffs, NJ: Prentice-Hall; 2000.

[9] Winston WL, Albright CS. Practical management science, second ed. Pacific Grove, CA: Duxbury; 2001.

SOME MIXED ALGORITHMS IN OPTIMAL CONTROL

Lado Lenart [1], Jan Babič [1], Janez Kušar [2]
[1] Jožef Stefan" Institute , Jamova 39, Ljubljana
[2] University of Ljubljana, Faculty of mechanical Engineering
lado.lenart@ijs.si; jan.babic@ijs.si, janez.kusar@fs.uni-lj.si

Abstract: The algorithms for optimal control system are generally split into two solution classes. The first class uses methods to directly cope with the Hamilton-Jacoby-Bellman equation (HJBE). The second class solves HJBE after transforming it into canonical system of ordinary differential equations (ODE) with split boundary values problem (BVP), also called dual point problem. Among a great number of direct methods the principles of collocation and Galerkin's error estimation principle are highly interesting. The canonical equations seem to be practically the more standard way of solution, even if one has to solve two-point problem.

Keywords: HJBE – equation, collocation method, Galerkin method, canonical equations

1. INTRODUCTION, GENERAL FORMALISM

The introductory part of the paper handles the common formalism in optimal control from the viewpoint of solution of partial differential equations (PDE), HJBE equation in particular. Because of this common view the general theory can be found in any book of PDE's, e.g. [1],[2],[3]. In the second part we are dealing with the collocation [4], [5], [6] and Galerkin method [7], [8] for solving HJBE . The last section is more standard again, the theoretical background can be found in [9], [10].

We will consider some questions in open and closed loop optimal control. The common expression for the optimal control is posed as a Bolza problem:

$$\begin{aligned} &J = \min a(x(T)) + \int_0^T f_0\left(x(t), u(t) \right) dt \qquad \text{, s.t.} \\ &\frac{d}{dt} x(t) = f_s\left(x(t), u(t)\right) \quad ; x(0) = x_0 \; ; b\left(x(T)\right) = 0 \end{aligned} \tag{1.1}$$

The following theorem is proven in [3]:

Theorem : If problem (1.1) is given and the functions a, f_0, f_s, b are continuously partially differentiable, let u^* be the optimal control and x^* the resulting trajectory. Let the matrix $b_x\left(x^*(T)\right)$ have full row rank. The linearized system:

$$\frac{d}{dt} x(t) = \frac{df_s}{dx}\left(x^*(t), u^*(t)\right)^T x(t) + \frac{df_s}{du}\left(x^*(t), u^*(t)\right)^T u(t) \tag{1.2}$$

shall be controllable. The Hamiltonian function is defined

$$H = f_s\left(x^*(t), u^*(t)\right)^T p^*(t) + f_0\left(x^*(t), u^*(t)\right)^T \tag{1.3}$$

Then a costate variable function exists p^* and vectors $(q_0)^*$ and $(q_T)^*$ such that the boundary value problem (1.4), (1.5), (1.6), (1.7) can be solved almost anywhere on $[0,T]$.

$$\frac{d}{dt} x^*(t) = \frac{\partial H}{\partial p}; \quad x^*(0) = x_0 ; \quad b\left(x^*(T)\right) = 0 \tag{1.4}$$

$$\frac{d}{dt} p^*(t) = -\frac{\partial H}{\partial x}; \quad p^*(0) = -q_0^* \tag{1.5}$$

$$p^*(T) = \frac{d}{dx}x^*(T) + \left(\frac{d}{dt}b\left(x^*\right) \right)^T q_T^* \tag{1.6}$$

$$0 = \left(\frac{d}{du} f_s\left(x^*(t), u^*(t)\right) \right)^T p^*(t) + \frac{d}{dt} f_0\left(x^*(t), u^*(t)\right) \tag{1.7}$$

Eq. (1.4) is the dynamic system for *x* to be controlled, (1.5) is the adjoint equation, (1.6) are the transversality conditions and (1.7) is the local Pontryagin maximum principle. If the Lagrangian function is set up for Bolza- problem in the form,

$$\begin{aligned} L\left(x,u,p,q_0,q_T\right) &= a\left(x(T)\right) + b\left(x(T)\right)^T q_T + \left(x(0) - x_0\right)^T q_0 + \\ &+ \int_0^T f_0\left(x(t),u(t)\right) + p(t)^T \left[f_s\left(x(t),u(t)\right) - \frac{d}{dt}x(t) \right] dt \end{aligned} \tag{1.8}$$

then (1.5), (1.6), (1.7) can be obtained from Frechet derivatives of (1.8) by *x* and by *u* , if they are set to be zero. Equations (1.4), (1.5), (1.6), (1.7) are canonical equations .

The same Bolza problem solved with DP delivers the Hamilton – Jacobi – Bellman equation (HJBE) in the form:

$$-\frac{\partial J}{\partial t} = \min_u \left(H\left(x,u,J_x,t\right) \right) \tag{1.9}$$

Hamiltonian *H* in (1.9) is the function (1.3), costate variables *p* are equal to J_x. But despite the fact, that both (1.9) and (1.3) equivalently describe the same problem, the numerical possibilities to solve either of them differ very much. If the equivalency of (1.3) and (1.9) shall be proved, then the transfer (1.9) to canonical form can be showed directly. We write the partial differential equation (PDE) of the first order in the form:

$$F(x_1, x_2, \ldots, x_n, \varphi, p_1, p_2, \ldots p_n) = 0; \quad p_i = \frac{\partial \varphi_i}{\partial x_i} \tag{1.10}$$

For (1.10) the system of characteristics can be constructed, the single characteristic is one-parameter space curve:

$$\frac{dx_i}{ds} = F_{x_i}, \quad \frac{d\varphi}{ds} = \sum_{i=1}^n p_i F_{x_i}, \quad \frac{dp_i}{ds} = -\left(F_\varphi p_i + F_{x_i}\right) \tag{1.11}$$

The function $F\left(x_i, u, p_i\right)$ is integral of (1.11), as (1.10) can be derivated by s :

$$\frac{dF}{ds} = \sum_{i=1}^n F_{x_i} \frac{dx_i}{ds} + \sum_{i=1}^n F_{p_i} \frac{dp_i}{ds} + F_\varphi \frac{d\varphi}{ds} = 0 \tag{1.12}$$

If equations from (1.11) are inserted into (1.12), the result is 0 again, then characteristics solve the PDE. If one returns back to HJBE (1.9), it can be seen, that the function *J* in it does not appear explicitly. For such PDE we can isolate one of the parameters, say $x_n = x$, and for simplicity reasons *n* shall be resized. DPE shall be resolved by the derivative of the solution by x

$$p + H\left(x_1, x_2, \ldots, x_n, x, p_1, p_2, \ldots, p_n\right) = 0; \quad p = \frac{\partial \varphi}{\partial x}; \quad p_i = \frac{\partial \varphi_i}{\partial x_i} \tag{1.13}$$

Then one has the characteristic equations (1.11), as $\partial F / \partial \varphi = 0$ and respecting (1.13) the characteristic equations (one of them degenerates into $dx / ds = 1$):

$$\frac{dx_i}{dx} = H_{p_i}; \quad \frac{dp_i}{dx} = -H_{x_i} \tag{1.14}$$

More else, the next equation is true:

$$\frac{d\varphi}{dx}=\sum_{i=1}^{n} p_i H_{p_i} - H \tag{1.15}$$

Herewith the equivalency: (1.9) $\Rightarrow$ (1.4) is proved.

The proof in the other direction can be shortly outlined as Cauchy problem: it is necessary to find the solution, which includes the known space curve *l.* Let for simplicity (1.10) be written for 2- dimensions in space (*x,y*) with derivatives (*p,q*).

$$p = f_c(x,y,\varphi,q); \text{ or } \quad F_c(x,y,\varphi,p,q)=0; \tag{1.16}$$

Space curve *l* is defined in the plane $x = x_0$ with the function

$$\varphi\,|_{x=x_0} = \psi\left(y\right) \tag{1.17}$$

Then the parametric presentation of l is: $x = x_0;\ y = y;\ u = \psi\left(y\right)$. But every point $\left(x_0, y_0, p_0, q_0\right)$ of *l* must satisfy the next pair of equations:

$$F_c\left(x_0, y_0, \varphi_0, p_0, q_0\right)=0; \quad \frac{d\varphi_0}{dt} = p_0\frac{dx_0}{dt} + q_0\frac{dy_0}{dt} \tag{1.18}$$

Because of (1.18) *l* must match the next two eqns.:

$$p_0 = f_c\left(x_0, y, \psi\left(y_0\right), q_0\right); \quad \psi'\left(y\right) = q_0 \tag{1.19}$$

Then p_0 and q_0 are fully determined along l and herewith the band of the solution plane. Then through every supporting point the characteristic can be drawn and herewith the complete surface is known.

2. DIRECT METHODS: OPTIMAL CONTROL IN PDE

The general formulation of BVP in PDE is to determine the function $\varphi\left(x_1,\ldots,x_n\right)$ of *n* independent variables which satisfies a PDE:

$$F\left(x_1,\ldots,x_n,\varphi,\varphi_1,\ldots,\varphi_n,\varphi_{11},\ldots,\varphi_{nn}\right)=0 \text{ in } B \tag{2.1}$$

The boundary conditions are expressed as:

$$V_\mu\left(x_1,\ldots,x_n,\varphi,\varphi_1,\ldots,\varphi_n,\varphi_{11},\ldots,\varphi_{nn}\right)=0 \text{ on } \Gamma \tag{2.2}$$

In (2.1), (2.2) *B* is the given region of the x – space, Γ_μ are the $(n-1)$ dimensional hyper-surfaces and φ_i, φ_{ij} are partial derivatives. All functions are assumed to be continuous. The PDE-s in control theory generally are linear or pseudo-linear and have the form:

$$\sum_{\alpha_1+\alpha_2\ldots+\alpha_n=m} A_{\alpha_1,\ldots,\alpha_n} \frac{\partial^{\alpha_1+\ldots+\alpha_n}}{\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}} = r \tag{2.3}$$

If boundary conditions are linear, they are written as:

$$U_\mu\left(x_1,\ldots,x_n,\varphi,\varphi_1,\ldots,\varphi_n,\varphi_{11},\ldots,\varphi_{nn}\right)=\gamma_\mu \text{ on } \Gamma_\mu,\ \left(\mu=1,\ldots,k\right) \tag{2.4}$$

From the numerous methods for solving (2.3), (2.4) in control problems, collocation and orthogonality method shall be mentioned.

In pure collocation method the first approximation of the solution is assumed to be dependent on a finite number of parameters $a_1, a_{2,}\ldots$. The error ε is made to vanish at collocation points, which shall to be distributed fairly uniformly over the region *B* or boundary surfaces Γ_μ. Let us follow [4] to illustrate collocation algorithm. Eqn. (1.9) is first split into two coupled eqns.:

$$-\frac{\partial J}{\partial t} - \nabla J.f\left(t,x,u^*\right) - L\left(t,x,u^*\right) = 0$$
$$u^* = \arg\sup_u \left[-\nabla J.f\left(t,x,u\right) - L\left(t,x,u\right)\right] \tag{2.5}$$

Then let $x_i^c, \quad x = 1,..M$ be a set of collocation points. The inverse multiquadric radial basis functions (RBF) are defined as:

$$\Phi_i\left(x\right) = \left(\sqrt{\| x - x_i \|^2 + c^2}\right)^{-1}, \quad i = 1,\dots M \tag{2.6}$$

c>0 is the shape parameter. Using (2.6) one defines

$$J_r\left(x,t\right) = \sum_{i=1}^{M} \alpha_i\left(t\right)\Phi_i\left(x\right) \tag{2.7}$$

Replacing J in (2.5) with J_r yields a linear system with M unknowns $\{\alpha_i\}$ and u^* at any t. These unknowns can be calculated at collocation points, the result is the system of ordinary differential equations in $\{\alpha_i\}$ and $u^*\left(x_i\right)$ for $i = 1,2,\dots M$.

Beside this spatial discretization the uniformly discretization in N subintervals is needed for time axis, the points are $t^n = 1 - n\Lambda t$. Applying the two level finite difference scheme to the previous system of ODE and decoupling one has the system of difference equations, which can be iteratively solved.

The second quite standard method for direct solving of PDE in control is the orthogonality method. To solve problem (2.1), (2.2) one chooses linearly independent functions g_ρ and requires error ε to be orthogonal to these functions in the region B, i.e.

$$\int_B \varepsilon . g_\rho dt = 0, \quad \rho = 1,2,.. \tag{2.8}$$

Functions g_ρ are often chosen to be the first functions of a complete system of functions in B. Boundary conditions are linear and of the form (2.4) , the PDE may still be non linear. One can therefore take the approximate solution w be a linear expression :

$$w = v_0\left(x_1, x_2, \quad, x_n\right) + \sum_\rho a_\rho v_\rho\left(x_1, x_2, \quad, x_n\right) \tag{2.9}$$

One has parameters a_ρ and v_0 satisfying the inhomogeneous boundary conditions and the v_ρ corresponding homogeneous conditions. Galerkin method is a special case of orthogonality method in which the functions v_ρ are used for the g_ρ in (2.8). If for the error one has the form:

$$\varepsilon\left(x_j, a_j\right) = F_G\left(x_j, w, w_1, \dots, w_{11}\dots\right) \tag{2.10}$$

then (2.8) can be read as:

$$\int_B F_G\left(x_1,\dots,x_n, w, w_1,\dots,w_n\dots\right) v_\rho\left(x_1,\dots,x_n\right) dt = 0 \tag{2.11}$$

The Galerkin method was demonstrated by [7] with the dynamic system:

$$\frac{dx}{dt} = f(x) + g(x)u(x) \tag{2.12}$$

and generalized HJBE:

$$\frac{\partial V}{\partial x}(f + gu) + l + \| u \|^2 = 0 ; \quad V\left(0\right) = 0 \tag{2.13}$$

Function l in (2.13) is stabilizing function. Then there exists coefficients b_j such that:

$$\| V(x) - \sum_{j=1}^{\infty} b_j \Phi_j(x) \| = 0 \tag{2.14}$$

One seeks an approximate solution with an error:

$$\text{error}_N = GHJB(\sum_{j=1}^{N} c_j \Phi_j(x); u) \tag{2.15}$$

The coefficients c_j are determined by setting the projection of the error (2.15) on the finite basis $\{\Phi_j\}_{j=1}^{N}$ to zero. Using (2.8) this expression reduces to the system of N equations in N unknowns.:

$$\sum_{j=1}^{N} \left\langle \frac{\partial \Phi_j}{\partial x}.(f + g.u), \Phi_n \right\rangle + \left\langle l + \| u \|^2, \Phi_n \right\rangle = 0 \tag{2.16}$$

(2.16) is in the Galerkin form and can be solved by successive approximations.

3. INDIRECT METHODS: CANONICAL EQUATIONS

The solution of canonical ODE system normally is regarded to be simpler as to cope HJBE solution. We will present the simple example, where even the commercial BVP solver is of limited use and then the problem is inverted into the variational problem. Let the dynamic system for single link manipulator be given in (3.1) Variables are angle Θ (between negative vertical axis and manipulator), angular velocity Ω, torque τ, dumping coefficient K_{12}, gravitational constant *g*, mass *m*, moment of inertia *I* and length l_2.

$$\begin{aligned}
&\frac{d\Theta}{dt} = \Omega; \quad \frac{d\Omega}{dt} = \frac{1}{I}\left(\tau - K_{12}\Omega^2 - mgl_2 \sin(\Theta)\right) \\
&J = K_r \Theta(T) + \int_0^T \tau^2 ds \\
&H = \tau^2 + p_1 \frac{d\Theta}{dt} + p_2 \frac{d\Omega}{dt} = \tau^2 + p_1 \Omega + p_2 \frac{1}{I}\left(\tau - K_2 \Omega^2 - mgl_2 \sin(\Theta)\right)
\end{aligned} \tag{3.1}$$

It happens that τ has the uniform analytic solution and can be directly expressed :

$$\frac{\partial H}{\partial t} = 0; \quad \Rightarrow \tau = -\frac{p_2}{2I} \tag{3.2}$$

If τ from (3.2) is inserted into the first line in (3.1) one gets the system of canonical equations:

$$\begin{aligned}
\frac{d\Theta}{dt} &= \Omega \\
\frac{d\Omega}{dt} &= \frac{1}{I}\left(-\frac{p_2}{2I} - K_{12}\Omega^2 - mgl_2 \sin(\Theta)\right) \\
\frac{dp_1}{dt} &= p_2 \frac{1}{I} mgl_2 \cos(\Theta) \\
\frac{dp_2}{dt} &= -p_1 + 2 p_2 \frac{K_{12}}{I} \Omega
\end{aligned} \tag{3.3}$$

The split boundary conditions are:

$$\Theta(0) = 0; \quad \Omega(0) = 0; \quad p_1(T) = \frac{\partial}{\partial \Theta}\left[K_r \Theta(T)\right] = K_r; \quad p_2(T) = 0 \tag{3.4}$$

Nevertheless dual point boundary problem must be resolved numerically. For our model the 'MATLAB' solver 'bvp4c', using the method of collocation, was successful only for small torques and angles. The numerical problem seems to be to guess the proper initial data for co-state variables. However, to prove the result, the calculus of variations was used. One writes the first line of (3.1) in the form:

$$\tau = I\frac{d^2\Theta}{dt^2} + K_{12}\left(\frac{d\Theta}{dt}\right)^2 + mgl_2\sin\left(\Theta\right) \tag{3.5}$$

(3.5) is conformal with the next basic formula in the calculus of variations:

$$I_c\left(y\right) = \int_{x_0}^{x_1} f\left(x, y, y', y'', \ldots, y^{(n)}\right)dx \overset{!}{=} \text{Extremal} \tag{3.6}$$

Then it follows for the Euler-Lagrange formula of manipulator model:

$$2\frac{d^2\Theta}{dt^2}K_{12} - mgl_2\cos\left(\Theta\right) = 0 \tag{3.7}$$

Taking into account the fact, that the object function for (3.7) is $\int \tau ds$ and for (3.3) $\int \tau^2 ds$, the results from solver 'bvp4c' and (3.7) are good comparable. It remains to square (3.5) and get the new variational problem, which is then fully compatible with (3.3) – and gives the initial data for direct integration of (3.3).

REFERENCES

[1] R.Courant,D.Hilbert, Methods of mathematical physics,Vol II.,MIR,Moscow,(1964)
[2] L.Collatz, The numerical treatment of differential equations, Springer, Berlin, (1966)
[3] J.Jahn, Introduction to the theory of nonlinear optimization, Springer, Berlin, (1996)
[4] C.S.Huang,S.Wang,C.S.Chen,Z.C.Li, A radial basis collocation method for Hamilton-Jacobi-Bellman equations, Automatica Vol.42,6 (2006),2201-2207
[5] M.Alamir, Solutions of nonlinear optimal and robust control problems via a mixed collocation/DAE's based algorithm, Automatica Vol. 37,7(2001),1109-1115
[6] T.Neckel,C.Talbot,N.Petit, Collocation and inversion for reentry optimal control problem, http://cas.ensmp.fr/~petit/papers.cnes03/main.pdf
[7] R.W.Beard,G.N.Saridis,J.T.Wen, Galerkin Approximations of the generalized Hamilton-Jacibi-Bellman equation, Automatica, Vol 33, 12 (1997) 2159-2177
[8] O.Lepsky,C.Hu, Analysis of the discontinuous Galerkin method for Hamilton-Jacobi equations, Applied Numerical Mathematics, Vol 33(2000),423-434
[9] F.Lewis,V.L.Syrmos, Optimal Control, Wiley-Interscience, New York (1995)
[10] J.Fox, L.Leslie, The numerical solution of two-point boundary problems in ordinary differential equations, Dover Publications, New-York, (1990).

A DECISION SYSTEM FOR VENDOR SELECTION PROBLEM

Tunjo Perić*, Zoran Babić**
* Pekarne Sunce d.o.o., 10431 Sveta Nedelja, e-mail: tunjo.peric1@zg.t-com.hr
** Faculty of Economics, 21000 Split, Matice hrvatske 31, Croatia, e-mail: babic@efst.hr

Abstract: One of the important tasks in the operation of every firm is the choice of suppliers. Vendor (supplier) selection problem is of vital importance for operation of every firm because the solution of this problem directly and substantially affects costs and profit. This paper develops the procedure of supplier selection by multicriterial analysis and gives an example of its application in a concrete bakery.

Keywords: supplier selection, Analytic Hierarchy Process, baker industry

1. Introduction

Vendor selection problem (or supplier selection as it is often called) is one of the most important tasks in every industry. Namely, costs of buying equipment from external vendors can have a significantly affect the company operation quality as well as its development and survival. In this paper the vendor selection problem is treated as a multicriteria problem because it covers various aspects of both qualitative and quantitative criteria. For example Weber et al. (7) identified 23 different criteria evaluated in the vendor selection process.
In principle there are two kinds of supplier (vendor) selection problem:

First, when in supplier selection there is no constraint or in other words all suppliers can satisfy the buyer's requirements of demand, quality, delivery etc. In this kind of supplier selection the management needs to make only one decision - which supplier is the best one.

Second, the other type of supplier selection problem is when there are some limitations on suppliers' capacity, quality and so on. In other words no supplier can satisfy the buyer's total requirements and the buyer needs to purchase some part of demand from one supplier and the other part from another to compensate for the shortage of capacity or low quality of the first supplier. The firm must decide which vendors it should contract and it must determine the appropriate order quantity for each vendor selected. This kind of model was discussed in (3).

In this paper we will discuss the first kind of supplier selection problem. Although the first type of model looks simpler, it requires quite a bit of consideration and communication with the decision-maker, especially in choosing selection criteria, and also in assessing their importance. Moreover, vendor selection includes various types of criteria (quantitative, qualitative, and subjective). Due to this, the appropriate method of vendor selection seems to be Analytic Hierarchy Process (AHP) which in almost every step uses the knowledge and experience of the decision-maker (manager) as it is based on pairwise comparison of all criteria and also on the comparison of each pair of alternatives (vendors) in each of the selected criteria. In any case vendor selection by AHP requires the analyst to have a good knowledge of the problem and also the decision-maker to have a basic knowledge of the analytical hierarchy process and methodology of decision-making by that model. The right choice of vendor, which means the choice of good production equipment (in our case the oven as the basic facility in production of bread and similar products) affects greatly the final result of the production process (product quality and profit).

For these reasons vendor selection problem (or supplier selection as it is often called) is one of the most important tasks in every industry.

2. Analytic Hierarchy Process

The Analytic Hierarchy Process (AHP) is one of the most outstanding multicriteria decision making approaches. It employs a method of multiple paired comparison of attributes (criteria) to rank order alternatives. The attributes themselves are decomposed into levels. The top level contains only one element which reflects the overall objective of the system. The lower levels usually contain a larger number of elements (criteria or subcriteria) which are thought to be independent of the elements at the same level. But these elements directly relate to, or influence, elements at the level below them. At the bottom level there are alternatives which are also compared in pairs to all of the criteria above.
The first step of the AHP approach is the formulation of a problem as a hierarchy.
The next step leads to the determination of the relative weights of the elements at each level. For this a method of multiple paired comparisons based on a standardized evaluation scheme (1 = equally important; 3 = slightly more important: 5 = much more important; 7 = very much more important; 9 = absolutely more important) is used.

The result of the pairwise comparisons of n elements can be summarized in a (nxn) evaluation matrix A in which every element a_{ij} is the quotient of weights of the criteria, e.g. $a_{ij} = w_i/w_j$, whereby small errors in consistency of judgments are acceptable.

In a further step the largest eigenvalue of the evaluation matrix has to be determined. If no errors in judgment exist, the relation $Aw = nw$, or $(A - nI) \cdot w = 0$, holds, where w is the vector of n evaluation weights w_j .

Small errors in judgment lead to small perturbations of the coefficients of the matrix A and its eigenvalues as well. The basic relation for the eigenvalue problem now becomes $A'w' = \lambda_{max} w'$, where λ_{max} is the largest eigenvalue of matrix A'. If the average deviation $(\lambda_{max} - n)/(n - 1)$ exceeds a predetermined value (e.g. 0.1) the evaluation procedure has to be repeated to improve consistency.

The next step leads to a combination of the priority weights of the various hierarchies in order to determine the overall priority weight of an alternative.

These composite weights are the final measure of importance for each alternative considered in the AHP evaluation process. The alternative with the highest total priority weight has therefore to be selected for decision making.

The calculations to be made for AHP studies will usually prove to be fairly complex and they will call for the use of special software packages. In this paper we will use Expert Choice 2000, one of the most valuable programs for analytic hierarchy process.

3. An example of decision system for supplier selection in baker industry

This paper deals with the problem of vendor selection for equipment in baker industry, more precisely the purchase of new ovens for bakery products. Such a decision can have far-reaching consequences on operation efficiency of a business system, not only in the short run but also in the long run, as the consequence of a poor decision can hardly be prevented.

In agreement with the decision-maker criteria for the final vendor selection were divided into three groups that represent the second level of hierarchy. They are: economic criteria, vendor quality criteria, and criteria referring to service and maintenance quality. The first hierarchy level is naturally the basic goal, i.e. vendor selection. At the third hierarchy level these three groups are divided into subcriteria in the following way:

Economic (cost) criteria

1. (C1) Purchasing price in thousands of euros (min)
1. (C2) Total consumption of gas per baking hour in cubic metres (min)
2. (C3) Consumption of electricity per baking hour in kwh (min)
3. (C4) Number of workers serving the ovens (min)
4. (C5) Required floor area in square metres (min)

Vendor quality criteria

5. (C6) Annual average of breakdowns (min)
6. (C7) Probability of oven loading system failure (min)
7. (C8) Guarantee term (max)
8. (C9) References (number of ovens installed) (max)
9. (C10) Duration in years based on daily 8-hour exploitation (max)
10. (C11) Quality of the obtained finished product measured by the subjective evaluation of decision-maker ranging from 5 to 10 (max)

Servicing and failure criteria

11. (C12) Price of annual obligatory service in euros (min)
12. (C13) Price of non-guarantee maintenance visit (min)
13. (C14) Service engineer's wages per hour during the warranty period and after the warranty period (min)
14. (C15) Annual cost in euros of keeping a spare oven in case of breakdown (min)

The final decision matrix, i.e. evaluation of all the vendors in terms of each criterion is shown in Table 1. All the data in that table refer to concrete vendors and only the criterion C11 is subjectively evaluated by the decision-maker.

The subsequent step is pairwise evaluation of importance of the three main groups of criteria. The results of this evaluation and the matrix of mutual comparisons based on Saaty's scale are shown in the Table 2. Obviously, the decision-maker's evaluation is that the first group of criteria is the most important (65.8 %) of the three. Now it is necessary to carry out an analogous evaluation of importance pairwise comparison in terms of each sub-criterion. This requires three more matrices of mutual comparisons of which the Table 3 shows only the first group of criteria. In this table it is obvious that for the decision-maker the most important criterion is the price (as would be expected) with the highest weight of 0.292. The price remains the most important criterion of all, which can be seen also in the Table 1 where the last column shows the final priorities among all the criteria obtained by the AHP.

The next step is evaluation of all the alternatives (5 competing vendors) – again it is pairwise comparison in terms of each of the 15 criteria. Fifteen matrices of mutual comparison are formed, of which only the first one is shown in the Table 4, which shows that in terms of the price criterion the first vendor is the most desirable one (weight 0.445). The final results obtained by the Expert Choice programme can be seen in the Figure 2, which shows the ranking of all the vendors in terms of each criterion separately and in the last column the final ranking of all the alternatives in terms of all the criteria taken together. The final ranking of vendors clearly shows that the first and the second vendor will be the best solution, even though V1 is a bit better than V2.

It is to be noted that evaluations are made quite consistently, which can be seen from the total consistency index, which is 0.04 (Figure 1).

Table 1. Decision matrix

	V1	V2	V3	V4	V5	Measure unit	Criterion type	Criteria priorities
C1	690	890	780	760	720	000 Euro	min	
C2	65	49	52	54	69	m^3	min	
C3	5,5	4,5	4,5	4,5	5,5	kwh	min	
C4	4	3	3	4	5	num.	min	
C5	120	110	100	90	160	m^2	min	
C6	0	2	1	2	0	num.	min	
C7	0.10	0,20	0,30	0,20	0,10	num.	min	
C8	2	1	1	2	3	years	max	
C9	2	10	8	6	6	num.	max	
C10	15	20	16	17	12	years	max	
C11	8	10	9	8	7	attrib.	max	
C12	5000	11000	12000	13000	6000	Euro	min	
C13	3000	170	3800	4500	100	Euro	min	
C14	70	45	70	70	35	Euro	min	
C15	0	20000	20000	20000	0	Euro	min	

Although the corresponding criteria values are expressed quantitatively, it is not recommendable to take the corresponding ratios as priority ratios (for instance for cost of service or price). Namely, the evaluation of advantage given to particular vendors in terms of a single criterion also depends on the attitude of the decision maker who is informed about a number of other elements of the given problem. Therefore it is justifiable to use Saaty's scale with the quantitative criteria as well.

Table 2. Evaluations of criteria weights (second level)

	Cost	Vendors	Service	Priorities
Cost	1	3	5	0.637
Vendors		1	3	0.258
Service			1	0.108

Table 3. Evaluation of criteria weights (third level)

	C1	C2	C3	C4	C5	Priorities
C1	1	3	5	3	7	0.292
C2		1	3	1/3	3	0.091
C3			1	1/4	3	0.050
C4				1	7	0.178
C5					1	0.027

Table 4. Evaluation of alternatives (fourth level)

	V1	V2	V3	V4	V5	Priorities
V1	1	7	5	4	2	0.445
V2		1	1/3	1/4	1/6	0.042
V3			1	1/2	1/4	0.086
V4				1	1/3	0.133
V5					1	0.294

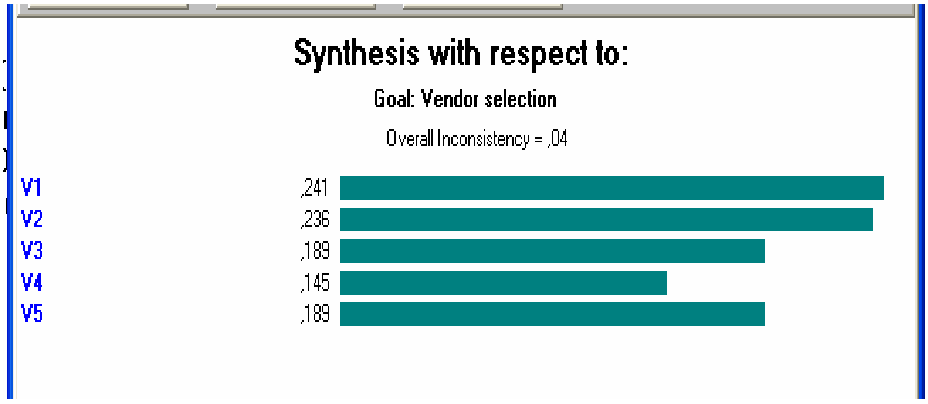


Figure 1. Final ranking of alternatives

Further analysis is performed by sensitiveness analysis (Figure 2) which allows the decision-maker an adequate forecast on what would happen if one criteria group (or a single criterion) changed its weight, i.e. its importance. It is obvious that an increase of importance intensity of the second group of criteria (Figure 3) to approx. 34% (from 26%) results in the change of final ranking, i.e. the vendor V2 takes the leading position. This shows how important it is to evaluate criteria weights in mutual comparison matrices, and it still allows the decision-maker an auto-correction in the final selection of the best vendor.

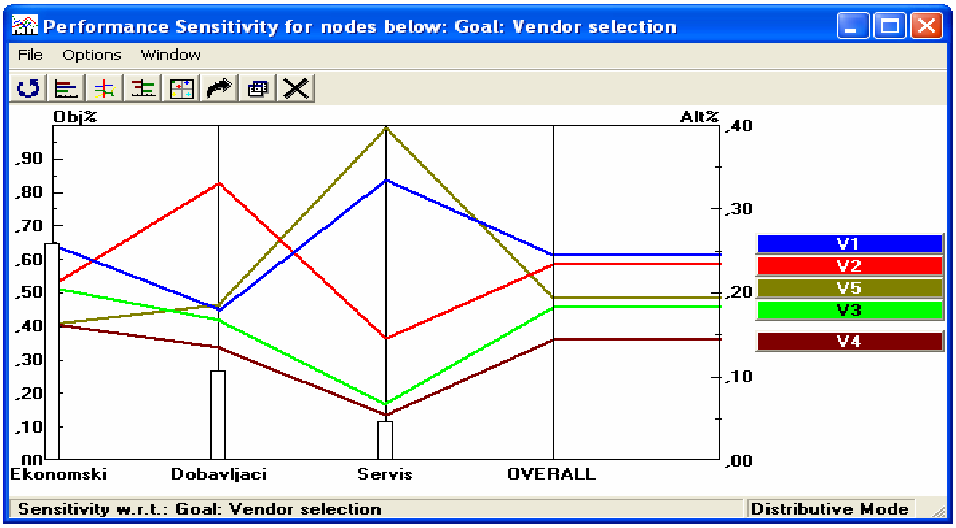


Figure 2. Sensitiveness analysis

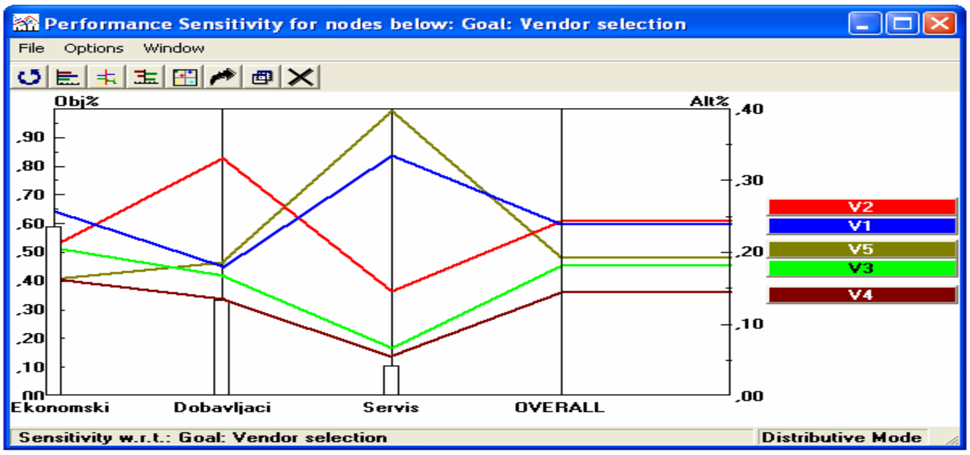


Figure 3. Sensitiveness analysis in alteration of criteria weights (vendors approx. 34%)

4. CONCLUSION

Since orders from external suppliers present a significant item for the majority of firms, supplier selection has a decisive influence upon firm competitiveness. Supplier selection is a long process not only because of many differences that exist among suppliers of the same item but also because of a number of various aims that a customer wants to achieve when selecting a supplier.

This paper presents a supplier selection quantitative model obtained by multicriteria analysis, especially AHP. A developed model can be successfully used in solving similar problems in practice that are dependent on several qualitative and quantitative criteria.

References:

1. Babić, Z., I.Veža (1999): *A Decision System for Supplier Selection in Virtual Enterprise*, Proceedings of the 3rd International Conference Enterprise in Transition, Šibenik, Croatia, p. 451-456.
2. Ghodsypour, S.H., C. O'Brien (1998.): *A Decision Support System for Supplier Selection Using an Integrated AHP and Linear Programming*, in: International Journal of Production Economics, Volume 56-57, Special Issue, pp. 199-212.
3. Jurun, E., Z. Babić, N.T. Plazibat (1999): *Supplier Selection Problem in City of Split Kindergartens*, Proceedings of the 5th International Symposium on Operational Research, Preddvor , Slovenia, p. 99-104.
4. Saaty, T.L.: *Decision Making for Leaders. The Analytic Hierarchy Process for Decision in a Complex World*, RWS Publications, Pitsburgh USA, 2001.
5. Saaty, T.L.: *Theory and Applications of the Analytic Network Process, RWS Publications,* Pitsburgh USA, 2005.
6. Veža.I., Z.Babić (1999): *Supplier Selection in a Virtual Enterprise by the Application of the VSP/CD Method*, Proceedings of the 5th International Scientific Conference on Production Engineering - CIM'99, Opatija, Croatia, pp. 1-10.
7. Weber, C.A.; Current, J.R.; Benton, W.C. (1991): *Vendor Selection Criteria and Methods, in: European Journal of Operational Research*, Vol. 50, No. 1, pp. 2-18.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 6
Location Theory and Transport

HOW DOES EDUCATIONAL POLICY INFLUENCE INTERREGIONAL DAILY COMMUTING OF STUDENTS?

Samo Drobne[*], Marija Bogataj[**] and Ludvik Bogataj[**]
[*] University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia, sdrobne@fgg.uni-lj.si
[**] University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia, {marija,ludvik}.bogataj@ef.uni-lj.si

Abstract
In this paper, we introduce an extended interregional gravity model of daily commuting students in Slovenia. The model is based on our previous investigation on daily commuting persons in employment. But, when we analyse the daily (or weekly) commuting of population in formal education we find out that economic coefficients, significant for daily commuting of working populations, are not significant for population in formal higher education.

Keywords: educational policy, regionalization, gravity model, daily commuting, Slovenia.

1 INTRODUCTION

Colleges and universities play a major role in the system of central places as well as in the regions which they serve. They provide high education and skills to citizens from the region where they are situated and also to the citizens from other regions. Especially for those who live within a certain distance estimated by gravity model. Studies lead to three, four or five-year degrees and allow students to transfer to four or five-year colleges and universities in other regions, having an old or Bolognas programs. The colleges and universities often provide also workforce training to regional businesses and industries, economic development services to both businesses and local (regional) authorities and cultural or sports events.

But, the very often overlooked role is economic one. Colleges and universities produce jobs, and its employees and students consume goods, utilize services, own or rent property, and invest financially in the community. Funds circulate throughout the local economy through college expenditures, purchases of goods and services, salary payments, and capital construction. These funds, in turn, stimulate the local economy, leading to new jobs and additional spendings. In short, the colleges and/or universities have a significant economic impact upon the region they service.

The model used most commonly to measure a college's economic impact was developed by John Caffrey and Herbert Isaacs in 1971 [10]. The model is based on the gravity theory which states that the amount of money spent for non-housing expenditures is inversely proportional to the square of the distance to the point of purchase. As Caffrey and Isaacs note, approximately 35 cents of every dollar spent by community residents in local businesses are returned to the spenders as income. The remaining 65 cents are spent by the businesses for supplies and services from other businesses locally, state-wide, and nationally. A portion of this, again, is spent on additional supplies and services, and this cycle continues, with diminishing returns each time, until eventually the income received by local residents from the initial dollar spent totals approximately 66 cents. The ratio of the total income, 66 cents, to the initial income received, 35 cents, is typically almost two to one, so if a college has a direct economic impact of, say, 1 Mio EUR, the indirect economic impact, using the multiplier of two, would be 2 Mio EUR.

Using the gravity model, we can estimate the economic impact in general and estimate the potential of Slovenian central places to attract the students which will influence urban and regional growth on behalf of other regions.

For anyone attempting to analyze the general process of regional change, an understanding of interregional migration and daily commuting is vital [1]. Cadwallader [9] has pointed out, that policy-makers have become increasingly aware of the role of migrations. These are migration of human resources for any production or services and migrations related to other socio-economic issues, especially as regional growth.

The growth of regions relates closely to population growth, which is mostly a result of migrations and daily commuting. The migration between regions can be slowed down by daily commuting, which is becoming a surrogate for migration, if the commuting is bringing higher social well-being. If the contacts between regions, because of improved transportation abilities and removed barriers, are becoming less expensive and easier, the inhabitants often prefer daily commuting [14].

The gravity models, which belong to the family of spatial interaction models, offer a framework for building integrated models of land use compared to the econometric models [8]. The Lowry model, designed in 1964 for the Pittsburgh metropolitan region [13] and revised several times later, is the basic in this group of models. However, Lowry-like models miss many other aspects of integration; the impact of the transportation network on the land use. This we emphasized in [6], where we calculated a Lowry-like interregional model of daily commuting for persons in employment. The model was then improved to analyse daily commuters between Slovenian and Croatian border's regions [12].

However, in our previous investigations [2,3,4,5,6,7,11], we proved that daily commuting also has an important role in the context of a socio-economic issue in Slovenia. In [6], we summarized all results. We investigated the main factors of interregional migratory and daily commuting flows of human resources in Slovenia, previously predicted in above mentioned papers. In this paper we analysed the daily or weekly commuting of population in formal education. We discuss the significance of socio-economic coefficients for commuting students.

2 THE METHODOLOGY

In this study, the gravity model is extended with coefficients indicated higher educational services. Data on Slovenian population in formal education as well as the number of external daily commuters in population in formal education were obtained from statistical data (Census 2002 and Statistical news [15]). Table 1 and Figure 1 show the number and flows of population in formal education – commuters by and between statistical regions of Republic of Slovenia in 2002.

The simple gravity model helps us to determine the expected number of daily commuters which originate in municipality i and terminate in municipality j. In our previous analysis of daily commuting of persons in employment $DC_{i,j}^{(emp)}$ from region i with population P_i to region j with population P_j, we analysed the following model [6,12]:

$$DC_{i,j}^{(emp)} = aP_i^{\alpha_i} P_j^{\alpha_j} d(t)_{i,j}^{\beta} K_{GDP,i}^{\gamma_1} K_{GDP,j}^{\gamma_2} K_{GEAR,i}^{\gamma_3} K_{GEAR,j}^{\gamma_4} K_{EMP,i}^{\gamma_5} K_{EMP,j}^{\gamma_6} K_{UEMP,i}^{\gamma_7} K_{UEMP,j}^{\gamma_8} \qquad (1)$$

for

$$K_{GDP,\circ} = \frac{GDP(\circ)}{GDP(SI)},\ K_{GEAR,\circ} = \frac{GEAR(\circ)}{GEAR(SI)},\ K_{EMP,\circ} = \frac{EMP(\circ)}{EMP(SI)}, K_{UEMP,\circ} = \frac{UEMP(\circ)}{UEMP(SI)}, \qquad (2)$$

where $(\circ)$ denotes region of origin i or region of destination j ($i = 1,2,...,12; j = 1,2,...,12$), GDP is Gross Domestic Product per capita in region, and in Slovenia (SI) respectively, $GEAR$ is an average gross earning per person in region, and in Slovenia (SI) respectively,

EMP is the number of persons employed in region, and in Slovenia (SI) respectively, and $UEMP$ is the level of registered unemployment in region divided by the level of registered unemployment in the country, and $UEMP(SI)$ is the level of registered unemployment in Slovenia. The model was tested for d being Euclidian distance, the shortest road distance, as well as for the quickest time-spending distances. However, the best results gave the quickest time-spending distances.

Table 1: The population in formal education – commuters by statistical regions of the Republic of Slovenia in 2002

	Region of destination												
Region of origin	Together	Pomurska	Podravska	Koroška	Savinjska	Zasavska	Spodnjeposavska	Jugovzhodna Slovenija	Osrednjeslovenska	Gorenjska	Notranjsko-kraška	Goriška	Obalno-kraška
SLOVENIJA	**21861**	**11901**	**38972**	**6615**	**27122**	**2555**	**6932**	**16166**	**59904**	**22414**	**3947**	**10973**	**11111**
Pomurska	**14102**	11542	2107	<3	75	-	6	<3	320	23	-	<3	25
Podravska	**32749**	338	30374	13	638	<3	8	<3	1101	110	74	10	71
Koroška	**8650**	5	1336	6255	588	<3	<3	6	360	53	7	20	17
Savinjska	**30459**	6	3316	336	24505	125	184	34	1675	180	6	15	77
Zasavska	**4917**	<3	192	<3	517	2256	32	7	1793	77	<3	6	22
Spodnjeposavska	**9209**	-	273	-	462	35	6469	1283	641	28	<3	<3	14
Jugovzhodna	**17964**	-	244	-	26	<3	143	14204	3109	142	41	<3	44
Osrednjeslovenska	**42579**	5	588	3	212	134	58	490	38305	2285	185	75	239
Gorenjska	**28169**	-	232	3	33	-	26	61	8371	19171	86	100	86
Notranjsko - kraška	**5713**	<3	48	<3	12	<3	-	47	1728	57	3174	65	576
Goriška	**13115**	-	99	<3	28	-	<3	16	1499	249	160	10349	709
Obalno-kraška	**10986**	-	163	-	26	-	-	6	1002	39	198	321	9231

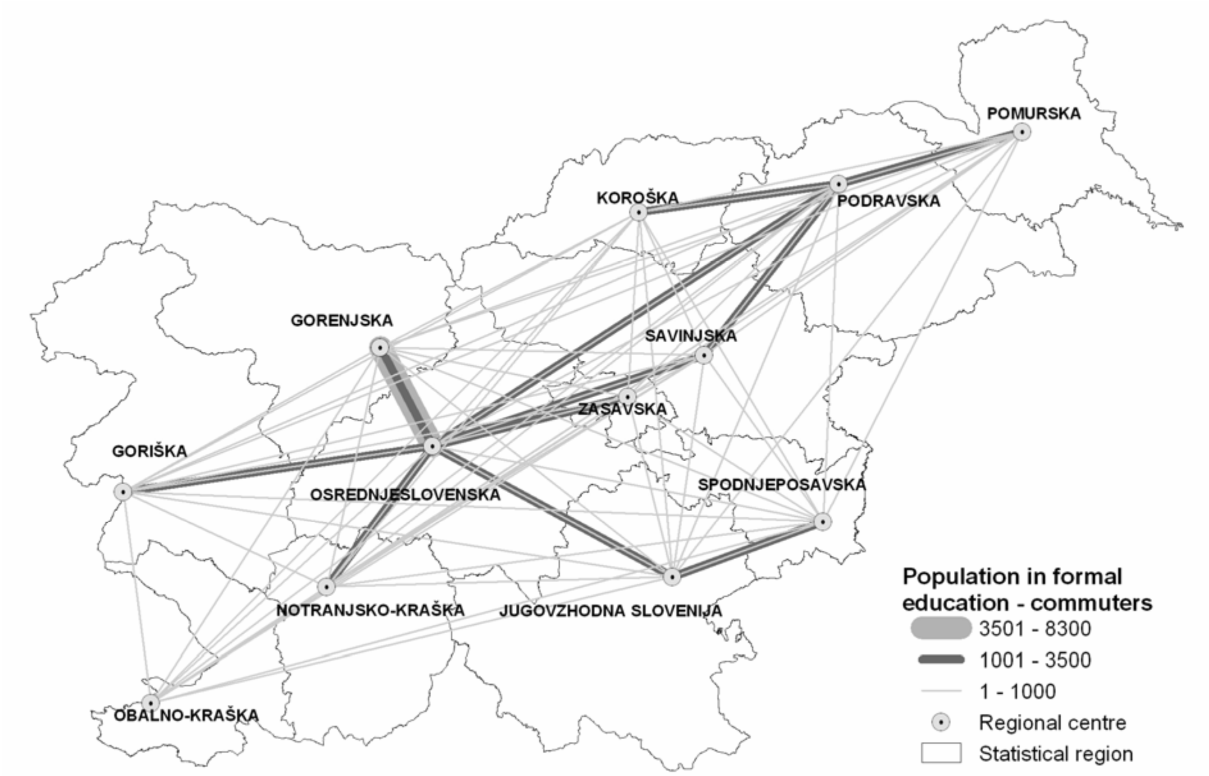


Figure 1: Interregional flows of population in formal education of the Republic of Slovenia in 2002

To analyse commuting flows for population in formal education between statistical regions in Slovenia, we extended the model (1) with coefficients C_i, C_j, U_i and U_j. Here the logarithms of coefficients C are equal to 1, if in region i or j there was at least one three-year college or single faculty in 2002, and 0 of there was no full-time college or single faculty education. In the same way, we introduce the coefficients U to describe if there is university in the region having more than one faculty educations (university study).

In the regression analysis of daily interregional commuting students, the following model was investigated:

$$DC_{i,j}^{(stud)} = aP_i^{\alpha_i} P_j^{\alpha_j} d(t)_{i,j}^{\beta} K_{GDP,i}^{\gamma_1} K_{GDP,j}^{\gamma_2} K_{GEAR,i}^{\gamma_3} K_{GEAR,j}^{\gamma_4} K_{EMP,i}^{\gamma_5} K_{EMP,j}^{\gamma_6} K_{UEMP,i}^{\gamma_7} K_{UEMP,j}^{\gamma_8} C_i^{\gamma_9} C_j^{\gamma_{10}} U_i^{\gamma_{11}} U_j^{\gamma_{12}} \tag{3}$$

But only the time-spending distance and following coefficients gave the results where P – value is smaller than 0.15:

$$DC_{i,j}^{(stud)} = aP_i^{\alpha_i} P_j^{\alpha_j} d(t)_{i,j}^{\beta} C_j^{\gamma_{10}} U_j^{\gamma_{12}} \tag{4}$$

3 THE RESULTS

We got the regression parameters for the interregional daily commuting flow equation, for 132 observations and where $d(t)$ is the road time-spending distance; see Table 2.

Table 2: Extended gravity model's coefficients and summary output for interregional daily commuting ($DC_{i,j}$) – students in 2002.

Name of Coefficients	*Value of Coefficients in (5)*	*Standard error*	*t Stat*	*P-value*
$Ln(a)$	-7.5069	4.9865	-1.5054	0.1348
α_i	0.6895	0.2822	2.4438	0.0159
α_j	1.3473	0.2820	4.7776	0.0000
β	-3.0343	0.2617	-11.5961	0.0000
γ_{10}	1.5389	0.3116	4.9382	0.0000
γ_{12}	0.7369	0.4845	1.5211	0.1308
Multiple R=0.87				

The interregional regression model of commuting students is then:

$$DC_{i,j}^{(stud)} = \frac{0.00058 P_i^{0.69} P_j^{1.35}}{d(t)_{i,j}^{3.03}} C_j^{1.54} U_j^{0.74} \tag{5}$$

4 DISCUSSION AND CONCLUSIONS

Using data from Census 2002, the results of regression analysis (5) show that in case of opening colleges in regional centres the number of population in formal education, which daily commute from other regions, would increase in average by coefficient 4.7 and in case that university would be opened in a region additionally the number of daily commuters from other regions would increase by coefficient 2.1.

In all this cases the reduction of flows by distance is higher than for workers; in [12] we calculated the interregional regression model of commuting population in employment:

$$DC_{i,j}^{(emp)} = \frac{2.13 \cdot 10^{-5} P_i^{0.95} P_j^{1.28}}{d(t)_{i,j}^{2.35}} K_{GEAR,j}^{5.48}$$

So, if time distance increase from 30 to 60 minutes, the percentage of daily commuters fall for nearly 90 %.

Therefore we can expect increase of full time students when colleges and universities will be opened in regional centres of Slovenia.

However, these results require further study of impact of the newly built Centres of higher education on increase of highly educated population in Slovenia. At the same time we need to follow Caffrey's and Isaac's study, estimating the impact of a college or university on the local economy to forecast the regional growth after dispersion of higher education in Slovenia.

References

[1] Anjomani, A., 2002: Regional growth and interstate migration, *Social-Economic Planning Science*, (36): 239–265.

[2] Bogataj, L., Bogataj, M., Drobne, S., Vodopivec, R., 2003: Management of investments in roads and in capacities of border regime to induce the flow of human resources in and out of region. In: Zadnik Stirn L., M. Bastič and S. Drobne (ed.), SOR'03 proceedings. International Symposium on Operational Research, Podčetrtek, 43–46.

[3] Bogataj, L., Bogataj, M., Drobne, S., Vodopivec, R., 2004: The influence of investments in roads and border crossing capacities on regional development after accesion. *Suvremeni promet*, 24(5/6): 379-387.

[4] Bogataj, L., Bogataj, M., Drobne, S., Vodopivec, R., 2006: Global business and economic developement management influenced by the investments in European corridors - the case of Slovenia. *Ekon. teme*, 44(1/2): 11-22.

[5] Bogataj, M., Drobne S., 1997: The influence of investments in highways on gravity and interaction in Slovenia. In: Rupnik, V., L. Zadnik Stirn and S. Drobne (ed.), SOR'97 proceedings, International Symposium on Operational Research, Preddvor, 55–60.

[6] Bogataj, M., Drobne S., 2005: Does the improvement of roads increase the daily commuting? Numerical analysis of Slovenian interregional flows. In: Zadnik Stirn, L., Indihar Štemberger, M., Ferbar, L., Drobne, S., *Selected Decision Support Models for Production and Public Policy Problems*, Slovenian Society Informatika, Ljubljana, 185-206.

[7] Bogataj, M., Drobne S., Bogataj L., 1995: The influence of investment and fiscal policy on growth of spatial structure, *Suvremeni promet*, Zagreb, 15(5):239–245.

[8] Briassoulis, H., 2000: *Analysis of Land Use Change: Theoretical and Modeling Approaches*. Regional Research Institute, West Virginia University.

[9] Cadwallader, M., 1992: *Migration and residential mobility: macro and micro approaches*. University of Wisconsin Press, Wisconsin.

[10] Caffrey, J., Isaacs, H. H., 1971: *Estimating the Impact of a College or University on the Local Economy*, Washington, D.C.: American Council of Education.

[11] Drobne, S., Bogataj, M., 2005: Intermunicipal gravity model of Slovenia. In: Zadnik Stirn, L. (ed.), Drobne, S. (ed.). SOR'05 proceedings. Ljubljana: Slovenian Society Informatika (SDI), Section for Operational Research (SOR), 207-212.

[12] Drobne, S., Bogataj, M., Bogataj, L., 2007: *Spatal interactions influenced by European corridors and the shift of the Schengen border regime*, KOI 2006 proceedings, Zagreb: Croatian Operational Research Society (CRORS), in print.

[13] Lowry, I. S., 1966, *Migration and metropolitan growth: two analytical models*. Chandler Publishing Company, San Francisco.

[14] Nijkamp, P., 1987: *Handbook of Regional and Urban Economics*. Vol. 1, Regional Economics, North – Holland.

[15] SURS 2005: Statistical Office of the Republic of Slovenia, URL: http://www.stat.si/eng/index.asp, date accessed: 20-June-2007.

ON OPTIMAL ORDERING AND TRANSPORTATION POLICIES IN A SINGLE-DEPOT, MULTI-RETAILER SYSTEM

Peter Köchel
Chemnitz University of Technology, Department of Informatics, D-09107 Chemnitz
pko@informatik.tu-chemnitz.de

Abstract: In the considered system two interdependent decision problems must be solved – the release of orders by the retailers and the allocation of a finite number T of transportation units to the orders by the depot. The retailers, which are faced with a random demand, follow an (s, nQ) ordering policy. To guarantee the stability of the system additional transportation units from outside the system can be rented. Thus a third decision problem – when to rent and how much – arises. For all three problems we consider simple algorithms, which are demonstrated by some numerical examples.

Keywords: single-depot, multi-retailer system, order policies, allocation of transporting resources

1. Introduction

The optimal design and control of logistic as well as inventory systems in the multi-location setting is an actual research field with growing importance both for practice and theory. Nevertheless, in the past were not so much publications dealing with combined logistic-inventory systems and models. The most important reason for that is the extreme complexity of such models. There are at least two approaches to cope with complexity – simulation and to put some structure into the system. Here we will follow the second way and to concentrate on systems, which in logistics are called hub-and-spoke systems and in inventory theory two-echelon systems. We assume a single depot as the hub and multiple retailers as the spokes. The central depot owns a fleet of trucks. These trucks are used to transport a single product to the retailers, which are faced with a random demand for that product. The decision problem is to find for all retailers such ordering policies and for the depot such a fleet size and allocation of trucks to retailer orders that optimise a given criterion function.

Special variants of the just formulated decision problem are investigated both in logistic and in inventory theory. For instance, in the fleet-sizing-and-allocation problem a central manager looks for an optimal fleet size and an optimal reallocation for empty fleet units. Instead of fleet unit we use the more general notion transportation unit (TU). Mostly the demand for a TU is assumed to be deterministic or known (see e.g. [5]). Some papers [1], [3] consider stochastic demand. But in most models the demand for TUs is directly given and not generated by the demand of customers at some spokes. The other decision problem to define optimal ordering policies for the retailers is well investigated in inventory theory (see [2]). However, all echelon models usually assume infinite transportation capacities.

In Chapter 2 we investigate a discrete time, multi location model. With respect to the ordering policies we concentrate on (s, nQ) policies, which are introduced in Chapter 3. We are dealing with the single location model and use results from [4]. Next, we develop an algorithm that calculates the optimal reorder point s^* for the (s, nQ) policy. An example demonstrates the algorithm. In Chapter 4 we translate the results for the single location model to the multi location model. The paper is finished with a brief summary.

2. Modelling of the decision problem

We construct now a single product, periodic review model with infinite planning horizon and stochastic demand. To this end we assume the following:

(1) There are M+1 locations, where location 0 represents a single central depot, the hub, and locations 1 to M the retailers, the spokes.
(2) The infinite planning horizon is divided into periods t, $t \in \mathbf{N} = \{1, 2, \ldots\}$.
(3) The hub owns an ample amount of a single product and a set of T homogeneous TUs. The capacity Q of a single TU is an integer multiple of product units.
(4) Each retailer can order product at the beginning of a period, whereby the order size must be an integer multiple of Q.
(5) The transport of product by TUs from the depot to a retailer needs a negligible time.
(6) Let $D_t = (D_{t1}, \ldots, D_{tM})$ denote the demand vector and D_{ti} the demand at retailer i in period $t \in \mathbf{N}$, $i = 1,\ldots, M$. We assume that $D_1, D_2, \ldots$ forms a sequence of independent, identical distributed (iid) random vectors with independent discrete components D_{ti} with distribution function $F_i(d) = \mathbf{P}(D_{ti} \le d)$, $d = 0, 1, \ldots$, and $E(D_{ti}) = \mu_i < \infty$, $t \in \mathbf{N}$, $i = 1,\ldots, M$.
(7) Demand unsatisfied in one period will be backordered.
(8) Cost arise for retailers only - ordering and transportation cost $K_i > 0$ for delivering a full TU to retailer i, and at the end of a period cost $h_i > 0$ for holding a product unit and $p_i > 0$ for shortage of a product unit, $i = 1,\ldots, M$.

Let x_t and y_t be the vectors of inventory positions (stock on hand plus stock on order minus backlogs) at the beginning of period t *before* respectively *after* ordering, $t \in \mathbf{N}$. Furthermore, let $n_{ti} \in \mathbf{N}_0 = \{0, 1, \ldots\}$ denote the number of quantities Q ordered at the beginning of period $t \in \mathbf{N}$ by retailer i, and $n_t = (n_{t1},\ldots, n_{tM})$ the *batch ordering vector*. Obviously, the set $\mathbf{N}_0(T)$ of admissible batch ordering vectors is given as

$$\mathbf{N}_0(T) = \left\{ n = (n_1,\ldots,n_M) : \sum_{i=1}^{M} n_i \le T, n_i \in N_0 \right\}. \quad (1)$$

For given x_t and $n_t \in \mathbf{N}_0(T)$ the expected total cost for period t can be expressed by

$$c_t(x_t, n_t) = \sum_{i=1}^{M} c_{ti}(x_{ti}, n_{ti}) = \sum_{i=1}^{M} \left[n_{ti} \cdot K_i + L_i(x_{ti} + n_{ti} \cdot Q) \right]. \quad (2)$$

Function L_i in (2) represents the expected holding and shortage cost for retailer i after ordering, i.e., $L_i(y_{ti}) = \mathbf{E}[h_i \cdot \max(y_{ti} - D_{ti}, 0) + p_i \cdot \max(D_{ti} - y_{ti}, 0)]$. Because of the assumption of discrete demand it follows for $y \in \mathbf{I} = \{0, \pm 1, \pm 2, \ldots\}$ that

$$L_i(y) = \begin{cases} (h_i + p_i) \sum_{d=0}^{y-1} F_i(d) + p_i \cdot (\mu_i - y), y > 0; \\ p_i \cdot (\mu_i - y), y \le 0. \end{cases} \quad (3)$$

It remains to introduce the notion of a policy and to choose a criterion function. Without going into detail a policy can be understood as a sequence $\pi = \{\pi_t, t \in \mathbf{N}\}$, where π_t defines a rule that chooses an admissible batch ordering vector n_t for period $t \in \mathbf{N}$. In dynamic programming with infinite planning horizon problems two criterions are common – the discounted cost criterion and the long-run average cost criterion (cp. [6]). We use here the latter. To explain this criterion we abstract from the concrete form of the cost function for period t and imagine that $\{c_t, t \in \mathbf{N}\}$ represents the sequence of incurred cost. Then the long-run average cost is defined as $C = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} c_t$. It is obvious that for our problem the long-run average cost is a function of Q, T, and the applied policy π, i.e., $C = C(\pi, Q, T)$. A policy π^* is average-optimal if for any policy $\pi \ne \pi^*$ holds $C(\pi^*, Q, T) \le C(\pi, Q, T)$. Now we can formulate our general decision problem as the problem to calculate for given Q and T

1. an average-optimal policy π^*, and
2. the minimal long-run average cost $C^*(Q, T) = C(\pi^*, Q, T)$.

We remark that the single period cost function c_t from (2) is separable. Thus we may think to consider each location independently and to reduce the decision problem to M classical newsboy problems, which are broadly investigated. However, we have two differences – we can not separate the locations because of they are coupled through the condition that n_t must be from the set $\mathbf{N}_0(T)$, and the order sizes must be integer multiples of quantity Q. We must say that up to now we do not have results neither on the optimal policy nor the minimal cost. Therefore in the subsequent chapters we will look for approximate solutions.

3. The single location model

Through the single retailer problem we hope to find candidates for good approximate solutions for $M > 1$. Thereby, to simplify matters, we omit the indexation of parameters and variables, i.e. we write $L(y)$ and $c(x, n)$ and so on. We start with a definition.

Definition 3.1. Let $s \in \mathbf{I}$ and $Q > 0$ given constants. If in each period the batch ordering number (we have $M = 1$) is chosen in accordance with the rule

$$n(x) = \begin{cases} 0 & , x > s; \\ min\{ n : x + n \cdot Q > s \} & , x \le s, \end{cases}$$

then such an ordering policy is called an (s, nQ) policy, where s is the reorder point and Q the base quantity.

Policies of the (s, nQ) type play an important role in inventory theory. First, they are optimal in several situations (cp. [4], [6]), and second, they are easy to implement. Since in [4] is shown the optimality of such a policy for the single-period case it makes sense in the infinite period model to concentrate on the class of stationary (s, nQ) policies. We have to answer two questions: Which is the optimal (s, nQ) policy in the infinite-period model? Which long-run average cost this policy will generate and how far is this cost from the optimal one?

To answer these questions let us briefly consider the *single-period model*. It is easy to show that $L(x+ \cdot Q)$ and $c(x, \cdot)$ are integer-convex functions of n for $\forall x \in \mathbf{R}$ and $\forall Q > 0$ (see [4]). To exclude the trivial case that a retailer will get no TUs in [4] is introduced

Assumption POI (Positive Optimal Inventory)**:** $\boldsymbol{Q \cdot p > K}$.

This assumption means that the cost for a shortage of one lot size Q is higher than the cost for delivering that lot size to the retailer. Obviously, from assumption POI follows $Q > K / p$, i.e., POI defines a lower bound for lot size Q . The main results for the single-period model we summarize in

Theorem 3.1. (cp. [4])

Let assumption POI be fulfilled for the single-period, single-location model. Then:

(I) The optimal order policy is an (s, nQ) policy with optimal reorder point $s^{(1)}$ as

$$s^{(1)} = min\left\{ s \in I : \sum_{d=s+1}^{s+Q} F(d) \ge \frac{Q \cdot p - K}{h + p} \right\}. \tag{4}$$

(II) For the minimum expected single-period cost holds $c(x) = n^{(T)}(x) \cdot K + L(x + n^{(T)}(x) \cdot Q)$, where $n^{(T)}(x) = \min(n^*(x); T)$ and $n^*(x)$ is defined through

$$n^*(x) = min\left\{ n \in N : \sum_{d=x+nQ}^{x+(n+1)Q-1} F(d) \ge \frac{Q \cdot p - K}{h + p} \right\}, \ x \in \mathbf{I}. \tag{5}$$

(III) The inventory level S^* that minimises function L is equal to

$$S^* = min\left\{ S \in N : F(S) \ge \frac{p}{h + p} \right\}. \tag{6}$$

We remark that $S^* > 0$ if and only if $F(0)\cdot(h+p) < p$, a condition that usually is fulfilled. In contrast to this $s^{(1)}$ must not be positive. However, $s^{(1)} +Q$ is positive under assumption POI.

Returning now to the infinite period case we have the problem that as the consequence of the finite number T of available TUs the inventory position before ordering can go to minus infinity. This means that steady-state regime may not exist and that we cannot apply in general the long-run average cost criterion. To prevent the drift of the inventory position away to minus infinity Köchel [4] quotes three variants – limit the number of backorders, define suitable conditions on the demand variables, or allow to rent additional TUs from outside the system. Since the first two variants are not easy to handle analytically and the third one is the most realistic one we follow [4] and introduce the following

Rental Assumption (RA)**:** In each period additional TUs can be rented for cost R with $R>K$.

Let $C(s, Q, T)$ denote the long-run average cost for fixed (s, nQ) policy and given number T. Köchel [4] has shown that

$$C(s,Q,T) = \frac{1}{Q}\cdot\left[\mu\cdot K + (R-K)\cdot\sum_{d=TQ}^{\infty}\overline{F}(d) + \sum_{y=s+1}^{s+Q} L(y)\right], \qquad (7)$$

where $\overline{F}(d)=1-F(d)$, $d \geq 0$. Formula (7) gives a partial answer to the second question formulated at the beginning of the chapter. To answer the first question we have to find that reorder point s^*, which minimises $C(s, Q, T)$, i.e., for which holds $C(s^*, Q, T) \leq C(s, Q, T)$ for any $s \in \mathbf{I}$. From (7) follows that s^* is minimising for

$$G(s) := \sum_{y=s+1}^{s+Q} L(y). \qquad (8)$$

Since $L(\cdot)$ is an integer-convex function (see [4]) function $G(\cdot)$ is also integer-convex in s. From the optimality conditions $G(s^*) \leq G(s^*\pm 1)$ it easily follows that

$$s^* = min\left\{ s \in I : \sum_{d=s+1}^{s+Q} F(d) \geq \frac{Q\cdot p}{h+p}\right\}. \qquad (9)$$

Condition (9) means that we need those Q consecutive values $F(s+1)$ to $F(s+Q)$ of the demand distribution, whose sum is the first time not smaller than $p\cdot Q/(h+p)$. From (6) and (9) and the non-decreasing property of distribution functions it follows by contradiction that

$$S^*- Q \leq s^* < S^*. \qquad (10)$$

Thus we start the search for s^* at $S^*- Q$ and use the following algorithm

Algorithm *Optimal reorder point* **ORP** {which calculates s^* from (9)}.

```
        Input:      S*, Q, p, h, {F(d), d∈N}, F(-Q) = …= F(-1) = 0;
        Output:     s*;
BEGIN
        s:= S*-Q;
        limit:= p · Q / (h + p);
        sum := F(s+1) + F(s+2) + … + F(s+Q);
        WHILE (sum < limit) DO
                BEGIN
                        sum := sum + F(s+Q+1) – F(s+1);
                        s:= s + 1
                END;
        s*:= s
END.
```

Let us remark that a similar approach is applicable also in case if we do not have linear cost functions as defined in assumption (8) in Chapter 2. We need only the quasi-convexity of function $L(\cdot)$. If that property is fulfilled the Q smallest consecutive $L(y)$-values can be defined starting with the point of the minimum and adding from the left and right neighbour points that one, which gives the smaller value of function L. After collecting Q points the process is stopped. However, here the computational effort may be much higher than for the above-formulated algorithm because of it is necessary to calculate the values $L(y)$.

We finish the present chapter with the simple **Example 3.1.** Let be Q =10, h =1, and p = 4. For the demand we assume a binomial distribution with parameters n = 20 and q = 0.5, i.e.,

$$F(d) = \sum_{k=0}^{d} \binom{n}{k} \cdot q^k \cdot (1-q)^{n-k} = \sum_{k=0}^{d} \binom{20}{k} \cdot 0.5^{20}, d = 0,1,...,20.$$

Table 1 contains for that distribution function the rounded off to four digits values.

d	1	2	3	4	5	6	7	8	9
F(d)	.0000	.0002	.0013	.0059	.0207	.0577	.1316	.2517	.4119
d	10	11	12	13	14	15	16	17	18
F(d)	.5881	.7483	.8684	.9423	.9793	9941	.9987	.9998	1.0000

Table 1. Values for the distribution function of the binomial distribution from Example 3.1.

Since $p/(h+p)$ = 0.8 we get from (6) and Table 1 that S^* = 12. Thus we start algorithm ORP with $s = S^*-Q$ = 12–10 = 2 and calculate from Table 1 the sum $F(3) + \ldots + F(12)$ as 3.0856, which is smaller than $Q\cdot p/(h+p)$ = 8. Next we take s = 3 and calculate $F(4) + \ldots + F(13)$ = 4.0266 < 8. We continue the same procedure until s = 8. Since $F(9) + \ldots + F(18)$ = 8.5309 the search process stops and the algorithm returns s^* = 8 as the optimal one.

As another demand distribution we take a discrete variant of the exponential distribution

$$\mathbf{P}(\text{demand} = d) = \int_{d}^{d+1} \lambda \cdot e^{-\lambda x} dx = e^{-\lambda d}(1-e^{-\lambda}), \; d = 0, 1, \ldots,$$

where $\lambda > 0$ is the parameter of the distribution. The expected demand is $1 / (e^{-\lambda} -1)$, which in case λ = 0.1 gives 9.508. From (6) we get S^*=16. Applying the algorithm ORP we can calculate for given Q the corresponding s^*. Some s^* values for different Q are given below:

Q	5	10	20	40	45	50
s^*	13	11	7	1	0	-1

For λ = 0.01 we get 99.5 as average demand, S^* = 160, and following results:

Q	10	50	100	200	300	400	450	480	490	495	500	600
s^*	155	136	114	76	45	19	8	2	0	-1	-2	-80

4. The multi-location model

It is obvious that the results of Chapter 3 hold for an arbitrary location. Especially, if all retailers will follow an (s, nQ) policy with reorder points $\mathbf{s} = (s_1, \ldots, s_M)$ then in analogy to formula (7) we get for the total long-run average cost $C(\mathbf{s}, Q, \mathbf{t})$ that

$$C(\mathbf{s},Q,\mathbf{t}) = \sum_{i=1}^{M} C_i(s_i,Q,t_i) = \frac{1}{Q} \cdot \sum_{i=1}^{M} \left[\mu_i \cdot K_i + (R_i - K_i) \cdot \sum_{d=t_i \cdot Q}^{\infty} \overline{F}_i(d) + \sum_{y=s_i+1}^{s_i+Q} L_i(y) \right], \quad (11)$$

where $\mathbf{t} = (t_1, \ldots, t_M) \in \mathbf{N}_0(T)$ denotes an allocation of the T TUs and $C_i(s_i, Q, t_i)$ the long-run average cost at retailer i under fixed (s_i, nQ) policy with t_i allocated TUs. But in the multi-location model we have two new problems. First, it is not clear how to allocate the T TUs to the M retailers. And second, formula (11) holds for a static allocation, i.e., the TUs are once for all allocated. A static allocation can not be optimal in the infinite horizon problem because of sufficiently often will occur a situation where one location does not use all allocated TUs whereas another location has to rent additional TUs. Thus formula (11) defines an upper bound for the minimal expected long-run average cost. Therefore we briefly investigate the optimal static allocation. Goal function, defined in (11), is separable with respect to the components t_i of the allocation vector $\mathbf{t}$. From (11) follows also that

$$\Delta C_i(t_i) := C_i(s_i, Q, t_i + 1) - C_i(s_i, Q, t_i) = -\frac{R_i - K_i}{Q} \sum_{d=t_i Q}^{(t_i+1)Q} \overline{F_i}(d) < 0 \qquad (12)$$

and

$$\Delta^2 C_i(t_i) := \Delta C_i(t_i + 1) - \Delta C_i(t_i) = \frac{R_i - K_i}{Q} \left[\sum_{d=(t_i+1)Q}^{(t_i+2)Q-1} F(d) - \sum_{d=t_i Q}^{(t_i+1)Q-1} F(d) \right] \geq 0,$$

i.e., function C_i is a decreasing and integer-convex function of t_i for arbitrary s_i and Q, $i = 1, \ldots, M$. These properties allow applying Marginal Analysis (see [4]), which means that the optimal static allocation $\mathbf{t^*}$ can be found by the

Algorithm *Optimal static allocation* **OSA**.
Input: T, $\{s_i^*, i = 1, \ldots, M\}$;
Output: $\mathbf{t^*}$;
BEGIN
$\mathbf{t}^{(0)} := (0, 0, \ldots, 0)$;
FOR k:=1 TO T DO $\mathbf{t}^{(k)} := \mathbf{t}^{(k-1)} + \mathbf{e}_i$;
$\{\mathbf{e}_i = (0,\ldots,0, 1, 0, \ldots, 0)$; i is that index, which minimises $\Delta C_i(t_i)$ from (12)$\}$
END.

Obviously the static allocation will be outperformed by a dynamic allocation, where the T TUs are newly allocated in each period. Which will be the optimal allocation in that case is not clear up to now. Therefore we apply the following policy: Chose in a given period that allocation, which minimises the expected cost for that period. We call such a policy the *myopic allocation policy*. To investigate the myopic allocation policy let

$$c_{ti}(x_i, t_i, r_i) = t_i \cdot K_i + r_i \cdot R_i + L_i(x_i + (t_i + r_i) \cdot Q) \qquad (13)$$

denote the expected cost at period t for location i, if the inventory position is x_i and if t_i own TUs and r_i rented TUs are allocated, $i = 1, \ldots, M$. Assume now that in a given period the allocation is such that there are two locations i and j with $r_i > 0$ and $t_j > 0$. Finally, let us assume that $R_i - K_i > R_j - K_j$. The last inequality suggests that a re-allocation of own TUs from location j to location i will decrease cost. If $m = \min(t_j, r_i)$ then the allocation $r_i' = r_i - m$, $t_i' = t_i + m$, $r_j' = r_j + m$, $t_j' = t_j - m$ and all other locations unchanged leads to a cost degree of $m \cdot (R_i - K_i + K_j - R_j) > 0$ (cp. Lemma 4.1 in [4]). Thus we get a very simple algorithm to find the optimal myopic allocation: The allocation of the own TUs starts with the location, which has the biggest cost difference between rented and own TUs. If all own TUs are allocated rented TUs must be taken.

Algorithm *Optimal myopic allocation* **OMA** {for given period $t \in \mathbf{N}$}.
Precondition: The M locations are ordered by decreasing differences $R_i - K_i$.
Input: T, $\{x_{ti}, i = 1, \ldots, M\}$, $\{s_i$ as solutions from (4), $i = 1, \ldots, M\}$.
Output: Optimal myopic allocation vectors $\mathbf{t}_t$ and $\mathbf{r}_t$.

```
BEGIN
    ownTU := T;
    FOR i := 1 TO M DO
    BEGIN
        n_i := min{n : x_ti + n · Q > s_i} ;
        t_i := min(n_i, ownTU);
        ownTU := ownTU – t_i;
        IF t_i < n_i THEN r_i := n_i – t_i;
    END
END.
```

We finish the present chapter with two remarks. First, if we take in the optimal myopic allocation algorithm the reorder points from (9) we get another policy and other costs. And second, because of we have no explicit expressions for the cost expectations to compare all policies we need simulation. We know only that the two myopic policies outperform the static allocation policy. But we don not know the cost differences and we do not know which of the two myopic policies with reorder points from (4) respectively from (9) is the best. In the future we will investigate this with the help of simulation.

5. Conclusion

We have formulated and investigated a complex model to solve two connected problems – the transportation resource allocation problem and the ordering problem from inventory theory. Based on results for the single period model we restrict ourselves to (s, nQ) ordering policies. To avoid instabilities in the infinite period model we introduced the possibility to rent additional resources from outside the system. An algorithm is given to find the optimal reorder point s of the ordering policy as well as a formula for the corresponding long run average cost. For the multi location situation we considered two classes of allocation policies – static and dynamic policies. In the class of dynamic allocation policies we restricted our investigations to myopic solutions, which optimise the cost for each actual period.. For myopic policies we got also a simple allocation algorithm, but no expression for the long run average cost. This can be done by simulation only and will be realised in the future. Another topic for future research is to solve the problem of defining an optimal fleet size T. And finally, it can be allowed that a single TU can deliver more than one location. In the latter case we have to choose in addition for each TU a corresponding route for the delivery of the locations. Simulation seems to be the most promising approach.

References

1. Du, Y.; Hall, R. (1997). Fleet Sizing and Empty Equipment Redistribution for Center-terminal Transportation Networks. *Man. Sci.*, 43, 145157
2. Federgruen, A.(1993). Centralized Planning Models for Multi-Echelon Inventory Systems under Uncertainty. *Handbooks in OR & MS*. (Graves, S.C., Editor), Elsevier Science Publishers B.V., Chapter 3
3. Köchel, P.; Kunze, S.; Nieländer, U. (2003). Optimal Control of a Distributed Service System with Moving Resources: Application to the Fleet Sizing and Allocation Problem. *International Journal of Production Economics*, 81-82, S. 443-459
4. Köchel, P. (2007). Order Optimisation in Multi-Location Models with Hub-and-spoke Structure. *International Journal of Production Economics*, 108, 368-387
5. Powell, W.B.; Carvalho, T.A. (1998). Dynamic Control of Logistics Queueing Networks for Large-Scale Fleet Management. *Transportation Sci.*, 32, 90-109
6. Veinott, A. (1965). The optimal inventory policy for batch ordering. *Operations Research*, 13, 424-432

THE REGIONALISATION OF SLOVENIA: AN EXAMPLE OF ADAPTATION OF POSTS TO REGIONS

Andrej Lisec
University of Maribor, Faculty of Logistics
Hočevarjev trg 1, SI – 8270 Krško, Slovenia
e-mail: andrej.lisec@uni-mb.si

Marija Bogataj
University of Ljubljana, Faculty of Economics
Kardeljeva ploščad 17, SI – 1000 Ljubljana, Slovenia
e-mail: marija.bogataj@guest.arnes.si

Anka Lisec
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
e-mail: anka.lisec@fgg.uni-lj.si

Abstract: This paper deals with the regionalisation and includes the case study how optimal location of regional postal centre coincides with political determined regional central place in Slovenia. We can see that the optimal distribution of parcels, which is based on a hierarchical postal network, requires Regional Parcel Centres mostly coincided with central places of statistical regions of Slovenia (NUTS 3 level).

Keywords: regionalization, logistics, postal services, regional parcel centre, parcel post, hub location problem.

1 INTRODUCTION

Logistic enterprises wish to supply customers in such a way that the positive difference between the revenue of the service and the operating costs will be the highest possible, however they are often limited by obligatory standards or the standards that the competitive companies guarantee. The decentralization of institutions and their activities plays an important role in more effective exploitation of logistics networks. The process of planning, implementing, and controlling the postal service, as an example of the logistic problem, has been faced by the reorganisation of the service in the most Central European regions and can be linked to the process of regionalisation, particularly in the case of Slovenia, where officially recognised regions still do not exist.

Slovenia has no historical tradition of regional government. The division of Slovenia into 12 statistical regions in the past was based on the social-geographic assumptions. They had no strong political or administrative function in the past. For several years the statistical regions have been the spatial units for Slovenian regional statistics, which holds an important function in supporting regional development. Regional statistics, referred to the statistical regions, presents a starting-point for regional policy planning and for measuring the effects of regional development.

Furthermore, Eurostat, the Statistical Office of the European Commission, initiated the Nomenclature of Territorial Units for Statistics (NUTS), which is a geocode standard for referencing the administrative regions of the EU member states for statistical purposes. There are three levels of NUTS defined. The whole territory of Slovenia corresponds to the one region on NUTS 1 level. On the NUTS 2 level, Slovenia is divided into two regions: Western Slovenia and Eastern Slovenia. The division on the NUTS 3 level on 12 regions derives from the statistical regions of the Republic of Slovenia [6].

Spatial hierarchy of postal services is more or less embedded in geographically and politically determined regionalisation which is now the top priority of the Slovenian government for establishing 14 new regions with political and administration functions. Numerous economic, administrative, geographic and other reasons justify the need to divide Slovenia into regions despite the small size of its territory. The fundamental goal of the regionalisation is efficient management with the aim to ensure quality services on the local and regional level. The importance of the regionalisation is obvious from the economic, social, political or administrative point of view.

Following the efforts of the post services in the Central European regions the Post of Slovenia, Ltd. would also like to improve its postal services. In our previous research articles we have presented an approach to the spatial optimization of postal services, particularly as applicable to the Post of Slovenia [2–5]. The Post of Slovenia has two Postal Logistics Centres (PLC). It has been found that having 2 of them coincides to regionalization of Slovenia on NUTS 2 level and is approved to be optimal if the total flow volume of parcels is high enough. According to dynamics of parcel post growth in the last 10 years, this volume is supposed to be achieved in the next two or three years.

Using the model of Bruns, Klose and Stahly [1], which was developed to restructure Swiss Parcel Delivery Services, we have reconsidered the decision whether two existing Postal Logistics Centres allocated in Slovenia are already optimally located. Simulations demonstrated that the PLC Maribor in addition to the PLC Ljubljana is acceptable if the variable costs of service from the PLC Maribor are lower or at least the same as the variable cost of PLC Ljubljana, and if the costs of services of both do not exceed a certain critical value [4]. A similar application has also been made for the covering service area of Postal Logistics Centre Ljubljana covering one of NUTS 2 region of Slovenia. For further study of hierarchy and required quality of services provided in lower level the hub location model is combined by Travelling Salesman Problem (see details in [5]), which is also used to get the required results for this paper.

Not only in Switzerland but also in other Central European Countries, the problem of postal hub location has been presented in some papers as being vital for efficient postal logistics. Wasner and Zapfel [7] have described the hub transportation network for parcel delivery service in Austria. According to them the problem of several parcel posts, their location and their coverage of area by the post connected in cycles is the basic problem on the lowest level.

The transportation network on four levels has to be built, which connect Posts, Parcel Posts, Regional Parcel Centres and Postal Logistics Centres, where the costs of daily transhipment of parcels would be minimal. This hierarchy could coincide with political determined central places on three or more levels. The criterion is that the total sum of logistic costs in this service of parcels should be minimal, often under certain capacity constraints. The Posts on the level of local communities, patronizing a certain area, have to be assigned to the proper Parcel Post.

2 APPLICATION

Our application is based on an analysis of national postal service of Slovenia. The country is divided according to NUTS system on NUTS 2 and NUTS 3 regions, patronizing 210 local communities – municipalities. In Slovenia, there are 556 posts. Today, the flows of parcels are directed from Post to Post until the truck is fully loaded and then sent to PLC Ljubljana or PLC Maribor and back. Potential Regional Parcel Centres are not open yet, but could be opened in Regional central places. They will patronize Parcel Post.

Figure 1 shows current parcel flows in Slovenia, where division of Slovenian territory into two regions is presented superficially, with the Postal Logistic Centres of Ljubljana and Maribor. Black thin arrows present local parcel flows (within the PLC territory), grey arrows are designated for the parcel flow between Logistic Centres, and black hatched are meant for the international post service, which is supported by the PLC Ljubljana.

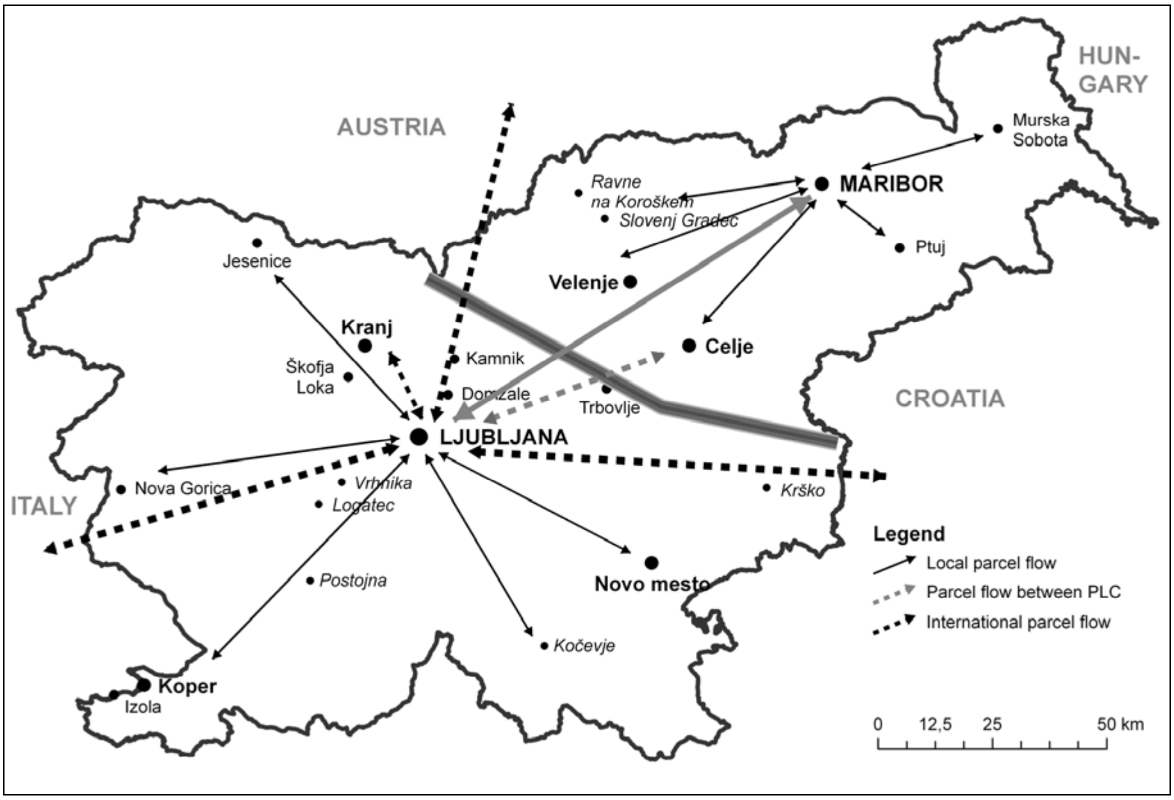


Figure 1: The present parcel flows in Slovenia.

For the purpose of our research we studied the parcel flows between eight regions (postal regional centres). The data about parcel flows from March 2005 on interregional level in the network of Post of Slovenia is presented in the following Table 1.

Table 1: The average daily number of parcels between regions in 2005.

Region	Ljubljana	Maribor	Celje	Kranj	Nova Gorica	Koper	Novo mesto	Murska Sobota	**Total**
Ljubljana	1.732	1.777	1.047	813	504	627	800	423	**7.723**
Maribor	2.277			857	535	449	708		**4.825**
Celje	320			151	92	88	150		**801**
Kranj	378	324	164	155	149	112	178	64	**1.524**
Nova Gorica	304	192	150	127	96	120	110	47	**1.147**
Koper	312	185	137	104	66	101	95	53	**1.052**
Novo mesto	269	210	139	111	68	78	118	50	**1.042**
Murska Sobota	73			28	15	21	40		**177**
Total	**5.665**	**2.686**	**1.636**	**2.347**	**1.526**	**1.597**	**2.198**	**636**	**18.291**

According to the results of our research, proper capacity and allocation of the Regional Parcel Centres and Parcel Posts should be assured. The discussing postal regions are spatially presented on Figure 2. On the base of data about the parcel flows between regions for the year 2005, as presented in the Table 1, we tried to optimize the hierarchical structure for picking process and delivery of parcels from Post to Post.

In our study the location (not capacitated) decision variables with the value of 0 or 1 (to establish or not to establish the logistics centres and regional Parcel Posts in the central places of NUTS 2 and NUTS 3 level) are limited as follows [3]:

- to have one or two Postal Logistics Centres;
- to have eight or fewer Regional Parcel Centres under the western NUTS 2 area, each of them comprising from three to twelve Parcel Posts;
- the Parcel Posts should patronize twenty or fewer Posts.

This heuristic reduces the problem of dimension.

The Posts are allocated to the Parcel Post on the criteria of minimum number of kilometres done by vehicles in the network and especially on the experience of daily transport from Post to Post to the Postal Logistics Centre. We took into consideration the advantages of experience of postal workers in Business Unit of Postal Logistics Centre Ljubljana with the main aim to reduce the admission solutions and combined hierarchical 4-level hub location problem with the method of the Travelling Salesman Problem. A new transportation network for the Post of Slovenia has been developed, with additional level of the Regional Parcel Centres and level of the Parcel Posts based on the criterion of 24-hour time window delivery.

The optimal structure of Regional Parcel Centres in the area covering the territory of Postal Logistics Centre Ljubljana is presented in Figure 2 [2]. The map (Figure 2) shows also the proposal for eight postal regions.

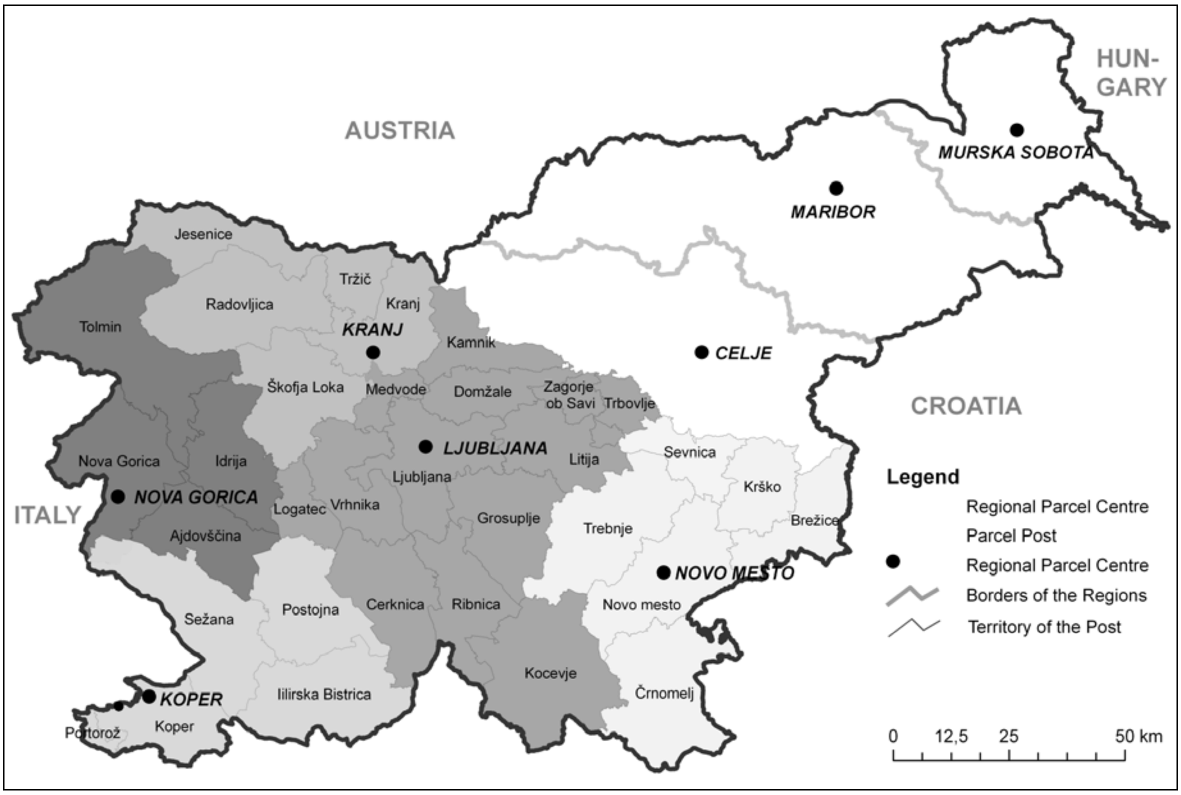


Figure 2: Regional Parcel Centres in Slovenia and Parcel Posts in the territory of Postal Logistics Centre Ljubljana.

Introducing 4-level hierarchical structure, transportation costs of the postal services have been reduced (reduction of the sum of transportation distances in kilometres per day) from 19.017 km today to 12.767 km. This reduction of the transportation costs can be achieved when new system with the Regional Parcel Centres and Parcel Posts starts to operate. The results of our study show that the total length of all routes of postal vehicles has been reduced by 32 percent.

In this case the optimal regionalization is as it is presented on Figure 1 and does not coincide with the political decisions which are the subject of debate in Parliament.

3 CONCLUSIONS

For the area covering PLC Ljubljana (NUTS 2 level) the optimal solution comprises four potential Regional Parcel Centres, which together would have 28 Parcel Posts. These results could also contribute to the optimal regionalisation of Slovenia, which is now the top priority of the Slovenian government.

Based on data of the parcel volume from March 2005 the suggested optimal decision reduces the total transportation distances by 32 percent and the total logistics costs by 20 percent. But we can expect that by changing political structure of central places the flow of parcels will change, but not very soon.

References

[1] Bruns, A., Klose, A., Stahly, P., 2000. Restructuring of Swiss Parcel Delivery Services, Operations Research – Spektrum, pp. 285–302.

[2] Lisec, A., 2006. Optimizacija logistike paketov v hierarhični zasnovi poštne mreže, Doctoral dissertation.

[3] Lisec, A., Bogataj, M., 2006. Combinatorial programming approach to postal systems: the case of parcel network in Slovenia, *Suvremeni promet*, 26/1-2, pp. 116–119.

[4] Lisec, A., Bogataj, M., 2005. Optimal allocation of postal logistics centres. Proceedings of the 10th International Conference on Operational Research - KOI 2004, pp. 35–40.

[5] Lisec, A., Bogataj, M., 2005. Traveling salesman problem at the Post of Slovenia. Nova Gorica, pp. 227–233.

[6] Regulations – Commission Regulation (EC) No 105/2007. European Commission.

[7] Wasner, M., Zapfel, G., 2004. An integrated multi - depot hub - location vehicle routing model for network planning of parcel service. Production Economics, 90/3, pp. 403–419.

THE IMPACT OF EXCHANGE RATES ON INTERNATIONAL TRADE IN EUROPE FROM 1960s TILL 2000 USING A MODIFIED GRAVITY MODEL AND FUZZY APPROACH

E. Oyuk[+], J. Crespo-Cuaresma[++], R. Kunst[+++], E Tacgin[+]
(+) International University of Sarajevo, eoyuk@ius.edu.ba, tacgin@ius.edu.ba
(++)University of Innsbruck, jesus.crespo-cuaresma@uibk.ac.at
(+++) University of Vienna, robert.kunst@univie.ac.at

Abstract: In this paper through the use of gravity model and cross sectional data for 41 pairs of EU15 countries, a significant negative impact of changes in exchange rates on international trade is found for the period from 1961 to 2000. Results illustrating the effects of exchange rates on bilateral trade are obtained by both a modified gravity model developed and using fuzzy approach. A remarkable match is observed between the two results.

Keywords: exchange rates, bilateral trade, cross sectional, gravity model, fuzzy.

A recent survey indicates that most countries abandon intermediate exchange rate regimes and instead prefer a purely floating or a purely fixed exchange rate. The percentage of fixed exchange rate regimes increased from 16% in 1991 to 26% in 1999 while percentage of the floating exchange rate regimes increased from 23 to 42% in the same years. On the other hand, the number of intermediate regimes declined from 62% in 1991 to 34% in 1999(Fischer, 2001). The increasing trend of fixing exchange rates between countries can be seen in the form of the common currency areas in the last years. An IMF study shows that 17.2% of fixed exchange rate regimes consist of currency unions (IMF, 2003).

Hoper and Kohlhagen (1978) analyzed the impact of exchange rate uncertainty on the volume of the US. – German trade between 1965 and 1975 and concluded that there was not any statistically significant effect. After a while, Gotur (1985) reached the same conclusion by analyzing the effects of exchange rate volatility on the volume of trade of the US, Germany, France, Japan and the UK. A famous IMF study (1984) summarized that the large majority of empirical studies could not establish a significant link between exchange rate variability and the volume of trade on the aggregated or bilateral basis. Literature was recently supported by a study carried out by Bacchetta and van Wincoop (2000) who stated that exchange rate uncertainty, or exchange rate systems do not have an impact on trade. On the other side, Ethier (1983) analyzed the effects of exchange rate uncertainty on the level of trade and found out that uncertainty of the future exchange rates will reduce the level of trade. Cushman (1983) estimated fourteen bilateral trade flows among industrialized countries and found a significant negative effect of exchange risk on trade quantity. This literature is supported by Akhtar and Hilton (1984) who established a significant negative effect of nominal exchange rate uncertainty on trade of Germany and the US. Kenen and Rodrik (1986) analyzed the effects of volatility in real exchange rates and concluded that volatility depresses the volume of trade. Another study, which is very similar to the current one was done by De Grauwe and De Bellefroid (1986) in which the authors used cross sectional techniques for the European Economic Community countries from 1960-69 and 1973-84, analyzed the effects of variability of exchange rates, especially of the real exchange rates. One of the studies that analyzed the effects of appreciation or depreciation of exchange rates on trade, done by Lanea and Milesi-Ferretti (2002) concluded that in the long run, larger trade surpluses are to be expected with more depreciated real exchange rates. Jenan-Marie Viaene and Casper G. de Vries (1992) analyzed this issue from a different perspective, by analyzing the effects of exchange rate volatility on the exports and imports separately and found out that exporters and importers are affected differently by the changes in exchange rates, because they are on the opposite sides of the forward market. Since most studies in the literature used time series methods, they were unable to analyze the effects of changes in variables, and changes in years on total trade properly.

Artificial Intelligence methods, like neural networks and fuzzy logic, are recently employed in econometric analysis, especially in time series analysis. Tseng et al. (2001) proposed a fuzzy model and applied it to the forecast of foreign exchange rates. Lee and Wong (2007) used an artificial neural network and fuzzy reasoning to improve the decision making under the foreign currency risk and analyzed the effect of trading strategy on the changes in exchange rates.

A Modified Gravity Model of Total Trade

According to the Gravity Model, trade flows between two countries depend on their income or GDPs positively and on the distances between them negatively. In this model, income of both countries has the same impact on total bilateral trade, therefore coefficient of each countries` income is equal. Gravity model was extended to catch other effects such as having a common language and common border or being in the same trade agreement that promotes bilateral trade. In our modified model, the gravity model is extended with additional variables, which are the population of both countries and changes in bilateral exchange rates. Another difference from the original model is that GDP of the first country and its pair has slightly different coefficients, and therefore they are not taken as products with the same coefficient. The same approach applies to the population of the pair countries where we have different coefficients for them. The proposed model that is used to capture the effects of variability of exchange rates is:

$$\ln T_{ijt} = \alpha + \beta_1 D_{ij} + \beta_2 \ln Y_{it} + \beta_3 \ln Y_{jt} + \beta_4 \ln Pop_{it} + \beta_5 \ln Pop_{jt} + \beta_6 d(\ln XR_{ijt}) + \varepsilon_{ijt}$$

where T_{ijt} represents total bilateral trade between country *i* and country *j* during time *t* which is calculated as the sum of exports and imports. Exports and imports are measured in nominal terms and then are converted to volumes by using GDP deflators for each country at time *t*. D_{ij} is the distance between capital cities of the country *i* and country *j* that is measured in kilometers. Two basic variables of gravity model are Y_{it} and Y_{jt}, real GDP of country *i* and *j* respectively. Pop_{it} and Pop_{jt} are the populations of country *i* and country *j* in time *t*. These variables are expected to have a negative sign because the higher is the population of a country, the less is its GDP per capita. XR_{ijt} represents official bilateral real exchange rate between the country *i* and country *j* in time *t*. Exchange rates in the data set were originally official exchange rates per US dollar. At the beginning, bilateral exchange rates for each pair was calculated using these exchange rates and then by means of GDP deflators, these nominal bilateral exchange rates were converted to real bilateral exchange rates.

Results of Modified Gravity Model

The sample period covers 40 years from 1961 to 2000. Countries included are EU15 countries where Belgium and Luxemburg are taken as one because of data availability. The sources for the data are World Bank`s World Development Indicators 2005, OECD`s International Trade by Commodity Statistics and IMF`s International Financial Statistics.

The model was estimated using bilateral trade flows among 15 EU countries from 1961 to 2000. With these15 countries, 91 bilateral trade flows were obtained during fixed, flexible and Euro periods. The equation is estimated by using bilateral trade volumes and results are shown in Table 1.

In this table, according to the gravity theory, income variable Y is expected to have a positive sign and it already has the expected sign. Here the difference from the previous studies arises with different coefficients for pair countries. Thus, when analyzing bilateral trade between pair countries we see that the contribution of both countries` income to bilateral trade is very different from each other. In this study, it is investigated that 1 percent higher income of the first country leads 0.11 % higher trade in the long run. On the other hand, 1% higher income of the second country results in an increase of 1.09 % higher trade.

Population variable POP has a negative sign for the second countries as expected, with different coefficients for both countries. The intuition behind this is that the higher population is assumed to decrease income per capita that leads to less specialization and less trade.

	Nominal Exchange Rates			Real Exchange Rates		
Variable	**Coefficient**	**t-Statistic**	**Prob.**	**Coefficient**	**t-Statistic**	**Prob.**
C	-2.467	-3.384	0.0000	-2.427	-3.295	0.0000
DISTANCE	-0.000639	-2.996	0.0000	-0.000644	-2.983	0.0000
LOG(Y1)	0.110	2.334	0.0196	0.010	0.228	0.8189
LOG(Y2)	1.096	3.164	0.0000	1.167	3.410	0.0000
LOG(POP1)	0.668	1.441	0.0000	0.761	1.659	0.0000
LOG(POP2)	-0.39	-1.062	0.0000	-0.472	-1.269	0.0000
DLOG(XR)	-1.88	-9.628	0.0000	**-0.601**	-2.777	0.0055

Table 1: The effects of nominal and real exchange rates on bilateral trade volumes

One of the basic elements of the gravity model is the distance between pair countries, which is on the denominator of the equation. Since it is on the denominator, it should have a negative sign with the assumption that higher distances tend to decrease trade between countries by increasing transportation costs and adding some additional difficulties and costs to international trade. Furthermore, it should be emphasized that the effect of distances on bilateral trade does not change depending on nominal or real exchange rates as can be seen from Table 1.

Lastly, exchange rate variable XR has the expected negative sign, which means higher changes in exchange rates lead to less trade between countries. When real exchange rates are taken into account, this effect tends to be smaller. This result is different from most of the previous studies, for example from the results obtained by De Grauwe, in which they have had higher effects when real exchange rate variability is used. Since people make their decisions according to real variables, estimation results under real exchange rates are considered more reliable. According to Table 1, when nominal bilateral exchange rate of the first country - in each pair- increases by 1 percent, which means a nominal appreciation of the currency in the first country, bilateral trade decreases by 1.88 percent. On the other hand, 1 percent real appreciation leads to a decrease of 0.60 percent in bilateral trade.

A Fuzzy Approach to Total Trade

Changes in bilateral exchange rates lead to changes in the volume of trade as explained and proved above. An appreciation of the currency of a country causes its total trade (its exports and imports) to decrease. If the currency of a country is more valuable, its goods will be more expensive abroad which leads to a decrease in exports and in its total trade as a result. In this part, the effects of changes in exchange rates on total trade will be analyzed using fuzzy reasoning. According to the results that are obtained using an econometric methods 1 percent real appreciation of the currency in a country causes its bilateral trade to decrease by 0.60 percent. Steps to be taken to apply a fuzzy approach to total trade are (i) setting the fuzzy decision table, and (ii) determining the change in total trade following a 1 percent change in real exchange rates. Figure 1 shows the partitioning of the universe of real exchange and that of total trade into six fuzzy sets; namely, very low, low-medium, medium, high medium, high, where membership values, μ, are set based on experience intuitively.

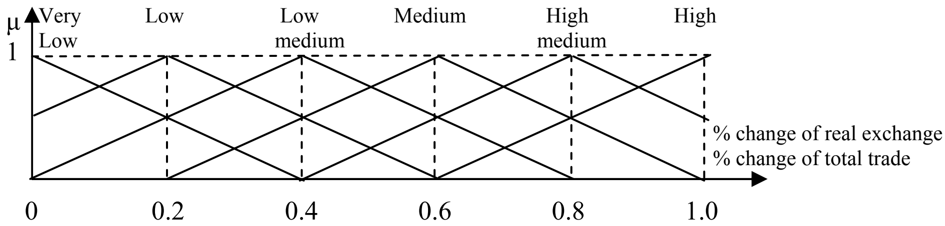


Figure 1: Changes in Real Exchange Rate Partitioning and those in Bilateral Trade Partitioning.

In economics real variables are considered as the most basic variables that have an effect on the decisions of individuals. Therefore, real exchange rates are considered to affect total trade or bilateral trade more than nominal exchange rates. Under fixed exchange rate periods, the impact of real exchange rates on total trade is high, because people expect that exchange rates will not change enormously. When there is a real change, then, the effect of this change will be high. According to econometric estimations, when the effects of real exchange rates on bilateral trade are investigated for long periods consisting of different exchange rate regimes, not only fixed exchange rate regimes, people do not expect such high changes. The fuzzy rule employed is constructed by considering these facts.

According to the fuzzy rule used (see Table 2), high changes in real exchange rates (1 percent) result in high medium (0.8 percent) changes in bilateral trade, while high-medium (0.8 percent) changes in exchange rates lead to medium (0.6 percent) changes in bilateral trade. Moreover, medium changes in real exchange rates cause low-medium changes, low and very low changes have a very low or zero influence. Given the conclusions obtained by the individual fuzzy rules, the overall fuzzy relation is obtained by taking the union of all individual effects.

	FUZZY RULE:
	IF change in XR is high (A_1); THEN change in Total Trade is high-medium (B_2)
ELSE	IF change in XR is high-medium (A_2); THEN change in Total Trade is medium (B_3)
ELSE	IF change in XR is medium (A_3) ; THEN change in Total Trade is low-medium (B_4)
ELSE	IF change in XR is low-medium (A_4); THEN change in Total Trade is low (B_5)
ELSE	IF change in XR is low (A_5) ; THEN change in Total Trade is very low (B_6)
ELSE	IF change in XR is very low (A_6) ; THEN change in Total Trade is very low (B_6)

Table 2: Fuzzy Rule employed for explaining the effects of exchange rates on bilateral trade

$$\widetilde{R} = \sum_{i=1}^{5} \widetilde{A}_i \times \widetilde{B}_{i+1} + \sum_{i=6}^{6} A \times B_i = (\widetilde{A}_1 \times \widetilde{B}_2) \cup (\widetilde{A}_2 \times \widetilde{B}_3) \cup (\widetilde{A}_3 \times \widetilde{B}_4) \cup (\widetilde{A}_4 \times \widetilde{B}_5) \cup (\widetilde{A}_5 \times \widetilde{B}_6) \cup (\widetilde{A}_6 \times \widetilde{B}_6)$$

where $\widetilde{A}_i$ *and* $\widetilde{B}_i$ are fuzzy sets and "x" denotes cartesian product. Using this fuzzy relation in matrix form, the impact of 1 percent change in real exchange rates on bilateral trade can be determined. 1 percent change in real exchange rates means the change is high according to the exchange rate partitioning which is defined by the membership function illustrated in Figure 2.

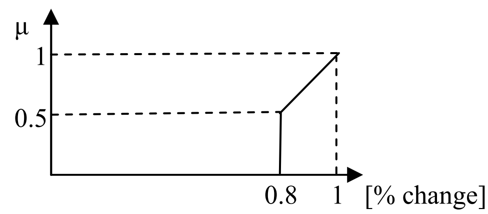


Figure 2: Membership function of 1 percent change in exchange rates

The effects of 1 percent change in exchange rates on bilateral trade can be obtained by composition product $\widetilde{B}^{`} = A \circ R$, as following.

$$\widetilde{B} = [0 \quad 0.25 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.75 \quad 0.75 \quad 0.75 \quad 1 \quad 0.75 \quad 0.5]$$

This is the fuzzified change in bilateral trade where each number is a weight factor between 0 and 1, corresponding to percentage values between 0 and 1 with an increment 0.1. The last step requires the defuzzification process, which converts the overall fuzzy conclusion into a real number that will represent the change in bilateral trade following a 1 percent change in exchange rates. When centroid method is employed in defuzzification process, it is obtained that: "***% Change in total trade* = 0.608 percent".** This result illustrates that 1 percent change in real bilateral exchange rates leads to 0.608 percent change in bilateral trade. It is evident that this result is in accordance with the one obtained in Table 1 -*0.601*- by using cross sectional methods.

$$\text{\% Change} = \frac{0x0 + 0.1x0.25 + 0.2x0.5 + 0.3x0.5 + 0.4x0.5 + 0.5x0.75 + 0.6x0.75 + 0.7x0.75 + 0.8x1 + 0.9x0.75 + 1x0.5}{0 + 0.25 + 0.5 + 0.5 + 0.5 + 0.75 + 0.75 + 0.75 + 1 + 0.75 + 0.5} = 0.608$$

Conclusion

In the first part of this study the effects of exchange rates on bilateral trade between EU15 countries is explained by using cross sectional methods. Considering data of 40 years we found a significant negative effect of changes in exchange rates on bilateral trade. Furthermore, a very close result is acquired by using a fuzzy approach to total trade. The key tasks of fuzzy approach were to set fuzzy decision rules describing the event, and to set membership functions to fuzzy sets intuitively based on experience. It should be emphasized that although the use of econometric methods is essential to obtain reliable results, employing a fuzzy intuitive approach can be useful for estimating first approximate results, especially in the absence of adequate data.

References

1) Gandolfo Giancarlo, 2004, Elements of International Economics, Springer Verlag, Berlin Heidelberg, page 37.
2) Fischer, S., 2001, "Exchange Rate Regimes: Is bipolar view correct?", High level seminar on Exchange Rate Regimes: Hard peg or free floating?, IMF Headquarters.
3) IMF, 2003, "Exchange Arrangements and Foreign Exchange Markets- Developments and Issues".
4) Hooper Peter and Steven W.Kohlhagen, 1978, "The effects of exchange rate uncertainty on the prices and volume of international trade", *Journal of International Economics* 8(4), 483-511.
5) De Grauwe, Paul, 1987, "International Trade and Economic Growth in the European Monetary System", *European Economic Review* 31, 389-398.
6) De Grauwe, Paul and Bernard de Bellofroid, 1986, "Long-run exchange rate variability and international trade", *NBER-AEI Conference on Real Financial Linkages in Open Economies.*
7) Gotur, Padma, 1985, "Effects of exchange rate volatility on trade", *IMF Staff Papers*, 32, 475-512.
8) Cushman, David O., 1983, "The effects of real exchange rate risk on international trade", *Journal of International Economics* 15(1), 45-63.
9) Baccheta Philippe and Eric van Wincoop, 2000, "Does exchange-rate stability increase trade and welfare?", *The American Economic Review*, 1093-1108.
10) Akhtar, M.A. and R.S. Hilton, 1984, "Effects of exchange rate uncertainty on German and US trade", Federal Reserve Bank of NewYork, *Quarterly Review* 9(1), 7-16.
11) International Monetary Fund, 1984, "Exchange Rate Volatility and World Trade: A Study by the Research Department of the IMF", *Occasional Paper* 28.
12) Kenen Peter B. and Dani Rodrik, 1986, "Measuring and analyzing the effects of short-term volatility in real exchange rates", *Review of Economics and Statistics* 68, 311-315.
13) Ethier Wilfred, 1973, "International Trade and the Forward Exchange Market", *American Economic Review* 63(3), 494-503.
14) Viane, Jean-Marie and de Vries, Casper G., 1992, "International trade and exchange rate volatility", *European Economic Review* 36(6), 1311-21.
15) Lane, Philip.R. and Gian Maria Milesi-Ferretti (2002), "External Wealth, the Trade Balance and the Real Exchange Rate," *European Economic Review* 46, 1049-1071.
16) Tseng Fang-Mei, Gwo-Hshiung Tzeng, Hsiao-Cheng Yu and Benjamin J.C. Yuan, 2001, "Fuzzy ARIMA model for forecasting the foreign exchange market", *Fuzzy Sets and Systems* 118, 9-19.
17) Lee Vincent C.S. and Hsiao Tshung Wong, 2007, "A multivariate neuro-fuzzy system for foreign currency risk management decision making", *Neurocomputing* 70, 942-951.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 7

Environment and Human Resource Management

SATELLITE SYSTEM FOR INTEGRATED ENVIRONMENTAL AND ECONOMIC ACCOUNTING

Draženka Čizmić
Faculty of Economics, University of Zagreb
Kennedyev trg 6, 10 000 Zagreb, Croatia
dcizmic@efzg.hr

Abstract: The use of the environment for economic purposes is not taken into account in the calculation of cost in the System of National Accounts (SNA) and is therefore not reflected in important aggregates of national accounts.
The System of integrated Environmental and Economic Accounts (SEEA) is satellite system of the SNA that comprises four categories of account. Very few countries have developed a broad range of accounts, and no country has yet developed the full set of accounts.

Keywords: satellite accounts, environment statistics, SEEA 2003, green GDP

1. Introduction

The discussion of environmentally sound and sustainable socio-economic development has received increased attention from the international community. The aim is to combine economic and social development while simultaneously protecting the environment. The purpose of environmental accounting is to measure the extent of natural resources, their flows and changes, the effects of human activity on the environment, i.e. the sustainability of development over time and space.

The use of the natural environment for economic purposes is not taken into account in the calculation of cost in the System of National Accounts (SNA)[1] and is therefore not reflected in important aggregates of national accounts. The GDP is thus meaningless as a general indicator of changes in economic. Nevertheless, it is definitely useful as an indicator of economic stability.

Satellite accounts generally stress the need to expand the analytical capacity of national accounting for selected areas of social concern in a flexible manner, without overburdening or disrupting the central system. One approach is to concentrate on one field to give a full picture of it.[2]

The SEEA is a satellite system of the SNA, which brings together economic and environmental information in a common framework to measure the contribution of the environment to the economy and the impact of the economy on the environment.

2. The Social Significance of Adjusted Aggregates

If we adopt the framework common to political economy since the 19th century, we may propose three broad classes of "funds" as important to social well-being:

- The stocks and infrastructures of produced economic capital
- The health of the population and the wider communal infrastructures

[1] The System of National Accounts consists of a coherent, consistent and integrated set of macroeconomic accounts, balance sheets and tables based on a set of internationally agreed concepts, definitions, classifications and accounting rules. It provides a comprehensive accounting framework within economic data can be compiled and presented in a format that is designed for purposes of economic analysis, decision-taking and policy-making.

[2] Such accounts are relevant for many fields, such as culture, education, health, social protection, tourism, environmental protection, research and development, development aid, transportation, data processing, housing and communications.

- The systems/funds of "natural capital", which are at the origin of direct delivery of many environmental amenities and life-support services as well as providing inputs and waste absorption services for production and consumption activities

These three categories all have important interfaces with each other. However, up until now the "green" extensions to national accounting systems have mostly focussed on the interface of economic and natural capital assets within the national territory. This includes depletion of stock resources and damages or depreciation to the national funds of environmental capital caused by certain forms of pollution. There has been relatively less systematic attention to the interfaces between economic and environmental funds, and "social capital".

2. The System of Integrated Environmental and Economic Accounts 2003

The System of integrated Environmental and Economic Accounts (SEEA) is satellite system of the SNA that comprises four categories of accounts. The first considers purely physical data relating to flows of materials and energy. The second category of accounts takes those elements of the existing SNA which are relevant to the good management of the environment and shows how the environmental-related transactions can be made more explicit. The third category of accounts comprises accounts for environmental assets measured in physical and monetary terms. The final category of accounts considers how the existing SNA might be adjusted to account for the impact of the economy on the environment.

As an integrated accounting system, the SEEA stands apart from individual sets of environmental statistics. While sets of environmental statistics are usually internally consistent, there is often no consistency from one set of statistics to another. The SEEA may stand apart from sets of environmental statistics, but it also relies upon them for the basic statistics required in its implementation. It is reasonable to expect that over time the implementation of the SEEA will result in changes to the way in which environmental statistics are collected and structured.

The interaction between the environment and the economy manifests itself in physical terms. Despite their strengths, physical accounts suffer from important limitations. One such limitation is the general lack of relative weights that could allow aggregation of measures expressed in physical terms. Purely physical accounts also suffer from a lack of economic context.

The use of relative prices to weight disparate measures in monetary accounts allows the compilation of aggregate measures. The monetary approach is not without limitations. In particular, it is empirically and conceptually challenging to implement. A great deal of data may be required and these data may not exist completely in many countries. In addition, the techniques can be controversial.

2.1 Physical and hybrid flow accounts

Often different data sets are collected and published for different sorts of environmental resources. The objective is to see extent to which the economy is dependent on particular environmental inputs and the sensitivity of the environment to particular economic activities.

Hybrid environmental accounting[3] is a means of confronting physical information about the use of environmental resources with information in both physical and monetary terms about the processes of economic production.

[3] It is the combination of different types of units of measure that leads to the name "hybrid" accounting.

The key sustainability policy goal to which hybrid accounts respond is the desire to maintain or improve economic performance while simultaneously reducing or eliminating the impact on the environment.

2.2 Economic accounts and environmental transactions

Activities are undertaken and products are made with the deliberate intention of relieving pressure on the environment. As well as using the hybrid accounting structure to examine where pressures exist, it is also desirable to identify where expenditure is undertaken to alleviate or rectify these pressures.

It is increasingly common for more environmentally friendly behaviour to be encouraged by means of economic instruments. These may be taxes to discourage consumption by increasing prices or they may be means of controlling property rights and access to environmental media by means of selling licences and permits.

What the accounts in this category do allow, is an assessment of the economic costs and benefits, including their sectoral impact, of reducing human impact on the environment.

2.3 Asset accounts in physical and monetary terms

Natural capital is generally considered to comprise three principal categories: natural resources, land and ecosystems. This category of the SEEA includes asset accounts in physical and monetary terms for each of these three broad categories.

Natural resources, land and ecosystems represent the stocks that provide the many environmental inputs required to support economic activity. If such activity is to be sustainable, the capacity of natural capital stocks to furnish these inputs must be maintained over time or the economy must find a substitute.

The weak sustainability viewpoint is one of technological optimism in which it is assumed that the economy will always find a substitute. The strong sustainability viewpoint takes position that it is imprudent to assume that the economy will always find a substitute. Whatever perspective one takes on weak and strong sustainability, the asset accounts of the SEEA are fundamental to understanding the evolution of sustainability.

2.4 Extending SNA aggregates to account for depletion, defensive expenditure and degradation

The final category of the SEEA deals with the extension of the existing SNA aggregates to account for depletion and degradation of natural capital, as well as for defensive expenditures related to the environment.

The use of resource functions raises the question of whether the resource is being depleted and if so whether the allowance in the economic accounts to maintain produced capital intact should be augmented by a term which might be called the consumption of natural capital.

Some of the expenditure in the economy relates to attempts to avoid using the sink function of the environment.

Like the asset accounts, the extended aggregates are highly relevant to the measurement of sustainability from the capital perspective.

3. Indicators of progress based on GDP corrections

Many people regard progress as synonymous with economic growth. Therefore, they implicitly or explicitly use GDP as an indicator of welfare and progress. Using the GDP as the single progress indicator implies that substitution of "nature" by "economy" is neglected, and that any shift from naturally and freely supplied goods and services to market goods and services is evaluated as "progress", irrespective of natural and environmental losses.[4]

A "green GDP"[5] has been proposed as a good indicator of progress. It is an adaptation of the regular GDP. In essence, all changes in capital need to be accounted for. This means that not only depreciation of economic capital needs to be included but also depreciation of natural resources.

There have been a number of efforts to generate alternative progress indicators. The most well-known recent alternative progress indicators building upon the GDP are: the Index of Sustainable Economic Welfare (ISEW) and Genuine Progress Indicator (GPI).

ISEW[6] is a monetary indicator of sustainable welfare. The starting point is the GDP which is adapted by considering non-market goods and services, defensive costs of social and environmental protection and repair, the reduction of future welfare caused by present production and consumption[7], the costs of effort to obtain the present welfare level, and the distribution of income and labour.

GPI[8] is also a one-dimensional indicator in monetary terms, based on adjusting the GDP by considering over twenty features of human life. These can be categorised as: crime and family breakdown, household and volunteer work, income distribution, resource depletion, pollution, long term environmental damage, changes in leisure time, defensive expenditures, lifespan of consumer durables and public infrastructure, and dependence on foreign assets.

The Human Development Index[9] is based on aggregating a number of other indicators. The subindicators are: life expectancy, adult literacy, combined first, second and third-level gross enrolment ratio, and adjusted real GDP per capita.

4. Environmental accounts in European Union

The domain "Environment Statistics" comprises ten collections: 1) land use, 2) air pollution /climate change, 3) waste, 4) water, 5) transport and environment, 6) environmental expenditure and environmental taxes, 7) agriculture, 8) regional environment statistics, 9) biodiversity, and 10) indicators on water.

The role of Eurostat is not itself to produce environmental accounts, but to encourage and coordinate production by the Member States in areas that correspond to EU and national policy needs. Eurostat proposal for environmental accounting is to define as main priorities at EU level the regular production of data through a Eurostat environmental data base, and a closer integration of environmental accounts, environmental statistics and Sustainable Development Indicators.

A large number of projects have been completed and a substantial number are ongoing. The Member States and Eurostat have been progressively successful in converting the results of these projects into regular production of environmental accounts results.

---

[4] For example, GDP grows when a forest is cut. Using the GDP as a progress indicator implicitly assumes that basic human conditions, such as space, direct access to resources and serenity, can be substituted by economic goods such as large apartments, roads and personal cars, water purification, sewage systems and expensive holidays.

[5] Costanza et alia (1997) argue that annual degradation of nature's services is in the order of 25% of GDP.

[6] It has been calculated for several European countries (Austria, Denmark, Germany, the UK, the Netherlands).

[7] loss of natural areas, loss of soil, depletion of non-renewable resources, air and water pollution, greenhouse effect

[8] GPI claims that America is "down" by 45% since 1970, while GDP is "up" by 50% at the same time.

[9] HDI drives Switzerland from its 4th place in terms of per capita GDP down to 16th.

The Environmental accounting team had to deal with internalising a great workload and therefore also to focus on the following core activities:

- Material Flow Accounts
- NAMEA[10] air emissions
- Expenditure Accounts
- Consolidation and data, exploiting the results achieved so far, and making them available to users (Database on Environmental Accounting)
- Assist new Member States

5. Environment statistics in Croatia

This domain comprises ten collections: 1) air and heavy metals, 2) substances which deplete ozone layer, 3) the red list of threatened plant and animal species of the Republic of Croatia, 4) protected areas of nature, 5) water, 6) quality of sea water along the beaches, 7) investment, 8) environmental accidents, 9) violations in the environment, 10) waste.

The Republic of Croatia is a counter-signatory of the Convention on Long-Range Transboundary Air Pollution (LRTAP) and the United Nations Framework Convention on Climate Change (UNFCCC) and is obligated to submit data on pollutants and greenhouse gases emission into the atmosphere. Data on emissions were calculated on the base of energy balance, statistical data, data on Cadastre of the emissions into the environment and other sources. Data on substances which deplete ozone layer are based on statistical data on imports and exports of substances which deplete ozone layer, as well as on export and import permits for the same substances.

Data on specially protected plant and animal species and protected areas of nature have been taken over from the State Institute for Nature Protection.

Data on public water supply and public sewage system are collected through regular annual reports. Data on the quality of sea water along the beaches are taken from the Ministry of Environmental Protection, Physical Planning and Construction.

Data on investments in environment protection are collected by the reporting method trough the Annual Report on investments in Environmental Protection.

Data on environmental accidents are available through the Environmental Impact Assessment and Emergency Planning Section of the Ministry of Environmental protection, Physical Planning and Construction.

Data on violations in environment are obtained from the Inspection Division office of the Ministry of Environmental Protection, Physical Planning and Construction. Data on wastes were collected through the Pilot Annual Report on Wastes.

Environmental statistics are often collected with a particular regulatory or administrative purpose in mind and the way in which they are structured is specific to this need.

The present data were not sufficient to meet the demands of completed environmental accounts.

6. Conclusion

Traditional national economic accounting system has played huge roles in the times when the resource and environmental problems have not affected the life quality and threatened social and economic sustainable development. However, with the rapid economic development and population growth, various resource and environmental problems, such as environmental pollution, ecological destruction, energy crisis and grain deficit, become more

[10] National Accounting Matrix including Environmental Accounts

and more outstanding. Under these circumstances, it is unreasonable to still continue to use traditional national economic accounting system to measure the economic development status.

The objectives of the SEEA are: 1) segregation of environmental information, 2) a data framework for the linkage of physical and monetary accounting statistics, 3) assessment of environmental costs and benefits, 4) accounting for the maintenance of natural wealth, and 5) environmentally adjusted ("green") indicators.

Very few countries have developed a broad range of accounts, and no country has yet developed the full set of accounts. There have been relatively few empirical exercises to calculate a green GDP. The size of a GDP depends on many assumptions regarding economic behaviour and environmental preferences and can therefore only be the result of model simulations.

The main reasons for the absence of comprehensive environmental accounting are the difficulties in describing the natural environment with its climatic, biological, physical and chemical changes within a generic model of complex interrelationships.

It is therefore necessary to concentrate first of all on improving basic environment statistics and to develop as a second step consistent systems for describing the natural environment.

References

(1) Bartelmus P., Greening the National Accounts: Approach and Policy Use, United Nations, 1999.
(2) Bergh J., Human Progress, Economic Growth and Transport Infrastructure, http://www.pangea.org
(3) Commission of the EC… (et al.), System of National Accounts 1993, Brussels/ Luxembourg…, 1993
(4) CROSTAT, Statistical Yearbook of the Republic of Croatia 2006, Zagreb 2006.
(5) European Commission, Towards a Typology of "Environmentally Adjusted" National Sustainability Indicators, Luxembourg, 2001.
(6) EUROSTAT, Environment statistics, http://europa.eu.int/comm/eurostat
(7) Sachs J. … (et al.), Global Initiative for Environmental Accounting, United Nations, New York; 2005.
(8) Schoer K., Sustainable Development Strategy and Environmental-Economic Accounting in Germany, Federal Statistical Office Germany, Wiesbaden, 2006.
(9) Schoer K., Policy use of Environmental-Economic Accounting in Germany, Federal Statistical Office Germany, Wiesbaden, 2006.
(10) Statistics Canada, National Accounts and the Environment, Papers and Proceedings from a Conference, Ottawa, 1997.
(11) Statistics Denmark, Ninth Meeting of The London Group on Environmental Accounting, Copenhagen, 2004.
(12) United Nations, Integrated Environmental and Economic Accounting, New York, 1993
(13) United Nations…(et al.), Integrated Environmental and Economic Accounting 2003, United Nations…, 2003

SPATIAL MULTI-ATTRIBUTE ANALYSIS OF LAND MARKET – A CASE OF RURAL LAND MARKET ANALYSIS IN THE STATISTICAL REGION OF POMURJE

Anka Lisec
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
e-mail: anka.lisec@fgg.uni-lj.si

Samo Drobne
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
e-mail: anka.lisec@fgg.uni-lj.si

Abstract: In the paper the spatial multi-attribute analysis is discussed in the context of land market analysis – a case study of rural land market analysis in the statistical region of Pomurje. The article focuses on two interrelated concepts of geographical data and multi-criteria analysis. From the problem point of view, the analysis is based on chosen legal and physical characteristics of land and its location, where accessibility is pointed out. The main stress is on spatial analytical tools in the GIS environment where there is more options to choose the appropriate distance function.

Keywords: land, land market, market value, multi-attribute analysis, GIS, location, accessibility.

1 INTRODUCTION

Market research is fundamental to economic decision making. Economics is concerned with choices made in a competitive environment under the constraint of limited resources [3]. Land is one of vital goods for human being from old. In the market-oriented economy land is considered as a fundamental source of capital [9]. Land market analysis is becoming of vital importance for social and economic development of the society, which represents together with the environmental development the main pillars of the sustainable development.

In a land context, market analysis examines the attributes of a land vis-à-vis the relationship of supply and demand, delineating the market in which the land (property) competes. Land has a number of characteristics, which make it different from other assets that may be traded on the market. Heterogeneity is a basic quality of land. Besides economic aspects – such as immovability, limited supply, planning regulations that affect the permitted land use, legal framework of the title transfer etc. – geographical location and accessibility to the supply centre influence the land value. Therefore, the use of spatial multi-attributes analysis methods has become a necessity in the land market analysis.

Geographical Information Systems (GIS) and multi-attribute analysis have developed largely independently, but a trend towards the exploration of their synergies is now emerging. More general, GIS as the environment for spatial analysis and spatial decision making is becoming more and more connected with the statistical and mathematical manipulation of spatial data in order to provide an advanced support for decision making in environmental, land-use and similar issues. According to Worrall (1991, In [7]), it is estimated that 80 % of data used by managers and decision makers is related geographically. Geographical or spatial data are defined as undigested, unorganized, and unevaluated material that can be associated with a location i.e. that is geo-referenced. For GIS environment, each entity is described by locational (coordinate) data and attributes. An attribute is a measurable quantity or quality of a geographical entity or a relationship between geographical entities. It can be defined as any property that distinguishes a geographical entity. The most essential property of the attributes is that their values vary over geographical space [7]. When data has a locational component two problems arises [4]:

- spatial dependence exists between the observations, and
- spatial heterogeneity occurs in the relationships we are modelling.

In the presented paper, the concept of multi-attribute analysis of spatial phenomena is presented, and a sample problem of rural land market analysis in Pomurje is discussed. Besides the attributes of land, that comprise legal and physical characteristics, the spatial component of land is emphasized. The location is discussed in the sense of accessibility, where travel time to the administrative centres in the chosen statistical region is determined. The delineated approach of the multi-attribute analysis of the rural land market is a simplified example. The main purpose of this article is to make an introduction to the methodological approaches that combine GIS and multi-criteria methodologies.

2 PROBLEM DEFINITION, METHODS AND MATERIALS

Multi-attribute analysis in GIS environment allows aggregation of geo-referenced data, involving a variety of both qualitative and quantitative dimensions. Data are of little value in and of themselves. To be useful, they must be transformed into information. When data are organized, presented, interpreted, and considered useful for a particular problem solving, they become information. The main role of multi-criteria analysis techniques in general is to deal with the difficulties in handling large amounts of complex information in a consistent way. The approaches differ in how the data are combined and how a weighting system for criteria is provided [6].

Analysis of land market is a multi-step study process. Market value of the land (v), as well as the land market activity, is a function of numerous attributes (*1, …, n*), referred to the land i.e. geographical location which we might label *i*. Formally, we can state (1):

$$v_i = f(x_{i1}, x_{i2}, \dots, x_{in}) \tag{1}$$

Land is a resource fixed in locational terms. Unlike labour and capital one unit of land is not directly substitutable for another because each unit is unique at least in terms of its geographical location [8]. In addition, spatial dependence can be taken into account. According to LeSage [4], spatial dependence in a collection of sample data observations associated with a location *i* depends on other observations at locations j ≠ i (2) [4]:

$$v_i = g(v_j), \qquad j = 1,\dots,k \quad j \neq i \tag{2}$$

The presented work is concentrated on multi-attribute analysis of the rural land market as formally presented with (1). Since there was a lack of data about land market activities for the individual land parcels, our study was carried out for a spatial unit of the cadastral community. The analysis of the rural land market in Pomurje was based on the transaction data acquired from the Surveying and Mapping Authority of the Republic of Slovenia for the period 2004–2006 [10]. Cadastral community is the elementary administrative unit in the Land Cadastre, which presents the elementary land information system in Slovenia.

To address the problem of multi-criteria analysis of the rural land market in the selected statistical region, we organised attributes hierarchically. The influences of two main groups of land properties on the rural land market were studied:

- attributes of the rural land, where chosen legal aspects of land were considered;
- location, which was considered on the base of the accessibility to the administrative centres.

2.1 Physical and legal characteristics of rural land

In statistical region of Pomurje in the north-eastern part of Slovenia, flat land with agricultural land use dominates with some deviation in the uttermost northern part and in the south-eastern part of the region with hilly landscape, where agricultural land use is combined with forestry land use. Figure 1 shows the main categories of land use in the region together with administrative units and its centres, which are also the largest cities in the region.

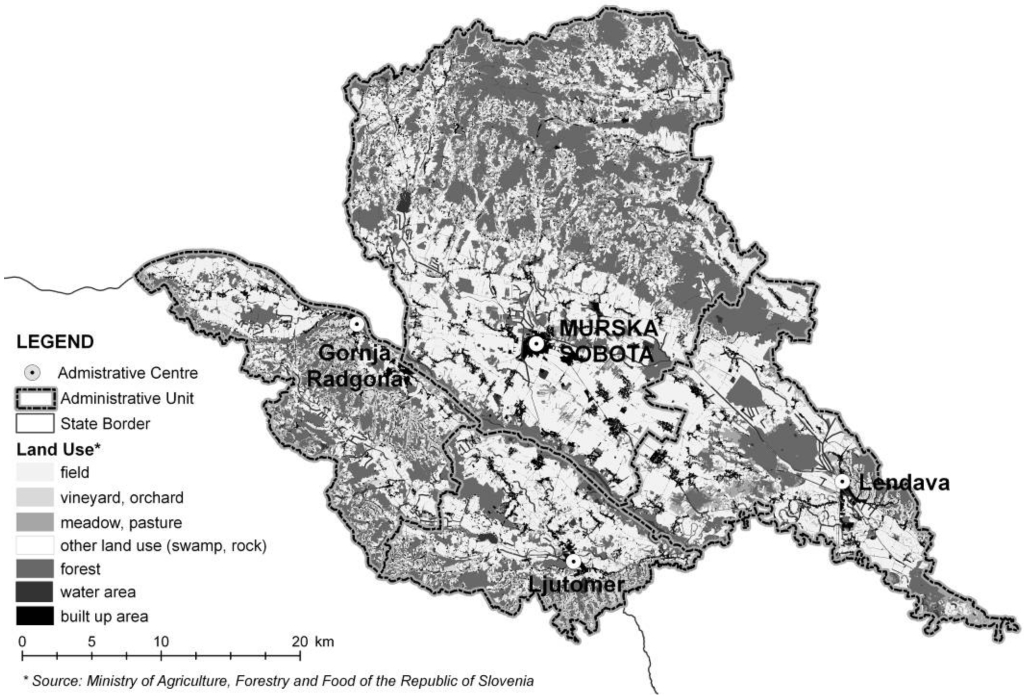


Figure 1: Land use and administrative units in the statistical region of Pomurje, north-east part of Slovenia.

The influence of land use on the rural land market could not be discussed in details on the base of available market data for the cadastral communities. Assuming, that the physical characteristics of rural land in the study area are comparable on the level of the cadastral community, we took into consideration, besides location, some legal characteristics of the land that may influence the market activity and land market value.

In this paper we discuss the influence of the protected areas i.e. pre-emption rights on the rural land market. Protected areas are often associated with special pre-emption rights which influences land market activities (See [5]). Our assumption was that in the protected areas (natural protected areas, water areas) the rural land market was less active and the average market value of the rural land was lower consequently. As our elementary spatial unit was cadastral community we also did not take into consideration soil quality and other physical characteristics of land.

2.2 Location – Accessibility to the administrative centres

The influence of location on the rural land market can be appreciated in terms of transportation facilities – accessibility. Accessibility can be measured in several different ways, such as composite measures, comparative measures, and the time-space approach based on determination of travel time [2]. In our case, the accessibility to the administrative centres in the Slovenian statistical region of Pomurje was based on travel time (by car) as modelled by Drobne [1].

The raster-based accessibility evaluation GIS methodology required layers describing the public road network, administrative regions and administrative centres (Figure 2). In the application [1], the vector layers were rasterized with the resolution of 100 m. Modelling accessibility was based on cost surfaces, whose evaluation required a friction surface that indicates the relative cost of moving through each cell. In the application, costs of movement were expressed as travel time, where they represent the time it would take to cross areas with certain attributes [1]. The cell crossing time in the road network was determined by average travelling speed for each category of road network. For every cell outside the road network the average driving speed was taken as a constant value. Each cell was determined by the time needed to travelling to the administrative centre (For details see [1]).

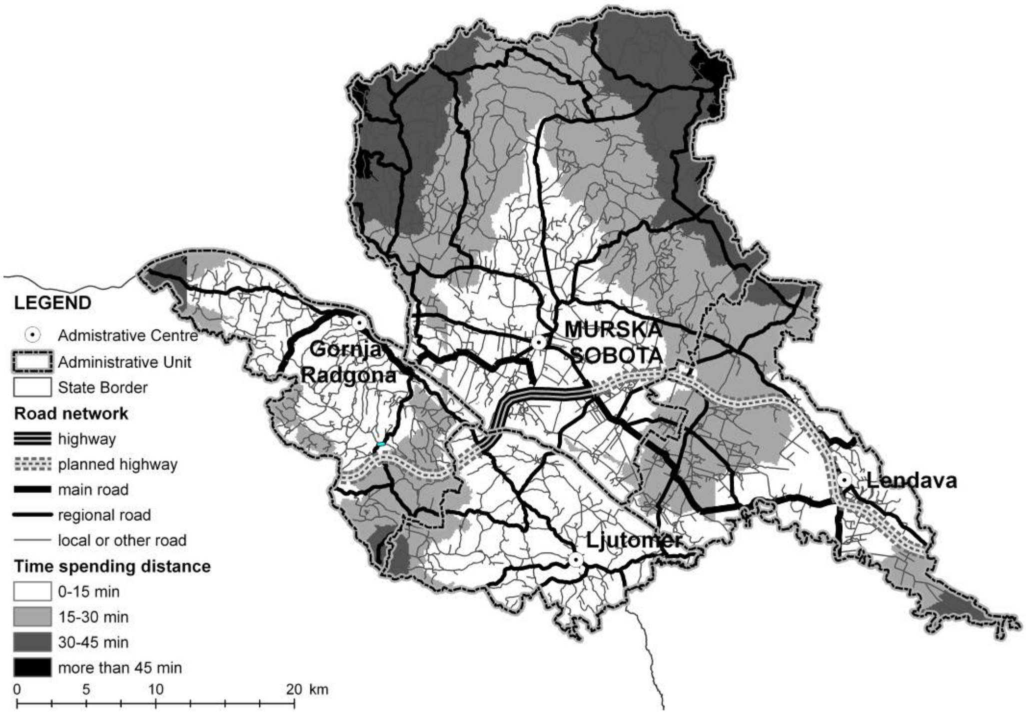


Figure 2: Road network and travel time (by car) needed to the administrative centres in Pomurje.

3 RESULTS AND DISCUSSION

3.1 The influence of legal regimes on the rural land market

Having denoted with T the total number of transactions of the rural land in the cadastral community, and with S the surface of the cadastral community, the market activity coefficient k for the cadastral community was defined as:

$$k = \frac{T}{S} \tag{3}$$

The study of the influence of the protected areas on the rural land market activity comprised the cases of natural protected areas and water protected areas. Figure 3 shows the activity of the rural land market in Pomurje according to the market data of the Surveying and Mapping Authority of the Republic of Slovenia for the period 2004–2006 [10]. Each cadastral community, with the exception of the cadastral communities with less then 2 transactions, is classified according to the market activity coefficient k. The classification was implemented on the base of quantile method, where each of five classes contains an equal number of the cadastral communities.

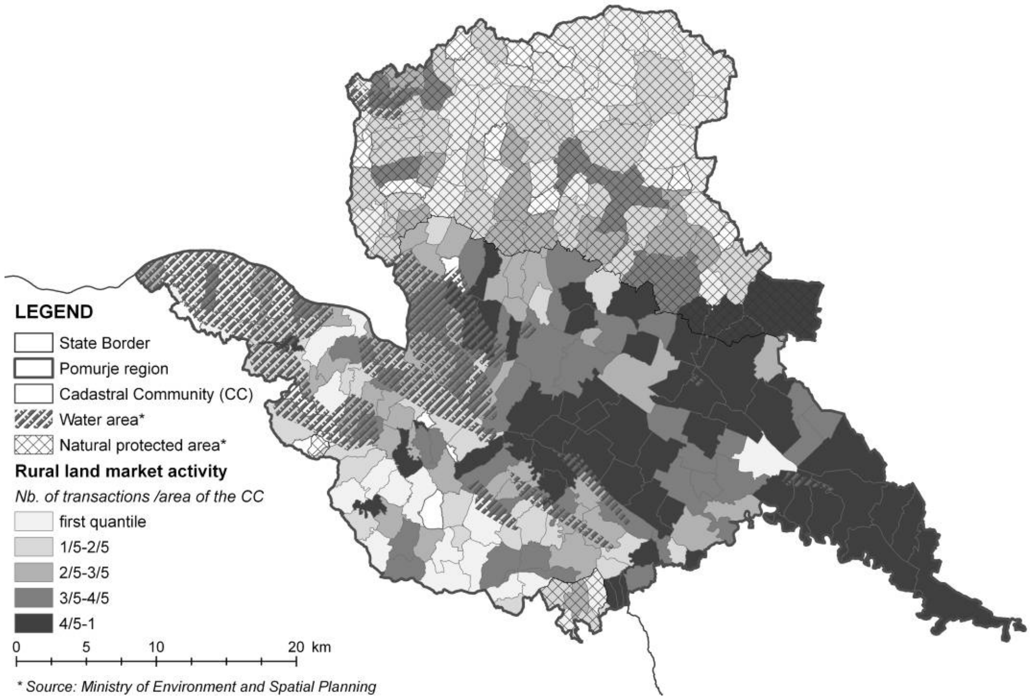


Figure 3: The protected areas and rural land market activity in the cadastral communities in Pomurje for the period 2004–2006 according to the data of the Surveying and Mapping Authority.

In the study period (2004–2006) the market activity was weaker in the northern part of Pomurje, more precisely in the natural protected area, known as Park Goričko. The same holds for the natural protected area in the utmost southern part of the region. In addition, special areas important for the water supply and water protection are denoted with weaker activity of the rural land market in the study period (Figure 3). Furthermore, when comparing the map of the market activity (Figure 3) with the land use map (Figure 1), it can be ascertained that the rural market was more active in the areas with the prevailing acre land in the flat areas of the Pannonian valley.

The influence of the natural and water protected areas is reflected also in the market price of the rural land. We limited our study on the transactions of the rural land with the transaction value between 0,5 and 5,0 EUR per square meter. On the thematic map (Figure 4) the average price is marked with the circle where the classification of the cadastral community was implemented as explained in the legend of the map. The average price of the rural land is shown only for the cadastral communities where at least 5 transactions with the price between 0,5 and 5,0 EUR per square meter ware carried out in the period 2004–2006 according to the transaction data from the Surveying and Mapping Authority of the Republic of Slovenia. In addition, borders of the administrative units and municipalities are shown on the map in order to introduce the spatial administrative structure of the region.

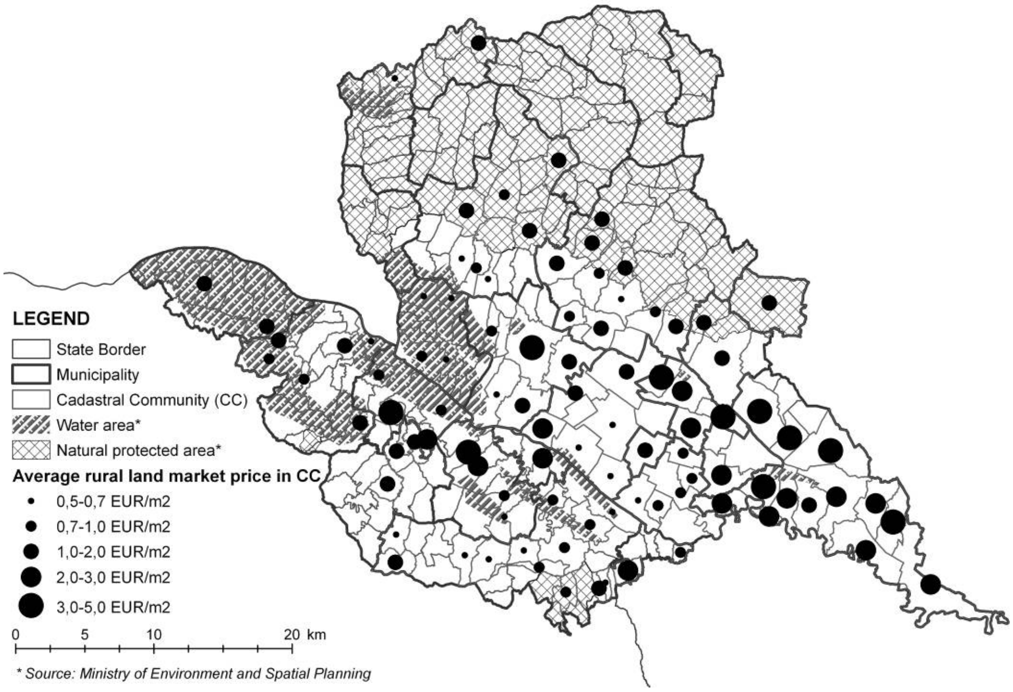


Figure 4: The protected areas and average transaction price of rural land in the cadastral communities in Pomurje for the period 2004–2006 according to the data of the Surveying and Mapping Authority.

Not only that the activity of the rural land market is weaker in the protected areas, but also the market value of the rural land differs between the protected areas and areas outside those territories. From Figure 4 it is evident that average market price of the rural land in the cadastral communities is higher in the areas outside the protected areas. Another fact is that the average market price of the rural land (and the market activity with the rural land) is higher in some municipalities, which can be correlated also with the accessibility to the administrative centres (See Figure 2). Furthermore, average higher market value of the rural land can be associated with the planned highway in this region – is better accessibility to the administrative centres in the future the reason for this phenomena or maybe the compulsory purchase? The protection of personal and tax data and consequently limited access to market data is the reason why the answer is not easy to find.

3.2 The influence of location on the rural land market

The influence of location on the rural land market has been already partly discussed. Having supposed that the administrative centres present the elementary supply centres for the farms, we analysed the market activity and the market price of the rural land in Pomurje with regards to the accessibility to administrative centres following the market data from the Surveying and Mapping Authority of the Republic of Slovenia [10].

Each cadastral community was classified according to the accessibility to the administrative centres. Figure 5 shows the classification of the cadastral communities according to determined prevailing accessibility in Pomurje. In addition, the number of transaction per area of the cadastral communities with above-average and under-average price of the square meter of land is presented – for the cadastral communities with at least 20 such transactions in the study period.

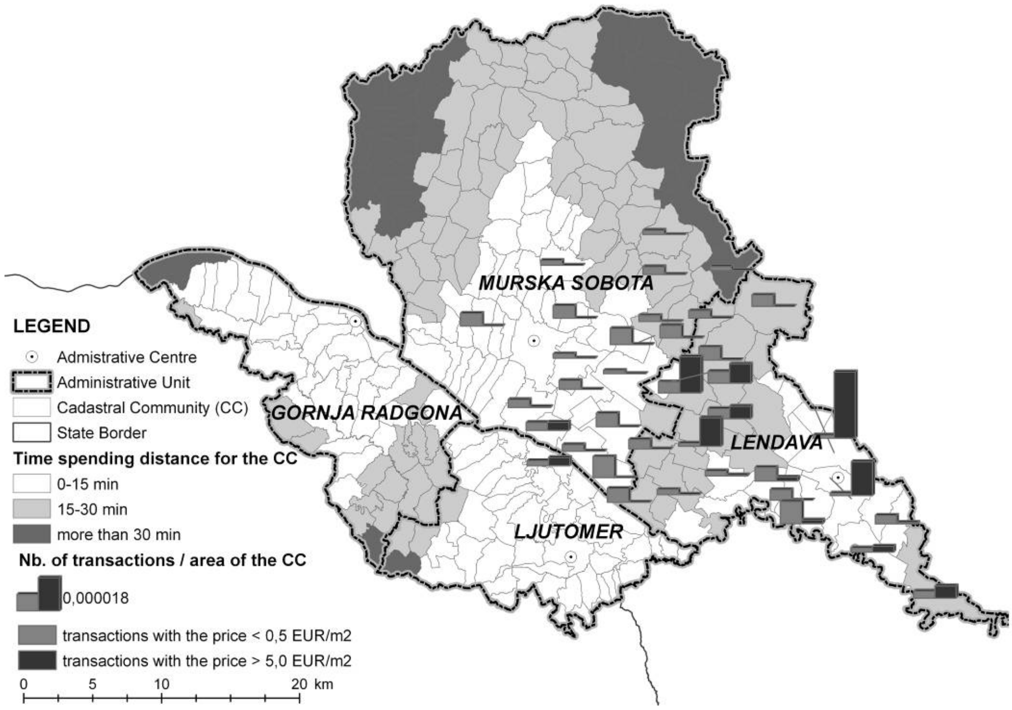


Figure 5: Time spending distance to the administrative centres for the cadastral communities and the transactions of the rural land with above-average and under-average transaction price.

According to Figure 5, the administrative unit of Lendava has in average more expensive rural land than other areas in the region. This can derive besides from the progressive agricultural sector also from the attractive area for economic activities, such as tourism. Furthermore, visual interpretation of the market activities shows that the rural land market is more active along planned highways where the market price of the land is higher as well. On the other side, the most of land transactions with the price under 0,5 EUR/m^2 are in the Pannonian valley, that might refer to the hired land, which is a form of land transaction.

Since standard statistical and spatial analysis are hidden behind the thematic mapping, which is a very useful tool for the visualisation and interpretation of spatial related data and the results of its analyses, the conventional methods for presentation of the results of analyses are still supported (Table 1).

Table 1: Numeric interpretation of land market activity in Pomurje with regards to accessibility and protected areas for the period 2004–2006 according to the data of the Surveying and Mapping Authority.

	Areas outside protected territories			***Protected areas***		
	Time spending distance			***Time spending distance***		
	0-15 min	*15-30 min*	*over 30 min*	*0-15 min*	*15-30 min*	*over 30 min*
$k_{average}$ **(**10^{-6}**)**	2,602	2,714	1,180	0,941	0,486	0,368
k_{under} **(**10^{-6}**)**	2,438	2,691	0,826	1,224	1,434	1,059
k_{above} **(**10^{-6}**)**	1,118	1,440	0,231	0,230	0,036	0,095
P_a ***(EUR/***^{m2}***)***	1,35	1,33	0,34	1,03	1,17	0,25

$k_{average}$ — *average value of the market activity coefficients in the cadastral communities for the transaction with the price between 0,5 and 5,0 EUR/*m^2*;*

$k_{under(above)}$ — *average value of the market activity coefficients in the cadastral communities for the transaction with the price lower than 0,5 or above 5,0 EUR/*m^2*;*

P_a — *average transaction price in the cadastral communities for the transaction with the price between 0,5 and 5,0 EUR/*m^2*.*

As already ascertained, the numerical presentation of market analysis (Table 1) shows weaker market activities in the protected areas, where the transaction price is lower as well. The influence of the accessibility to the administrative centres does not get out for the areas within 30 minutes travel time. On the other side, the areas that are more than 30 minutes far-away from the administrative centres are strongly affected by the weaker market activity and lower transaction price of the rural land.

CONCLUSION

The results of our study confirmed the anticipation that spatial component is of vital importance in the land market analyses. For the case of the rural land market in the statistical region of Pomurje, it has been evidenced, that protected areas and the accessibility to the administrative centres affect land market activity as well as market price. The limitation of our study derives from the limited accessibility to the data about land market. The elementary spatial unit of the land market analysis was a cadastral community, which can be treated as too generalised unit in comparison with the land parcel as the elementary spatial unit in the land market. Land parcel is the elementary unit of the Land Cadastre, which is the fundamental land evidence in Slovenia. In that respect, GIS can provide a very useful tool for detailed analysis of land market as soon as there is appropriate market data available.

A contribution of this work can be seen in the wider aspect. Today, there is a complex spatial data available for different spatial analysis and GIS can provide a useful support for the multi-attribute analysis and multi-criteria decision making relating to the environmental and human resources problems. Spatial statistics in the GIS environment can be adopted across a range of problem solving areas from infrastructure and logistics through environmental, economic and social fields, where large amounts of data are brought together, many of which include a geographical component.

References

[1] Drobne, S., 2005. Do Administrative Boundaries fit Accessibility Fields in Slovenia? In: Cygas, D., Fröhner, K. D. (eds.), Environmental Engineering: the 6th International Conference, Selected papers, Vilnius, pp. 537–542.

[2] Drobne, S., Bogataj, M., Paliska, D., Fabjan, D., 2005. Will the Future Motorway Network Improve the Accessibility to Administrative Centres in Slovenia? In: Zadnik Stirn, L., Drobne, S. (eds.), Proceedings of the 8th International Symposium on Operational Research SOR'05, Slovenian Society Informatika, Ljubljana, pp. 213–218.

[3] Fanning, S. F., 2005. Market Analysis for Real Estate: Concepts and Applications in Valuation and Highest and Best Use. Appraisal Institute, Chicago, 543 p.

[4] Le Sage, J.P., 1999. Spatial Econometrics. University of Toledo, Toledo: 279 p.

[5] Lisec, A., Ferlan, M., Šumrada, R., 2007. UML Notation for the Rural Land Transaction Procedure. In: Geodetski vestnik, Vol. 51, Nb. 1, pp. 11–21

[6] Lisec, A., Zadnik Stirn, L., 2005. Multi-atribute Utility Theory in Sustainable Rural Land Management. In: Zadnik Stirn, L., Drobne, S. (eds.), Proceedings of the 8th International Symposium on Operational Research SOR'05, Slovenian Society Informatika, Ljubljana, pp. 337–342.

[7] Malczewski, J., 1999. GIS and Multicriteria Decision Analysis. John Wiley & Sons, Inc., New York etc., 393 p.

[8] Schiller, R., 2001. The Dynamics of Property Location. Spon Press, London, New York, 240 p.

[9] Soto, H., 2001. The Mystery of Capital: Why Capitalism Triumphs in the West and fails everywhere else. Black Swan, London, 275 p.

[10] The Evidence of the Real Estate Transactions EPN, 2004–2006. Surveying and Mapping Authority of the Republic of Slovenia, Ljubljana.

BEST TRAINING PROPOSAL SELECTION BY COMBINING PERSONAL BELIEFS WITH ECONOMIC CRITERIA

Dubravko Mojsinović
Consule d.o.o, Dr. Franje Tuđmana 8, 10 434 Strmec Samoborski, Croatia
dmojsinovic@consule.hr

Abstract: Company wants to provide education for its employees. Since the company has no experience in education it announces its need and gets the offers from potential vendors. The work focuses on the preparation phase in which vendor selection is done. The methodology is a multiple criteria decision making and comprises individual beliefs of management about training and economic criteria in terms of costs. Among four training proposals, one proposal was selected and the education project was successfully completed. In addition consistency check with AHP is shown.

Keywords: decision making, trainer selection, AHP

1. INTRODUCTION

The company plans to invest HRK 200.000 (without VAT) in education during the next year. Management decided that the highest priority education is communication training. Management believes that improving written and oral skills will add to the company value. In addition to this in previous year some errors were done which originated, it seems, from inadequate communication within the company. Therefore the members of Board are highly committed to carry on the project called communication training. Management board appointed two employees to the project. It wants these employees to provide communication training with high degree of quality and to be rational with expenses and other hidden costs.

Project team members created a rough version of the project plan which is shown in Table 1.

Table 1 Project plan

Id	Activities	Start	End	Duration	Resources
1	Define training expectations	1. week	1. week	1 week	Team members, Board
2	Define elements of offers	1. week	1. week	1 week	Team members
3	Find potential trainers	1. week	2. week	2 weeks	Team members
4	Obtain offers and select the most favourable one	3. week	7. week	5 weeks	Team members, Board
5	Further activities	8. week	14.week	7 weeks	Team members, Board, Selected trainer, other employees

Team members very soon realized that the Management board decision was not enough to carry on the project, because there were a lot of uncertain details. Therefore they decided to define training expectations by putting the questions on paper and checking them with management.

- *What do we want to achieve with the communication training?* People haven't participated on such organized trainings until now. They will feel rewarded. They will gain similar level of knowledge in written and spoken skills. People should be aware that the information has to be shared. Also we will increase the level of proficiency of all of our employees in written and oral communication skills.
- *Are there competent people in the company to execute the training?* No.
- *Should it be a customized training?* Project team members thought that this kind of training is a fairly standard one. Different consulting companies provide it in the standard form and the prices are competitive. However management board member strongly wants a custom made training, ability to look into and change the program and so on. This proved as valuable information which was not considered before.

- *To what extent are PC-s being used within the company?* All of our employees use PC-s.
- *Are all of the employees on the same location?* Most of our employees are in Zagreb. Some are frequently traveling as sales agents.
- *Is there a need to communicate in foreign languages?* No.
- *Where and when should the training take place?* The best is that it is performed outside office during weekends in Zagreb. We think that 3 weekends are enough. We are opened for any suggestions regarding this.
- *What methodology should be used during training courses?* We want that attendees have homework and that they analyze business cases. In order to do that trainer will spend some time in our office as a preparation phase. We would like camera to be used to film some parts of the training. We also want training materials to be prepared in advance.

Team members then defined elements of offers. Luckily the company already has had a document describing what an offer should include. Of course, for this specific need some more accurate definitions should have been made. Offer should include: total price without VAT, total price with VAT, session list for each day including duration in hours and breaks in minutes, topics covered in each session, payment schedule proposal, minimum and maximum number of attendees in a group, trainers' CV, something about the school and references and optionally offer can include more than one training scenario, but then all items should be shown separately. After this team members searched internet and yellow pages in order to find trainers. They also asked people they knew about potential trainers. They already knew one business school which could have been considered. Team members found 11 potential offers. They agreed that contacting and getting offers from all of them was too much. They selected three schools to which they sent a letter with question are they competent and ready to provide communication training for employees. One school has had a really good reputation and they agreed to include it. One school seemed OK and one school was picked randomly. All three schools replied that they are interested and ready to provide communication training. Of course, they asked for more information regarding scope of the training and other details. Most of these questions have already been a part of the training expectations which team members have already prepared. Meeting with trainers was organized. Both team members took part in the meeting. During meeting team members gave some general information about the company. Also they informed schools on what an offer should include. They agreed on the deadlines for sending offers. A few days later one of the schools asked for some additional data in order to prepare an offer. Other two schools were able to prepare their offers on the basis of inputs provided at the meeting.

2. EDUCATION PROPOSALS EVALUATION

After meetings team members evaluated potential trainers on the basis of their first impression. At that point no offers were obtained and evaluation was subjective. Team members added different criteria which seemed to be important for future cooperation with trainers. Then they gave their impressions in terms of grades. They agreed to give them equal importance for simplicity reasons. Grades were given from 1 to 5 where 1 stands for bad and 5 for excellent.

Table 2 Impressions

Criteria			Trainer 1		Trainer 2		Trainer 3	
Id	Name	Definition	Member 1	Member 2	Member 1	Member 2	Member 1	Member 2
1	Reliability	Trainer seems a person who keeps what he/she promises.	4	4	3	3	5	5
2	Selfconfidence	Trainer seems a person who knows what he/she is doing.	5	5	4	4	4	4
3	Carisma	Trainer is able to lead students to change something in their behaviour and to learn.	5	5	4	3	4	4
4	Good rolemodel	Trainer is successful in his/her own business and private life and students will trust him/her.	5	5	4	5	4	5
5	Long term relationship	Trainer is able to continue business relationship with the company even when this training is over and is able to give more favourable terms in the future.	4	3	4	3	5	5
6	Competency	Trainers knowledge seems adequate for communication training.	5	5	4	4	4	4
7	Systematic	Trainer is accurate and precize and will provide high quality materials dedicating enough time for preparation.	5	5	4	4	5	5
8	Communication	Trainer is undersandable and gets to the point.	5	5	5	3	4	4
9	Trustworthy	Trainer will use company information only for the purposes of training and will not misuse them.	5	4	4	3	5	5
Total			43	41	36	32	40	41
Average			4,67		3,78		4,50	
Impression			BEST		INADEQUATE		NEAR BEST	

As it can be seen in Table 2 trainer 1 left the best impression on team members. All trainers got grades 3 and more. This means that impressions can be affirmative ones. What does this mean? If, for example, trainer one got 2,8, trainer two 1,6 and trainer three 2,6 they would have been in the same order, but considering absolute terms all of them would be ranked as bad. Trainer one would be almost as bad as trainer three, trainer two the worst and trainer three bad. The differences between team member opinions are not high. Team member two in total graded trainers one and three the same, but the best impression got trainer one due to team member one. Let us discuss stability issue. Minimum average grade that could have been reached was 1 and maximum 5. Let us suppose that trainers two and three were underestimated in the second criteria and their grades were increased by 1. In this case the following result would occur (Table 3).

Table 3 Sensitivity of impression result

Criteria			Trainer 1		Trainer 2		Trainer 3	
Id	Name	Definition	Member 1	Member 2	Member 1	Member 2	Member 1	Member 2
1	Reliability	Trainer seems a person who keeps what he/she promises.	4	4	3	3	5	5
2	Selfconfidence	Trainer seems a person who knows what he/she is doing.	5	5	5	5	5	5
3	Carisma	Trainer is able to lead students to change something in their behaviour and to learn.	5	5	4	3	4	4
4	Good rolemodel	Trainer is successful in his/her own business and private life and students will trust him/her.	5	5	4	5	4	5
5	Long term relationship	Trainer is able to continue business relationship with the company even when this training is over and is able to give more favourable terms in the future.	4	3	4	3	5	5
6	Competency	Trainers knowledge seems adequate for communication training.	5	5	4	4	4	4
7	Systematic	Trainer is accurate and precize and will provide high quality materials dedicating enough time for preparation.	5	5	4	4	5	5
8	Communication	Trainer is undersandable and gets to the point.	5	5	5	3	4	4
9	Trustworthy	Trainer will use company information only for the purposes of training and will not misuse them.	5	4	4	3	5	5
Total			43	41	37	33	41	42
Average			4,67		3,89		4,61	
Impression			BEST		INADEQUATE		NEAR BEST	

As it can be seen the conclusion still stands. It is obvious that another increase of grades for trainer number 3 will change the result. Trainer 2 will still remain inadequate.

"Best" and "Near best" qualifications verbally provide a soft distinction which takes into account the insignificancy of their difference. If all three impressions were similar then they would have been classified as acceptable or not acceptable.

Four offers came. Trainer 3 gave two offers and other trainers 1. The problem with offers was that although the offer content was agreed, some differences harden the comparison. For example, one of offers included free catering with no mention as to how much does it cost. For some offers hour was equivalent to 45 minutes and for some it was 40 minutes. One of the offers didn't include information on how long was their "hour". Additional inquiry was made to trainers in order to get the adequate data. CV-s proved to show that all trainers have a respectable experience in training courses of this kind.

Table 4 Adding economic criteria to impressions

Offers	Trainer 1	Trainer 2	Trainer 3	
			Scenario 1	Scenario 2
Average	4,67	3,78	4,50	
Impression	BEST	INADEQUATE	NEAR BEST	
Cost in HRK without VAT	52.560,00	18.396,00	21.024,00	29.200,00
Cost in HRK with VAT	64.123,20	22.443,12	25.649,28	35.624,00
Maximum number of employees in a group	12	10	15	15
Number of days	3	2	3	5
Number of hours per day	8	7	8	8
Number of minutes in an hour	45	40	45	45
Cost without VAT per employee	*4.380,00*	*1.839,60*	*1.401,60*	*1.946,67*
Effort as total number of minutes	1.080	560	1.080	1.800
Cost without VAT per minute	*48,67*	*32,85*	*19,47*	*16,22*
Advice	**REJECT**	**REJECT**	**ACCEPT**	**REJECT**
Explanation	There is just a little difference in impression from trainer 3 and a big difference in price	It is just a little bit cheaper than first next, the trainer effort is much smaller and the impression is inadequate.	There is a good impression, a lot of effort and a good price	There is a lot of aditional effort and a significantly greater price than the price quoted for scenario 1.

As it can be seen in Table 4 scenario 1 submitted by trainer 3 was chosen. There was no problem in agreeing to reject trainer 2 because the price was just a little bit less, but the impression was inadequate. Trainer 1 looked very attractive due to the fact that the company wanted the best. However, financial terms in this case proved to be a reason for rejecting it. The budget constrained mentioned was around 200.000 HRK. For one group it could work, but the company has around 150 employees and this is roughly 10 groups. In this case it wouldn't fit the budget. To select between 2 scenarios was also difficult, but the argument of lower costs was in favour of Scenario 1. The proposal to accept Trainer 3 Scenario 1 was submitted to the Management board. It was accepted. The Management board member had a meeting with the trainer chosen and negotiated some changes regarding payment terms and some other aspects of offer. Letter was sent to schools which were rejected with explanation. The Communication project was carried on. There was a meeting with trainer were contract was defined and a more detailed Project plan created. The implementation then started. Detailed program was written by the trainer. It was updated by the company. Training materials were distributed and training sessions held. Payments were done. Training satisfaction was measured. Long term training results will be monitored. The training material will be in future given to all new employees.

3. VERIFICATION WITH AHP-ANALYTICAL HIERARCHY PROCESS

The same procedure is carried out with AHP. The goal is to select one of four options for training. The goal consists of impressions and economic criteria. Bellow is a table showing the relationships between all of the criteria. All impression criteria are equally weighted. In project economic criteria were not explicitly defined as well as the relationship between economic and impression criteria. Since AHP demands it, it was added. Impression criteria in total constitute ¼ of total weight (Table 5).

Table 5 Comparison of criteria for AHP

Id	Criteria name	1	2	3	4	5	6	7	8	9	10	11	Weight
1	Reliability	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
2	Selfconfidence	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
3	Carisma	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
4	Good rolemodel	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
5	Long term relationship	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
6	Competency	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
7	Systematic	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
8	Communication	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
9	Trustworthy	1	1	1	1	1	1	1	1	1	0,06	0,11	0,03
10	Cost in HRK without VAT	18	18	18	18	18	18	18	18	18	1	2	0,50
11	Effort as total number of minutes	9	9	9	9	9	9	9	9	9	0,50	1	0,25
													1,00

Proposals are compared with respect to each criterion. Their comparison is consistent to evaluation carried on using the original approach. For example if Trainer 1 got grade 5 and trainer 2 4 then one is 5/4 stronger than two and so on. Concerning criteria 10, less is better so transformation x=1/y is used. If two team members graded differently the same trainer, average grade is used (Table 6).

Table 6 Comparison of proposals for each criterion in AHP

Reliability

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,33	0,80	0,80	0,24
2	Trainer 2	0,75	1,00	0,60	0,60	0,18
3	Trainer 3-scenario 1	1,25	1,67	1,00	1,00	0,29
4	Trainer 3-scenario 2	1,25	1,67	1,00	1,00	0,29
						1,00

Selfconfidence

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,25	1,25	1,25	0,29
2	Trainer 2	0,80	1,00	1,00	1,00	0,24
3	Trainer 3-scenario 1	0,80	1,00	1,00	1,00	0,24
4	Trainer 3-scenario 2	0,80	1,00	1,00	1,00	0,24
						1,00

Carisma

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,43	1,25	1,25	0,30
2	Trainer 2	0,70	1,00	0,88	0,88	0,21
3	Trainer 3-scenario 1	0,80	1,14	1,00	1,00	0,24
4	Trainer 3-scenario 2	0,80	1,14	1,00	1,00	0,24
						1,00

Good rolemodel

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,11	1,11	1,11	0,27
2	Trainer 2	0,90	1,00	1,00	1,00	0,24
3	Trainer 3-scenario 1	0,90	1,00	1,00	1,00	0,24
4	Trainer 3-scenario 2	0,90	1,00	1,00	1,00	0,24
						1,00

Long term relationship

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,00	0,70	0,70	0,21
2	Trainer 2	1,00	1,00	0,70	0,70	0,21
3	Trainer 3-scenario 1	1,43	1,43	1,00	1,00	0,29
4	Trainer 3-scenario 2	1,43	1,43	1,00	1,00	0,29
						1,00

Competency

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,25	1,25	1,25	0,29
2	Trainer 2	0,80	1,00	1,00	1,00	0,24
3	Trainer 3-scenario 1	0,80	1,00	1,00	1,00	0,24
4	Trainer 3-scenario 2	0,80	1,00	1,00	1,00	0,24
						1,00

Systematic

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,25	1,00	1,00	0,26
2	Trainer 2	0,80	1,00	0,80	0,80	0,21
3	Trainer 3-scenario 1	1,00	1,25	1,00	1,00	0,26
4	Trainer 3-scenario 2	1,00	1,25	1,00	1,00	0,26
						1,00

Communication

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,25	1,25	1,25	0,29
2	Trainer 2	0,80	1,00	1,00	1,00	0,24
3	Trainer 3-scenario 1	0,80	1,00	1,00	1,00	0,24
4	Trainer 3-scenario 2	0,80	1,00	1,00	1,00	0,24
						1,00

Trustworthy

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,29	0,90	0,90	0,25
2	Trainer 2	0,78	1,00	0,70	0,70	0,19
3	Trainer 3-scenario 1	1,11	1,43	1,00	1,00	0,28
4	Trainer 3-scenario 2	1,11	1,43	1,00	1,00	0,28
						1,00

Cost in HRK without VAT

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	0,35	0,4	0,56	0,12
2	Trainer 2	2,86	1,00	1,14	1,59	0,35
3	Trainer 3-scenario 1	2,50	0,88	1,00	1,39	0,31
4	Trainer 3-scenario 2	1,80	0,63	0,72	1,00	0,22
						1,00

Effort as total number of minutes

Id	Proposal	1	2	3	4	Local priority
1	Trainer 1	1,00	1,93	1,00	0,60	0,24
2	Trainer 2	0,52	1,00	0,52	0,31	0,12
3	Trainer 3-scenario 1	1,00	1,93	1,00	0,60	0,24
4	Trainer 3-scenario 2	1,67	3,21	1,67	1,00	0,40
						1,00

Finally (Table 7) local priorities are multiplied with weights. Third proposal is the best. The conclusion is the same as the one originating from the project.

Table 7 AHP ranking of proposals

Criteria groups	impressions									economic		Global
Criteria	1	2	3	4	5	6	7	8	9	10	11	priority
Weight	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,03	0,50	0,25	1,00
Trainer 1	0,24	0,29	0,30	0,27	0,21	0,29	0,26	0,29	0,25	0,12	0,24	**0,19**
Trainer 2	0,18	0,24	0,21	0,24	0,21	0,24	0,21	0,24	0,19	0,35	0,12	**0,26**
Trainer 3-scenario 1	0,29	0,24	0,24	0,24	0,29	0,24	0,26	0,24	0,28	0,31	0,24	**0,28**
Trainer 3-scenario 2	0,29	0,24	0,24	0,24	0,29	0,24	0,26	0,24	0,28	0,22	0,40	**0,27**
	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Since the above AHP verification is not the adequate one due to problems with economic criteria inclusion, the partial verification can be done using only impression criteria. AHP results comparable to impressions originating from the project are shown in Table 8. Equal weights are assigned to each impression criterion.

Table 8 AHP results only for impression criteria and comparable to project methodology

Criteria groups	impressions									economic		Global
Criteria	1	2	3	4	5	6	7	8	9	10	11	priority
Weight	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,11	0,00	0,00	1,00
Trainer 1	0,24	0,29	0,30	0,27	0,21	0,29	0,26	0,29	0,25	0,00	0,00	**0,27**
Trainer 2	0,18	0,24	0,21	0,24	0,21	0,24	0,21	0,24	0,19	0,00	0,00	**0,22**
Trainer 3-scenario 1	0,29	0,24	0,24	0,24	0,29	0,24	0,26	0,24	0,28	0,00	0,00	**0,26**
Trainer 3-scenario 2	0,29	0,24	0,24	0,24	0,29	0,24	0,26	0,24	0,28	0,00	0,00	**0,26**
	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00			1,00

It can be seen in Table 8 that trainer one is the best, trainer two inadequate and trainer three near best. The methodology for getting impression grade is completely consistent with AHP leading to the same results.

4. CONCLUSION

The work presented the methodology for carrying on education project within the company. It focused on the preparation phase of the project and selecting the best offer. It showed how multiple criteria quantitative methods can fit into the training project.
Among four proposals, scenario one from trainer three was selected. This was due to the fact that trainer three left a good impression on team members. Also trainer three presented an offer with good price and enough effort to justify it.
AHP verification could not have been done properly because during the course of the project the selection and weighting of economic criteria was not done. However estimated result shows that the selected proposal was not a bad decision after all. The partial AHP result shows consistency with project methodology.

BIBLIOGRAPHY AND REFERENCES

- Analyticial Hierarchy Process, http://www.icaen.uiowa.edu/~coneng/lectures/AHP.pdf
- Howe G., McKay A.: Combining Quantitative and Qualitative Methods in Assessing Chronic Poverty: The Case of Rwanda, World Development, Vol 35, No. 2, Feb. 2007, Elsevier, 2006, page 203.

RANKING OF THE MECHANISATION WORKING UNITS IN THE FORESTRY OF CROATIA

Ksenija Šegotić [a] Mario Šporčić [b], Ivan Martinić [b],
[a] Department of process techniques, Faculty of Forestry University of Zagreb, Croatia
[b] Department of forest engineering, Faculty of Forestry University of Zagreb, Croatia
segotic@sumfak.hr

Abstract: In this article the two multi-criteria decision making methods, AHP and DEA are used with regard to ranking the mechanisation working units in forestry. The efficiency of the working units was estimated taking into consideration their business results as well as quantities of hazardous waste produced during their operations. Mathematical models may be a very powerful support in planning and decision making in forestry.

Keywords: DEA, AHP, forestry, efficiency, environment.

1. INTRODUCTION

Nowadays, forestry operations are highly complex due to multiple aims of forest management. The principle of sustainable development implies management and use of forests and forest land aimed at preserving their biological diversity, productivity, capability of regeneration, vitality and potential so that forests can fulfil, now and in future, their significant economic, ecological and social function. The above requirements impose increasingly demanding conditions on forestry operations, and cause the management of organisational units to perform continuous analysis of all relevant indicators of business efficiency. From the standpoint of complexity of the present business environment, imperative of ecological acceptability and business success, it is necessary to use new models and more precise methods of business analyses.

The issues of ecological efficiency of mechanisation in performing forest operations were studied by many authors (Bojanin 1997, Sabo 2003, Martinić et al 1999.), while the issues of ecological standards in maintaining numerous forestry mechanisation in Croatia have not been so far the subject of professional discussions or research. This was the reason for establishing the quantities of hazardous waste produced in maintaining forestry mechanisation within the research of the ecological aspect of planning and performing forest operations.

Adverse ecological effects of irresponsible and inappropriate disposal of hazardous waste are almost immeasurable. There are many proves of serious contamination of water, soil and air by automotive waste disposed of without control.

The complexity of today's business environment, as well as the imperative of ecological acceptability and business success, imposes the necessity of continuous analysis of all relevant factors of business efficiency in the management of forestry organisational units. Under such circumstances it is difficult to assess business success by traditional approaches. This paper deals with additional techniques of efficiency assessment applicable when comparing the environmental management organisations, where their successfulness is not only determined by financial profit but also by the ecological aspect of their business operations.

2. METHODOLOGY

The company „Hrvatske šume" ltd. Zagreb (hereinafter: CF – Croatian Forests) manage the state-owned forests of the Republic of Croatia. CF mostly rely on their own capacities for felling, processing, skidding/forwarding and transportation of wood, as well as for the construction of forest roads. These capacities are organised in mechanisation working units (hereinafter MWUs) within CF.

In order to determine the awareness of general issues related to waste disposal, a Hazardous Waste Disposal Questionnaire was completed in MWUs. The questionnaire was developed by the Department of Forest Engineering of the Faculty of Forestry in Zagreb. Collecting of data was carried out in late 2004. After the data were collected, the questionnaires were verified by responsible persons in MWUs, and then submitted to the Faculty of Forestry where they were processed. Additionally, MWUs'annual business reports for the year 2004. were analysed. The data are in Table 1.

Table 1. Data set of results for input and output factors regarding different Working Units

MWU						
			Unscaled data		Scaled data	
	Employees N	Means of work, N	Financial results, 1000 kn;	Hazardous waste (ton)	Financial results, 1000 kn *(+3.500)*	Hazardous waste (ton) *(1/t x10^{-1})*
1	**2**	**3**	**4**	**5**	**6**	**7**
Delnice	106	58	-3.359	11	141	0,909
Đurđevac	95	48	561	5	4.061	2,000
Bjelovar	88	42	-53	18	3.447	0,556
Ogulin	95	29	-124	27	3.376	0,370
Senj	58	28	4.409	23	7.909	0,435
Gospić	42	22	1.841	7	5.341	1,429
Vinkovci	62	20	-3.355	15,5	145	0,645
Kutina	38	19	622	12,5	4.122	0,800
Požega	46	15	2.631	8	6.131	1,250

To rank these 9 MWUs we have used two multicriteria decision making methods, AHP and DEA.

Analitical Hierarchy Process, (AHP) [Saaty (1980)] was used in order to show the differing degrees of importance given to the criteria and to rank MWUs by all the four criteria together. The AHP uses a hierarchical model comprised of a goal, criteria, perhaps several levels of subcriteria and alternatives for each problem. An AHP evaluation is based on the decision maker's judgements about the relative importance of each criterion in terms of its contribution to the overall goal, as well as their preferences for the alternatives relative to each criterion. First we set up the decision hierarchy and then generated the input data consisting of comparative judgement (i.e. pairwise comparisons) of decision elements. A mathematical process (eingenvalue method) was used to calculate priorities of the criteria relative to the goal and priorities for the alternatives to each criterion. These priorities were then synthesized to provide a ranking of the alternatives in terms of overall preference.

Our model was constructed with four criteria: number of employees, number of means of work, quantity of hazardous waste and financial results. MWUs were used as alternatives. The importance of criteria was determined by the forestry experts from the Department of Forest Engineering of the Faculty of Forestry in Zagreb. The matrices of relative importance of the alternatives for the individual criteria were filled on the basis of the Table 1.

Data Envelopment Analysis', developed by Charnes et al. (1978), is a well-known non-parametric method for the assessment of relative efficiency of comparable entities/decision making units (DMU) with different inputs and outputs (Cooper et al. 2003). By linear programming, DEA models determine empiric efficiency frontier (frontier of production possibilities) based on data of used inputs and achieved outputs of all decision making units. Efficiency level is calculated for each production unit, and consequently efficient and inefficient units may be differentiated. The best practice units, those that determine the efficiency frontier, are rated '1', while the degree of technical inefficiency of other decision making units is calculated based on the difference of their input-output ratio with respect to the efficiency frontier (Coelli et al. 1998).

In this paper, the basic CCR model was applied. DEA Excel Solver software was used for solving the problem.

In order to determine MWU efficiency by the application of DEA models, it is necessary to define inputs and outputs, to be used as the input for the analysis. Two variables are selected for both inputs and outputs. The number of employees and the number of means of work are entered into the model as inputs. Outputs are represented by the quantity of hazardous waste generated in maintenance of mechanisation and by the value of monetary gain/loss incurred by MWUs in the year concerned. Hazardous waste includes the quantities of waste tyres, solid waste and waste oils. The value of monetary gain/loss expresses the financial result of business activities of individual working units.

3. RESULTS

In the AHP model we used the eigenvalue method to obtain criteria weights: number of employees, 0.112; number of means of work, 0.067; quantity of hazardous waste, 0.305; financial results, 0.517. Based on this and on data from Table 1. the MWUs were ranked (Figure 1).

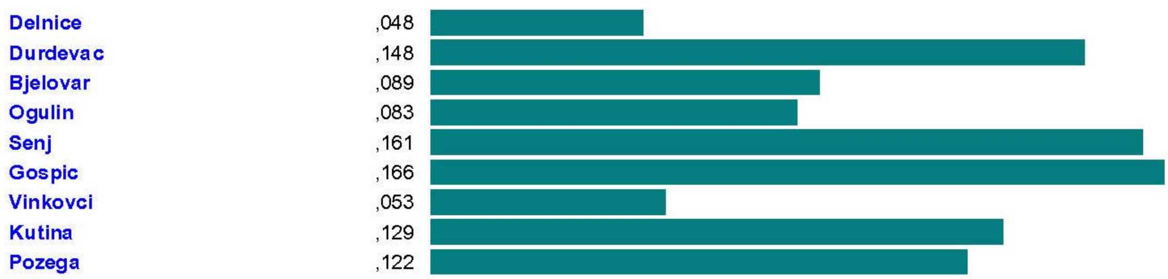


Figure1: Ranking of the MWUs with AHP

The results of the determination of MWU efficiency by the basic DEA model are presented in Table 2. These results show that the average CCR efficiency achieved in 2004 was 0.629. This means that the average (assumed) MWU, if it wishes to conduct business at the efficiency frontier, has to produce 59% more outputs with the used input level, i.e. achieve proportionally lower quantity of waste and higher profit.

Table 2 Results of CCR output oriented models

	CCR model
number of DMUs	9
efficient DMUs, N	3
efficient DMUs, %	33,3%
relative efficiency, E	0,629
maximum	1,000
minimum	0,252
DMUs with efficiency lower than mean efficiency, N	4

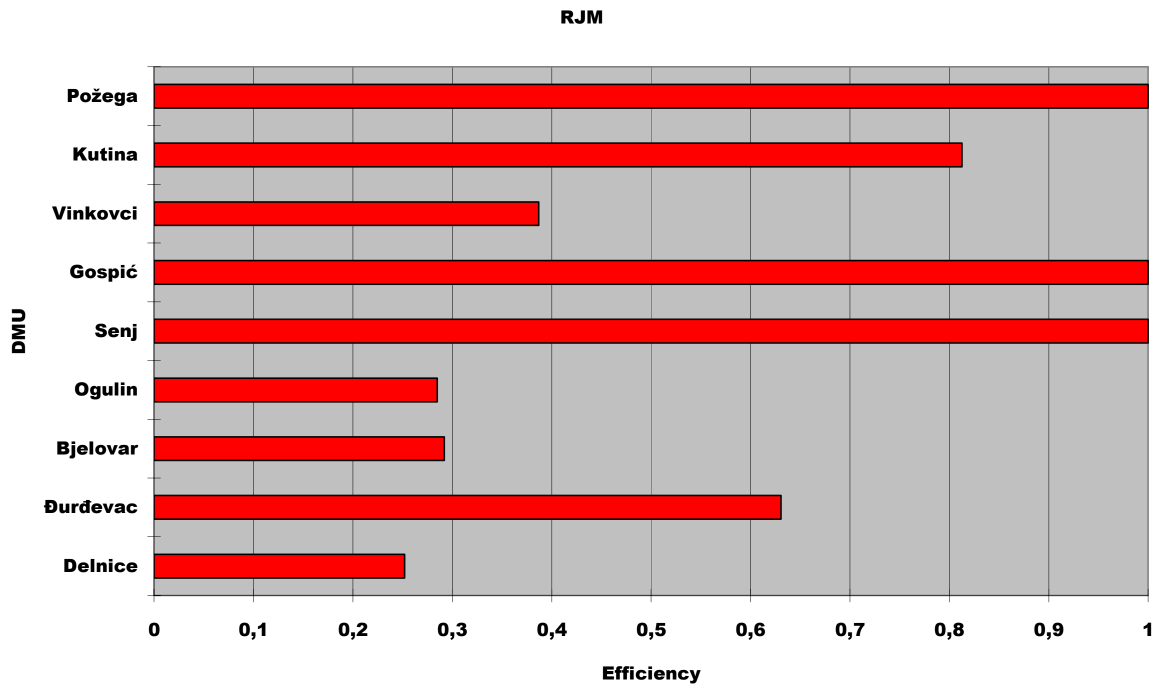


Figure 2: Efficiency of DMUs according to CCR model

According to CCR model (Figure 2.) Po`ega, Gospi} and Senj MWUs were efficient.

4. CONCLUSION

This paper provides insight into additional techniques of efficiency assessment applicable in comparing organisations dealing with environmental management, where their success is not only determined based on financial profit but also based on ecological aspect of business operations. The possibility of application of 'Data Envelopment Analysis' and AHP is presented from the standpoint of multi-criteria evaluation of environmental and financial efficiency of forestry organisational units. On the example shown in this paper, based on the actual data, we have assessed the ecological aspect of mechanisation maintenance and the result of business activities of the working units operating within Hrvatske šume ltd Zagreb.

Based on the obtained results we can see that both mathematical methods show similar results. Because of the different importance of the criteria in AHP there are some difference in ranking.

REFERENCES

- Bojanin, S., 1997: Stanje šumske mehanizacije i struktura šumskog rada u eksploataciji šuma, s obzirom na terenske uvjete, te način gospodarenja u šumama Hrvatske. Mehanizacija šumarstva 22 (1): 19-35.
- Charnes, A., Cooper, W.W., Rhodes, E., 1978: Measuring the efficiency of decision making units. European Journal of Operational research 2, 429–444.
- Coelli, T., Prasada Rao, D.S., Battese, G.E., 1998: An introduction to efficiency and productivity analysis. Kluwer academic publishers, Massachusetts.
- Cooper WW, Seiford LM, Tone K, 2003: Data Envelopment Analysis – A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Kluwer Academic Publishers, p. 1–318 + XXVIII.
- Martinić, I., Jurišić, M., Hengl, T., 1999: Neke ekološke posljedice uporabe strojeva u šumarstvu. Strojarstvo 41 (3-4): 123-129.
- Saaty T.L.(1980) *The Analytic Hierarchy Process*, McGraw-Hill, New York
- Sabo, A., 2003: Oštećivanje stabala pri privlačenju drva zglobnim traktorom Timberjack 240C u prebornim sastojinama. Šumarski list 127 (7-8): 335-347.

DEEPING INSIGHTS OF STAKEHOLDERS' PERCEPTIONS REGARDING FOREST VALUES

Lyudmyla Zahvoyska
Institute of Ecological Economics, Ukrainian National Forestry University,
103 Gen. Chuprynky, Lviv, Ukraine 79057,Tel.: + 38-032-2339678; fax: +38-032-2970388.
E-mail address: zld@forest.lviv.ua (L. Zahvoyska)

Abstract: Misunderstanding among different stakeholders seems to be an obstacle in transition towards sustainable forest management (SFM). Multifaceted and comprehensive picture of stakeholders' perceptions regarding forest values will provide forest planning and decision-making processes with relevant information, which can light searching compromises among discrepant stakeholders' interests for the sake of maximizing common benefits from sustainable forest resource use. A conceptual content cognitive mapping technique and non-parametric statistic methods are used in this study for discovering stakeholders' preferences regarding forest values.

Key words: sustainable forest management, collaborative decision-making, forest values, conceptual content cognitive mapping technique and non-parametric statistic analysis.

1. INTRODUCTION

In the spirit of post-normal science and economic theory of sustainable development a modern scientific inquiry reorients from searching an optimal solutions for value-free puzzle-solving exercises to discovering possible consequences of different scenarios of further development, searching consensus, and mitigating conflicts among stakeholders through a dialog, co-operative learning and co-management (Söderbaum, 2001). Transforming forest management paradigm from a sustainable timber harvesting to integral management of ecosystem resources and services (Bengston, 1994, Kant, 2003) forces us to discover a whole picture of attitude, values and preferences regarding forest resources from multi-stakeholders perspective.

An identification and further organization of revealed forest values from different stakeholders' perspective can appreciably light searching compromises among discrepant stakeholders' interests for the sake of maximizing common benefits from sustainable forest resource use. An understanding of personal values, re-understanding values of own and others' groups will support a lot a co-learning process, which will result in comprehending matters and roots of usual conflicts in forest decision-making and will assist in achieving fruitful compromises for all interested parties.

The main purpose of this paper is to deep insights of Ukrainian stakeholders' perceptions regarding forest values to avoid misunderstanding and mismanagement of such crucial part of natural capital as forest ecosystems resources and services in condition of transformation to a market economy.

2. METHODOLOGY

A large body of modern environmental economic literature does exist on environmental knowledge, valuation and attitudes regarding forest values (Dubgaard, A., 1998, Gregersen et al., 1998, Kristrom, 1990, Sisak, 2004). But a dominant part of them deals with monetary appraisals of forest ecosystems resources and services, which provides with impersonal and substitutable forest values, that is not true. S. Kant and S. Lee justly pointed out that use of market analogies for forest valuation "will not only be deceptive but also fully erroneous" (Kant, 2003, Kant and Lee, 2004). They consistently proved that individual's preferences for the social states of the forest "can be determined through non-market-oriented stated preferences and/or preferences revealed through mechanisms other than the market". In addition to these reflections, Canadian scientists produce another six arguments in favor of non-market-oriented techniques and conclude that the emergence of SFM paradigm itself is a proof of the limitations of the market and market signals (Kant and Lee, 2004).

To capture the breadth of forest values we used in this study the conceptual content cognitive mapping (3CM) technique (Kaplan, 1973, Kearny et. al, 1999, Kant and Lee, 2004). This values identifying and organizing technique serves as a universal and useful instrument of eliciting and analyzing perceptions regarding resources and services in question in all their versatility. This method allows respondents to verbalize their attitudes using relevant concepts, which expose role and function of forests from social-economic and environmental perspectives. The next very important step is identified values organizing: respondents create groups of values and name them. In this way respondents organize forest value universe. Indeed, each stakeholder has own set of values and, no doubt, own ranking. To capture the all the breadth of stakeholders' preferences we propose them to express both their individual' and group' (own or other's ones) priorities. In this way respondents generated a continuum of cognitive maps of preferences regarding resources and services provides by forest ecosystems.

Fulfilling 3CM task respondents assigned a values like 1 – 'the first' (the most important forest value), 2 – the second (less important then the first but more important then rest and so on. Therefore data, collected and organized using 3CM, were considered as ordinal data and were examined by non-parametric statistical methods. To test the statistical significance of similarities and differences in the generated cognitive maps we used the Friedman and the sign test (Newbold, 1995).

The Friedman test was used to check at a 5% significance level a presence of significant differences in preferences regarding forests for each stakeholder. In other words can we state with 95 % probability that for each stakeholder group has own preferences, i.e., some values are more important then others, for instance ranking of environmental values statistically differs than ranking recreational ones?

The real order of preferences was checked using the sign test. This test let us to elicit (at 5 % significance level) a relevant ranking of forest values, their relative importance for each stakeholder group. Results of this test enable us to record stakeholders' cognitive maps of preferences, which reflected individuals' or groups' attitudes.

Hypothesis of this paper was that in case if transition economy each interested party (each stakeholder) has own set of forest values like we see it in case of the Northwestern Ontario (Kant and Lee, 2004) and the Pacific Northwest (Kearney et al, 1999). To capture a multifaceted picture of preferences and attitudes we identified interested parties using set of criteria, formulated by Hotulyeva *et al.* (2006), which consider responsibility, influence, relationship, dependence, representation, and relevance. The 3CM task was done for each target group – local population, forest industry, environmental non-governmental organizations (ENGOs), and city population. 25 representatives of each group were asked to identify values they associate with forest, to indicate individual and own group preferences regarding values in question as well as own consideration concerning other stakeholders' attitudes.

The study area for this research is Lviv region consisting of Zhovkva, Mykolaiv, Yavoriv, and Pustomyty administrative districts (West part of Ukraine).

3. RESULTS OF INVESTIGATIONS

Examining groups of forest values identified and labeled by respondents we developed forest values universe, which consist of 9 dominant themes and 37sub-themes (table 1). This universe shows that respondents indicated the breadth of environmental and socio-economic values, which they associate with forests.

Table 1. Values stakeholders associate with forests (Zahvoyska and Bus, 2007)

Dominant Themes	Sub-themes
Environmental	Air purification and oxygen supply; Climate regulation; Biodiversity; Water regulation; Nutrient cycling
Recreational	Rest; Hiking; Picnics; Pastime with friends
Economic	Income and benefits from forest industry spin off; Timber and other marketed wood products; Employment and relevant satisfaction; Options for tourist business development
Local values	Non-wood forest products; Meat and furs of wild animals; Firewood; Fodder
Educational	Education and training; Science and research; Observations and monitoring; Educational actions
Health care and recovery	Health improving; Medical herbs; Vitamins; Relaxation
Tourism	Hunting; Rock-climbing; Tourism; Sport competitions
Aesthetic	Picturesque landscapes; Birds and other animals watching and listening; Decorative articles; Odors and sounds
Cultural and Emotional	Spiritual and historical values; Quietness, insouciance, solitariness; Inspiriting, stimulation creative ability; Relations with wildlife

Results of the Friedman test for checking statistical confidence of forest values differentiation by each stakeholder group proved the fact of existence of priorities regarding forest values. In case of nine themes for a 5% significance level for each stakeholder calculated Friedman statistics values are much greater then the relevant critical value χ^2 (Zahvoyska and Bus, 2007). This means that at a 5% significance level we can state that each of four groups of stakeholders has different attitude regarding identified nine dominant themes, some of them respondent consider more important then others. True rankings of the dominant themes for a 5% significance level was ascertain using the sign test. Developed integrated map of stakeholders' preferences regarding forest values is presented in Table 2.

This table shows us individuals' and corresponding group representatives' ranking of forest values (in numerator) as well as opinion of other stakeholders regarding corresponding group preferences (in denominator). For example, talking about attitude of local population regarding environmental values we can say, that this group representatives set environmental values on the first position in individual preferences while in the group's map it was indicated on the fourth place. Forest industry, City inhabitants, and Environmental NGO think that from Local population perspective the environmental values should be set on the third, fourth and third position accordingly.

As it is seen from the first sight, this integrated map of preferences is not homogeneous: on the one hand there are examples of full agreement among individuals', own group's and other stakeholders' opinion regarding particular values (for instance role of Economic and Local values for Local population) and on the other hand there are examples of discrepancy and incomprehension among individuals' and groups' rankings (for instance the Environmental values for Local population or Economic values for forest industry).

Let start analyses of the developed maps of preferences from **individual preferences**. As one can see from table all respondents set the Environmental values on the first position, Recreational values were set on the second place and Economic values seems to be the third one. Cultural and Emotional ones follow them. Health care, Educational and Tourist values bring up the rank. And Local values look as most misunderstanding one: for local population they the most important but other stakeholders set them on the last positions In their maps of individuals' preferences.

Table 2. Cognitive map of preferences regarding forest ecosystem services (Bus, 2007)

Groups of stakeholders \ Themes	Environmental	Recreational	Economic	Local values	Educational	Health care	Tourism	Aesthetic	Cultural and Emotional
Local population	1 / 4 3, 4, 3	2 / 4 5, 4, 3	2 / 2 2, 2, 2	1 / 1 1, 1, 1	5 / 6 6, 6, 7	4 / 4 6, 5, 6	5 / 5 6, 5, 6	4 / 3 4, 3, 4	3 / 3 4, 3, 4
Forest industry	1 / 3 3, 2, 4	2 / 1 2, 2, 3	3 / 1 1, 1, 1	5 / 6 6, 6, 6	5 / 5 6, 5, 5	3 / 3 5, 3, 4	5 / 4 2, 3, 3	5 / 6 4, 4, 2	4 / 2 4, 4, 2
City inhabitants	1 / 1 1, 2, 1	1 / 2 1, 1, 3	4 / 4 4, 4, 5	6 / 6 6, 6, 6	4 / 5 5, 6, 6	2 / 1 2, 1, 2	5 / 3 4, 5, 5	3 / 3 3, 3, 2	1 / 2 2, 3, 2
Environmental NGO	1 / 1 1, 1, 1	1 / 2 3, 3, 3	2 / 4 2, 4, 2	5 / 6 5, 6, 6	3 / 4 1, 1, 1	4 / 3 3, 3, 3	6 / 5 4, 5, 4	4 / 1 1, 2, 2	2 / 2 3, 2, 2

Explanation to data in cells:

- in numerator: individuals' / groups' ranking;
- in denominator: opinion of other stakeholders regarding appropriate group's ranking, namely:
  - Local population: Forest industry, City inhabitants, Environmental NGO;
  - Forest industry: Local population, City inhabitants, Environmental NGO;
  - City inhabitants: Local population, Forest industry, Environmental NGO;
  - Environmental NGO: Local population, Forest industry, City inhabitants.

Further let make a glance on the attitudes stated by respondents about **own groups' preferences** (the figure after the slash in numerators). As a common feature of al stakeholders' preferences we can state that Environmental, Recreational, Economic, Cultural and Emotional values all respondents mentioned as the most important. As one can see City inhabitants and ENGO confirm environmental and recreational values as their main points, local population concentrated their attention on the local values, Forest industry shifted its interest to Economic and recreational ones. Forest industry, City inhabitants, and Environmental NGO set Cultural and Emotional on the second place and only Local population put them on the third place. Tourism and Educational themes were mentioned nearer to the end of the list, in most cases the fourth or the fifth accordingly. Correspondingly to opinion of three stakeholders Local values should be the last theme, the sixth, but Local population set the first and this controversy is not accidental, as we will see it later.

And next view let make from **other stakeholders points of view.** Interest of the group Local population seems to be the most understandable for all other groups, but nevertheless they do not accept the crucial role of Local values in their own groups' maps (as it is in Local population's map). The highest number of misunderstanding features Forest industry group. All other stakeholders think that Tourism and Aesthetic values should be more important for Forest industry, but both individual and group statement do not meet these expectations. In the same time Forest industry's interest to Recreational, Cultural and Emotional themes is a bit unexpected for other groups too. Also it is interesting to note that preferences of ENGO are not clear for other stakeholders too. In particular, other stakeholders count that Educational values will have the first place in ENGOs' maps, but they are set on the third and fourth places in individuals' and groups' maps accordingly. Also other stakeholders think that Economic and Tourist values should be more important for ENGOs too. City inhabitants are more interested in Tourist values then other groups generally assume it.

4. CONCLUSIONS

Even underestimated accounts show that global overshoot, growth beyond Earth's carrying capacity has been occurring. Humanity's ecological footprint overall exceeded the worldwide biological productive capacity by over 20 % (Lawn, 2006). This circumstance challenges the post-Brundtland society to tackle present state of art. To avoid uneconomic growth (Daly and Farley, 2004) and to turn to sustainable natural resource us we have to understand inherent motives drive different stakeholders to particular model of production / consumption behavior.

The 3CM technique enables us to collect and organize data regarding implicit values and attitudes and further to verbalize them in form of values universe. Non-parametric statistical methods help us to introduce a continuum of values as a cognitive and comprehensive maps of preferences associated with natural resource in question. Such maps could be treated, analyzed and compared to provide decision makers with relevant information for planning and collaborative management for common benefits.

Using 3CM technique as open-ended data collecting and organizing technique and non-parametric statistical methods we received forest values universe and integral cognitive maps of four stakeholders' preference, which at 5 % significance level describe attitudes of Ukrainian stakeholders to forest resources and services.

Our universe of forest values consists from 9 dominant themes and 37sub-themes. The dominant themes are Environmental, Recreational, Economic, Local, Educational, Health care and recovery, Tourism, Aesthetic, Cultural and Emotional. Comparing with forest values universe, developed to Northwestern Ontario we can state that both universes are quite similar and some minor differences can be explained by methods applied for data analysis. In both cases respondents (indeed, with different fullness) captured almost all the breadth of goods and services provided be ecosystems (Costanza et al, 1997, Daly and Farley, 2004).

Revealed values, preferences and underestimations, common understanding and some misunderstanding create robust background for planning and decision-making process in the context of arising paradigm of sustainable forest management.

5. REFERENCES

1. Bengston, D. 1994. Changing forest values and Ecosystem management: A Paradigmatic Challenge to Professional Forestry. Society and Natural Resources. - 7(6). - P. 515-533.

2. Bus, T., 2007. Investigations of Lviv Region Population's Preferences regarding Forest Ecosystems Services. M. Sc. Thesis, Institute of Ecological Economics, Ukrainian National Forestry University (Ukr.).
3. Costanza, R., D'Aarge, R., De Groot, R., Farber, S., Grasso, M , Hannon, B., Limburg, K., Naeem, S., O'Neill, R., Parvelo, J., Raskin. R., Sutton, P., and van den Belt, M., 1997. The Value of the World's Ecosystem Services and Natural Capital // Nature. - P. 253-260.
4. Daly, H. and Farley, J., 2004. Ecological Economics. Principles and Applications. – Island Press, Washington.
5. Dubgaard, A., 1998. Economic Valuation of Recreation Benefits from Danish Forest: the economics of Landscape and Wildlife Conservation. CABINTERNATIONAL.
6. Gregersen, H., Lundgren, A., Arnold, J.E.M., and Contreras, A., 1998. Valuing Forests: Context, Issues and Guidelines. FAO Forestry Paper 127.
7. Kant, S., 2003. Extending the boundaries of forest economics // Forest Policy and Economics. – № 5. – P. 39-56.
8. Kant, S., Lee, S., 2004. A social choice approach to sustainable forest management: an analysis of multiple forest values in Northwestern Ontario // Forest Policy and Economics. — № 6. – P. 215-227.
9. Kaplan, R., 1973. Prediction of environmental preference: Designers and "clients". In Environmental design research, Preiser, W.F.E. (ed). Dowden, Hutchinson & Ross, Stroudsburg, PA.
10. Kearney, A., Bradley, G., Kaplan R., and Kaplan, S., 1999. Stakeholder perspectives on appropriate forest management in the Pacific Northwest // Forest Science. – 45(1). – P. 62-73.
11. Kristrom, B., 1990. Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis. Univ. of Umee, Sweden.
12. Lawn P. (Ed.), 2006. Sustainable Development Indicators in Ecological Economics. Edward Elgar, Cheltenham.
13. Newbold P., 1995. Statistics for Business and Economics. Fourth ed. Prentice Hall, NJ.
14. Sisak, L. Application and prospects of the CVM in forest recreational service valuation in the Czech Republic. In Scasny, M., Melichar, J., 2004. Development the Czech Society within the European Union. Sbornik z mezinarodni conference. – Prague: Charles University Environment Center. – P. 273-277.
15. Söderbaum, P., 2001. Ecological Economics. Earthscan, London.
16. Hotulyeva M. *et. al*., 2006. Strategic environmental assessment for regional development and municipal planning. Ekoline, Moscow. (Rus).
17. Zahvoyska, L., Bus, T., 2007. Discovering values and stakeholders' preferences regarding forest ecosystem services. Proc. of IUFRO conf. in Lviv (to be published).

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 8

Duration Models

EFFECTS OF THE EDUCATION LEVEL ON THE DURATION OF UNEMPLOYMENT IN AUSTRIA[1]

Bernhard Boehm
Institute of Mathematical Methods in Economics
University of Technology
Argentinierstr.8/105-2, A-1040 Wien, Austria
bernhard.boehm@tuwien.ac.at

Abstract: Using data of the annual "micro-census" sample survey of the Austrian statistical office the influence of the level of education obtained on the duration of unemployment is investigated. Because of changes in the sampling methodology we use two data sets, one gathered between 2000 and 2003 and the second set collected in 2004 and 2005. Information on the personal positions in the labour market helps to identify censured observations. To estimate the effect of the level of education the proportional hazard model of Cox has been used.

Keywords: Duration of unemployment, Cox proportional hazard model, censored data

1. Introduction

The present paper is a first attempt to use subsets of the micro-census database of the Austrian statistical office for an investigation into the determinants of unemployment duration. Due to the use of only a relatively small amount of data it is a tentative exploratory study. A more exhaustive study is planned to provide results later in the year. Up to the present there is no publicly available individual statistic data on the duration of unemployment. The data bank of the labour market service contains individual data of dependent employment, but is not made available even for research. The other alternative data bank contains the micro-census data of Statistics Austria which conducts quarterly surveys of the conditions on the labour market amended by programs on specific questions of interest. Several sets of those sample data are made available for research. It is this body of information that has been reviewed in order to extract the relevant individual data on unemployment duration and appropriate covariates.

The current study starts with a description of the data sources. The information retrieved to determine whether unemployment duration data are right censored is discussed and the final data sets are characterised. Next, the Kaplan – Meier estimator is presented for the two main data sets. The major question on the effects of education levels on the length of unemployment spells is approached by the proportional hazard model of Cox (1972), [1]. The model is estimated for both data sets and the results are compared. The effects are visualised by presenting estimated survivor functions for the different covariates.

2. The data base

Statistics Austria provides anonymous individual data from their regular labour market survey within the program of micro-census sample surveys. These samples provide a rich set of information on most relevant economic, social and cultural issues concerning Austrians. For research sub samples of the full sample are made available (see [5]). They represent roughly three percent of the full amount of data. So far these data have not yet been used for

[1] This research was supported by a grant from the Austrian Science and Liaison Offices Ljubljana and Sofia. The paper reflects only the authors' views. The Austrian Science and Liaison Offices are not liable for any use that may be made of the information contained therein.

the analysis of unemployment duration. In fact it seemed questionable whether they could be used at all for this objective.

Micro-census data for the 1st quarters of the years 2000, 2001, 2002, 2003, and for the whole year of 2004 and 2005 were downloaded from the web page of Statistik Austria. To meet the requirements of the European Union labour force survey the micro-census had to be completely reorganized in the beginning of 2004 (cf. [4]). A new legal basis and a change in the design of the survey as well as in its organisation started in January 2004. Therefore the data between 2000 and 2003 have been combined in one data file and those of 2004 and 2005 in another. Altogether about 20000 individual data sets are available for the latter years, and about 8000 for the former ones. More than 200 different items are reported, beginning with personal information and the household in which the individual is living, the detailed conditions of the working or the unemployment situation, and the housing conditions. From these fields all data sets of individuals have been selected which showed information on job search and unemployment duration. This information has been combined with information about the last and current job, and with the current position in the labour market according to the definitions of the labour force concept. This was particularly relevant for deciding whether the duration data are right censored. Generally all observations were considered as right censored if the person in question was classified as unemployed (i.e. all people between 15 and 74 who are not working, can start work within the next two weeks and have been actively searching for work during the past four weeks). Among these there are some who will start work soon after the interview time and who are thus classified as uncensored. All other individuals who have either left the search for work because they have either found a position (and are thus recorded as economically active) or have withdrawn from the labour market (and recorded as no longer economically active) are giving rise to uncensored observations of the duration data. Whenever it was possible to distinguish between the duration of job search and the duration of unemployment the latter information has been used. This issue could be resolved in many cases where the date of the end of the last job was recorded.

Table 1: Characteristics of data sets

Characteristics	**2000-2003**	**2004-2005**	**Characteristics**	**2000-2003**	**2004-2005**
Male	203	322	Education class 1	112	213
Female	195	364	class 2	180	221
Total	398	686	class 3	28	83
of which married	189	250	class 4	24	44
Austrian nationality	345	577	class 5	24	57
Foreign nationality	53	109	class 6	6	2
Average age	36.56	33.66	class 7	8	14
Average duration, months	8.49	11.45	class 8	16	48

From the filtered data the following characteristics have been obtained: AGE (as continuously measured variable), sex (FE=1 for female), marital status (MARR=1 for married), nationality (FOR=1 for non-Austrian), and education. The latter was measured as highest education level achieved and grouped into eight classes: no formal education or only compulsory school, apprenticeship (EDU2=1), vocational middle school (EDU3=1), general high school (EDU4=1), vocational high school (EDU5=1), college and special courses, university level courses, university degree. The last three groups were combined into one (EDU678=1). The two data files obtained in this way contained altogether data on 1084 individuals, 398 for the years 2000-03 (175 of which are censored) and 686 for 2004-05 (of which 468 are censored). Table 1 gives an impression of the data.

3. A survey of the results of applying different estimation methods

Denoting the probability of survival until time t or longer, i.e. the survival function, by S(t) = P(T ≥ t) = 1 – F(t) with F(t) the distribution function P(T < t) of survival time T, the hazard function is usually defined by $\lambda(t) = \lim_{\delta \to 0} \frac{P(t \le T < t + \delta \mid T \ge t)}{\delta}$ measuring the risk of the event happening at time t conditional upon reaching that duration. It is usually estimated by the number of events occurring at duration t divided by the number at risk at that duration. The integrated hazard $\Lambda(t) = \int_0^t \lambda(u)du$ is related to the survival function by –log S(t) = Λ(t).

A convenient and popular method to estimate the survival function is the Kaplan-Meier estimator. In essence this method sets the estimated conditional probability of completing a spell at t equal to the observed relative frequency of completion at t. We have applied the Kaplan-Meier estimator to both data sets. Because of space limitations only the results (survival function, hazard rate, and integrated hazard) of the 2004/05 sample are presented (cf. fig. 1-2). The hazard rate shows a slightly upward trend which is dominated by a large hazard at the longest spell observed. But those values for the larger spells are not precisely estimated at all with so few observations. The concave part of the integrated hazard would however conform to expectations as it implies negative duration dependence. Especially for short durations the precision of the estimate is much better. The shapes of the survival functions for the two sample periods do not differ very much. In both cases a sharper drop in the functions appears around the duration of 50 months although at different probabilities which reflects the tighter labour market in the more recent periods.

This feature can also be noted from a parametric approach using the exponential distribution. The choice of this function may be justified by recalling that the hazard function is constant and the integrated hazard is a straight line in this case. As a very rough approximation this may do in our cases. The ML estimate of the exponential model with $\lambda(t, \gamma) = \gamma$ and $\Lambda(t) = \gamma t$ yields the result given in table 2. The implied expected duration of an unemployment spell is 7.3 months in the first sample with an estimated lower bound of 6.4 and an upper bound of 8.4 months. For the years 2004-05 we obtain a significantly longer expected duration of 11.3 months with bounds of 9.9 and 13.1 months. This matches well with the increase in the average unemployment rate of more than one percentage point between these two periods.

Table 2: ML estimates of the exponential model

Period	γ	Var(γ)	Std.dev. (γ)	Lower bound	Upper bound
2000-2003	0.1369	8.4035E-05	0.009	0.1185	0.1552
2004-2005	0.0885	3.5907E-05	0.006	0.0765	0.1004

Since the parametric approach makes rather strong assumptions it seems preferable to resort to a semi-parametric method which permits the analysis of the effects of covariates (given by the columns of matrix x) on the hazard rate. The well known and popular proportional hazard model of Cox (1972) (cf. [1],[2] or [3]) specifies the hazard rate typically as $\lambda(t; x, \beta) = \exp(x'\beta)\, \lambda_0(t)$ with $\lambda_0(t)$ the baseline hazard which is an individual specific constant. Because $\partial \log \lambda(t; x, \beta)/\partial x = \beta$ the coefficients β can be interpreted as the constant proportional effect of x on the conditional probability of completing a spell. The survivor function for t is given by $S(t) = \exp(-\Lambda_0(t) \exp(x'\beta))$ with $\Lambda_0(t)$ the integrated baseline hazard. We calculate the baseline hazard relative to an observation with predictors equal to the means of the columns of x.

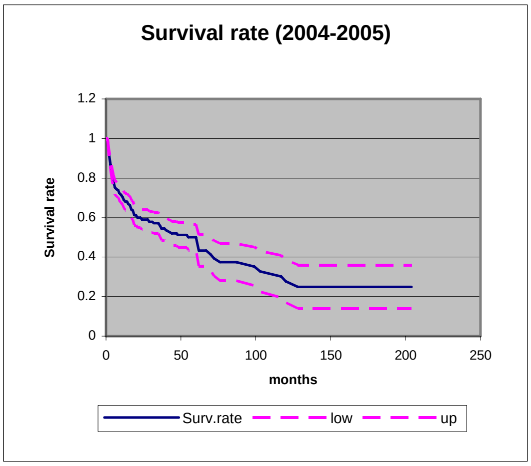


Figure 1: Survival rate estimate for 2004/05 data set

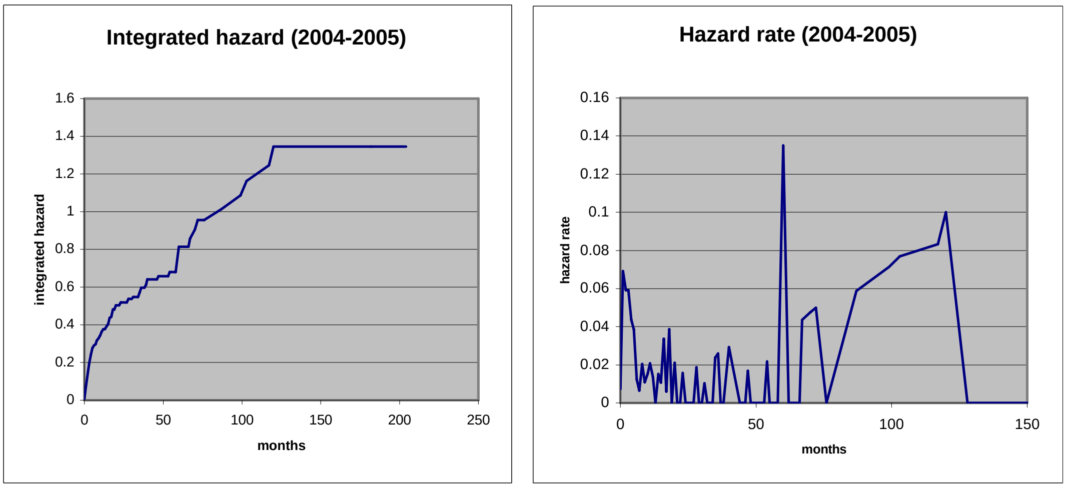


Figure 2: Hazard rate and integrated hazard estimates for 2004/05 data set

The estimates of the Cox-model can be found in Table 3 for the 2000-03 data and in Table 4 for 2004-05. For both periods we find relatively similar results with the exception of the gender variable. Women seem to have significantly bigger chances than men to end their unemployment spell in 2000/03. The opposite seems to be the case in 2004/05 but this effect is not significant at the 5% level. Marital status and citizenship have no noticeable effect and may be dropped from the equations without loosing anything essential. In both periods the influence of age is significant and shows the expected sign. The higher the age the smaller are the chances of ending the status of being unemployed.

Interesting and quite consistent effects across the periods can be expected from the education variables. We assume that the base case is a male unmarried Austrian person with no education beyond obligatory school. The university level education combines courses at universities with and without degrees in variable EDU678. Individuals with such an education have a significantly larger chance to quit the unemployment status quickly. Still higher chances are found for persons with a special education at a higher vocational school. They have twice as big a chance to end unemployment than a person without additional qualifications. This result also squares well with the often mentioned need for qualified personnel in many professions. The tendency towards specialisation can also be inferred from the consistent significance of the apprentice group in both periods, while the effect of general high school is weakest among all groups. Even vocational middle schools are showing higher significance levels than these.

Table 3: Cox Proportional Hazard Model for the 2000-2003 data

Variable	Coefficient	Std. Err.	b/St.Err.	P[|Z|>z]	Haz. ratio
AGE	-0.032309	0.0065283	-4.949	7.4581E-07	0.9682
FE	0.29031	0.14093	2.06	0.0394	1.3368
MARR	0.072998	0.15068	0.48445	0.62807	1.0757
FOR	-0.31308	0.23512	-1.3316	0.18299	0.7312
EDU2	0.39122	0.18199	2.1497	0.031578	1.4788
EDU3	0.5243	0.27748	1.8895	0.058823	1.6893
EDU4	0.30077	0.27869	1.0792	0.28048	1.3509
EDU5	0.83666	0.27394	3.0542	0.0022568	2.3086
EDU678	0.61148	0.25808	2.3693	0.017823	1.8431

logL =-1133.24577 restr.logL=-1158.3256 chi2 Test = 50.1596656 P-val = 1.0053E-07

Table 4: Cox Proportional Hazard Model for the 2004-2005 data

Variable	Coefficient	Std. Err.	b/St.Err.	P[|Z|>z]	Haz. ratio
AGE	-0.024331	0.0067838	-3.5866	0.00033501	0.9759
FE	-0.26114	0.13891	-1.8798	0.060129	0.7702
MARR	0.069499	0.16075	0.43233	0.6655	1.0720
FOR	-0.19932	0.21417	-0.93065	0.35203	0.8193
EDU2	0.42707	0.18704	2.2833	0.022414	1.5327
EDU3	0.36477	0.23526	1.5505	0.12103	1.4402
EDU4	0.51196	0.27974	1.8301	0.067235	1.6685
EDU5	0.54915	0.25488	2.1546	0.031196	1.7317
EDU678	0.50499	0.25333	1.9934	0.046215	1.6569

logL = -1242.89766 restr logL = -1254.9698 chi2= 24.1442738 P-val = 0.00407895

In order to demonstrate the differences in survival probability we present estimated survival functions for different categories. We shall first look at age and compare a 50 year old person with the average (36.5 years) and a 20 year old one for the situation between 2000 and 2003. The difference in the probability of staying unemployed is quite striking. The survivor functions differ already at relatively short spells and tend to narrow with spells of about 4 to 5 years.

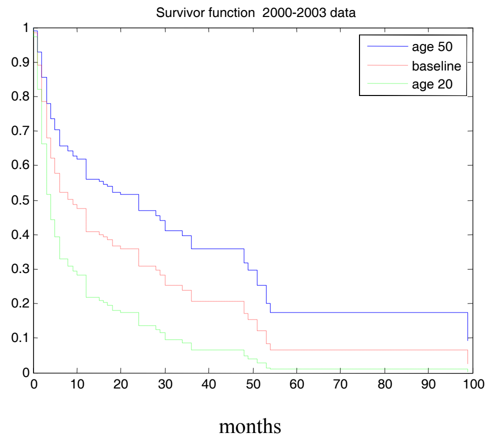


Figure3: Survivor functions for different ages based on Cox proportional hazard model

Differences between the genders are not the same during the two periods. Only for 2000-03 do we find a significantly lower probability of women to stay unemployed. These differences are not as large as for the age gap but again tend to be wider for shorter spells than for longer ones.

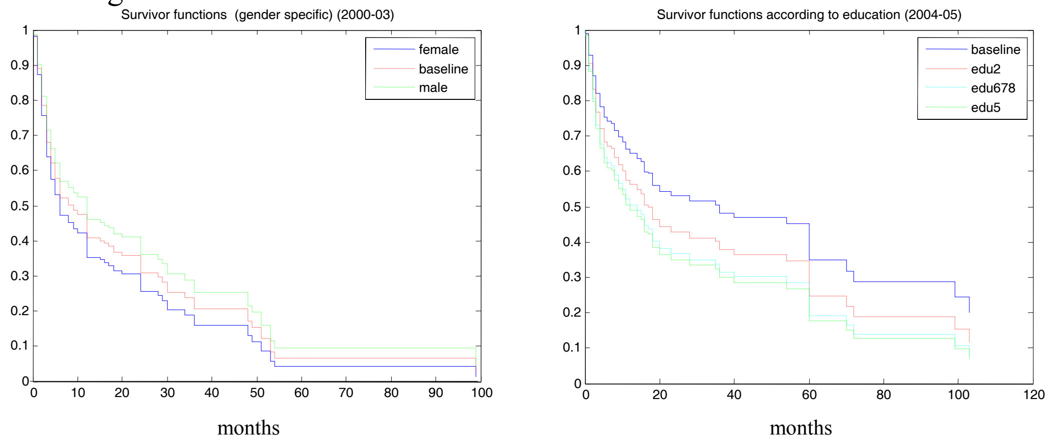


Figure 4: Gender and education specific survivor functions based on Cox model

Turning to survivor functions according to education levels it is immediately seen that the gap between baseline (no or only basic education) and apprenticeship is larger than between the latter and higher education (here taken as higher vocational and university). In fact the differences between the higher levels of education, as long as they are more specialised, are not large at all. They all show an approximately more than 50% larger probability to exit from unemployment than an unskilled person. This result confirms many statements about the desired skill level of the working force. The fact of getting unemployed as well as the duration to stay unemployed both has to do with educational deficiencies.

4. Conclusion

The data for this investigation derive from the labour force survey sample of the Austrian statistical office. While not ideally suited for this kind of duration study, they permitted construction of a sub-sample. Changes in the sample survey made it preferable to work with data from two periods. The methods applied have been the standard ones for duration studies. Starting with the traditional Kaplan-Meier estimate and an attempt to parametrically estimate the duration of unemployment, the results of a proportional hazard model have indicated that, apart from age and possibly gender, the level of education achieved plays a significant role for the duration of unemployment. All human capital investment thus should pay off in case one gets unemployed. In this respect this tentative study is able to confirm studies in other countries which with other data and at other periods have also provided evidence about the role of education for the length of unemployment spells. A further study based on a larger and more exhaustive sample should corroborate the findings.

References

[1] Cox D. R. (1972), Regression Models and Life Tables, *Journal of the Royal Statistical Society*, Series B, 34, 187-220

[2] Greene W. H. (2003), *Econometric Analysis*, (5th ed.) Prentice Hall

[3] Kiefer N. M. (1988), Economic Duration Data and Hazard Functions, *Journal of Economic Literature*, XXVI, 2, June, 646-679

[4] Kytir J., B. Stadler (2004), Die kontinuierliche Arbeitskräfteerhebung im Rahmen des neuen Mikrozensus. Vom „alten“ zum „neuen“ Mikrozensus. *Statistische Nachrichten*, 6/2004, 511-518

[5] Statistik Austria, Mikrodaten für Forschung und Lehre, Standardisierte Datensätze, http://www.statistik.gv.at/_institution/forschung/forschung_standard.shtml

ESTIMATING DETERMINANTS OF UNEMPLOYMENT SPELLS IN CROATIA[1]

Darja Boršič
Alenka Kavkler
Faculty of Economics, University of Maribor, Razlagova 14, 2000 Maribor, Slovenia
alenka.kavkler@uni-mb.si, darja.borsic@uni-mb.si

Ivo Bićanić
Faculty of Economics, University of Zagreb, Trg J.F.Kennedya 6, 10000 Zagreb, Croatia
ibicanic@efzg.hr

Abstract: In this paper duration data techniques are applied to estimate the effect of age, gender, region and level of education on duration of unemployment spells in Croatia for the period 2002–2005. Based on a extensive dataset Cox proportional hazards model and Cox regression with time-dependent covariate have shown that the chances of finding employment are lower for women and older unemployed. The best off are those unemployed who have obtained doctorate or university degree and are from Istarska region, while the worse off are those who have elementary school and are from Karlovačka region.

Key words: unemployment, duration models, hazard ratio, Croatia

Introduction

Unemployment in Croatia was increasing throughout the 1990s and reached its peak in 2002 with registered rate of 21% (Bićanić and Babić 2006). In the last years it has fallen to 17%. On the other hand the ILO unemployment rate is considerably lower (12.7% in 2005). Women are in a disadvantageous position with unemployment rate of 14%. The long term unemployment rate was 7.4% in 2005, indicating that nearly half of the unemployed are searching for a new job for more than a year. Thus, this paper attempts to estimate how different factors, such as age, gender, region and level of education determine the duration of unemployment spells in Croatia.

The paper begins with description of the database, which has been used in the presented empirical study. It is followed by a brief overview of methodology. Next, the results are presented and discussed. Finally the paper is concluded with a short summary of most important findings.

Data description

Data for this analysis were obtained from Employment Office of Croatia. The database is composed of the unemployment spells completed between January 2002 and December 2005 and all on going spells in December 2005. Since the data about individual unemployed is not allowed to be revealed, a personal identifying number was provided to enable identification of repeated spells. For each unemployment spell we have information about the start and end date of registering, gender, age, statistical region and level of education. The database consists of

[1] This research was supported by a grant from the Austrian Science and Liaison Offices Ljubljana and Sofia. The paper reflects only the authors' views. The Austrian Science and Liaison Offices are not liable for any use that may be made of the information contained therein.

1,408,596 unemployment spells out of which 316,567 (22.5%) are censored. Descriptive statistics for non-censored data can be found in Table 1.

Table 1: Descriptive statistics for duration of unemployment in Croatia (in days)

	N	*Mean*	*Std. Dev*
Total	1090964	1090964	683.98833
Factor: Sex			
Male	542753	413.7723	635.18861
Female	548211	496.3462	726.76216
Factor: Age group			
18 years or less	34204	510.7670	654.00642
Over 18 till 25 years	403686	367.6294	586.10293
Over 25 till 30 years	189106	399.4250	646.14524
Over 30 till 40 years	229934	521.8075	768.52098
Over 40 till 50 years	163513	573.2188	781.73590
Over 50 till 60 years	66427	596.0027	706.61742
60 years and over	4094	480.4924	453.10599
Factor: Education			
Without education	456	248.8268	226.39511
Up to 4th grade	1067	284.1593	264.56165
5th to 7th grade	2333	279.5486	252.86965
6 months training without elementary school	1030	475.7078	750.79630
Elementary school	184350	534.7035	782.76468
3-year vocational education	167260	555.4050	811.77646
Vocational secondary school	255338	433.3164	661.23788
Training after secondary school	15582	450.8371	649.76354
Secondary school of more than 4 years	308248	422.7682	607.55067
Gymnasium	36823	480.6032	679.06675
Higher professional education	48474	355.0520	564.66362
University degree	69324	295.9605	481.10936
Master’s degree	663	359.8763	579.92323
Doctorate	16	113.3750	87.85888
Factor: Region			
Zagrebačka	60519	419.8190	580.36912
Krapinsko-Zagorska	27371	402.4571	603.51223
Sisačko-Moslavačka	50682	477.7332	648.11411
Karlovačka	34727	579.3251	815.03268
Varaždinska	40532	411.4060	661.29278
Koprivničko-Križevačka	26002	421.4772	608.28192
Bjelovarsko-Bilogorska	40086	481.4325	748.54494
Primorsko-Goranska	69039	413.8406	636.38986
Ličko-Senjska	12106	473.8970	666.10391
Virovitičko-Podravska	30098	447.3899	640.35259
Požeško-Slavonska	24008	439.9995	674.80278
Brodsko-Posavska	49801	514.3362	794.61498
Zadarska	45877	502.6989	775.61682
Osječko-Baranjska	97284	480.1167	676.19988
Šibensko-Kninska	37352	501.6974	691.60728
Splitsko-Dalmatinska	127594	531.0578	814.79701
Istarska	45875	300.3864	505.92896
Dubrovaćko-Neretvanska	34551	416.3338	649.78742
Međimurska	27491	366.5183	571.47186
Grad Zagreb	149311	393.4675	575.69819
Vukovarsko-Srijemska	60147	525.8154	751.43002

The empirical analysis was conducted by SPSS 14.0 software. The factors were coded as follows: male (1), female (2). The factor age indicates the age of the unemployed at the beginning of unemployment spell. We obtained information for 21 statistical regions: Zagrebačka (1), Krapinsko-Zagorska (2), Sisačko-Moslavačka (3), Karlovačka (4), Varaždinska (5), Koprivničko-Križevačka (6), Bjelovarsko-Bilogorska (7), Primorsko-Goranska (8), Ličko-Senjska (9), Virovitičko-Podravska (10), Požeško-Slavonska (11), Brodsko-Posavska (12), Zadarska (13), Osječko-Baranjska (14), Šibensko-Kninska (15), Splitsko-Dalmatinska (16), Istarska (17), Dubrovačko-Neretvanska (18), Međimurska (19), Grad Zagreb (20) and Vukovarsko-Srijemska (21). The dataset provides also information about the education of the unemployed at the onset of unemployment, which is divided into 14 levels: without education (1), up to 4th grade (2), 5th to 7th grade (3), 6 months training without elementary school (4), elementary school (5), 3-year vocational education (6), vocational secondary school (7), training after secondary school (8), secondary school of more than 4 years (9), gymnasium (10), higher professional education (11), university degree (12), master's degree (13) and doctorate (14).

Methodology: Duration data analysis

In this paper we apply survival analysis. According to Therneau and Grambsch (2001) and Klein and Moeschberger (2005) the random variable T denotes the *survival time*. The equation $F(t) = P\left(T < t\right)$ is the distribution function of T and determines the probability that an event will last up to time t. The *survival function* $S(t) = P(T \geq t) = 1 - F(t)$ measures the probability that an event will survive until time t or longer. The limit $\lambda(t) = \lim_{\delta \to 0} \frac{P\left(t \leq T < t + \delta \mid T \geq t\right)}{\delta}$ represents the risk or proneness to death at time t and is called the *hazard function* or the *failure rate.*

A semi-parametric method for estimating the impact of different covariates on the hazard function is *Cox proportional hazards model* (Kleinbaum 2005 and Hosmer and Lemeshow 2003). The model can be written as: $\lambda_i(t) = e^{x_i'\beta} \cdot \lambda_0(t) = c_i \cdot \lambda_0(t), \quad i = 1, 2, \ldots, n$, where x_i is the vector of k covariate values for the individual i, β is the vector of regression coefficients, $\lambda_i(t)$ is the hazard function of the individual i and $\lambda_0(t)$ is *the baseline hazard* consistent with an observation with $x_i = 0$. The ratio $\lambda_i(t)/\lambda_0(t)$ is equal to the constant c_i, thus the impact of individual factors on hazard function does not depend on time. Hence, the configuration of the hazard function is set by baseline hazard. According to Greene (2003) the ratio of the hazard functions of the individuals i and j, is called the *hazard ratio*. When the value of the ratio is lower than 1, it indicates decreased risk or less chances of finding a new job in our case. While a ratio higher than 1 denotes increased risk or better chances of re-employment.

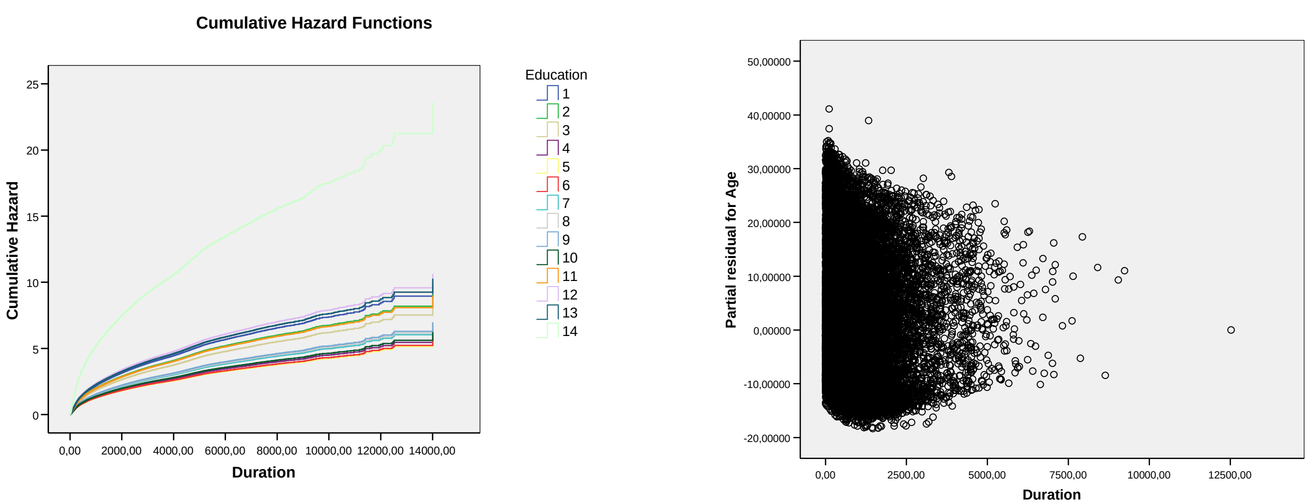
Results

The results of Cox proportional hazard model are presented in the first part of Table 2 and cumulative hazard functions are shown in the left panel of Figure 1. By default in SPSS the reference category of a covariate is the last category. The baseline hazard is the hazard for a female from Vukovarsko-Srijemska region with doctorate.

All four variables are highly significant. Column Exp(B) presents the hazard ratio of a given category with the reference (last) category. Each year of unemployment a chance of being re-employed is decreased for 2.4%. Men have 32% better chances of getting a new job than women. As for the level of education, the hazard ratios are relatively low implying that unemployed with all other levels of education are much worse off than those unemployed with doctorate. The highest hazard ratio is 0.451 for university degree, followed by 0.435 for master's degree. That means that those unemployed with university degree have 54.9% less "risk" of re-employment than those with doctorate. It is interesting to note that relatively high hazard ratio was estimated for those unemployed with no education at all (0.421). The worst off are those unemployed with finished elementary school (level 5) and with 3-year vocational education (level 6). The comparison of regions shows a clear advantage of those who live in Istarska region, where the hazard ratio is considerably higher than others (1.954). This region is followed by Međimurska, Krapinsko-Zagorska region and Požeško-Slavonska regions, with hazard ratios of 1.303, 1.274 and 1.274, respectively. On the other hand, unemployed from Karlovačka, Splitsko-Dalmatinska and Vukovarsko-Srijemska region are the worst off among the unemployed in Croatia with hazard ratios of 0.917, 0.979 and 1, respectively.

Since the proportional hazards assumption is crucial for the Cox regression approach and is often violated, we tested it graphically by partial residuals. The right panel of Figure 1 reveals positive correlation between partial residuals and time implying that the proportional hazards assumption does not hold. This result was confirmed by a separate model including the covariate and an interaction term between time and the covariate (age). The estimated coefficient of the interaction term was proven to be highly significant, which again violates the proportional hazards assumption.

Figure 1: Cumulative hazard functions (left) and partial residuals (right)



In order to circumvent the violation, Cox regression with time-dependent covariate was estimated. The results of this model are presented in the second part of Table 2. Factors age and sex are again highly significant. The differences among men and women are slightly less pronounced. Men have 30.7% higher "risk" of re-employment than women. Table 2 shows that the factor education is significant but taking a look at different levels of education reveals that the hazard ratios are higher than before but none of the education levels are significant.

Table 2: Results of Cox models

	Cox proportional hazard model -2 Log Likelihood=28862268; Chi-square= 134861				Cox model with time-dependent covariate -2 Log Likelihood=2382559; Chi-square= 13658			
	B	SE	Sig.	Exp(B)	B	SE	Sig.	Exp(B)
Age	-,024	,000	,000	,976	-,027	,000	,000	,973
Age*Time					,0000071	,000	,000	1,0000071
Sex	,277	,002	,000	1,320	,268	,006	,000	1,307
Education			,000				,000	
Education(1)	-,865	,254	,001	,421	-,617	1,010	,541	,539
Education(2)	-,953	,252	,000	,385	-,632	1,005	,529	,531
Education(3)	-1,039	,251	,000	,354	-,756	1,002	,451	,470
Education(4)	-1,360	,252	,000	,257	-1,009	1,005	,315	,365
Education(5)	-1,421	,250	,000	,242	-1,180	1,000	,238	,307
Education(6)	-1,403	,250	,000	,246	-1,175	1,000	,240	,309
Education(7)	-1,259	,250	,000	,284	-1,018	1,000	,308	,361
Education(8)	-1,236	,250	,000	,290	-,963	1,000	,336	,382
Education(9)	-1,218	,250	,000	,296	-,971	1,000	,332	,379
Education(10)	-1,331	,250	,000	,264	-1,081	1,000	,280	,339
Education(11)	-,967	,250	,000	,380	-,736	1,000	,462	,479
Education(12)	-,797	,250	,001	,451	-,567	1,000	,571	,567
Education(13)	-,832	,253	,001	,435	-,525	1,007	,602	,591
Region			,000				,000	
Region(1)	,201	,006	,000	1,223	,190	,018	,000	1,209
Region(2)	,242	,007	,000	1,274	,225	,023	,000	1,252
Region(3)	,008	,006	,168	1,008	,034	,019	,070	1,035
Region(4)	-,087	,007	,000	,917	-,116	,021	,000	,890
Region(5)	,224	,006	,000	1,251	,219	,020	,000	1,245
Region(6)	,124	,007	,000	1,132	,130	,024	,000	1,139
Region(7)	,060	,006	,000	1,062	,059	,020	,004	1,061
Region(8)	,226	,006	,000	1,254	,200	,018	,000	1,222
Region(9)	,136	,010	,000	1,146	,125	,032	,000	1,133
Region(10)	,069	,007	,000	1,071	,066	,023	,003	1,068
Region(11)	,243	,008	,000	1,274	,268	,024	,000	1,307
Region(12)	,024	,006	,000	1,024	,024	,019	,204	1,025
Region(13)	,181	,006	,000	1,198	,155	,020	,000	1,168
Region(14)	,027	,005	,000	1,027	,015	,016	,346	1,016
Region(15)	,121	,007	,000	1,128	,116	,021	,000	1,123
Region(16)	-,021	,005	,000	,979	-,034	,016	,032	,967
Region(17)	,670	,006	,000	1,954	,666	,020	,000	1,946
Region(18)	,207	,007	,000	1,230	,182	,021	,000	1,200
Region(19)	,264	,007	,000	1,303	,275	,023	,000	1,317
Region(20)	,165	,005	,000	1,179	,162	,015	,000	1,176

Factor region is statistically significant with an exception of Brodsko-Posavska region and Osječko-Baranjska region. The results for region are similar to those from the previous model. Again, the best off are the unemployed in Istarska and Međimurska region, while the worst off are those from Karlovaška and Splitsko-Dalmatinska region.

If values of covariates sex, region and education are equal for two individuals and the age of i is one year higher than of j, the hazard ratio equals to $\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{e^{x_i'}\lambda_0(t)}{e^{x_j'}\lambda_0(t)} = e^{b_1 + b_2 \cdot T}$. Table 2 reveals that b_1=-0.027 and b_2=0.0000071. Thus, after one year of unemployment (T=365), hazard ratio equals to 0.976, which means that the risk is reduced with increasing age of the unemployed by 2.4% each year. After two years of unemployment (T=730) hazard ratio equals to 0.978, meaning that the hazard is reduced by 2.2% each year of unemployment. This implies that the differences in the "risk" of re-employment are diminishing when the duration of unemployment is increasing.

Conclusion

Cox proportional hazards model and the more appropriate Cox regression with time-dependent covariate have yielded similar results. It has been shown that men and younger unemployed are better off in the labour market in Croatia. The latter model has also revealed that differences among age groups of the unemployed become smaller as duration of unemployment spells increases. For the unemployed from Istarska region it takes the least time to find a new job. Istarska region is followed by Međimurska, Krapinsko-Zagorska and Požeško-Slavonska region. Regarding the level of education, the best chances of getting re-employed have those who have obtained doctorate, university degree, master's degree or have no education at all. On the other hand, it takes the longest time to find a new employment for those who have finished elementary school or have finished 3-year vocational school.

References

1. Bićanić, Ivo, and Zdenko Babić. (2006). *Survey of the Croatian Labour Market with Special Reference to Unemployment Related Issues of Human Capital Endowed Youth*. Manuscript: ASO project report.
2. Greene, William H. (2003). *Econometric Analysis*. New York: Prentice – Hall.
3. Hosmer, David H., and S. Lemeshow. (2003). *Applied Survival Analysis: Regression Modeling of Time to Event Data.* New York: Wiley-Interscience.
4. Klein, John P., and Melvin L. Moeschberger. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer Verlag.
5. Kleinbaum, David G. (2005). *Survival Analysis: A Self-Learning Text.* New York: Springer Verlag.
6. Therneau, Terry M., and Patricia M. Grambsch. (2001). *Modelling Survival Data: Extending the Cox Model*. New York: Springer Verlag.

MODELLING TIME OF UNEMPLOYMENT – A COX ANALYSIS APPROCH

Daniela-Emanuela Dănăcică
Ana Gabriela Babucea
Faculty of Economics, "Constantin Brâncuşi" University of Târgu-Jiu, Romania
danutza@utgjiu.ro
babucea@utgjiu.ro

Abstract: The aim of this paper is to present the results of ASO project "The Role of Education for the Duration of Unemployment" for one county of Romania, Gorj County. Using techniques to estimate models for duration data, like the Kaplan Meier method and Cox's proportional hazard model, we analyzed the influence of the level of education, age and gender for the duration of unemployment in Gorj County.

Key words: *unemployment, education level, gender, age, survival analysis*

Acknowledgments: In this paper are presented the results of the research within ASO grant "The Role of Education for the Duration of Unemployment", 2-36-2006, founded by the Austrian Science and Liaison Offices Ljubljana and Sofia on behalf of the Austrian Federal Ministry for Education, Science and Culture; it reflects only the author's view and the ASO Ljubljana and ASO Sofia are not liable for any use that may be made of the information contained hereby**.**

1. Introduction

Factors influencing the time of unemployment in Gorj County, Romania are analyzed in this paper. Using techniques to estimate models for duration data, like the Kaplan Meier method and Cox's proportional hazard model, we tried to answer to the following question: does the level of education, age and gender influence the duration of unemployment in Gorj County?

The empirical analysis is based on data offered by the National Agency for Employment of Romania (NEA). Although the Romanian research team filed an application to NAE in June 2006, in order to obtain data for the whole country, at the end of August 2006 we received only the database for Gorj County.

The paper is organized as follows: (1) Introduction, (2) Database description, (3) Kaplan Meier results, (4) Cox results and (5) Conclusions.

2. Database description

Our database has individual information about all the subjects registered at NAE during January 1st 2002- August 31st 2006.

The sample contains 80961 registrations, with information concerning the start date and the end date of unemployment spells, gender, age, and level of education and the reason of unemployment leaving for each registered person. Among the 80961 subjects, 33270 are women (41.1%) and 47691 men (58.9%). The minimum duration of unemployment spells is of 0 months and the maximum duration is of 57 months, with an average of 8.8 months and a median of 6 months. The coresponding distribution of the duration of unemployment spells is asymetrical, positively skewed with a skewness of 2.192 and kurtosis of 5.652. 53.6% of the total of registered persons (with the date of unemployment end) were in short and average duration of unemployment and 34.3% of the registered persons being in long duration of unemployment.

Regarding the *factor gender*, from the analysis of the distribution by gender and by duration of unemployment we noticed that the male unemployment in Gorj County for the analyzed period is higher than the female unemployment, and for the unemployed men it lasts longer than for unemployed women. Taking into account the fact that the number of women in Gorj County **who** are able to work is higher than the number of men, we draw the conclusion that differences between the number of women and men registered as unemployed are a direct consequence of the continuous reorganization, after 1992, of the mining sector, thermo energetic and oil tanker in Gorj County area, with negative effects on men belonging to all educational levels, employed in this sector.

Regarding the *factor age*, the average age of the persons registered in the database is of 32.58 years, and the median is of 32 years. Most of the unemployed registered in the database are aged between 15-35 years; the youngest subject is 15 years old and the oldest is 62. The high number of young unemployed registered in Gorj County shows that young people cannot find a job after finishing their studies, as the labor market in the county is not ready to receive them. The age distribution is positively skewed. The highest number of unemployed is represented by the young people aged between 15-34 years, representing 60.40% of the total unemployment (for whom the unemployment end is known) are young people aged between 15-34 years. The young graduate people cannot find a job after finishing their studies and become unemployed, but most of them stay unemployed for up to 6 months**.** Persons aged over 35 years are prone to long duration of unemployment. We can notice a positive correlation between age and duration of unemployment from table 1.

As for the factor level of education, 5.9% are university graduates, 0.5% are High School graduates, 2.4% graduated from post high schools, 20.2% graduated from special high schools, 15.0% graduated from theoretical high schools, 0.3% are special education graduates, 24.5% graduated vocational schools, 4.8% graduated from foremen schools, 5.5% are apprenticeship complementary education graduates, 18.1% graduated only secondary schools, the educational level for 2.1% is unfinished secondary schools, and 0.6% are without education. In data processing we have grouped persons by their educational level in 5 groups: group 0 - without graduated school, group 1- unfinished secondary school, secondary school, vocational school, apprenticeship complementary education, special education, with the maximum number of 10 years of study, group 2- theoretical high school, special high school, with 12 respectively 13 years of study, group 3 – foremen school and post high school with 14 years of study and group 4 corresponding to university education, (with short form – *college*), with 15, 16 and respectively 17 years of study. Unfortunately the received data do not provide information about the registered unemployed post university education graduates, (master's or doctorate graduates). Analyzing the distribution of registered unemployed person from our database we noticed a negative correlation between the variable level of education and duration of unemployment.

Table 1: *Descriptive statistics for the duration of unemployment spells (in months)*

	N	MEAN	STD. DEV.	95% CONFIDENCE INTERVAL FOR THE MEAN
Total	71145	8.82	8.74	(8.75, 8.88)
Factor Sex				
Male	47691	9.32	9.56	(9.23-9.41)
Female	33270	8.03	7.17	(7.94, 8.11)
Factor: Education				
Level 0 –	440	12.78	9.26	(11.92-13.65)
Level 1	35683	9.16	8.53	(9.08-9.25)
Level 2	25456	8.77	9.01	(8.66-8.88)
Level 3	5012	9.69	10.16	(9.41-9.97)
Level 4	4554	5.05	5.44	(4.89-5.21)
Factor: Age				
15-24 years	24015	6.03	6.27	(5.95-6.11)
25-34 years	18960	9.30	9.38	(9.17-9.44)
35-44 years	15338	10.53	9.31	(10.38-10.68)
45-54 years	11727	11.17	9.46	(11-11.35)
55-64 years	1105	12.47	10.56	(11.84-13.05)

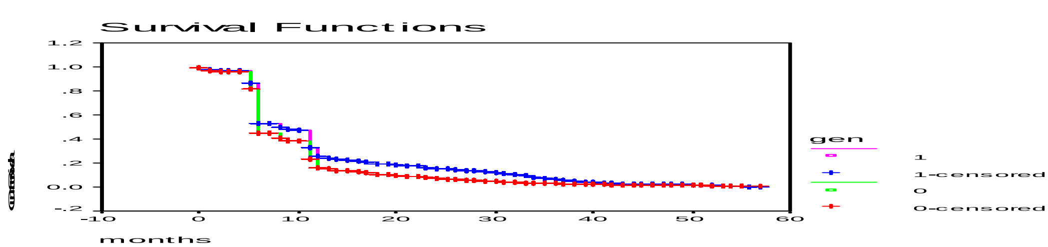
The result of the Kruskal –Wallis test allowed us to reject the null hypothesis. The differences noticed for each of the levels of the factors gender, age and level of education, regarding the mean duration of unemployment spells are statistically significant.

3. Kaplan Meier results

For our survey the pre-established event is employment, this event being ascribed the value 1; 61592 subjects from our database either did not achieve the event, or their track has been lost (they don't have the date of unemployment leaving), they have been censored at the right side, being ascribed the value 0.

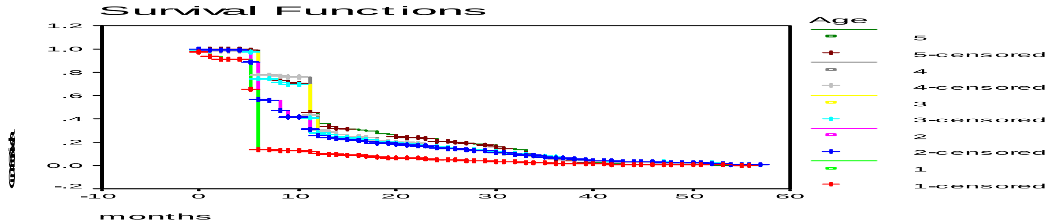
In figure 1 there is presented the survival curve for the women (0) and men (1) in the database. The results suggest a significant difference in probabilities of remaining unemployed between female and male; the median unemployment duration for female is 10 months and for male is 13 months. After 40 months the curves coincide.

Figure 1: *Survival function estimates for male and female unemployed*



In figure 2 there is presented the survival curve for the age groups 15-24 years, 25-34 years, 35-44 years, 45-54 and 55-64 years. Applying Kaplan-Meier analysis we have:

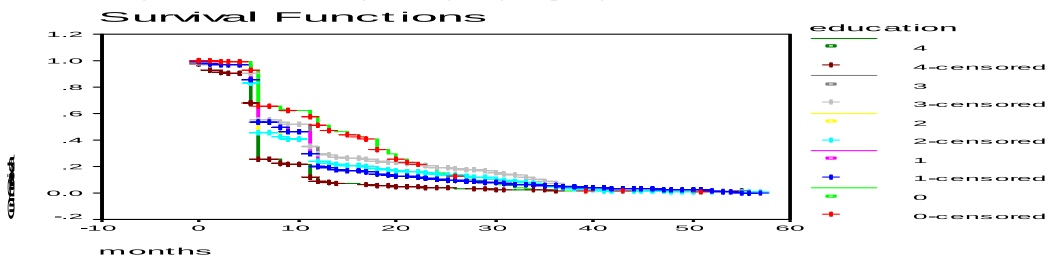
Figure 2**:** *Survival function estimates for the age groups 15-24 years, 25-34 years, 35-44 years, 45-54 and 55-64 years*



We can notice that the probability of remaining unemployed increased with age. The older persons are disadvantage on the labor market of Gorj County. The median unemployment duration for the age group 15-24 years is 6 months; for the age group 24-34 years is 8 months, for the age group 35-44 years is 11 months, for the age group 45-54 is 11 months and for the age group 55-64 is 11 months. The differences observed are statistically significant.

In figure 3 there is presented the survival curve for the level of education. Applying Kaplan-Meier analysis we have:

Figure 3: *Survival function estimates for the five groups of education*



We can notice that the probability of remaining unemployed is higher for the persons without education, followed by the persons with foremen school and post high-school and the lowest probability of remaining unemployed is for the persons with university education. We can also notice that after 40 months unemployment curves start to coincide and the educational level no longer influences the probability of finding a job. Testing the statistical signification for Kaplan Meier method presupposes the choice of one of the two hypotheses: the null hypothesis, which supposes that curves should be the same for two or several levels of a specified factor, or the alternative hypothesis, which supposes that they should be different. The result of the log rank test with Chi-Squared distribution under the null for all three factors, confirm the results derived graphically from the Kaplan-Meier estimates of the survival functions.

4. Cox results

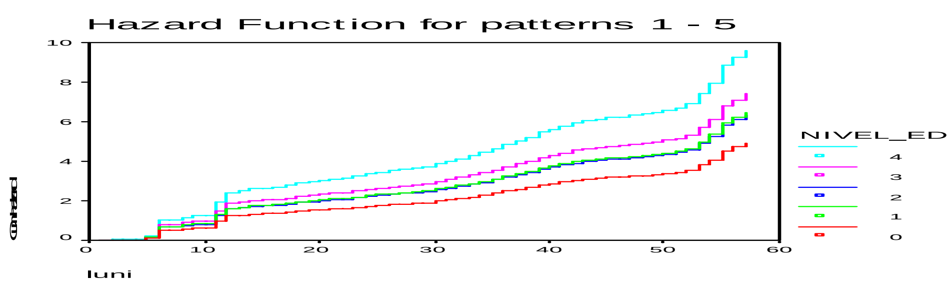
For our Cox analysis we used the SPSS 10.0 package. The reference category of a covariate was the last category, and the Enter method was selected. The results of the Omnibus test of the model coefficients allow us to reject the null hypothesis. In Table 2 are presented the results of the Cox regression analysis: B is the estimate vector of the regression coefficients, Exp(B_p) is the predicted change in the hazard for each unit increase in the covariate.

Table 2: *Variables in the equation*

	B	SE	WALD	DF	SIG.	EXP(B)	95,0% CI FOR EXP(B)	
							Lower	Upper
Age	-,002	,009	0.041	1	,000	,998	,981	1.016
Sex	-,151	,021	54,002	1	,000	,860	,826	,895
Education			428,441	4	,000			
Education(0)	-1,284	,143	80,382	1	,000	,277	,209	,367
Education(1)	-,745	,038	387,847	1	,000	,475	,441	,511
Education(2)	-,748	,039	368,430	1	,000	,473	,438	,511
Education(3)	-,701	,051	191,203	1	,000	,496	,449	,548

As we can notice from table 2 the hazard for the unemployment spell to end is 14% lower for the unemployed female that for the unemployed male. With increased age, the hazard is reduced by 0.2% each year. All levels of education have significant hazard ratios of less than 1; the hazard ratio is the lowest for the level 0 - without education - 0.277 and the highest for level 3 - foremen school and *post high school* (0.496). As we expected, the hazard ratio increased with higher levels of education. We can notice the fact that the hazard ratio for the level 1 - unfinished secondary school, secondary school, vocational school and apprenticeship complementary education, special education is slightly higher than for the level 2- theoretic high school, special high school. The cumulative hazard functions for different levels of education are presented in Figure 4.

Figure 4: *Cumulative hazard functions for different levels of education*



After we performed the log-minus-log plot and the partial residual plot we noticed the fact that the baseline hazards are not proportional and R squared linear indicates a positive correlation between partial residual and time, therefore the proportional hazard assumption does not hold. For the next step we used a model that includes the covariate age and the interaction term between time and age (Table 3). The results of the omnibus test for the model coefficients allow us to reject the null hypothesis. In table 4 are presented the estimates of the Cox model with time-dependent covariate.

Table 3: *Cox model with time-dependent covariate age*time*

	B	SE	WALD	DF	SIG.	EXP(B)
Age	-,006	,001	52,6782	1	,000	,999
Age*T	,001	,000	3,959	1	,000	1,006

Table 4: *Variables in the Equation*

	B	SE	WALD	DF	SIG.	EXP(B)	95,0% CI FOR EXP(B)	
							Lower	Upper
Age	-,006	,001	52,6782	1	,000	,999	,999	1,000
Age*T	,001	,000	3,959	1	,000	1,006	1,004	1,009
Sex	-,149	,021	52,369	1	,000	,862	,828	,897
Education			434,214	4	,000			
Education (0)	-1,296	,143	81,848	1	,000	,274	,207	,362
Education (1)	-,750	,038	394,014	1	,000	,473	,439	,509
Education (2)	-,752	,039	371,588	1	,000	,472	,437	,509
Education (3)	-,700	,051	192,057	1	,000	,497	,450	,548

We can notice from the table that the estimates for all the variables are almost similar to the Cox proportional hazards model from Table 2, the hazard ratios for the levels of education are slightly lower than before. The comparison between Cox proportional hazards model and Cox regression model with time-dependent covariate gives similar conclusions for all the three factors, sex, age and level of education.

5. Conclusions

Our analysis regarding the duration of unemployment spells gives the following results:
In respect of the duration of unemployment, persons with university education level remain unemployed for 5 months on the average, unlike persons without education, who remain unemployed for 13 months on the average, and persons with maximum 10 years of study, who remain unemployed for 9 months on the average. As for age, young people aged between 15-24 years remain unemployed for 6 months on the average, unlike the group 45-54 year or 55-64 year who remain unemployed for 11 respectively 13 months on the average. Regarding the variable gender, of 33270 women registered in our database 19.21%, leave unemployment by becoming employed and of 47691 men registered 27.21% leave unemployment by becoming employed. But the duration of unemployment is smaller for women with about a month on the average.

References

Chan Y.H (2004). *Biostatistics 203. Survival Analysis.* Singapore Med J 2004 Vol. 45(6): 249.
Greene, William H. (2003). *Econometric Analysis.* New York: Prentice-Hall.
National Agency for Employment (2006). *Statistics.* http://www.anofm.ro/
Kavkler Alenka, Borsic Darja (2006). *The Main Characteristics of the Unemployed in Slovenia*, Nase Gospodarstvo, Vol. 52, No.3-4.
Popelka John (2004). *Modelling Time of Unemployment via Cox Proportional Model*, paper presented at Applied Statistics 2005 International Conference, http://ablejec.nib.si/AS2005/Presentations.htm.
Zeileiss, Achim (2002). *Slides for the lecture Biostatistics.* www.ci.tuwien.ac.at/~zeileis/teaching/Biostatistics/.

DETERMINANTS OF UNEMPLOYMENT SPELLS IN SLOVENIA: AN APPLICATION OF DURATION MODELS[1]

dr. Alenka Kavkler, dr. Darja Boršič
Faculty of Economics and Business, University of Maribor, Slovenia
Razlagova 14, 2000 Maribor
alenka.kavkler@uni-mb.si, darja.borsic@uni-mb.si

Abstract: The paper shows how different factors influence the duration of unemployment spells in Slovenia. Significant effects of most of the factors were found by duration models such as Cox proportional hazard model and Cox regression with time-dependent covariate. It takes longer for women and older unemployed to get a job. An unemployed from Gorenjska or Goriška with higher professional education or university degree is the best off. While, those unemployed who live in Pomurska or Savinjska and have only elementary school have the worst chance of getting a new job.

Keywords: unemployment, survival analysis, Cox proportional hazards model, Cox regression model with time-dependent covariate.

1 Introduction

Survival analysis and duration models originate in biostatistics, where the survival time is the time until death or until relapse of an illness. During recent years these techniques have gained popularity in social sciences to model the length of unemployment spells and strike duration.

This paper studies the impact of the level of education on the length of unemployment spells in Slovenia, after adjusting for the factors age, sex and region. The data for our empirical investigation were obtained from the Employment Office of the Republic of Slovenia. The database consists of the unemployment spells completed between January 1st, 2002 and November 18th, 2005 and all of the ongoing spells on November 18th, 2005. For each of the unemployment spells, the start and end date and the variables sex, age, level of education and statistical region were made available to us. 442703 unemployment spells with positive duration are included in our database, out of which 94422 (21.3%) are censored. The maximal length of an unemployment spell is 13547 days. The empirical analysis was performed with the SPSS 13.0 program package.

2 Methodology: Basic notions

A comprehensive overview of the methods and models used in survival analysis is given by Therneau and Grambsch (2001) and by Klein and Moeschberger (2005). Let the random variable T denote the *survival time*. The distribution function of T is defined by the equation $F(t) = P(T < t)$ and measures the probability of survival up to time t. Since T is a continuous random variable, its density function $f(t)$ can be computed as the first derivative of the distribution function. The *survival function* S(t) denotes the probability to survive until time t or longer and is given by $S(t) = P(T \geq t) = 1 - F(t)$.

[1] This research was supported by a grant from the Austrian Science and Liaison Offices Ljubljana and Sofia. The paper reflects only the author's views. The Austrian Science and Liaison Offices are not liable for any use that may be made of the information contained therein.

The limit $\lambda(t)=\lim\limits_{\delta\to 0}\dfrac{P\left(t\le T<t+\delta\,|\,T\ge t\right)}{\delta}$ represents the risk or proneness to death at time t. The function $\lambda(t)$ is called the *hazard function* or the *failure rate* and measures the instantaneous death rate given survival until time t. Larger values of the hazard function can also be interpreted as higher potential for the event to occur. By integrating the hazard function over the interval $\left[0,t\right]$ one obtains the so-called *cumulative hazard function* $\Lambda(t)=\int_0^t \lambda(u)du$. It is easy to see that $-\log S(t)=\int_0^t \lambda(u)du$, therefore $S(t)=e^{-\int_0^t \lambda(u)du}$.

3 Cox proportional hazards model

The so-called *Cox proportional hazards model* is a semiparametric method of analyzing the effects of different covariates on the hazard function (Kleinbaum 2005, Hosmer and Lemeshow 2003). Assuming n individuals under observation, the Cox proportional hazards model is of the form $\lambda_i(t)=e^{x_i'\beta}\cdot\lambda_0(t)=c_i\cdot\lambda_0(t),\quad i=1,2,\ldots,n$, where $x_i=\left(x_{i1},x_{i2},\ldots,x_{ik}\right)'$ is the vector of k covariate values for the individual i, $\beta=\left(\beta_1,\beta_2,\ldots,\beta_k\right)'$ is the vector of regression coefficients, $\lambda_i(t)$ is the hazard function of the individual i and $\lambda_0(t)$ is *the baseline hazard*, which corresponds to an observation with $x_i=0$. The effect of the covariates on the hazard function does not depend on time, since the ratio $\dfrac{\lambda_i(t)}{\lambda_0(t)}$ is equal to the constant c_i. Consequently, the baseline hazard determines the shape of the hazard function. The ratio of the hazard functions of the individuals i and j, namely $\dfrac{\lambda_i(t)}{\lambda_j(t)}$, is called the *hazard ratio*. This quotient equals to $\dfrac{\lambda_i(t)}{\lambda_j(t)}=\dfrac{e^{x_i'\beta}\cdot\lambda_0(t)}{e^{x_j'\beta}\cdot\lambda_0(t)}=e^{(x_i-x_j)'\beta}$. Since the hazard ratio is independent of time, this is called *the proportional hazards assumption*. The interpretation of the hazard ratio is similar to the odds ratio interpretation for logistic regression. If a hazard ratio is lower than 1, it indicates decreased risk. While a ratio higher than 1 denotes increased risk. Suppose that the vectors of covariates x_i and x_j differ only in the value of the p-th covariate and only for one unit. In this case, the hazard ratio $\dfrac{\lambda_i(t)}{\lambda_j(t)}=e^{\beta_p}$ measures the change of the hazard function for a unit change in the p-th covariate (if the covariate is a numerical variable). The hazard ratio is said to be statistically significant at the given level, when its confidence interval excludes 1. In this case the null hypothesis that the variable is not related to survival can be rejected. This is the basis for the interpretation of the Cox regression results. By using the *Cox's partial likelihood estimator*, it is possible to estimate the parameter vector β without specifying and estimating the baseline hazard (see Greene (2003) for details).

3.1 Interpreting the Cox regression results

The Enter method was selected and all of the predictors were specified in the model simultaneously. The results of the omnibus tests of the model coefficients are given in Table

1. The score chi-square statistic (41685.761) and the likelihood ratio statistic given by -2 log likelihood (8356390.81) are asymptotically equivalent tests for the null hypothesis $H_0 : \beta = 0$, which has to be rejected in our case (df=23). The baseline hazard has been set to the hazard for female individuals with doctorate from Obalno-kraška region.

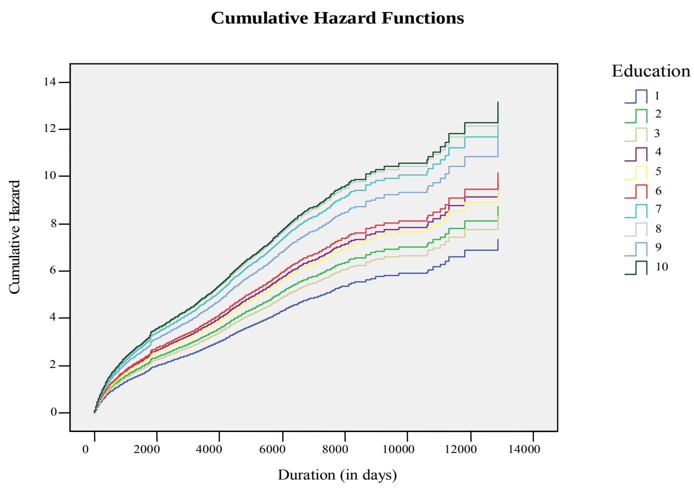
The results of the Cox regression analysis are presented in Table 1. All four variables are highly significant. The estimate of the regression coefficients vector β is denoted by B. From the column with the Exp(B) values, one can see that the hazard for the unemployment spell to end is 20.8% higher for the male unemployed than for the female unemployed. With increasing age, the hazard is reduced by 2.4% each year. The hazard rate for the higher professional education, university degree and master`s degree with the reference category doctorate is not significant at the 5% level. All other levels of education yield significant hazard ratios of less than 1 with a decreased risk for the unemployment spell to end. The hazard ratio is the lowest for the elementary school (0.558) and the highest for post-secondary vocational education (0.769).

Table 1: Results of Cox proportional hazards model

							95.0% CI for Exp(B)	
	B	SE	Wald	df	Sig.	Exp(B)	Lower	Upper
Age	-.025	.000	24847.951	1	.000	.976	.975	.976
Male	.189	.003	3021.850	1	.000	1.208	1.200	1.216
Education			8125.285	9	.000			
Elementary school	-.584	.070	68.970	1	.000	.558	.486	.640
2-year lower vocational education	-.412	.071	34.111	1	.000	.662	.577	.761
3-year lower vocational education	-.461	.072	41.016	1	.000	.630	.547	.726
Middle vocational edudation	-.298	.070	17.960	1	.000	.742	.647	.852
Secondary education	-.324	.070	21.257	1	.000	.723	.630	.830
Post-secondary vocational education	-.262	.071	13.603	1	.000	.769	.669	.884
Higher professional education	-.050	.071	.487	1	.485	.952	.828	1.094
University degree	-.011	.071	.026	1	.871	.989	.861	1.135
Master's degree	-.124	.082	2.316	1	.128	.883	.752	1.036
Region			5763.064	12	.000			
Pomurska	-.288	.010	895.972	1	.000	.750	.736	.764
Podravska	-.238	.009	774.366	1	.000	.788	.775	.802
Koroška	-.155	.012	179.799	1	.000	.857	.837	.876
Savinjska	-.283	.009	1010.164	1	.000	.753	.740	.766
Zasavska	-.225	.013	313.076	1	.000	.799	.779	.819
Spodnjeposavska	-.162	.011	204.528	1	.000	.850	.832	.869
JV Slovenia	-.224	.011	450.723	1	.000	.799	.783	.816
Osrednjeslovenska	-.118	.009	181.690	1	.000	.889	.874	.904
Gorenjska	.157	.009	273.670	1	.000	1.170	1.148	1.192
Notranjsko-kraška	-.095	.014	48.248	1	.000	.910	.886	.934
Goriška	-.096	.012	67.275	1	.000	.908	.888	.930

For regions, the results are always highly significant. The hazard for Gorenjska region is 1.17 times that of Obalno-kraška region. All other hazard ratios are less than 1, indicating decreased risk for the unemployment spell to end. In the most disadvantageous position in the labour market are the unemployed from Pomurska and Savinjska region with the hazard ratios 0.750 and 0.753, respectively. The cumulative hazard functions for different levels of education are given in Figure 1.

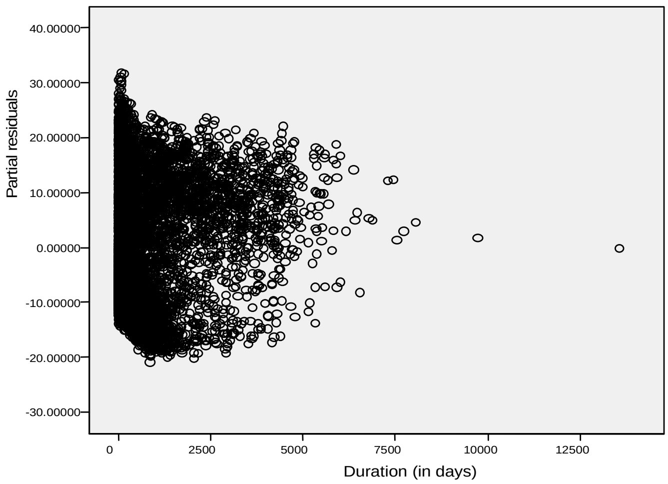
Figure 1: Cumulative hazard functions for different levels of education



4 Cox regression model with time-dependent covariate

The proportional hazards assumption is crucial for the Cox regression modelling approach. It can be examined graphically or by performing suitable statistical tests (Norušis 2005, Therneau and Grambsch 2001). It implies that the survival curves for different groups of individuals (with the same covariate values inside every group) do not cross. The hazard ratio should be the same for the unit change in the given covariate, independently of the initial covariate value. The proportional hazards assumption was examined graphically by scatterplot of the partial residuals (Figure 2). The partial residual for a given covariate at the k-th event is the difference between the observed value of the covariate at the case experiencing the k-th event and the conditional expectation of the covariate based on the cases still under observation when the k-th case fails. No patterns should be observed in this plot. The regression line that was added to the plot indicates a positive correlation between partial residuals and time, therefore the proportional hazards assumption does not hold.

Figure 2: Partial residuals for the variable age



One of the statistical tests for proportional hazards was performed in the time-dependent covariates setting. For each covariate, a separate model is fitted that includes the covariate and an interaction term between time and the covariate under inspection. If the proportional hazards assumption holds, the estimated coefficient of the interaction term in the obtained model with time-dependent covariate should not be significantly different from zero. The results are given in Table 2. The Cox proportional hazards model is not well-suited, as the interaction term is highly significant.

Table 2: Cox model with time-dependent covariate age*time

	B	SE	Wald	df	Sig.	Exp(B)
Age	-.035	.001	1706.785	1	.000	.966
Age*Time	.00002384	.000	479.295	1	.000	1.00002383

Consequently, an interaction term with time (age*time) was introduced into the equation and the obtained Cox regression model with a time dependent covariate was estimated. According to the omnibus tests of the model coefficients (the score chi-square statistic equals 41685.761 and the likelihood ratio statistic given by -2 log likelihood equals 8356390.81), the null hypothesis that all of the model coefficients are equal to 0 has to be rejected .

From Table 3 one can see that the estimates for the variables sex and region are similar to the Cox proportional hazards model. The hazard ratios for the levels of education are much lower than before, indicating that the unemployed with higher levels of education are in a much better position in the labour market. The interpretation of the results for the time-dependent variable age is different in this setting. If the age of the individual i is one year higher than for the individual j, while the values of the covariates sex, region and education are the same for both individuals, then the hazard ratio is equal to $\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{e^{x_i'\beta} \cdot \lambda_0(t)}{e^{x_j'\beta} \cdot \lambda_0(t)} = e^{b_1 + b_2 \cdot T}$. In our case, $b_1 =$ -0.035 and $b_2 = 0.00002248$. This means that after for example 1 year of unemployment ($T = 365$) the hazard ratio is eqal to $\frac{\lambda_i(t)}{\lambda_j(t)} = e^{-0.035+0.00002248 \cdot 365} = 0.974$ and after 2 years of unemployment ($T = 2 \cdot 365 = 730$) $\frac{\lambda_i(t)}{\lambda_j(t)} = e^{-0.035+0.00002248 \cdot 730} = 0.982$. Thus, the hazard ratio is time-dependent, since it increases with time. After 1 year of unemployment the hazard is reduced with increasing age of the unemployed by 2.6% each year and after 2 years of unemployment by 1.8% each year. In other words, the longer the unemployment spell lasts, the less pronounced are the differences between different age groups.

Table 3: Cox regression with time dependent covariate

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Age	-.035	.001	1628.590	1	.000	.966	.964	.967
Age*Time	.0000225	.000	420.696	1	.000	1.0000225	1.00002033	1.000
Male	.157	.016	102.209	1	.000	1.170	1.135	1.206
Education			406.346	9	.000			
Elementary school	-.927	.334	7.703	1	.006	.396	.206	.762
2-year lower voc. educ.	-.720	.335	4.626	1	.031	.487	.252	.938
3-year lower voc. educ.	-.892	.342	6.802	1	.009	.410	.210	.801
Middle vocational ed.	-.631	.334	3.576	1	.059	.532	.276	1.023
Secondary education	-.667	.334	3.992	1	.046	.513	.267	.987

Post-secondary voc. ed.	-.570	.337	2.855	1	.091	.566	.292	1.095
Higher professional ed.	-.467	.337	1.919	1	.166	.627	.324	1.214
University degree	-.340	.335	1.025	1	.311	.712	.369	1.374
Master's degree	-.626	.385	2.642	1	.104	.535	.251	1.138
Region			294.441	12	.000			
Pomurska	-.262	.044	35.470	1	.000	.769	.706	.839
Podravska	-.196	.039	24.839	1	.000	.822	.761	.888
Koroška	-.132	.052	6.344	1	.012	.876	.791	.971
Savinjska	-.283	.041	47.878	1	.000	.753	.695	.816
Zasavska	-.191	.057	11.180	1	.001	.826	.739	.924
Spodnjeposavska	-.174	.052	11.275	1	.001	.840	.759	.930
JV Slovenia	-.198	.048	17.210	1	.000	.820	.747	.901
Osrednjeslovenska	-.088	.040	4.748	1	.029	.916	.847	.991
Gorenjska	.182	.043	17.575	1	.000	1.199	1.102	1.306
Notranjsko-kraška	-.040	.061	.421	1	.516	.961	.853	1.083
Goriška	-.055	.053	1.074	1	.300	.946	.853	1.050

5 Conclusion

The results show that it takes longer time for women and older workers to get a job. The difference between Pomurska, Podravska and Savinjska region on one hand and Gorenjska and Obalno – kraška region on the other hand is obvious. The regions Gorenjska and Goriška are the most advantageous in the labour market. The unemployed from Pomurska and Savinjska region are in the worst position. Unemployed with higher levels of education are in a better position in the labour market. The risk of re-employment is the lowest for the unemployed with only elementary school, whereas the unemployed with higher professional education, university degree, master`s degree and doctorate have significantly higher hazard function values. The comparison of the Cox proportional hazards model and the Cox regression model with time-dependent covariate reveals similar conclusions. The model with time dependent covariate seems to be more appropriate when studying the impact of the level of education on the length of unemployment spells. Namely, this model sets more emphasis on obtaining a higher level of education, which on average guarantees relatively short unemployment spells.

References

1. Greene, William H. (2003). *Econometric Analysis*. New York: Prentice – Hall.
2. Hosmer, David H. and S. Lemeshow. (2003). *Applied Survival Analysis: Regression Modeling of Time to Event Data.* New York: Wiley-Interscience.
3. Klein, John P. and Melvin L. Moeschberger. (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer Verlag.
4. Kleinbaum, David G. (2005). *Survival Analysis: A Self-Learning Text.* New York: Springer Verlag.
5. Norušis, Marija J. (2005). *SPSS 14.0 Advanced Statistical Procedures Companion.* New York: Prentice Hall.
6. Therneau, Terry M. and Patricia M. Grambsch. (2001). *Modelling Survival Data: Extending the Cox Model*. New York: Springer Verlag.

ELABORATION OF THE UNEMPLOYMENT IN THE REPUBLIC OF MACEDONIA THROUGH DURATION MODELS

Dragan Tevdovski
Katerina Tosevska
Faculty of Economics – Skopje
University Ss. Cyril and Methodius – Skopje
Blvd. Krste Misirkov bb Skopje, Macedonia
e-mail: dragan@eccf.ukim.edu.mk
katerina@eccf.ukim.edu.mk

Abstract: The aim of this paper is to present some consideration on the influence of the level of education on the unemployment in the Republic of Macedonia. The analysis of the unemployment in the Republic of Macedonia is done on a dataset from the Employment Agency of the Republic of Macedonia, complying 422 527 observation in the period between January 1st 2002 and December 30th 2005. For the analysis we applied Kaplan-Meier and Cox duration models.

Keywords: unemployment duration, level of education, survival analysis, Kaplan – Meier model, Cox model

Introduction

The Republic of Macedonia with 36% rate of unemployment belongs to the group of countries with highest unemployment rates in Europe[1]. The unemployment in the Republic of Macedonia has structural characteristics, with considerable high rate of long-term unemployment and low level of education of the unemployed. The low level of economic growth in the last two decades and the structural inadequacy of the economic sectors have been the main reasons for high unemployment in the country. The central problem, actually, has been the lack of labor demand in the formal sector of the economy.

In this paper we are trying to determine the influence of several variables: the level of education, sex, age and region, on the duration of unemployment. The models used in this analysis are duration models, the nonparametric Kaplan – Meier and the parametric Cox models.

The Data

The empirical analysis in this research paper was done using data from the Employment Agency of the Republic of Macedonia. The period under consideration is between January 1st 2002 and December 30th 2005. We must point out that we have observed the unemployment spells and not the unemployed persons. The reason for this is because a certain person can enter and exit the unemployment status several times during the observation period.

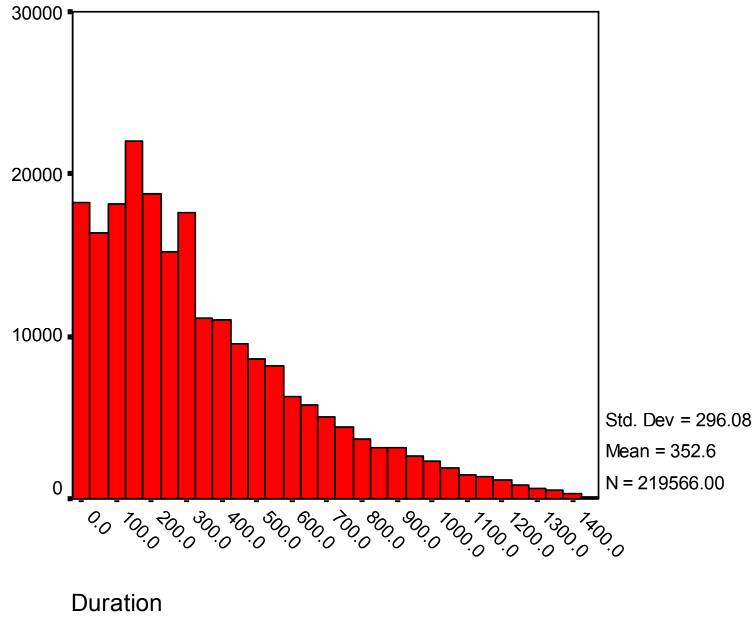
The characteristics that we have for the unemployment spells under analysis are:

- the duration of unemployment (start and end date of unemployment);
- the reason for ending the unemployment;
- the gender;
- the age;
- the level of education;
- and the statistical region.

[1] State Statistical Office of the Republic of Macedonia, www.stat.gov.mk

The total number of unemployment spells is 422,527, from which 219,566 spells have ended by December 30th, 2005. The rest of 202,961 unemployment spells are censored. Since the end of the unemployment spells has not occurred till the end of the study, it is only possible to estimate the lower bound of the survival time. The proportion of the ongoing unemployment spells as of December 2005 is significant (48.04%) in relation to the total number of unemployment spells occurred in the period under observation.

Figure 1: Histogram of the duration of unemployment spells (in days)



The histogram for the duration of unemployment spells is given in Figure 1. The average duration of the unemployment spells is 352.63 days. It indicates that the average time for waiting for a job is approximately one year. The median indicates that half of the unemployment spells have duration of unemployment lower than 277 days. High fluctuation around the mean duration of spells is presented by the dispersion measures (with skewness of 1.102 and kurtoisis of 0.742). The distribution of the duration of unemployment spells has positive asymmetry, with a long right tail.

Survival Analysis in Brief

In this paper we are trying to achieve the influence of several variables on the duration of unemployment spells. In order to determine our goal we are using survival analysis models. The modeling approach of this type of models answers the question, how the survival experience of a group of persons depends on the values of one or more explanatory variables. The *survival time* is the duration of unemployment. The dependent variable is the length of unemployment spell defined as the number of days between the starting date of job search to the date of its end. The *specific event* under observation is the end of an unemployment spell.

Survival analysis deals with the problem that often end of an unemployment spell is not observed, either because it is an event that does not occur in all cases or because observation time is limited. Therefore it is only possible to estimate the lower bound of the survival time. This type of censoring is called *right censoring*. Cases where the specific event is not observed are called censored observations. However, they must not be omitted from the analysis. The analysis uses the information that at least until the end of our observation period, indeed no “end of” unemployment occurred “in” some of the cases. Thus, for all cases, the analysis needs at least two variables:

- A *time variable* indicating how long the individual case was observed, and
- A *status variable* indicating whether unemployment duration case terminated with or without end of unemployment.

The duration models can be divided in two groups: non parametric and parametric. Non parametric model that we used in our analysis is the Kaplan – Meier model and parametric is the Cox model.

Kaplan – Meier Model

Basic element in the Kaplan – Meier model is the *survival function* $S(t)$. The survival function is defined as:

$$S(t)=\frac{\text{number of the unempl. spells surviving until time t or longer}}{\text{total number of unempl. spells observed}}$$

The survival function $S(t)$ denotes the probability of unemployment duration until time t or longer and is given by

$$S(t)=P(T\geq t)=1-F(t),$$

where T denotes survival time – duration of unemployment spell, and $F(t)$ is distribution function of T. $F(t)$ measures the probability time of survival – unemployment duration up to time t.

The product limit method of Kaplan and Meier is used to estimates S:

$$\hat{S}(t)=\prod_{t_i\leq t}\left(1-\frac{d_i}{n_i}\right),$$

where t_i is the survival time – duration of unemployment spell at the point i, d_i is the number of ends of unemployment spell up to time t_i and n_i is the number of cases of unemployment spells *at risk* just prior to t_i. The survival function is based upon probability that a case of unemployment spell survives at the end of a time interval, on the condition that the individual was present at the start of the time interval. The survival function is the product of these conditional probabilities.

The method is based on three assumptions:

- Censored cases of unemployment spells have the same prospect of survival as those who continue to be followed. This can lead to a bias that artificially reduces survival function.
- Survival prospects are the same for early as for late observations.
- The specific event – end of unemployment spell happens at the specified time.

In Figure 2, the Kaplan-Meier survival function estimates for the unemployed without education, for the four years secondary education and for university level education are displayed. We found these levels of education as the best representatives of low, medium and high level of the factor level of education. The probability to exit from unemployment decreases with educational level increases. Or, with other words the exit from unemployment increases with obtaining higher level of education. The median unemployment duration for the four years secondary education is 26,16% higher then the median unemployment duration for the university level education. The median unemployment duration for the ones without education is 35,98% higher then the median unemployment duration for the university level education.

Figure 2: Survival function estimates for the university level education, four years secondary education and without education

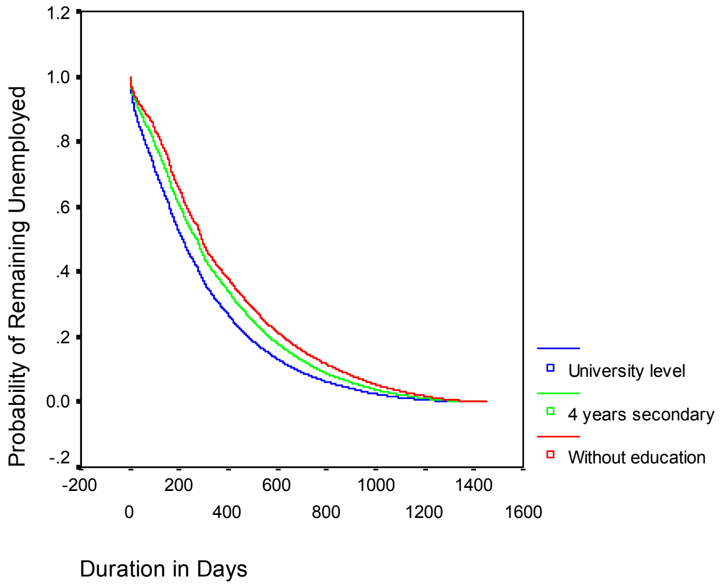
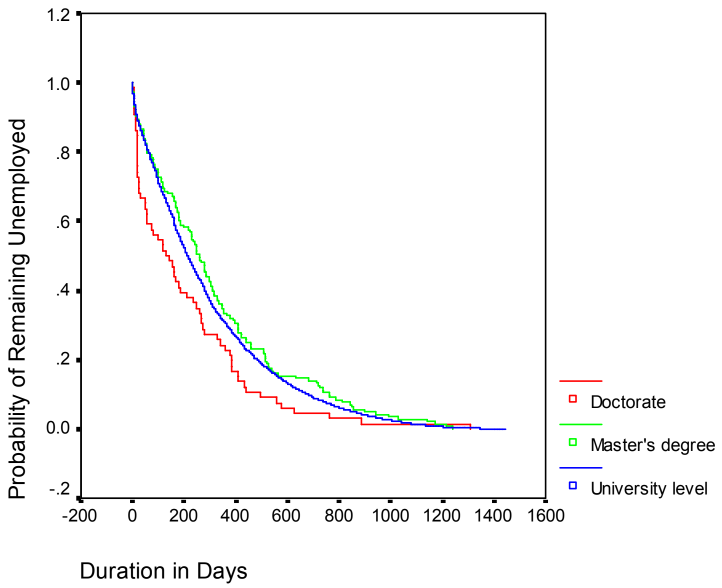


Figure 3: Survival function estimates for the university level education, the Master's degree and Doctorate



The effects to the unemployment duration of different levels of university education are presented in Figure 3. The Doctorate is found with the lowest probability of remaining unemployed in comparison to every other educational level. But, there is one more very important conclusion. The persons with Master's degree have worse position on the labor market than the ones with university level, which is actually one educational level lower. The probability of remaining unemployed is higher for the persons with Master's degree than for the one with university level education.

Cox Model

The regression method introduced by Cox is used to investigate several variables at a time[2]. It is also known as proportional hazard regression analysis. Cox's method does not assume a particular distribution for the survival times, but rather assumes that the effects of the different variables on survival are constant over time and are additive in a particular scale[3].

The limit

$$h(t) = \lim_{\delta \to 0} \frac{P(t \le T < t + \delta \mid T \ge t)}{\delta}$$

represents the hazard function. The hazard function is the probability that the employment will occur within a small time interval, given that the unemployment spell has lasted up to the beginning of the interval. It can therefore be interpreted as the risk (hazard) of employment at time t.

We determine the influence of the level of education on the length of unemployment spells in Macedonia, after adjusting for the factors age, sex and region. We express the hazard of employment at time t as:

$$h(t) = h_0(t) \exp\left(b_{sex}.sex + b_{edu}.education + b_{reg}.region\right)$$

The quantity $h_0(t)$ is the baseline or underlying hazard function, and corresponds to the probability of employment when all the explanatory variables are zero. In our analysis these are the male sex, Stip community and the doctorate for the factor level of education. The baseline hazard is thus the hazard for male individuals with doctorate from Stip community.

The regression coefficients: b_{sex}, b_{edu}, and b_{reg} give the proportional change that can be expected in the hazard, related to changes in the explanatory variables. They are estimated by a statistical method called maximum likelihood, using the computer program SPSS.

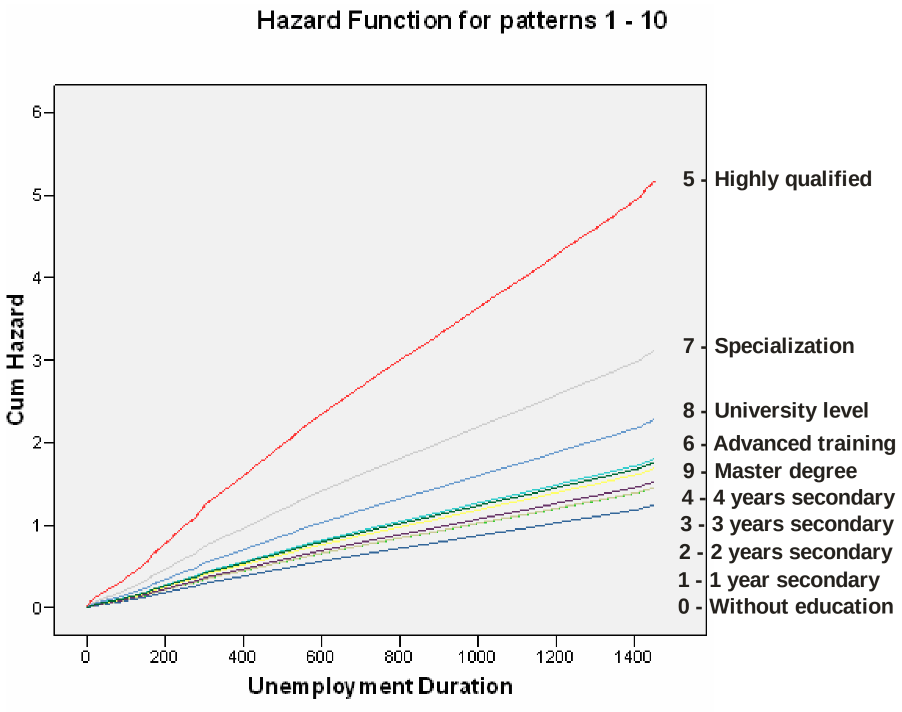
We perform Omnibus tests and we found that all model parameters are significant.

The hazard rate for the two years of secondary education, three years of secondary education, four years of secondary education, university level and master degree with the reference category doctorate is not significant at the 5% level. The hazard ratio is the lowest for the level one year of secondary education (0.707) and significantly highest is the level specialization (2.937). Generally, the hazard ratio increases with higher levels of education. But, we must stress that specialization as vocational education has advantage against all others academic levels of education on the Macedonian labor market. Very interesting is the fact that the hazard for the unemployment spells with master degree is lower than the hazard for the spells with university level, having in mind that master degree is one degree higher than the university level. The cumulative hazard functions for different levels of education are given in Figure 4.

Figure 4: Cumulative hazard functions for different levels of education

[2] Cox D. Regression Models and Life Tables, *J Roy Statist Soc B, 34*, 1974, pp.187-220

[3] Walters S. J., *What is a Cox Model?*,, www.evidence-based-medicine.co.uk, Vol. 1, No. 10, 2001, p.3



References:

1. Altman D. G. (1991) *Practical Statistics for Medical Research*, London: Chapman & Hall
2. Collett D. (1994) *Modelling Survival Data in Medical Research*, London: Chapman & Hall
3. Cox D. (1974) Regression Models and Life Tables, *J Roy Statist Soc B, 34*
4. Kavkler A. and Borsic D (2006) The Main Characteristics of the Unemployed in Slovenia, *Our Economy*, Vol. 52, No. 3-4, Faculty of Economics and Business, Maribor
5. Klein, J.P. and M. L. Moeschberger (1998) Survival analysis: Techniques for Censored and Truncated Data, New York: Springer Verlag.
6. Lacombiez L., Bensimon G., Leigh P. N. et al. (1996) Dose-ranging study of riluzole in arnyotrophic lateral sclerosis, Lancet
7. Nivorozhkina, L., E. Nivorozhkin and A. Shukhmin (2002) Modeling Labor Market Behavior of the Population of a Large Industrial City: Duration of Registered Unemployment, *EERC Working Paper* No. 01-08.
8. Peterson, A.V. Jr. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association;* 1977; 72: 854-858.
9. Stetsenko, S. (2003) On the Duration and the Determinants of Ukrainian Registered Unemployment. A Case Study of Kyiv, Master of Arts Thesis (EERC, Kiev).
10. Esser, M. And J. Popelka (2003) Analysis of Factors Influencing Time of Unemployment Using Time Analysis, Zbornik 12, medzinarodneho seminara V ypoctova statsitika, SSDS, Bratislava.
11. Walters S. J. (2001) *What is a Cox Model?*,, www.evidence-based-medicine.co.uk, Vol. 1, No. 10

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 9

Finance and Investment

COMOVEMENTS OF PRODUCTION ACTIVITY IN EURO AREA AND CROATIA

Nataša Erjavec*, Boris Cota* and Josip Arnerić**
*University of Zagreb, Faculty of Economics, Trg J.F.Kennedya 6, 10000 Zagreb, Croatia
**University of Split, Faculty of Economics, Matice Hrvatske 31, 21000 Split, Croatia
nerjavec@efzg.hr, bcota@efzg.hr and jarneric@efst.hr

Abstract: This paper tries to give an answer on the current degree of comovements of production activity in euro area and Croatia or in other words on business cycle synchronization in Croatia to the euro area cycle. It is known that participation in a currency union may itself lead to greater synchronization of business cycles. If the business cycles are sufficiently synchronized then the new member(s) can easy give up monetary and exchange rate policy independence.

Keywords: comovements, business cycle, VAR model, cointegration, impulse response function

1. INTRODUCTION

The common economic movements (comovements) that are occurring at the same time in different countries have received the attention of economic research for many years. The presumption of positive comovements can be higher degree of openness of economies, the integration of different economies in economic union and the deregulation of financial markets and the liberalization on international capital movements. Economic comovements are also important for an economic policy. If European economies are fairly synchronized, they are in position to suffer very little with a common economic policy. However if there are strong divergences, then different economic policies would be needed by the different countries.[1]

The purpose of this paper is to assess the current degree of comovements of production activity in euro area and Croatia. In fact we are interested in business cycle synchronization in Croatia to the euro area cycle. The benefits and costs of a currency union have been extensively analyzed in the literature. It is well known that participation in a currency union may itself lead to greater synchronization of business cycles.[2] The question also has to be asked whether the business cycles are sufficiently synchronized so that the new members can comfortably give up monetary and exchange rate policy independence. Therefore, when considering the appropriate timing of entry into the euro zone, satisfying the Maastricht criteria of nominal convergence of inflation, long term interest rates, fiscal deficit, public debt and exchange rate stability within ERM II are only one set of factors to be taken into account.

Artis, Kontolemis and Osborn (1997) found a strong association between the business cycles regimes in several European countries. Using a panel of thirty years of data for twenty industrial countries, Frankel and Rose (1998) find a strong positive relationship between trade integration and business cycle correlation. Therefore, to the extent that participation in a currency union increases trade integration, membership in a currency union will lead to more highly correlated business cycles. Rose (2000) finds that currency unions increase trade substantially and hence concludes that a country is more likely to satisfy the criteria for entry into a currency union *ex post* than *ex ante*.

---

[1] In that case independent monetary policies or independent exchange rate policies could be necessary to stabilize domestic economy.

[2] This is referred to as the endogeneity of the optimum currency area properties. The optimal currency area theory postulates that the benefits of a currency union depend on whether the countries contemplating to form a monetary union share certain common characteristics, called the optimum currency area properties.

However it should be pointed out that in the study we did not try to investigate the sources of shocks and the channels of transmission of business cycles from one country to another. Identifying the sources of shocks is important, because monetary policy can not deal with all types of shocks similarly, but if business cycles are synchronized, it is most likely that the countries are not subject to significant asymmetric shocks. The empirical evidence discussed in the literature shows that openness, trade integration and similarity of economic structures have a strong effect on international comovements.

The paper is organised as follows. Next section presents the methodology employed in the study. Section 3 gives data description and time series properties of the variables. The analysis of the short-term responses of the Croatian industrial production to the shocks to euro area production is presented in section 4. The final section concludes.

2. METHODOLOGY

First we examined time series properties of the variables included in the analysis. In order to find out if there are stochastic trends in the data, ADF (Dickey-Fuller, 1979) and KPSS (Kwiatkowski *et al.,* 1992) unit root tests were performed.[3]

After that, the existence of cointegration relationship between variables was tested using vector autoregressive (VAR) methodology proposed by Johansen (Johansen, 1988, and Johansen and Juselius, 1990). However, the existence of a long-term relationship between Croatian industrial production and industrial production in the euro area does not give sufficient information about the correlation of short-term cyclical movements. It is analysed on the basis of two variable VAR model (with the possibility of cointegration relationship included) through the effects of shocks in euro area production on the production in Croatia.

3. DATA

Data on industrial production indicator for the euro area (variable ***euroind***) and Croatia (variable ***croind***) was provided by International Financial Statistics. In the study we used monthly data from June 1994 to December 2005. The beginning of the empirical period has been chosen due to the fact that effects of the stabilization program brought in Croatia by the end of 1993 started to show only by the mid of 1994. The original series were rebased to be 100 in 2000, seasonally adjusted and ln- transformed. The results of ADF and KPSS unit root tests are presented in Table 1. The top part of table reports tests of stationarity of the levels of the variables and the bottom half of their first differences. The variables used in this study are given in the first column. Columns two to four contain test values for ADF tests with the information about adding a constant term or/and a deterministic trend to the model. The fifth column contains KPSS test values for testing trend stationarity of the variables. The sixth column gives KPSS test values for testing stationary around level. For each test the length of included lags is given in the square brackets after the test value. The appropriate number of lagged differences was determined by adding lags until a Lagrange Multiplier test fails to reject no serial correlation at 5% significance level.

[3] The difference between the tests is in the specification of the null hypothesis. The null of ADF test is nonstationarity and of KPSS test it stationary of the variable. KPSS test is usually used to confirm the conclusion suggested by other unit root tests.

Results suggest that variables contain unit root while their first differences are stationary. Therefore the series must be differenced once to obtain stationarity.[4] We proceed with the analysis by treating both variables as being I(1), *i.e.* integrated of order one.[5]

Table 1: Variables and unit root tests

Variable	ADF value Constant and trend included	ADF value Constant included	ADF value	KPSS value H_0 trend stationary	KPSS value H_0 stationary around a level
croind	-5,9561*(0)	-0,6673(3)	2.7630 (3)	0,56875**	3,4072**
euroind	-1.7452(4)	-1.0786(4)	2.2390(4)	0,5102**	2,6716**
First differences:					
Δcroind	-9,3026**(2)	-9,3468**(2)	-8,7048** (2)	0,0369	0.0421
Δeuroind	-4,6566**(3)	-4.6573**(3)	-4.0115**(3)	0,0668	0,1670

Notes: Δ is the first difference operator. One (two) asterisk(s) indicates a rejection of the Null at 5% (1%) significance level. The critical values for ADF tests were taken from Hamilton (1994) and for KPSS tests from Kwiatkowski and al. (1992).

Summary statistics on differenced variables are given in Table 2. As it was expected, the volatility of euro area is much smaller then that for Croatia.[6]

Table 2: Summary statistics on differenced variables

	Δcroind	***Δcroind***
Mean	0.003676	0.001653
Median	0.003104	0.001991
Maximum	0.086022	0.021385
Minimum	-0.070392	-0.017805
Std. Dev.	0.026536	0.007872
Skewness	-0.047705	-0.066860
Kurtosis	3.810099	2.776238
Jarque-Bera	3.825841	0.390715
Probability	0.147649	0.822541
Sum	0.507287	0.228161
Sum Sq. Dev.	0.096467	0.008489

Correlation matrix:

Δcroind
Δcroind

Δcroind
1.000000
0.124393

Δcroind
0.124393
1.000000

We continued the analysis by testing for the presence of cointegrating relationships between euro industrial production and production in Croatia. For the bivariate VAR model, the cointegration test based on Johansen's maximum likelihood procedure was performed, Table 3. The maximum and the trace eigenvalue tests indicate no cointegration at 10% significance level, *i.e.* a linear combination between variables that is stationary does not exist. This speaks against including an error correction term in the VAR estimated in the next section.

[4] In the case of Croatian industrial production the null of trend nonstationarity is rejected at 5% significance level. However, additional testing, as well as confirmation of a unit root hypothesis by KPSS tests, leads to the conclusion that the variable has a unit root.

[5] The variable is integrated of order ***d***, $X \approx I(d)$, if it needs to be differenced ***d***-times to become stationary.

[6] Lower overall volatility in a large economic area means that developments in industrial countries would tend to offset each other to some extent. Croatia is a small economy and its industrial base is not very well diversified. Therefore its volatility is expected to be higher than in larger economies (Korhonen, 2003).

Table 3: Johansen's test for the number of cointegrating vectors

H_0: r =	p-r	λ_{max}	λ_{trace}	λ	λ_{max} - 10% critical value	λ_{trace} - 10% critical value
0	2	2,47	2,90	0,0183	12,07	13,33
1	1	0,43	0,43	0,0032	2,69	2,69

Notes: Critical values for Johansen's test were taken from Osterwald-Lenum, (1992).

4. SHORT-TERM RESPONSES

In this section we analyse the short-term responses of the Croatian industrial production to shocks to euro area production. It is obvious that country have less to lose by joining the monetary union if the propagation shocks are similar to that in the euro area.

The impulse response functions (IRFS) were generated on the basis of two-variable VAR model in first differences. The lag length of the VAR model was determined by starting from *k*=24 and dropping lags sequentially until further reduction of the model was rejected by LR test. As a result, the optimum lag showed to be 11. The impulse response functions are the dynamic responses of each endogenous variable to a one-period standard deviation shock to the system. In this study we were interested in the responses of the indices of both the euro area and Croatia to a one standard deviation shock to the euro area index. Correlations of the resulting impulses were calculated for different time horizons. Euro area industrial production was ordered first in calculating the impulse responses because it is natural to assume that shocks to euro area production influence production in Croatia, not vice versa. The OLS estimates for Croatia indices (obtained from VAR model) are given in Table 4.

Table 4: Estimated OLS regression for Croatian industrial production (VAR model)

Variable	Coefficient	Variable	Coefficient
Constant	**0.0078**(0.0089)**	$euroind_{t-1}$	-0.1989 (0.5046)
$croind_{t-1}$	**-0.4335**(0.0000)**	$euroind_{t-2}$	-0.0363 (0.9097)
$croind_{t-2}$	**-0.4732**(0.0000)**	$euroind_{t-3}$	0.2175 (0.4904)
$croind_{t-3}$	**-0.1996 (0.0592)**	$euroind_{t-4}$	**0.6894* (0.0345)**
$croind_{t-4}$	**-0.3417**(0.0013)**	$euroind_{t-5}$	0.1088 (0.7418)
$croind_{t-5}$	-0.1135 (0.2848)	$euroind_{t-6}$	0.3499 (0.2855)
$croind_{t-6}$	-0.1765 (0.0878)	$euroind_{t-7}$	0.0450 (0.8902)
$croind_{t-7}$	-0.0037 (0.9713)	$euroind_{t-8}$	**0.8810** (0.0077)**
$croind_{t-8}$	-0.0764 (0.4492)	$euroind_{t-9}$	**-0.5871 (0.0673)**
$croind_{t-9}$	**0.23458* (0.0163)**	$euroind_{t-10}$	**-0.7204** (0.0222)**
$croind_{t-10}$	-0.02453 (0.7955)	$euroind_{t-11}$	**-1.0191** (0.0006)**
$croind_{t-11}$	0.16750 (0.0603)	RSS	0.045988
R-squared	0.444211	Adj. R-squared	0.326640

*Δy_{t-i} denotes the differenced series at lag i. Significance is reported in parentheses. ** indicates significance at 1% level and * at 5% level.*

The Croatian industrial production is influenced by its own lags in the previous four months while lags of euro area index needs some time (around eight months) to have an impact on Croatian index. Additionally, F-test for exclusion of lags of euro area index variable can not be rejected (F-statistic equals 2,6353 with p-value of 0,0052) which implies that euro area production is useful in predicting the Croatian production.

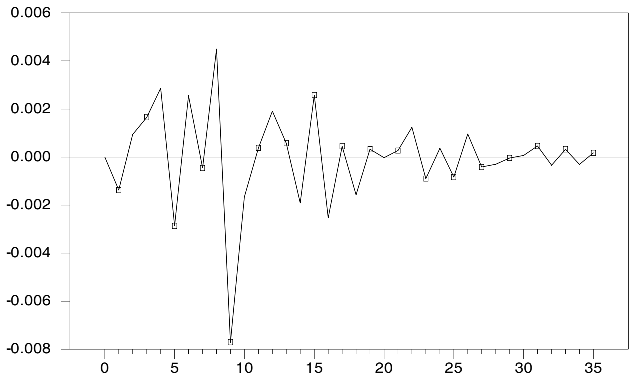
Table 5 gives some indicators for correlation of short-term business cycles in the euro area and Croatia. First, the table reports the correlation of impulse responses for the first 36 months. Although the effects had died out mostly after 18 to 20 months, we extended the period to three years. To remove the possibility of large outliers, we additionally calculated correlations for three-month moving average responses. In the last column the speed of adjustment coefficients are reported. It shows how fast the shock in euro area is transmitted in Croatian industrial production, *i.e.* how much of the 36-month accumulated shock has already been transmitted in 6, 12 and 24 months.

Table 5: Correlation of business cycles in Croatia

Correlation of impulse responses		**Speed of adjustment**	
Correlation	**0,1394**	**6 months**	1,8567
Correlation of MA impulse	**0,2915**	**12 months**	1,7889
		24 months	1,1805

Correlation coefficients show that effects of shocks are in the same direction and they are of moderate size. On the other hand, speed of adjustment coefficients indicate that there is a significant initial overshooting of the impulse responses. The effect gradually dies out and the most of it is transmitted within two years.[7] The graph of impulse responses of Croatian industrial production to one standard deviation shock to euro area production, Figure 1, supports the conclusion.

Figure 1: Impulse responses of Croatian industrial production to one standard deviation shock to euro area production



The results of variance decomposition for defined VAR model, Table 6, show how much of the forecast error variance of Croatian production index is explained by innovations in the euro area index at different forecast horizons.[8]

Table 6: Variance decomposition in % for Croatian production index

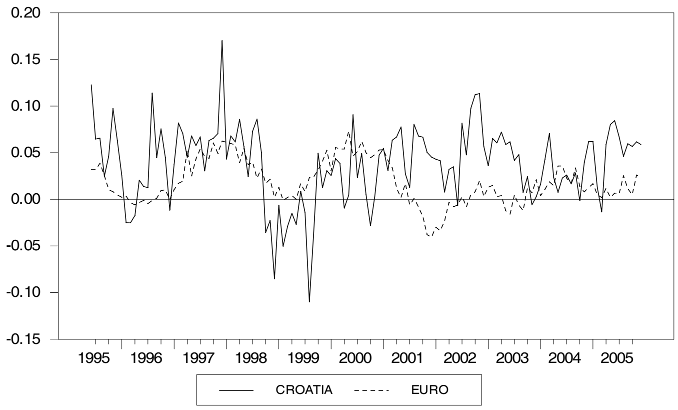
6 months	**12 months**	**24 months**	**36 months**
4,103	15,170	17,333	17,545

[7] The same situation is reported by Korhonen (2003) for the smallest accession countries in that time, namely; Estonia, Lithuania and Slovenia.
[8] Choleski decomposition was used with euro area production ordered first.

As it can be seen, the share of forecast error variance explained by euro area production is quite high. Innovations in the euro area index account for 15,17% of the forecast variance after one year, which increases to 17,545% in the next two years.

To assess the comovements of industrial production in euro area and Croatia we also plot the 12-month differences of the industrial indices, which correspond to annual growth rates, Figure 2. As it can be seen, peaks and troughs coincide more or less, although absolute changes are larger in Croatian industrial production.

Figure 2: Annual changes in industrial production, 1994:6-2005:12



5. CONCLUSION

The empirical findings suggest that the lower persistence of cycles in Croatia would have led to higher cyclical volatility and sensitivity to foreign shocks. Short-term responses of Croatian industrial production to shocks to euro area production show that euro area production is useful in predicting the Croatian production and that the share of forecast error variance of Croatian production index explained by euro area production is quite high. The obtained results also support the thesis of the high level of synchronization of Croatian industrial production with the euro area production. Namely, positive correlations of three-month moving averages of impulse shocks indicate increased correlation of business cycles. It is expected as we know that increased trade linkages would increase business cycle correlation. Industry in Croatia generates a large proportion of foreign trade, which is one of the main channels through which synchronization can occur.

REFERENCES:

Artist, M.J., Z.G. Kontolemis and D.R. Osborn (1997). Business cycles for G7 and European countries, *Journal of Business*, 70(2), 249-279.

Dickey, D. A. and W. A. Fuller (1979). Distributions of the estimators for autoregressive time series with unit root, *Journal of the American Statistical Association* 74, 427-431.

Frankel, J.A., Rose A. K. (1998). The endogenity of the optimum currency area criteria, *The Economic Journal* 108, 1009-1025.

Hamilton, J. D. (1994). *Time series analysis*, Princeton: Princeton University Press.

Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231-54.

Johansen, S. and K. Juselius (1990). Maximum likelihood estimation and inference on cointegration–with application to the demand for money, *Oxford Bulletin of Economics and Statistic* 52, 211-244.

Korhonen, I. (2003). Some empirical tests on the integration of economic activity between the euro area and the accession countries, *Economics in Transition*, 11(1), 177-196.

Kwiatkowski D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationary against the alternative of a unit root, *Journal of Econometrics* 54, 159-178.

Osterwald-Lenum, M. (1992). A note with fractals of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: Four cases, *Oxford Bulletin of Economics and Statistics* 54, 461-472.

Rose, A. K. (2000) One money, one market: Estimating the effect of common currencies on trade, *Economic Policy* 30, 7-33.

DIVERSIFICATION OF INVESTMENT IN BRANCHES

Roman Hušek, Václava Pánková[1]
husek@vse.cz, pankova@vse.cz
Univ. of Economics, Winstona Churchilla 4, 130 67 Praha 3, Czech Republic

Abstract: Uncertainty about future rewards from the investment generally have a negative effect on investment. Nevertheless, impacts of monetary uncertainties can differ according to the type of industry what is shown by anticipating monetary uncertainties as its permanent part. The fifteen branches of the Czech industry exhibit different responses to common macroeconomic determinants. To evaluate the branches exhibiting investment under/over the country´s average, an effectiveness measurement is proposed and performed in this paper by the help of a value added.

Keywords: investment, monetary uncertainties, effectiveness, frontier production function, panel data

1. Introduction

Optimization of an investment decision always involves uncertainty about future rewards, as a consequence of the fact that investment is sensitive to volatility and uncertainty over the economic environment. Usually, there is also an influence of irreversibility of investment decisions and of opportunity cost of possibility to wait rather then to invest. An economy which is mainly a capital acceptor usually operates under an investment – supporting policy. Nevertheless, though the important variables as inflation or exchange rate are controlled by a National Bank, their future values are not known. Monetary uncertainties rising from a volatility of relevant variables influence expected rewards from the investment and a negative effect of uncertainties on an investment inflow is generally assumed. But, a hypothesis is widespread that the impacts of monetary uncertainties can differ according to the type of industry.

Analyzing relevant data of the fifteen branches of the Czech industry, different investment - responses to common macroeconomic determinants were found. Formulating inflation, respective exchange rate uncertainties by the help of the concept of permanency supposed to subject an adaptive expectation process and using a panel data technique, a hypothesis of different impact of monetary uncertainties into Czech industrial branches was validated. The results are given as a part of Table 1 in which positive / negative deviation from the country´s average is represented by + / -.

There is a high demand for home and foreign investment in the Czech Republic the government of which performs an investment – supporting policy for years, that is why the minus results may imply a negative image and there is a question, how to evaluate such branches. As a tool for such a comparison, an effectiveness measurement is proposed and performed in this paper. A value added as a response to past investment is compared among the fifteen Czech industrial branches, both variables related to the number of employed persons. A technique of frontier production functions is used and the results are found showing that there are no straightforward consequences between investment and economic performance.

[1] Financial support of GACR 402/07/0049 and GA CR 402/06/0190 is gratefully acknowledged by the authors.

2. Investment and economic uncertainties

The investment behavior of a firm is supposed to be formalized as an optimizing problem of maximizing a firm´s value subject to a creation of wished capital stock. Optimizing of an investment decision also involves uncertainty about future rewards from the investment as its implicit constraint. As a consequence, there is an evidence that investment is sensitive to volatility and uncertainty over the economic environment. Usually, uncertainties are rising in monetary characteristics as inflation, interest rate or exchange rate. As the uncertainties in the economic environment are important determinants of investment, their nature and impact are in focus of recent studies.

Capital as one of the most important productive inputs can be characterised by a certain capital mobility, a degree of which is influencing an economic growth. A more open capital account shows out a higher productive performance than economies with restricted capital mobility. A small open economy with transition characteristics tends to be an acceptor of capital and aspires to attract big investments from abroad, that is why inflation uncertainty and / or exchange rate uncertainty can play an important role. Though both this variables are controlled by a National Bank, their future values are not known. It is generally assumed that such uncertainties have a negative effect on investment. Nevertheless, there is also an influence of facts as irreversibility of investment decisions and opportunity cost of possibility to wait rather then to invest. So, apart from transaction motives, a speculative motive can also take place here. That is why impacts of monetary uncertainties can differ according to the type of industry.

Theoretically, an uncertainty can be understood as a temporary component of relevant variable, the other component being its permanent part. An evidence of different effects from permanent and temporary changes is referred e.g. in [1]. An alternative approach introducing an uncertainty as a discount factor of future prices is given e.g. in [4].

3. Permanency as a part of a model

Monetary uncertainties are rising from a volatility of relevant variable, a value of which, though not observable, can be anticipated as its permanent part. The permanency is supposed to subject an adaptive expectation process, details e.g. in [3].

A variable X is supposed to split in two unobservable parts: a permanent one and a temporary one

$$X_t = X_t^P + X_t^T .$$

The permanent value is anticipated to subject an adaptive expectation process with a parameter λ as

$$\Delta X_t^P = X_t^P - X_{t-1}^P = \lambda(X_t - X_{t-1}^P) \quad \text{supposed} \quad 0 \le \lambda \le 1 .$$

It means

$$X_t^P = \lambda X_t + (1-\lambda) X_{t-1}^P$$

with the following interpretation. In year t a permanent value is a weighted average of an actual one and a previous permanent value. The previous permanent value follows the same schema, so

$$X_{t-1}^P = \lambda X_{t-1} + (1-\lambda) X_{t-2}^P \quad \text{a. s. o.}$$

By a substitution we then have

$$X_t^P = \lambda X_t + \lambda(1-\lambda)X_{t-1} + \lambda(1-\lambda)^2 X_{t-2} + \lambda(1-\lambda)^3 X_{t-3} + ... \quad (1)$$

what means that a current value has the greatest weight and the weights decline steadily by going back in the past.

Then, we can estimate a model

$$Y_t = \beta_0 + \beta_1 X_t^P + u_t \quad (2)$$

with parameters β and a disturbance u in variants. Constructing (1) under different choice of λ between zero and one (λ = 0.1, 0.2, …, 0.9), we compute (2). We than choose such a λ which produces a best fit of (2) according to the *R*-squared.

4. Application to the Czech industry

The fifteen branches of the Czech industry are studied. After a formalization of permanent inflation, respective exchange rate, their influence on investment in CR is estimated. As a common scheme

$$I = \beta_0 + \beta_1 X_1^P + ... + \beta_j X_j^P + \beta_{j+1} W_{j+1} + ... + \beta_k W_k + u \quad (3)$$

can be written with *j* permanent values of monetary variables with uncertainties in question and *k-j* other relevant exogenous variables as e.g.. level of wages, GDP per capita, a.s.o. In (3), *I* as an investment is an endogenous variable, β_0 is a constant and β_i´s are parameters of an econometric model. To demonstrate an influence of the common economic environment on different industrial branches the seemingly unrelated regression will be appropriate here to get individual sets of parameters under an assumption of correlated disturbances. Thus, an eventual diversity of monetary uncertainties impacts could be proved. Unfortunately, only four years of data (1999 – 2003) observations were available in the sources of ČSÚ (Czech Statistical Bureau), that is why the model was dramatically restricted and an other estimation method used. So, existing in the same economic environment .the investment in an industrial branch is exposed by the same but only one permanent value of an X^P variable

$$I = \beta_0 + \beta_1 X^P + u$$

and W´s are dropped. A technique of panel data (pooled regression, 60 observations, 1999 – 2002) was used which allows at least for distinguishing in a constant β_0.

For a quick survey, directions of deviations from a mean (in parentheses) are given in Table 1. Constructing permanent exchange rate CZK/EUR according to (1) with five lags, λ = 0.5 was found as giving optimal results (highest R- squared by valid t – tests). Repeating the same principal by using permanent inflation as an exogenous variable, λ =0.7

5. Efficiency measurement

Technical efficiency refers to maximizing of output from a given input vector or to minimizing of input subject to a given output level. Having a production function $Y = f(K,L)$, we explain an amount of production Y by the help of input factors capital *K* and labor *L*. As an actual production should be compared with a feasible technological maximum, a concept of a frontier production function is a useful tool allowing for a 'best practice' technology quantification.

Using a production function $Y = f(K,L)$ we understand the technical efficiency TE_i of the *i*-th subject as an output oriented measure defined by the relation

$$TE_i = y_i / f(K_i, L_i)$$

where y_i is current output of the subject and $f(K_i, L_i)$ is feasible technological maximum represented by frontier production function of the group of units to be compared. Evidently, $TE_i \le 1$.

Relevant frontier production function can be estimated by the help of the corrected ordinary least squares (COLS) method which is to be performed in two steps. In case of a Cobb – Douglas form which will be used in the further text, $Y = f(K,L) = AK^{\alpha}L^{\beta}$, first ordinary least squares (OLS) method is used to obtain consistent and unbiased estimates of the parameters α, β and consistent but biased estimate of the constant parameter, $\gamma = ln\, A$ in our case. Second, the biased constant γ is shifted up to bound all the observed data from above what is done by setting $\hat{\gamma}' = \gamma + max_i\{\hat{u}_i\}, \hat{u}_i$ being residuals from the OLS regression. The production frontier estimated by COLS represents in fact the 'best practice' technology. (For details see e.g. [5] or [2]).

Now, we have $y_i = \hat{y}_i\, exp(\hat{u}_i)$ and $f(K_i, L_i) = \hat{y}_i\, exp(max_i\{\hat{u}_i\})$. So

$$TE_i = \frac{\hat{y}_i\, exp(\hat{u}_i)}{\hat{y}_i\, exp(max_i\{\hat{u}_i\})} = exp(\hat{u}_i - max_i\{\hat{u}_i\}).$$

6. Application to the Czech industry

As a formalization of the proposal given in Paragraph 4, a relation

$$\frac{VA}{NP} = \alpha \left(\frac{IN_{-2}}{NP} \right)^{\beta} u \tag{4}$$

was estimated with *VA* for value added, *IN* for investment (mil. of CZK), *NP* means number of employed persons (in thousands). Having data related to 2001 – 2004, investment two years lagged, four panels with fifteen units were available. Using pooled regressions we have

$$\hat{\beta} = 0.144804 \quad (\text{st.er. } 0.06353) \quad \text{with} \quad F(1,58) = \quad 5.195\ [0.026]^{*} \tag{5}$$

what means that the F – test is valid only on 5%, not on the 1% significance level. Nevertheless, in the context of some parallel searching for a validating of expected consequences as e.g. growing investment means growing production or industrial production index, (5) was an encouraging finding.

Computing technical efficiency *TE* according to the Paragraph 2 instruction, four panels were averaged, that is why the most effective branch has not *TE* = 1 (as it will be naturally expected. Graphical results are given by the Figure 1. The x – axis represents industrial branches according to a official enumeration which corresponds with Table 1 in the further text.

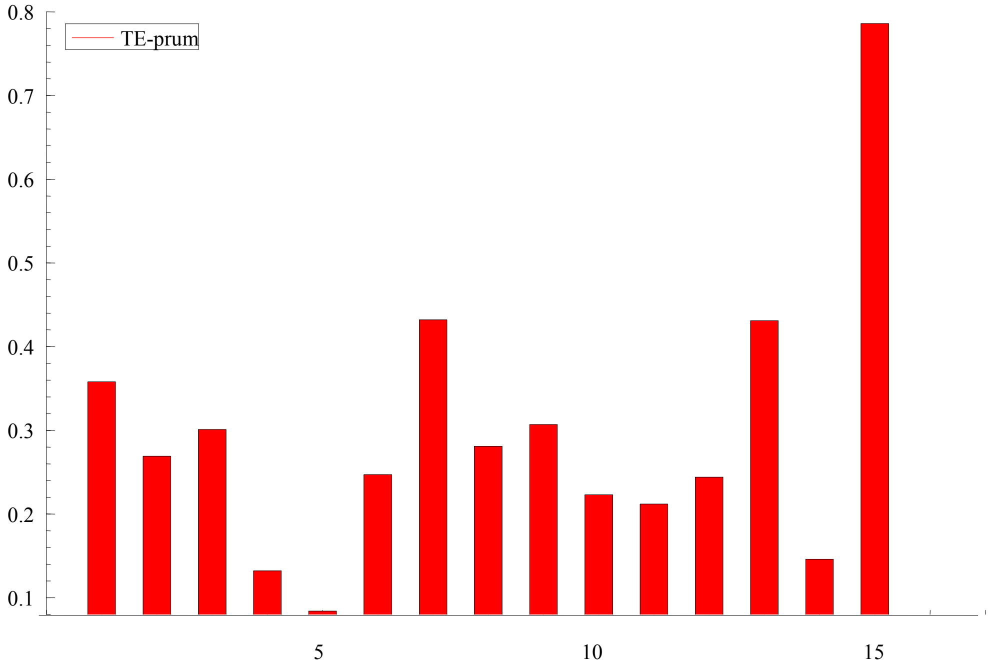


Figure 1

Enumerating the succession of the effectiveness and completing the Table 1 of relations to average investment, an evidence appears that there are no straightforward consequences between investment and economic performance. E.g. 11th row, machinery and equipment industry, has over – average investment according to inflation rate uncertainties, under – average investment when regarding with respect to exchange rate uncertainties and the branch exhibits a rather low effectiveness (12th place from 15) of investment when measured by VA created in the branch.

Conclusions

Investments in the Czech industry, especially foreign investments, are not coming equally to all industrial branches. It can be taken for granted, that the investors use to study the economic conditions and make their expectations about economic environment. Also their timing is well-considered. A hypothesis of different impacts of monetary uncertainties into industrial branches, which theoretically should be a consequence of such a behaving, seems to be validated for the Czech industry when followed in the beginning of this decade.

Looking for an effectiveness of investment per thousand of employed persons when measured by the help of following value added per thousand of employed persons, we can see that above-average investment need not be accompanied by an output oriented technical effectiveness. Electricity, gas and water supply branch is a very impressive example.

	industry	*Inflation rate (3433.26)*	*Exchange rate (2719.13)*	*TE*
1	Mining a. quarrying of energy prod. materials	+	-	**4**
2	Mining and quarrying except energy producing	-	-	**8**
3	Food products, beverages, tobacco	+	+	**6**
4	Textile and textile products	-	-	**14**
5	Wood and wood products	-	-	**15**
6	Pulp. paper and paper products, printing	-	-	**9**
7	Chemicals and chemical products	-	-	**2**
8	Rubber and plastic products	+	+	**7**
9	Other nonmetallic and mineral products	-	+	**5**
10	Basic metals and fabricated metal products	-	+	**11**
11	Machinery and equipment	+	-	**12**
12	Electrical and optical	+	+	**10**
13	Transport equipment	+	+	**3**
14	Manufacturing n.e.c.	+	+	**13**
15	Electricity, gas a. water supply	-	-	**1**

Table 1.

References

[1] Byrne, J. P., Davis, E.P.: Permanent and temporary inflation uncertainty and investment in United States, Elsevier, Economic Letters 85, pp 271 – 277, 2004

[2] Dlouhý, M., Pánková, V.: Hospital Performance and Trends. Vienna. In: RAUNER, Marion Sabine, HEIDENBERGER, Kurt (ed.). *Quantitative Approaches in Health Care Management.* Frankfurt am Main : Peter Lang, 2003, s. 189–199. ISBN 3-631-39009-2.

[3] Dougherty, Ch.: Introduction to Econometrics, Oxford Univ. Press, 1992

[4] Hallert, A. H., Peersman, G., Piscitelli, L.: Investment under Monetary Uncertainty: A Panel Data Investigation, Bank of England Working Paper, 2003

[5] Kumbakar, S.C. And C.A.K.: Lovell (2000) Stochasic Frontier Analysis, Cambridge Univ. Press

Data: ČSÚ - Czech Statistical Bureau

EXPECTED TRANSACTION COSTS AND THE TIME SENSITIVITY OF THE DELTA

Miklavž Mastinšek
Faculty of economics and business
University of Maribor
e-mail: mastinsek@uni-mb.si

Abstract : The paper deals with the problem of reducing and minimizing the expected proportional transactions costs. Higher order approximations of transaction costs are considered. The optimal hedge ratio is obtained and its dependence on the time sensitivity of the delta is given. The order of the hedging error is preserved.

Keywords: delta hedging, transaction costs

Introduction

The option valuation problem with transaction costs has been considered extensively in the literature. In many papers on option valuation with transaction costs the discrete-time trading is considered by the continuous-time framework of the Black-Scholes-Merton partial differential equation (BSM-pde) ; see e.g. [Le], [BV], [AP], [To]. Since in continuous-time models the hedging is instantenuous, hedging errors appear when applied to discrete trading.

It is known that transaction costs can be included into the Black-Scholes-Merton equation by considering the appropriately adjusted volatility; see e.g. [Le], [AP], [To], [Ma]. When the hedging is in discrete time, then over the time interval (t, t+Δt) the number of shares N is kept constant while at the time point t+Δt the number of shares is readjusted to the new value N'. Over that period of time the value S of the underlying changes to S+ΔS. The proportional transaction costs depend on the difference |N'-N| which is usually approximated by the gamma term, in general the largest term of the associated Taylor series expansion. In the case when other partial derivatives of delta are not small compared to the gamma, higher order approximations can be considered.

We will show that for a suitable choice of N which incorporates the time sensitivity of the delta, the expected proportional transaction costs can be reduced and minimized while the order of the hedging error can be preserved.

1.Transaction costs

We will assume (as in the above cited papers) that the number of shares N' at the point t+Δt is approximately equal to the Black-Scholes delta $N' = V_S(t+\Delta t, S+\Delta S)$. If N is also given by $N = V_S(t,S)$, then the proportional transaction costs at rehedging t+Δt are equal to:

$$TC = \frac{k}{2}|N' - N|(S+\Delta S) = \frac{k}{2}|V_S(t+\Delta t, S+\Delta S) - V_S(t,S)|(S+\Delta S) \qquad (1.1)$$

where k represent the round trip transaction costs measured as a fraction of the volume of transactions; for the details see e.g. [Le], [AP].

The absolute value of the difference ΔN=|N'-N| is usually approximated by $|V_{SS}\Delta S|$, in general the largest term of the Taylor series expansion.

If S=S(t) follows the geometric Brownian motion, then over the small noninfinitesimal interval of length Δt its change can be approximated by:

$$\Delta S = S(t+\Delta t) - S(t) \approx \sigma S Z\sqrt{\Delta t} + \mu S \Delta t, \qquad (1.2)$$

where Z is normally distributed variable with mean zero and variance one; in short Z~N(0,1), for the details see e.g. [Hu]. In that case the first order approximation of ΔN is given by the gamma term:

$$\Delta N = \left|N' - N\right| = \left|V_{SS}(t,S)\sigma S Z\sqrt{\Delta t}\right| \tag{1.3}$$

see e.g. [Le], [AP].

When other partial derivatives of the delta are not small compared to the gamma, then the following higher order approximation can be considered:

$$\Delta N = \left|N' - N\right| = \left|V_{SS}(t,S)\Delta S + V_{St}(t,S)\Delta t + \tfrac{1}{2}V_{SSS}(t,S)\Delta S^2 + O(\Delta t^{3/2})\right| \tag{1.4}$$

In that case the expected value of ΔN and thus the expected proportional transaction costs depend on other derivatives as well. When the delta $V_S(t,S)$ is more sensitive with respect to the time variable, the partial derivative $V_{St}(t,S)$ may be absolutely much higher than the gamma $V_{SS}(t,S)$.

We will show that for an adequate choice of N the expected transaction costs can be reduced or minimized while the order of the hedging error can be preserved.

Therefore our objective is to consider the discrete time adjusted hedge of the form:

$$N = V_S(t,S) + \alpha V_{St}(t,S)\Delta t \qquad 0 \le \alpha \le 1 \tag{1.5}$$

In this case the proportional transaction costs are equal to:

$$\begin{aligned} \Delta N &= \left|N' - N\right| = \\ &= \left|V_{SS}(t,S)\Delta S + (1-\alpha)V_{St}(t,S)\Delta t + \tfrac{1}{2}V_{SSS}(t,S)\Delta S^2 + O(\Delta t^{3/2})\right| \end{aligned} \tag{1.6}$$

For simplicity of exposition we assume that μ=0. Then ΔN can be approximated by:

$$\Delta N \approx D = \left|V_{SS}(t,S)\sigma S\sqrt{\Delta t}Z + (1-\alpha)V_{St}(t,S)\Delta t + \tfrac{1}{2}V_{SSS}(t,S)(\sigma S Z\sqrt{\Delta t})^2 Z^2\right|$$

We rewrite D more clearly as :

$$D = b\left|aZ + (1-\alpha)c + Z^2\right| \tag{1.7}$$

where

$$b = \tfrac{1}{2}V_{SSS}(t,S)(\sigma S Z\sqrt{\Delta t})^2 \qquad a = \frac{V_{SS}(t,S)\sigma S\sqrt{\Delta t}}{\tfrac{1}{2}V_{SSS}(t,S)(\sigma S\sqrt{\Delta t})^2}$$
$$c = \frac{V_{St}(t,S)\Delta t}{\tfrac{1}{2}V_{SSS}(t,S)(\sigma S\sqrt{\Delta t})^2} \tag{1.8}$$

The parameters a,b,c depend on S, σ, Δt, time to expiry T. However in most practical cases where Δt is small, the gamma term in (1.6) is predominant so that |a| is much larger than 1.

For instance let V(t,S) denote the value of a European call option. In that case from the BSM formula we get:

$$a = \frac{-2\sqrt{T}}{\sqrt{\Delta t}(d_1 + \sigma\sqrt{T})} \qquad d_1 = \frac{\ln S/S_0 + (\tfrac{1}{2}\sigma^2 + r)T}{\sigma\sqrt{T}} \tag{1.9}$$

where S_0 is the strike price,σ annual volatility, r the interest rate and T time to expiry.

Hence, when Δt is relatively small $|a|$ is usually relatively large. (especially when the options time to expiry T is not very small).

Example: when $\sigma=0.20$, $\Delta t=0.01$, $T=0.1$ $r=0$, $0.95<|S/S_0|<1.05$, then $|a|>7.3$
when $\sigma=0.20$, $\Delta t=0.01$, $T=0.04$ $r=0$, $0.95<|S/S_0|<1.05$, then $|a|>3.1$.

The terms with V_{St}, V_{SSS} in (1.6) are of the same order so that c is independent of Δ. If r=0, then

$$c = \frac{V_{St}(t,S)\sqrt{\Delta t}}{\frac{1}{2}V_{SSS}(t,S)(\sigma S\sqrt{\Delta t})^2} = \frac{-d_1+\sigma\sqrt{T}}{d_1+\sigma\sqrt{T}} \tag{1.10}$$

With the exception at $d_1=\sigma\sqrt{T}$ we thus have $c\neq 0$.

In order to reduce the expected value $E(\Delta N)$ of ΔN and thus to reduce the expected transaction costs (1.1) , the following minimization result will be considered:

Proposition 1 *: If $|a|>3$ and $c\neq 0$, then*

$$\min_{\alpha} E\left|aZ+(1-\alpha)c+Z^2\right| \tag{1.11}$$

is obtained, when $\alpha\approx 1$.

Proof If we introduce a new variable $Y=aZ+Z^2$, the minimization problem can be written as:

$$\min_{y} E\left|Y-y\right| \tag{1.12}$$

As known from stochastic analysis its solution is given by the median y_m of Y:

$$P(Y<y_m)=0.5 \tag{1.13}$$

In that case the following relation for Z holds: $P(z_1<Z<z_2)=0.5$,
where z_1 and z_2 satisfy the equation: $z^2+az-y=0$
Its solutions are approximately equal to :

$$z_1=-a-\frac{y_m}{a} \quad , \quad z_2=\frac{y_m}{a} \quad , \tag{1.14}$$

so that the value y_m can be readily obtained from the cumulative normal distribution function of Z. Since by assumption $|a|>3$, we find that $|y_m|<0.01$.
This means, if $|a|>3$ the minimum of (1.11) is obtained when $|(1-\alpha)c|\approx 0$. Hence if $c\neq 0$, the minimum can be achieved when $\alpha\approx 1$. □

Remark 1 In the case of lower $|a|$ the smaller value of α would be appropriate. For example if $|a|>2$, then $|y_m|<0.1$.
Remark 2 By the analogous analysis the case where $\mu\neq 0$ can be considered. The optimal value is then given by $N=V_S(t,S)+V_{St}(t,S)\Delta t+\mu V_{SS}(t,S)S\Delta t$.

Let us consider now the hedging error for the case where $|a|>3$ and $\alpha\approx 1$, so that $N\approx V_S(t,S)+V_{St}(t,S)\Delta t$
We will show that in this case the order of the hedging error is preserved .

2. The hedging error

We will consider now more closely the change of the value of a portfolio Π over the time interval $(t,t+\Delta t)$. Suppose that the portfolio Π at time t consists of a long position in the option and a short position in N units of shares with the price S :

$$\Pi = V - NS \tag{2.1}$$

We assume that the equivalent amount to the portfolio value can be invested in a riskless asset. Let us define the hedging error ΔH as the difference between the return to the portfolio value $\Delta\Pi$ and the return to the riskless asset.
By assumption the price of the underlying follows the geometric Brownian motion so that (1.2) holds. Then the following result can be obtained:

Proposition 2 *Let σ be the annualized volatility and r the annual interest rate of a riskless asset . Let V(t,S) be the solution of the Black-Scholes-Merton equation:*

$$V_t(t,S) + \frac{1}{2}\sigma^2 S^2 V_{SS}(t,S) + rSV_S(t,S)) - rV(t,S) = 0 \ , \tag{2.2}$$

If the approximate number of shares N held short over the rebalancing interval of length Δt is equal to:

$$N(t) = V_S(t+\Delta t, S) \approx V_S(t,S) + V_{St}(t,S)\Delta t \ , \tag{2.3}$$

then the mean and the variance of the hedging error is of order $O(\Delta t^2)$.

Proof Let us consider the return to the portfolio Π over the period $(t, t+\Delta t)$, $t\in[0, T_0-\Delta t]$, where T_0 is time at option expiry. By assumption over the period of length Δt the value of the portfolio changes by:

$$\Delta\Pi = \Delta V - N\Delta S \tag{2.4}$$

as the number of shares N is held fixed during the time step Δt.
First we consider the change ΔV of the option value V(t,S) over the time interval of length Δt. By the Taylor series expansion the difference can be given in the following way:

$$\begin{aligned} \Delta V &= V(t+\Delta t, S+\Delta S) - V(t,S) = \\ &= (V(t+\Delta t, S+\Delta S) - V(t+\Delta t, S)) + (V(t+\Delta t, S) - V(t,S)) = \\ &= V_S(t+\Delta t, S)(\Delta S) + \frac{1}{2}V_{SS}(t+\Delta t, S)(\Delta S)^2 + \frac{1}{6}V_{SSS}(t+\Delta t, S)(\Delta S)^3 + \\ &\quad + V_t(t+\Delta t, S)(\Delta t) + O(\Delta t^2) \end{aligned} \tag{2.5}$$

Note that the time change of the delta is implicitly included in (2.5):

$$V_S(t+\Delta t, S) = V_S(t,S) + V_{St}(t,S)\Delta t + O(\Delta t^2) \tag{2.6}$$

Hence by (2.4) it follows:

$$\begin{aligned} \Delta\Pi &= \Delta V - N(t)\Delta S = V_t(t+\Delta t, S)(\Delta t) + \left[V_S(t+\Delta t, S) - N(t)\right](\Delta S) + \\ &\quad + \frac{1}{2}V_{SS}(t+\Delta t, S)(\Delta S)^2 + \frac{1}{6}V_{SSS}(t+\Delta t, S)(\Delta S)^3 + O(\Delta t^2) \end{aligned} \tag{2.7}$$

When the number N of shares is equal to:

$$N = V_S(t+\Delta t, S) \approx V_S(t,S) + V_{St}(t,S)\Delta t \quad , \tag{2.8}$$

the ΔS term in (2.7) is eliminated completely. Hence we get:

$$\Delta\Pi = V_t(t+\Delta t,S)(\Delta t) + \frac{1}{2}V_{SS}(t+\Delta t,S)(\Delta S)^2 + \frac{1}{6}V_{SSS}(t+\Delta t,S)(\Delta S)^3 + O(\Delta t^2) =$$

$$= V_t(t+\Delta t,S)(\Delta t) + \frac{1}{2}V_{SS}(t+\Delta t,S)(\sigma^2 S^2 Z^2 \Delta t + 2\sigma\mu S^2 Z \Delta t^{\frac{3}{2}}) + \qquad (2.9)$$

$$+ \frac{1}{6}V_{SSS}(t+\Delta t,S)\sigma^3 S^3 Z^3 \Delta t^{\frac{3}{2}} + O(\Delta t^2)$$

By assumption Z~N(0,1) so that E(Z)=E(Z^3)=0 and E(Z^2)=1. Hence the expected value of $\Delta\Pi$ is equal to:

$$E(\Delta\Pi) = V_t(t+\Delta t,S)\Delta t + \frac{1}{2}V_{SS}(t+\Delta t,S)(\sigma^2 S^2 \Delta t + O(\Delta t^2) \qquad (2.10)$$

By assumption the amount Π can be invested in a riskless asset with an interest rate r. Thus over the rehedging interval of length Δt the return to the riskless investment is equal to:

$$\Pi r\Delta t = (V(t,S) - NS)r\Delta t =$$

$$= [V(t+\Delta t,S) - V_S(t+\Delta t,S)(S)]r\Delta t + O(\Delta t^2) = \qquad (2.11)$$

The hedging error is equal to: $\Delta H = \Delta\Pi - \Pi r\Delta t$ By (2.10) and (2.11) the expected value of ΔH is equal to:

$$E(\Delta H) = E(\Delta\Pi - \Pi r\Delta t) = V_t(t+\Delta t,S)\Delta t + \frac{1}{2}V_{SS}(t+\Delta t,S)(\sigma^2 S^2 \Delta t) -$$

$$-[V(t+\Delta t,S) - SV_S(t+\Delta t,S)]\rho\Delta t + O(\Delta t^2) \qquad (2.12)$$

Therefore, when V(t,S) satisfies at t+Δt the BSM equation, the hedging error can be written as:

$$\Delta H = \frac{1}{2}V_{SS}(t,S)(\sigma^2 S^2 (Z^2 - 1)\Delta t + 2\sigma\mu S^2 Z \Delta t^{\frac{3}{2}}) +$$

$$+ \frac{1}{6}V_{SSS}(t,S)\sigma^3 S^3 Z^3 \Delta t^{\frac{3}{2}} + O(\Delta t^2) \qquad (2.13)$$

Hence the mean and the variance of ΔH are zero to the order of $O(\Delta t^2)$:

$$E(\Delta H) = O(\Delta t^2) \quad \text{and} \quad V(\Delta H) = O(\Delta t^2). \quad \square \qquad (2.14)$$

References

[AP] Avellaneda M. and Paras A., »Dynamic hedging portfolios for derivative securities in the presence of large transaction costs«, Appl. Math. Finance 1 (1994), 165-194.

[BS] Black F. and Scholes M., »The pricing of options and corporate liabilities«, J. Pol. Econ. 81, (1973) ,637-659.

[BE] Boyle P. and Emanuel D., »Discretely adjusted option hedges«, J. Finan. Econ. 8 (1980), 259-282.

[BV] Boyle P. and Vorst T., »Option replication in discrete time with transaction costs«, J. Finance 47 (1992), 271-293.

[Hu] Hull J.C., Option, Futures & Other Derivatives, Prentice-Hall, New Jersey, (1997).

[Le] Leland H.E., »Option pricing and replication with transaction costs«, J. Finance 40 (1985), 1283-1301.

[Ma] Mastinšek M. "Discrete-time delta hedging and the Black-Scholes model with transaction costs", Math. Meth. Oper. Res. 64 (2006), 227-236.

[Me] Merton R.C., »Theory of rational option pricing«, Bell J. Econ. Manag. Sci. 4 (1973), 141-183.

[To] Toft K.B., »On the mean-variance tradeoff in option replication with transactions costs«, J. Finan. Quant. Analysis, Vol. 31, 2 (1996), 233-263.

THE MODEL FOR OPTIMAL SELECTION OF BANKNOTES IN THE ATMs

Gregor Miklavčič[1], Marko Potokar[2] and Mirjana Rakamarić Šegić[3]
[1]Bank of Slovenia, Slovenska 35, 1000 Ljubljana, Slovenia
E-mail: gregor.miklavcic@bsi.si
[2]Bankart d.o.o., Celovška 150, 1000 Ljubljana, Slovenia,
E-mail: marko.potokar@bankart.si
[3]Politechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia
E-mail: mrakams@veleri.hr

Abstract: Cash is still the most important and popular payment instrument in Slovenia and in EU, as well. Slovenia successfully introduced the euro at the beginning of this year and in relation to this, the adaptation of the ATMs (Automated Teller Machine) played a very important role. In Slovenia, the ATMs are currently operating with €10 and €20 banknotes. The purpose of this paper is to reconsider this variant and also include other variants with denominations from €5 to €100. Hence, we built a model for optimal selection of banknotes in the ATMs. All the calculations are based on the real (life) data. According to the results the authors suggest inserting €50 banknotes in the ATMs in Slovenia and some other practical improvements, as well.

Key words: ATMs, banknotes, modelling, optimal quantity breakdown.

1. INTRODUCTION

In this article we are presenting the model for optimal selection of banknotes in the ATMs and the pros and cons of introducing the €50 banknotes in the ATMs in Slovenia. Currently, we have €10 and €20 banknotes in the ATMs in Slovenia and in the future we should also consider the possibility of putting the €50 banknotes in the ATMs. In the paper we will present the results of four different variants with €10, €20 and €50 banknotes, as well. We built this model on the empirical experiences. The main goal of this model is to assess the different variants in the ATMs and try to find the best solution for Slovenia. The model is universal which means that it can be applied in different countries.

2. PRESENTATION OF THE MODEL FOR OPTIMAL SELECTION OF BANKNOTES IN THE ATMs

Before we introduce the quantitative model, we will present three views on the basis of which we can decide, which denominations we will put in the ATMs (there could be even more views; Drehmann, 2002).

The first view is the **bank's view**. The commercial banks wish to insert denominations with a high face-value and the algorithm, which minimises the total number of banknotes issued via ATMs. The reason for this is to minimise the cost of filling the ATMs.

The next view is the **central bank's view**. The goal of the Bank of Slovenia is that the quantity breakdown of issued banknotes via ATMs is similar to the "optimal" quantity breakdown calculated for Slovenia, in order to have a more rational supply of banknotes in Slovenia.

The third view is the **economy's viewpoint** (e.g. supermarkets, petrol stations, hotels, restaurants, …). The economy wishes that the ATMs issue neither denominations with too high face-value nor with too low face-value.

The article deals only with the first two views (bank's view and central bank's view).

We developed the model for optimal selection of banknotes in the ATMs using Microsoft Excel software and programming language Visual Basic. When defining the model we have to take into consideration the following assumptions:

1. We exclude the €5 banknote from the model, due to some improper technical properties of the banknote.
2. "Optimal" quantity breakdown of euro banknotes is given.
3. The probability mass functions of the amounts of withdrawals are given.
4. The number of boxes in the ATMs is between 2 and 4.
5. The amounts of withdrawals from the ATMs are distributed in the interval from 10 to 500 EUR.

"Optimal" quantity breakdown of euro banknotes is given and it is presented in table 1.

Table 1: "Optimal" ratios between individual banknotes from the central bank's view

€10 : €20 = 1 : 1.33
€10 : €50 = 1.55 : 1
€20 : €50 = 2.06 : 1
€10 : €20 : €50 = 1 : 1.33 : 0.645

Source: Miklavčič, 2006

The probability mass functions of the amounts of withdrawals are based on the empirical data. In table 2 we gathered the data about the total number of withdrawals and the average withdrawal from the ATMs for the period from 2000–06. We extrapolated the data and assessed that the average withdrawal from the ATM in 2007 will be approx. 70 EUR. The calculated values will be used in the next step when the probability mass functions of the amounts of withdrawals from the ATMs will be assessed.

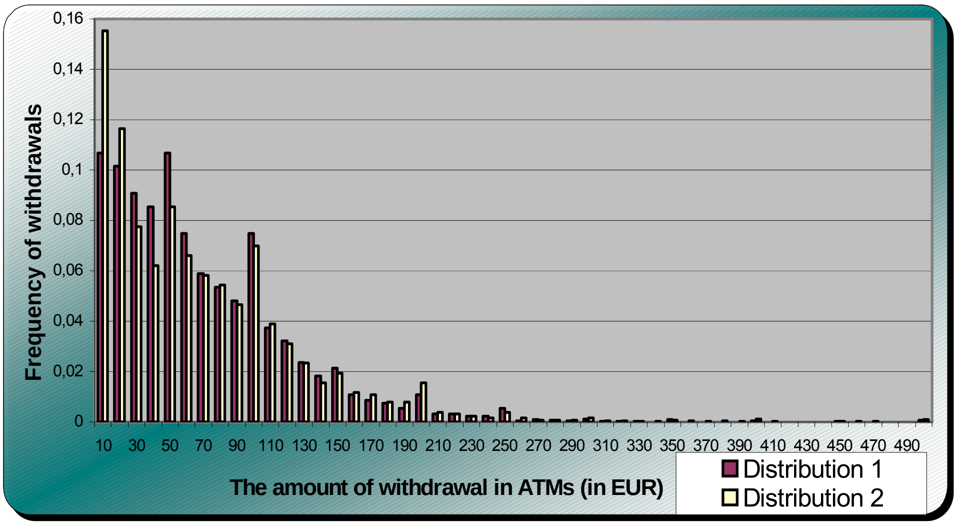
Table 2: Data on the average withdrawal and the number of withdrawals in Slovenia

Year	Number of ATMs	Number of withdrawals (in 1.000)	Value of withdrawals (in million SIT)	Average withdrawal (in EUR)
2000	865	41,048	425,016	**43.21**
2001	1,027	46,734	566,099	**50.55**
2002	1,095	52,160	642,742	**51.42**
2003	1,240	58,736	770,682	**54.75**
2004	1,389	63,700	892,207	**58.45**
2005	1,490	66,485	983,024	**61.70**
2006	1,522	64,160	1,010,028	**65.69**

Source: Bank of Slovenia, 2007

The probability mass functions of the amounts of withdrawals from the ATMs are shown in figure 1 and will be used later in the model. The average withdrawal from the ATMs is 70 EUR and is considered in both probability mass functions (Jamnik, 1987).

Figure 1: Histogram of probability mass function of the amounts of withdrawals from ATMs



Source: own calculations, 2007

The model for optimal selection of banknotes in the ATMs from bank's viewpoint and for four variants should be written as follows (Miklavčič, 2006):

$$f_{i,j,v} = \sum_{m=1}^{50} \left(g_{j,v}(10m) \times h_v(10m) \times p_{i,v}(10m) \right), \qquad (1)$$

where:

$f_{i,j,v}$ = expected total number of banknotes, required for distribution i (i = 1 and 2), denomination j (j = 10, 20 and 50) and variant v (v = 1, 2, 3, 4) for paying the amounts $10m$ (m = 1, 2, …, 50),

$10m$ = the amount of withdrawal from the ATMs in euro,

$g_{j,v}(10m)$ = the share of each denomination j for individual withdrawal of $10m$ and variant v,

$h_v(10m)$ = the minimum number of banknotes, required for the withdrawal of $10m$, and variant v,

$p_{i,v}(10m)$ = probability mass function of the distribution i, withdrawal $10m$ and variant v,

$h_v(10m) \geq 1$, $0 \leq p_{i,v}(10m) \leq 1$, $0 \leq g_{j,v}(10m) \leq 1$,

$g_{10,v}(10m) + g_{20,v}(10m) + g_{50,v}(10m) = 1$.

We can write down the ratios between different denominations of euro banknotes, standardised with regard to $f_{1,10,4}$, as follows (taking into consideration the assessed total number of banknotes, required for the first distribution, fourth variant with the €10, €20 and €50, from the bank's viewpoint and summed up for all withdrawals from the ATMs):

$$f_{1,10,4} : f_{1,20,4} : f_{1,50,4} = 1 : 1.74 : 2.53 = 19 : 33 : 48. \qquad (2)$$

In the next section we will look at the results of the model for all four variants.

3. THE RESULTS OF THE MODEL FOR FOUR VARIANTS

With the denomination of €10, €20 and €50 banknotes we formed the following four variants: (1) €10 and €20, (2) €10 and €50, (3) €20 and €50 and (4) €10, €20 and €50. We will present the results in this section with the model for optimal selection of banknotes in the ATMs.

3.1 First variant with the denominations of €10 and €20

As table 3 shows, the expected total number of issued banknotes for €10 and €20 via ATMs in 2007, in the case of the first distribution and from the **bank's view** (algorithm that minimises the total number of issued banknotes for given distribution *i*) is 303.0 million banknotes. The ratio between the denominations equals $f_{1,10,1} : f_{1,20,1}$ **= 14 : 86 = 1 : 6.1**. In the case of the second distribution we get quite similar results, where the expected total number of issued banknotes is 303.8 million banknotes, the ratio between the denominations remains the same. As we could see later, the difference between the first and the second distribution does not have any significant effect on the results of the model.

The second view is the **central bank's view**. The expected total number of issued banknotes in 2007 and in case of the first distribution is 361.6 million banknotes, the ratio between denominations equals **€10 : €20 = 44 : 56 = 1 : 1.3**. In order to achieve this ratio, we have to change the algorithm of issuing banknotes. The ATMs pay out the amounts from 10 EUR to 50 EUR only with €10 banknotes (e.g. the amount of 50 EUR is paid out with five banknotes for €10). For the amounts that are higher than 60 EUR we applied the principle of minimum number of issued banknotes (**bank's view**). The presented option is only one of many possible options, but it shows the difference in the number of issued banknotes from different views. So, the difference equals 58.6 million banknotes, which represents an increase of 19 %.

We calculated the ratios between both denominations, taking into account the restrictions from the number of boxes in the ATMs, as well. See table 3 for details (columns 6–11). The interpretations of the results are the same as before.

Table 3: The results of the model with €10 and €20 banknotes (in million banknotes)

	BANK'S VIEW		CB VIEW		TWO BOXES		THREE BOXES		FOUR BOXES	
	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio
Distr. 1	303.0	14:86	361.6	44:56	377.1	50:50	340.9	35:65	318.5	23:77
Distr. 2	303.8	14:86	359.3	43:57	377.8	51:49	336.4	32:68	319.5	23:77
Desired ratio €10 : €20				**43 : 57**		**50 : 50**		**33 : 67**		**25 : 75**

Source: own calculations, 2007

3.2 Second variant with the denominations of €10 and €50

The next variant is with the denominations of €10 and €50. In table 4 we can see that from the bank's viewpoint in 2007 we require in total 230.2 million banknotes and the ratio between the denominations is $f_{1,10,2} : f_{1,50,2}$ **= 64 : 36 = 1 : 0.6**. If we compare both distributions, we may see once again that there is practically no influence on the calculated ratio between the two banknotes.

Table 4: The results of the model with €10 and €50 banknotes (in million banknotes)

	BANK'S VIEW		CB VIEW		TWO BOXES		THREE BOXES		FOUR BOXES	
	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio
Distr. 1	230.2	64:36	230.2	64:36	IMPOSSIBLE		230.2	64:36	288.9	76:24
Distr. 2	226.6	63:37	226.6	63:37	IMPOSSIBLE		226.6	63:37	275.4	74:26
Desired ratio €10 : €50				**61 : 39**		**50 : 50**		**67 : 33**		**75 : 25**

Source: own calculations, 2007

In case of two boxes and both distributions, it is impossible to reach the ratio of **€10 : €50 = 50 : 50**, because we have already maximised the number of issued €50 banknotes in bank's view in order to minimise the total number of issued banknotes (see table 4, columns

"BANK'S VIEW" and "TWO BOXES"). The highest share of €50 banknotes that we can reach at a given distribution is 37 % of all issued banknotes.

The results of the model are identical from both the bank's and CB view and the ATMs that have two or three boxes, as well (see table 4). In case of four boxes, the ratio between **€10 : €50 = 76 : 24 = 1 : 0.3** and the expected total number of issued banknotes equals 288.9 million pieces. (first distribution).

3.3 Third variant with the denominations of €20 and €50

First of all, we have to adapt the probability mass function, because the ATMs are not able to pay out the amounts of 10 and 30 EUR (all others they can pay out), and the algorithm of issuing banknotes, as well.

Table 5: The results of the model with €20 and €50 banknotes (in million banknotes)

	BANK'S VIEW		CB VIEW		TWO BOXES		THREE BOXES		FOUR BOXES	
	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio
Distr. 1	187.1	65:35	187.1	65:35	IMPOSSIBLE		187.1	65:35	205.2	74:26
Distr. 2	187.5	64:36	187.5	64:36	IMPOSSIBLE		187.5	64:36	204.4	73:27
Desired ratio €20 : €50			**67 : 33**		**50 : 50**		**67 : 33**		**75 : 25**	

Source: own calculations, 2007

Similarly as in the previous variant, the results of the model are identical from both the bank's and CB view and the ATMs that have two or three boxes (in case of two boxes the ratio is impossible to reach, but anyway it is still the best possible solution). The expected total number of issued banknotes in 2007 is 187.1 million pieces, which corresponds to the ratio $f_{1,20,3} : f_{1,50,3}$ **= 65 : 35 = 1 : 0.5**. In case of four boxes, the expected total number of issued banknotes is approx. 205 million pieces.

3.4 Fourth variant with the denominations of €10, €20 and €50

The last variant includes all three denominations. According to this variant, we issue minimum number of banknotes in 2007 (i.e. 173.5 million banknotes from the bank's viewpoint) and the ratio between the denomination equals $f_{1,10,4} : f_{1,20,4} : f_{1,50,4}$ **= 19 : 33 : 48 = 1 : 1.7 : 2.5**. From the central bank's view the "optimal" ratio is **€10 : €20 : €50 = 34 : 44 : 22 = 1 : 1.3 : 0.6**, the expected total number of issued banknotes increases to 237.3 million pieces, which is 37 % more than from the bank's view.

Table 6: The results of the model with €10, €20 and €50 (in million banknotes)

	BANK'S VIEW		CB VIEW		TWO BOXES		THREE BOXES		FOUR BOXES	
	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio	Nu. ban	Ratio
Distr. 1	173.5	19:33:48	237.3	33:44:23	IMPOSSIBLE		217.4	35:33:32	223.0	25:49:26
Distr. 2	173.5	21:31:49	230.5	32:42:26	IMPOSSIBLE		210.1	35:31:34	222.0	24:51:26
Desired ratio €10:€20:€50			**34 : 44 : 22**		/		**33 : 33 : 33**		**25 : 50 : 25**	

Source: own calculations, 2007

In case of three boxes and first distribution we will issue in total 217.4 million banknotes and the ratio between different denominations will be **€10 : €20 : €50 = 35 : 33 : 32 = 1 : 0.9 : 0.9**. In case of four boxes and first distribution, 223.0 million banknotes will be issued by the ATMs and the ratio will be **€10 : €20 : €50 = 25 : 49 : 26 = 1 : 2.0 : 1.0**.

4. AN OVERVIEW OF THE QUANTITY AND VALUE BREAKDOWN OF EURO BANKNOTES IN CIRCULATION IN SLOVENIA

The total value of euro banknotes in circulation in Slovenia on 31st May 2007 was 390.2 million EUR, which represents 21.1 million pieces of banknotes.[1] The most widely used denomination in circulation is banknote for €20, because it is placed in the ATMs in Slovenia. At the end of May 2007 there were 18.5 million pieces of €20 banknotes or 88 % of all banknotes in circulation. This represents almost 370 million EUR, or 95 % of all banknotes in circulation in terms of value. In case of €50 banknotes we have a negative circulation in Slovenia (see table 7). This means that in the Bank of Slovenia more banknotes returned than were issued in the first five months (net inflow). Something similar is also happening in the other countries of the Eurosystem, especially in Austria, where they have a negative circulation in the case of €20 and €50 banknotes. The most important reason for negative circulation in Austria and Slovenia is that ATMs do not issue €50 banknotes. In Slovenia the ATMs are issuing €10 and €20 banknotes and in Austria €10 and €100 banknotes.

In conclusion, the quantity breakdown of euro banknotes in Slovenia is not appropriate. The reason for this is that around 90 % of circulation is represented by €20 banknotes in terms of quantity and value. The most important factor that determines the quantity breakdown of banknotes in circulation is the very selection of banknotes in the ATMs and the algorithm of issuing banknotes.

Table 7: Quantity and value breakdown of banknotes in circulation in Slovenia (31.5.2007)

Denomination (in €)	**Quantity (in pieces)**	**Share (in %)**	**Value (in EUR)**	**Share (in %)**
500	**156,616**	**0.7**	**78,308,000**	**20.1**
200	**47,575**	**0.2**	**9,515,000**	**2.4**
100	**138,563**	**0.7**	**13,856,300**	**3.6**
50	**-2,381,796**	**-11.3**	**-119,089,800**	**-30.5**
20	**18,468,916**	**87.7**	**369,378,320**	**94.7**
10	**3,035,339**	**14.4**	**30,353,390**	**7.8**
5	**1,585,193**	**7.5**	**7,925,965**	**2.0**
Total	**21,050,406**	**100.0**	**390,247,175**	**100.0**

Source: Bank of Slovenia, 2007

5. THE SELECTION OF THE OPTIMAL VARIANT FOR THE ATMS

In Slovenia, we are currently issuing banknotes according to the bank's view (see table 3). The ratio between denominations for €10 and €20 equals 1 : 6.1, which is the same as the quantity breakdown of banknotes in circulation in Slovenia at the end of May 2007 (€10 : €20 = 3,035,339 : 18,468,916 = 1 : 6.1). This proves our hypothesis about an important influence of ATMs on the quantity breakdown of banknotes in circulation.

Based on the results of the model, the best variant for Slovenia is with the €10, €20 and €50 banknotes. Also from the practical point of view, the easiest way would be to insert

[1] As a matter of fact, there are even more euro banknotes in circulation in Slovenia. The figures do not include banknotes that were in circulation before the cash changeover in Slovenia and banknotes that were issued by other central banks of the Eurosystem (migration of banknotes).

boxes for €50 banknotes instead of the boxes for €10 banknotes. We should insert boxes for €50 banknotes in the ATMs gradually, starting with the ATMs that have the highest average withdrawal. The average withdrawal from the ATMs last year was approx. 66 EUR, in 2007 it is estimated at around 70 EUR. Since the amounts of withdrawals from the ATMs are constantly increasing, we would have to place €50 banknotes in our ATMs.

Furthermore, in the case of three denominations a customer would have a wider variety of banknotes. ATMs should also be adapted in a way that they would ask a customer which denomination he/she would like to receive. For instance, in case of withdrawing 100 EUR, the first option is 2 banknotes for €50, the second option is 1 banknote for €50, 2 banknotes for €20 and 1 banknote for €10, the third option is 5 banknotes for €20, … ATMs should also be more user-friendly, especially for foreigners (i.e. operate in more foreign languages, like German, Italian, Croatian, Hungarian, …; Bounie, 2003).

One of the most important reasons for inserting €50 banknotes in the ATMs is to minimise the costs of filling up the ATMs. We would reduce the total number of issued banknotes via ATMs by 25 %, if €50 banknotes were inserted in the ATMs (from around 300 million banknotes to 220 million banknotes annually; see tables 3 and 6). In addition, this would also reduce the transportation costs and costs of sorting banknotes. For the Bank of Slovenia this decision would improve the quantity breakdown of banknotes in circulation and reduce the costs due to the migration of banknotes on the Eurosystem level.

On the other hand, there are also some disadvantages. Firstly, commercial banks need to buy new boxes for €50 banknotes. Secondly, the additional costs are related to the adaptation of the software and higher costs of insurance. Despite these additional costs, the savings are still much higher and hopefully sooner or later we may see €50 banknotes in our ATMs in Slovenia.

6. CONCLUSION

In this paper we presented the model for optimal selection of banknotes in the ATMs. Then, we presented the results of this model and on this basis we decided that at the medium term the best solution for Slovenia is the variant with €10, €20 and €50 banknotes. We also have to be aware that the results of the model strongly depend on the data used.

The main advantages of inserting €50 banknotes in the ATMs are: reduction of costs for filling up the ATMs by 25 %, reduction of transportation cost and cost of sorting banknotes, improvement of the quantity breakdown of banknotes in circulation and reduction of the cost related to the migration of banknotes and finally customers would have a wider variety of banknotes.

On the other hand, additional costs are: purchase of new boxes for €50 banknotes, adaptation costs of the software and higher costs of insurance.

REFERENCES

1. Banka Slovenije: Bilten. Ljubljana: Geodetski inštitut Slovenije. Maj 2007. Leto 16, št. 5. 153 str.
2. Benjamin J. R., Cornell C. A.: Probability, Statistics and Decision for Civil Engineers. New York: McGraw-Hill Book company, 1970. 684 str.
3. Bounie D., Soriano S.: Cash versus e-cash: A new estimation for the Euro Area. Paris: GET/ENST, Department of Economics, 2003. 20 str.

4. Drehmann M., Goodhart C.; Krueger M.: »The challenges facing currency usage: Will the traditional transaction medium be able to resist competition from the new technologies?«, Economic policy, 17(34) 2002, str. 193–227.
5. Internal documents of the Bank of Slovenia and the Bankart.
6. Jamnik R.: Verjetnostni račun. Ljubljana: Društvo matematikov in fizikov in astronomov SRS. 1987. 276 str.
7. Miklavčič, G.: Določitev optimalne količinske strukture evrogotovine ob vstopu Slovenije v EMU. Magistrsko delo. Ljubljana: Ekonomska fakulteta, 2006. 77 str.
8. Winston L. Wayne: Operations research. Belmont (Cal.): PWS Publishers, 1997. 1312 str.

TAXATION MODELS FOR THE GAMING INDUSTRY AS A TOOL FOR BOOSTING REVENUES FROM TOURISM

M. Sc. Boris Nemec, HIT d.d. Delpinova 7a, 5000 Nova Gorica
Boris.Nemec@hit.si

Abstract: The article is dealing with casino gaming-tax systems and regulations, gaming and concession tax models and VAT in the tourism industry. Three basic tax models are presented: progressive, proportional and digressive taxation. Discussion about the public interest on different models is done and some suggestions for a new, more development-oriented taxation of the gaming industry is recommended. Some new ideas to lower taxes on the casino industry's gross revenues are explained to gain support for the growth of the tourism industry and for the benefit of public finances.

Keywords: taxes, gaming tax, taxation model, VAT, casino gaming industry, tourism services

1. INTRODUCTION

The world's leading countries in demanding technologies, products and services also support the development of tourism, in spite of the fact that it generates less added value than many other producing or servicing activities. There are many reasons for such behaviour. Production is being more and more automated and robotised. Thus the quality of the products is increased and the per-unit production costs cut, as well as the need for workforce. The released workers need to be re-qualified for a different type of production or, even more often, for rendering new services. All developed countries support the tourism industry, since it can offer employment to many workers with different qualifications and preferences. Revenues from tourism are even more welcome in small countries, since this industry is exporting goods and services, for which the tourists are paying local taxes on goods and services (VAT, excise duties), while traditional exports are free of such levies. Foreign tourists generate opportunities for many other activities that would be otherwise not competitive enough to export (local agriculture and craftsmanship) or services that cannot be exported at all (cultural and natural heritage, free-time and recreational services, cuisine and many others (1).

In Slovenia, a small country neighbouring richer countries, the gaming industry has established itself as a niche opportunity, generating today over 25% of Slovenia's revenues from tourism, while gaming remains a trifling activity in the leading touristic countries. It is therefore important for Slovenia to understand whether gaming could be used to significantly increase revenues from the tourism industry. One of the opportunities for improving the competitiveness of the tourism industry, particularly in the case of small countries, is to have a relatively low value-added tax (VAT). In countries such as Luxembourg, Malta and Cyprus, the VAT rate is at the lowest levels allowed by the EU, i.e. a 15% standard rate and a 5% reduced rate. Switzerland has a 7.5% standard VAT rate and a reduced rate of only 2.7%. Smaller EU countries inside Schengen borders using euro currency have an extraordinary opportunity to increase their revenues through gaming tourism, since gaming services are excluded from the unified taxation of services in the EU single market. This means that small countries may choose gaming as one of their priorities to boost foreign tourism. The basic strategy is to select the lowest possible tax burden on gaming in order to attract large and comprehensive tourist gaming investments, thus ensuring that gaming and other tourists stay longer in the country. All the goods and services are encumbered by VAT as the only tax on consumption paid by tourists in the country they visit. Thus gaming tourism and a favourable taxation model for gaming revenues could represent a huge opportunity for small countries to significantly increase their revenues from the tourism industry.

Commercial games of chance are one of the oldest human activities and, as such, the business was controlled and taxed as a monopolistic service by the incumbent government. Such historical reasons and traditions preserved a whole range of different taxation models for the gaming services, sometimes serving the interests of the legislator and sometimes being absolutely unsuitable for the development goals pursued. The selection of the taxation model in a comprehensive regulation system of gaming activities is therefore an extremely important decision to be taken by the state, regional or municipal authorities regulating the activities of gaming.

2. REGULATION SYSTEMS AND TAXATION MODELS

All jurisdictions have their own regulation systems in accordance with the wishes, needs and goals of those who have licensed the monopolistic activity of gaming to authorised operators on the territory of such jurisdiction. The various systems may be divided in three basic groups:

1. systems prohibiting gaming or limiting it to foreign nationals
2. systems limiting gaming as a business for all visitors
3. systems supporting gaming as a business that boosts tourism.

The increase in gaming revenues can be limited with various restrictive measures (monopolies and restriction of the offering) or stimulated through incentives (promotion, rewarding of customer-loyalty).

This paper is dealing with the regulation systems that restrict or stimulate gaming and analyze the most effective regulation tools, such as the various models of taxation of gaming revenues.

Gross gaming revenue (GGR) is the difference between the casino's gaming wins and losses and is used as the tax base for calculating tax liabilities (2).

Most often jurisdictions charge levies on GGR through a gaming tax and a concession fee, with the same tax base (GGR) used for calculating both.

Let $t(x)$ be the function of tax rate depending on the tax base (GGR), expressed as variable x and both values are non-negative.Tax amount or tax revenue function $y(x)$ is determined by the expression: $y(x) = \int_0^x t(u)du$. Additionally, let us determine the average tax rate $\overline{t}(x)$, defined by the expression: $\overline{t}(x) = \frac{1}{x}\int_0^x t(u)du$

The taxation models for taxing gross gaming revenue (GGR) are grouped in three basic clusters, according to the behaviour of the function $\overline{t}(x)$ or its derive $\overline{t}'(x)$:

1. Progressive models, if $\overline{t}'(x) \succ 0$, or $\overline{t}(x_2) \succ \overline{t}(x_1)$ for $x_2 \succ x_1$

2. Proportional models, if $\overline{t}'(x) = 0$, or $\overline{t}(x_2) = \overline{t}(x_1)$ for $x_2 \succ x_1$

3. Digressive models, if $\overline{t}'(x) \prec 0$, or $\overline{t}(x_2) \prec \overline{t}(x_1)$ for $x_2 \succ x_1$

Progressive models are the most demanding ones and have only appeared recently in history, since they cannot be efficiently applied without thorough supervision GGR.

Proportional models are the simplest ones and have been used for many centuries.

Digressive models may be extremely efficient tax-regulation tools in properly regulated jurisdictions, as I already stated in 1999 (2) and in 2001 (3). The digressive models appeared in 2005 in the Hungarian Gaming Tax Act and in the Slovenian draft amendments to gaming regulations in 2007.

All these models have been used so far to govern tax revenue function $y(x)$ depending on GGR as x.

The proposed digressive taxation model is a novelty, since the tax-rate function $t(x)$ depends on overnight stays or the number of hotel rooms and not on gaming revenues. The proposal has been presented in the Notes to the Law Amending the Slovenian Gaming Act, which I intend to discuss below.

3. PROGRESSIVE MODELS

The basic feature of progressive models is that the tax-rate function $t(x)$ is a steadily increasing function based on the GGR as a tax base x. In a progressive taxation model, the tax revenue function $y(x)$ is a continuous and accelerative increasing function (Chart 1).

Tax jurisdictions have simplified the tax rate function $t(x)$ to make it easier to understand and comparable with other tax rates applicable to other business activities. They are using a gradually increasing function t(x), as shown in Chart 2 below. Tax rates are thus increased progressively in line with the gaming revenue generated. With the graduated function t(x) the legislator had two arbitrary options to increase or reduce tax rates on the ordinate (y-axis) and to widen or narrow the tax brackets on the abscissa (x-axis).

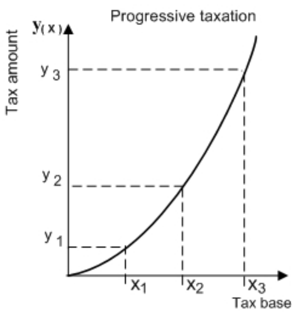


Chart 1

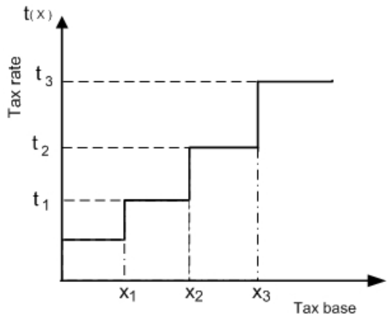


Chart 2

The progressive gradual function is displayed in the articles of a gaming law by means of a relatively simple table, while the tax revenue function $y(x)$ is similar as the function on Chart 1, with the only difference that it is a linear and continuous function within single segments with increasing slope coefficients. Tax revenue is defined as the area between the tax function $t(x)$ and the abscissa (x-axis) or a determined integral in the area $[0, x]$.

The progressive taxation model is very efficient in restricting the gaming business. The advantage of such model consists in the fact that gaming is allowed by the law, although very limited, so that countries avert illegal gambling, which usually comes as the consequence of an outright ban on gaming. Casinos are thus visited by those who would otherwise find a legal way abroad or an illegal one in their own country to participate in commercial games of chance. If the jurisdiction ensures a sufficient number of casinos on its territory, particularly in tourist destinations, the goal of uniformly spread offering for tourists and residents is achieved. The progressive taxation model prevents any casino from growing into a large operator. If the owner of all the casinos is a single company or a larger hotel chain, the progressive taxation model works extremely well. The best example of such regulation is France.

Progressive taxation, however, is not the best solution for small countries, such as Slovenia, which selected such a system by amending the Gaming law in 1993 for pure political reason. An extremely steep gradual function of tax rates $t(x)$ represented such a tax burden for the biggest company HIT that in the previous decade it suffered losses or managed to achieve trifling profits (4, page 264). The low initial tax rates and tax brackets allowed all the other smaller casinos (Bled, Maribor and Ljubljana) to fare relatively well. Good standing was also achieved by the somewhat larger Casino Portorož. Although the government rapidly collected large tax revenues, it hindered development for many years and prevented Hit from becoming Slovenia's first gaming multinational in Europe, since it missed the opportunities in the Eastern European countries that decided to allow private initiatives in gaming. The remaining smaller state casinos were not stimulated to increase their gaming revenue, since this would have entailed taxes and costs exceeding their overall revenues. If Slovenian casinos were visited mostly by local people, as is the case in Germany, a progressive taxation model would be justified. However, since most of the visitors come from abroad, a progressive model is detrimental to development and to the state coffers, as well (5, 6). With the introduction of VAT in mid-1999 the Slovenian Ministry of Finance eased the irrationally high tax burden to some extent. An 18% gaming tax has been introduced along with a concession fee, payable on the basis of the same tax base according to the progressive scale (4, page 265).

A concession fee was also charged for the gaming tax already paid. The system is still in use today and it is not in line with a development-oriented tax policy.

Additionally, there is another fiscal solution that hinders development. Since casinos are not liable to VAT, which has been replaced by gaming tax, the casinos are not allowed to deduct input VAT from the output VAT as all the other businesses liable to VAT do. Thus, all casino investments in non-gaming developments, such as hotels, are sanctioned with a 20% standard input VAT. Such tax is therefore an additional burden for the casinos, by which the legislator is actually "punishing" the investments in tourism made by the casinos, in spite of the fact that general statements about Slovenia's development strategy for the tourism industry expects casinos to invest in the overall tourism products .

In the paper (4) is exposed the government's anti-development attitude and suggested that the government replace the progressive rates of 20% with a digressive taxation model using decreasing rates. By doing so, we would have a progressive-digressive taxation model tailored to suit Slovenia's needs (4). Progressive taxation with low initial tax rates would guarantee the survival of small state casinos, while ensuring high tax revenues from only one casino with large income (HIT). As the gaming revenues exceed a certain amount, the tax-rate function would be inverted and the casinos stimulated to increase their turnover from visiting alien citizens. The purpose would be achieved with hotel guests from remote locations, which entail higher costs to the casino. Although the government would charge taxes on such additional income at lower rates, the moneys collected would be additional tax revenue. The government has got the message, but the 'innovation' was not accepted, because there were no comparable examples in the world. So, the Government of the Republic of Slovenia opted for a traditional approach instead and in the period 2001-2005 gradually abolished all the three tax rates above 20% as shown in (4, page 266). The casinos were thence enabled to increase their profits, but the model was not requiring casinos to invest in integrated tourist products. The portion of non-gaming revenues at HIT therefore gradually fell from 13% in 2002 to 11% in 2006.

In the past few years, Slovenia has opened the doors wide to gaming saloons, mostly catering to local residents (7, 8, 9). If the authorities wanted to have different methods for steering gaming consumption in the case of residents and in the case of gaming tourists, the progressive-digressive model would be an excellent choice: tax rates would be increased or decreased on the basis of the ration between the numbers of domestic and foreign guests. In Europe, unlike in the United States, registration at the entrance to any casino is compulsory. This requirement offers wide opportunities for tax regulation on the basis of the guest structure, as well as for recognising problematic gamblers or addicts. The governments IT system has real-time connections with computers registering the entrance of guests in casinos and gaming saloons. In future, such systems will be used to exchange information on gamblers within Europe. It means that there are concrete possibilities to prevent addiction, but the government is not too keen to renounce taxes that are also levied from addicts.

4. PROPORTIONAL MODELS

The main characteristic of proportional models is that the tax-rate function is constant, the tax revenue function $y(x)$ is linear (Chart 3) and increase in proportion to the increase of GGR as tax base x. The model should be named 'constant model', since the tax-rate function $t(x)$ is constant or flat rate (Chart 4), but more common. is named as proportional model due to linear growth of taxes to be paid..

Jurisdictions have simplified the function $y(x)$ to the maximum extent, so that the taxation model is comparable with the models applied to other business activities. The tax rate remains the same regardless of the GGR generated. The function t(x) = c leaves less freedom of manoeuvre to the legislator than in the case of the progressive model. The legislator can only choose between a higher or lower common tax rate.

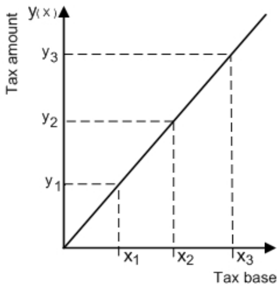


Chart 3

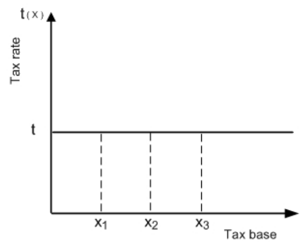


Chart 4

Tax revenue is represented by function $y(x)$, a linearly rising function, meaning that tax revenues increase in accordance with the tax base.

German states decided to apply extremely high flat tax rates of 80%. In this way, the German jurisdictions would like to discourage people from gambling as much as possible and therefore gaming companies must act very rationally. Thus, the German states created an environment, in which gambling is legal for the 'addicts', while illegal gambling is highly improbable to appear there. Today, internet gambling is seriously threatening their current solutions (10, 11, 12).

The other extreme use of the proportional model is Nevada. The tax rate there amounts to 7.5% and licences are granted to all who are willing to obey the Gaming law (13). Casinos raked in huge profits, thanks to low tax rates, only in the early days and only seemingly, because very rapidly the competition became extremely harsh and the quest for guests required that the casinos widen their offering. The competition war prompted the construction of huge hotels to attract guests from all the countries of the American continent and from all around the world. The State of Nevada achieved increased tax revenues in spite of the low gaming-tax rate, simply because all the other taxes on consumption are charged at usual taxes for goods and services.

Singapore decided to use gaming as a tool for developing the tourism industry. The government published an international call for tenders for two projects, each worth in excess of \$3 billion. The legislator there offered a proportional taxation model with a 15% common tax rate and only 5% for the 'premium' guests. All the local guests will be asked to pay a \$100 entrance fee for each visit and that is another source of tax revenue.

In the past few years, more and more governments have begun to realise that foreign tourism can be boosted by applying low tax rates to GGR and decisions of that kind are being taken by an increasing number of countries (Macao, Castilla-La Mancha in Spain).

5. THE DEGRESSIVE MODEL

The main characteristic of digressive models is that the tax-rate function $t(x)$ is decreasing on the basis of the increasing tax base x, the tax revenue function $y(x)$ is continuous and decelerative increasing (the first derivative being positive but below 1), see Chart 5.

Tax jurisdictions have simplified the function $t(x)$ to make it easier to understand and comparable with other tax rates applicable to other business activities. They have used a gradually decreasing function t(x), as shown in Chart 6. Tax rates are thus decreased digressively in line with the gaming revenue generated. With the digressive method using the graduated function t(x), the legislator has two arbitrary options to increase or reduce tax rates on the ordinate (y-axis) and to widen or narrow the tax brackets on the abscissa (x-axis). The digressive graduated function is displayed in the articles of a law by means of a relatively simple table.

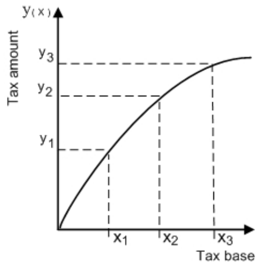


Chart 5

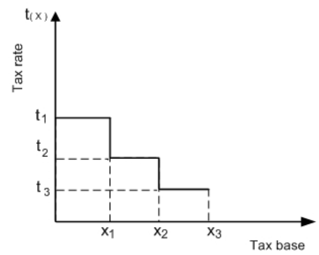


Chart 6

Digressive taxation is a good solution for small countries having a high tax rate imposed on gaming and wanting to considerably increase gaming revenues from foreign gaming tourists. The first to introduce such a model were the Hungarians with their new law of 2005. It should enable them to build EuroVegas not far from the international airports. This year, Slovenia has made a similar proposal which is discussed as a Slovenian case bellow.

Some jurisdictions not capable to supervise company GGR, are using digressive models too. Such countries demand the payment of a fixed monthly or annual gaming tax (Chart 7), regardless of the actual GGR over single periods of time. Thus, the jurisdictions are satisfied with advance tax, and it is then up to the casino operator to achieve lower average tax rate with higher GGR. The larger the turnover, the lower the average tax rate (Chart 8).

Some jurisdictions require a fixed annual fee, alongside with the proportional taxation of the casinos turnover. Such a scheme is also, in practice, a digressive taxation model (Charts 9 and 10).

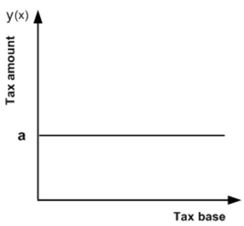


Chart 7

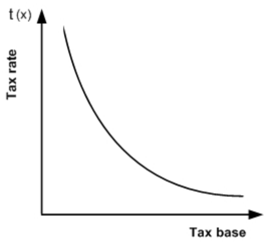


Chart 8

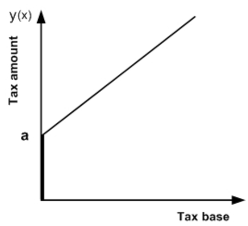


Chart 9

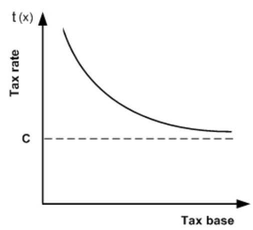


Chart 10

6. THE CASE OF SLOVENIA

In its draft amendments to the Gaming Act of July 2007, Slovenia presented a digressive taxation scheme (14), which would allow for the construction of a large gaming-and-entertainment resort with the participation of a strategic investor.

The proposal is inadequate, since it prompts the foreign partner to cannibalise its Slovenian partner, which has a comparable annual turnover (15, 16). The proposal would still be unacceptable if the digressive taxation model would apply to both partners having similar gaming revenues. The draft would stimulate an all-out war in a limited market, since the partner that increased its turnover to the detriment of the other partner, would be awarded lower tax rates, while the loosing partner would pay gradually increasing tax rates due to its decreasing turnover. With only one winner left, the state coffers would also suffer a loss due to the lower tax revenues caused by the decreasing tax rates. The proposal would be effective only until the two partners continued to grow in a rapidly growing gaming market, which is not the case for local markets of Western Slovenia and Eastern Italy.

The legislator should offer a digressive taxation method on the basis of tourist turnover, measured in overnight stays or number of rooms and other official indicators. For these reasons, we have proposed that the amendments to the Gaming Act be changed as shown in (17). The country's reduced gaming-tax revenues would turn out to be, over time, the best financial investment for much larger tax revenues from VAT on tourist services and all the related activities.

The civil society criticised the government's two-facedness and accused the state of being only interested in taxes and failing to address the negative consequences of gaming, such as problem gambling, addiction to games of chances and personal bankruptcy. As already mentioned, in Europe the governments have real-time control of all the people registered at the entrance of casinos. These data could be used to prevent people from visiting casinos too often. In such cases, personal freedom is justly limited as in other cases of addiction, such as alcoholism. One of the ways of restricting excessive gambling is the progressive taxation of casino entrance-fees (17), which flow directly to the state coffers.

7. CONCLUSION

Casino games are still considered in many countries as a special activity on the edge of legality or even prohibition. Several academic studies, articles and papers discuss the social aspects of gaming, such as addiction and other negative impacts. There are many conventions and meetings of lawyers associations addressing different aspects addressed by legislators. Conventions, symposiums and meetings have been used in recent years to discuss internet gambling, with the main questions being the prevention of money laundering and access by minors, the restriction of credit card use and similar topics. There are but a few academics, who study the economic impact of gaming on society and the ways to include the activity of gaming in the free-time industry for the benefit of the entire society. One of the reason is the fact that all the countries have their own specific, monopolistic solutions, which have almost nothing in common with scientific findings. One exception is the University of Nevada. The Faculty of Economics in Ljubljana, Slovenia, has also made a few important studies on the economic impacts of gaming, commissioned by the government, local communities, the Association of Gaming Companies at the Chamber of Commerce of Slovenia and by Hit. Two recently founded faculties in Nova Gorica: the European Faculty of Law and the Faculty of Applied Social Studies, along with their institutes are showing a growing interest in researching gaming-related issues. Slovenia therefore has a critical mass of knowledge to have an impact on politics and society and thus lead to selecting the best solutions so that gaming could support the development of tourism and related activities to the maximum extent, as well as to minimise the negative consequences of gaming. This paper is also intended to be a contribution to the increasing number of findings for a more effective regulation of gaming-taxation systems and models so that better economic and social effects can be realistically achieved.

8. REFERENCES:

(1) Nemec Boris: Slovenian Tourism Development Strategy, A Response to the Globalisation Process, Encuentros 2005, Nova Evropa: Nova turistična destinacija.

(2) Nemec Boris: Development Supported Taxation model in Gaming and Entertainment Industry in Slovenia, Proceedings of the 5th International Symposium on Operational Research, SDI-SOR, Ljubljana, 1999, p. 129-133

(3) Nemec Boris: Goods and Services Taxation Models and Optimum Solutions for Gaming Services Taxation, Proceeeding of the 7th International Symposium on Operational Research, SDI-SOR, Ljubljana, 2003, p. 133-138

(4) Nemec Boris:Strategic Dilemmas Regarding the Development of the Slovenian Entertainment and Gaming Industry, Proceeeding of the 6th International Symposium on Operational Research, SDI-SOR, Ljubljana, 2001, p. 263-268

(5) Zakon o posebnem prometnem davku od posebnih iger na srečo, Official Gazette of the RS, No. 67/1993, p. 3311-3312

(6) Zakon o igrah na srečo (ZIS), Official Gazette of the RS, No. 27/1995, p. 1909-1920

(7) Prašnikar Janez and others: Ekonomska podlaga nove družbene pogodbe med podjetjem HIT d.d. Nova Gorica in Republiko Slovenijo, study, Ekonomska fakulteta Univerze v Ljubljani, 2002

(8) Bole Velimir, Jere Žiga: Trg igranja na igralnih avtomatih (Segment v primorsko-kraškem področju), EIPF Ljubljana, 2004

(9) Prašnikar Janez, Pahor Marko, Ljubica Knežević: Analiza vpliva igralniške dejavnosti na gospodarsko in družbeno okolje v občini Nova Gorica, Ekonomska fakulteta UL, Ljubljana 2005

(10) Swiss Institute of Comparative Law: Study of Gambling Services in the Internal Market of the European Union, European Commission, 2006

(11) Nemec Boris: An Alternative Approach to Internet Gambling, 12th International Conference on Gambling & Risk-Taking, Vancouver, 2003

(12) Swiss Institute of Comparative Law: Cross-Border Gambling on the Internet, Challenging National and International Law, Schulthess, 2004

(13) Christiansen Eugen, Christiansen Capital advisers LLC: The Impacts of Gaming Taxation in the United States

(14) http://www.mf.gov.si/slov/zakon/predlogi_igre_sreca.htm

(15) Jaklič Marko, Zagoršek Hugo, Pahor Marko, Ljubica K. Cvelbar: Analiza upravičenosti spremembo obdavčitve posebnih iger na srečo v Sloveniji, 2006

(16) Jaklič Marko, Zagoršek Hugo, Pahor Marko, Ljubica K. Cvelbar: Dopolnitev študije Analiza upravičenosti spremembe obdavčitve posebnih iger na srečo v Sloveniji, 2006

(17) Nemec Boris; Pripombe na predlog Zakona o spremembah in dopolnitvah Zakona o davku od iger na srečo (ZDIS), www.fzg.si

BANKING SECTOR PROFITABILITY ANALYSIS: DECISION TREE APPROACH

Mirjana Pejić Bach*, Ksenija Dumičić* and Nataša Šarlija**
* University of Zagreb, Faculty of Economics, Trg J.F.Kennedy 6, Zagreb
{mpejic,kdumicic}@efzg.hr
** University of Osijek, Faculty of Economics, Gajev trg 7, Osijek, natasa@efos.hr

Abstract: The paper deals with problem of analyzing the profitability of the banking sector in Croatia. In our research, profitability is measured by the return on average assets (ROAA). The aim of the paper is to design a model which would forecast the profitability of banks by their characteristics and the environment factors in order to maintain the stability of the banking sector. The decision tree has been developed using C&RT algorithm. The results have shown that ratio of capital and assets, market share and loan to assets ratio have the positive influence on the profitability of the banks.

Key words: profitability of the banks, forecasting profitability, decision tree

1. Introduction

Current banking sector profitability analyses have been targeted to forecast bankruptcy of the bank and authors have used various methods. Barr, Seiford and Thomas (1994) tried to predict bankrupts using a non-parametric frontier estimation approach.. Lane, Looney and Wansley (1998) used the Cox model, and other researchers used neural networks (Tam et.al., 1992; Salchenberger et.al., 1992). These studies are based on the classification approach, according to which in the past banks have been classified as bankrupted or not. On the other side, Nuxoll (2003) proposes the benchmarking approach, which is based on the preposition that best results are achieved if banks follow the financial structure of the best banks, or in other words by benchmarking best banks.

The goal of this study is to design a model which would forecast the profitability of banks by their characteristics and the environment factors. All this is used to maintain the stability of the banking sector. A forecasting model like this would be of great use to the Croatian National Bank, as to all the Boards of Directors. The study consists of the following parts. In the second part various ways of banking sector profitability analysis are enlisted. The third part describes the methodology used in the study (the decision tree). The fourth part encloses the results and the fifth part comprises final thoughts.

2. Banking sector profitability analysis

Banking sector profitability is measured by the return on average assets (ROAA), return on average equity (ROAE) and the net interest margin (NIM). Based on these profitability indicators, recommendations to the boards of directors can be made. In this study we will try to express profitability with one value, which follows the duPont procedure of business activity estimation (Pavković, 2004).

In this study we will concentrate only on the return on average assets (ROAA), which is calculated as a ratio of profit and average assets. Hence, it is the banking profit gained for one Kuna (local currency) of assets.

Factors of banking sector profitability can be divided to two basic groups: characteristics of a specific bank and environment factors, and the selected profitability factors were used in a research by Demirguc- Kunt and Levine (2001). Characteristics specific for a bank are: market share, ratio of capital and assets of the bank, ratio of loans and assets of the bank, ratio of overheads and assets of the bank and the ratio of non-income assets and total assets of the bank.

The market share should have a positive effect on banking sector profitability indicators. Different hypotheses on the functioning of the market in various ways explain this fact. According to the relative market power hypothesis only monopolistic companies with high market shares and highly differentiated products can acquire above average profit margins. The efficient structure hypothesis claims that banks with effective asset structure achieve highest market shares. Berger (1995) tested these two hypotheses in the financial market and proved that the size of the bank is connected with profitability, which was also proved by Frame and Kamerschen (1997). On the other side, Smirlock (1985) shows that concentration isn't prior connected to superior performance of the leading banks, but the efficient banks become bigger and gain bigger market shares.

Share of capital in the assets is positively correlated with ROAA. Banks with high shares of capital in overall assets have lower costs of financing which effects higher profitability and lower probability of bankruptcy.

Ratio of loans and assets is also positively correlated with profitability indicators. A bank which approves more loans for a unit of assets with the same interest rate, acquires higher profit because it earns more on the interest rates. Let us just emphasize that the growth of profit is not proportional to the growth of approved loans if this is too risky.

The share of non-income assets in the bank assets is negatively correlated with profitability indicators, although there are exceptions. For example, a bank can transfer the costs of its non-income assets to its clients, and a bank that pays rent for real estate can have higher costs than the bank that has its own facility.

The ratio of overheads and bank assets is negatively correlated with profitability indicators.

The values of these indicators for banks from the sample are shown in Table 1. The average market share of banks has been decreasing in the past five years. The ratio of capital and assets of the bank has also been decreasing, but it is still high. That means that banks have been decreasing the share of capital in the assets, but are still very cautious because of the suspicion in the stability of the banking sector. The ratio of loans and banks assets is increasing. The ratio of non-income assets and bank assets and the ratio of overheads and bank assets do not show a visible trend.

Table 1. Bank activity indicators

Year	Market share	Ratio of capital and assets of the bank	Ratio of loans and assets of the bank	Ratio of non-income assets and assets of the bank	Ratio of overheads and assets of the bank
1999	3,45%	53,96%	24,29%	1,86%	2,68%
2000	3,45%	23,01%	50,97%	2,76%	5,00%
2001	3,13%	17,37%	50,41%	2,03%	4,15%
2002	3,33%	15,24%	55,52%	2,29%	3,74%
2003	3,13%	15,44%	57,65%	2,02%	4,19%

Source: The Scope

Environmental factors are: GDP real growth rate, inflation rate, average exchange rate, GDP per capita. Web pages of the Croatian National Bank were used as a source of data about the macroeconomic indicators (www.hnb.hr). Values of the indicators are shown in Table 2.

Table 2. Characteristics of the environment as a factor of profitability

	Inflation rate	GDP per capita (EUR)	Growth of GDP	Exchange rate HRK: EUR
1999	4 %	4102	-0,9 %	7,5796
2000	4,6 %	4560	2,9 %	7,635
2001	3,8 %	4998	4,4 %	7,469
2002	1,7 %	5451	5,2 %	7,4068
2003	1,8 %	5747	4,3 %	7,5634

Source: www.hnb.hr

The growth of GDP should have a positive effect on the profitability of the bank. The inflation rate can have a positive and negative effect on the profitability, depending on the capability of the management of the bank to effectively conduct the resources of the bank during inflation. Finally, the exchange rate should be negatively correlated with the profitability of the bank, which is explained by the following. In the case of a strong HRK, Croatian companies are less competitive on the world market, which decreases the GDP and this way has a negative impact on profitability.

3. The decision tree

The decision tree can be used for classification and regression problems, and unlike neuron networks, the decision tree generates a model which can explain the mutual influence of input and output variables by a set of rules. The generated rules can be expressed like SQL commands and can simply be built in to the program solution.

For a problem to be appropriate for solving it with a decision tree, it has to have the following characteristics (Mitchell , 1997): (1) The data has to be described in a form of a final number of attributes, for example there are attributes for every bank; (2) The number of attributes is known in advance, for example it is well known how many attributes one bank can have; and (3) Every part of data should belong to only one category.

The decision tree is a classification algorithm which has a structure of a tree (McLahlan, 1992). There are two types of nodes connected with branches: leaf node which is the end of a particular branch, and the decision node which defines a certain condition in a form of a value described with attributes. It is made by searching for patterns with the algorithm , and the most famous algorithm are Chaid, exhaustive Chaid, C&RT (Breiman et.al, 1984) and Quest (Loh et.al., 1997).

The algorithm is made of a selection of attributes for generating the decision nodes, with all data sorted to one group in the beginning. Data are then divided to branches according to all possible criteria, and the criterion chosen is the one that divides data to groups that are more homogenous that the initial group of data. When the data can no longer be divided into groups that are more homogenous than the initial data, the tree is finished. Entropy is used as a measure of data group homogenousity.

4. Results

In order to forecast profitability of the banks C&RT and CHAID decision trees have been developed. All methods are processed with StatSoft Statistica 7.1. Results of C&RT algorithm is shown here as it gave better model.

Accuracy of the prediction is analyzed. Measures that are usually used are mean absolute deviation (MAD) and root mean-squared error (RMSE). The lower prognostic errors mean the higher accuracy of the model. According to both criteria C&RT has been shown as method which generates the lowest errors (Table 3). This method can be shown in a form of SQL statements which enables what-if scenario where the aim is to analyze what could happen if the characteristics of the banks and the environment factors are changed. Figure 1 shows the structure of the decision tree.

Table 3. Measures of accuracy prediction for the return on average assets (ROAA) for C&RT and CHAID

Measures of accuracy prediction	ROAA	
	C&RT	CHAID
MAD	0,74	0,85
RMSE	1,03	1,18

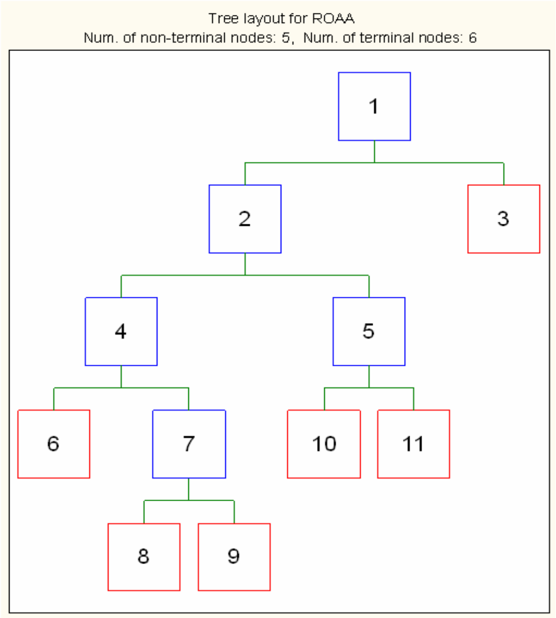


Figure 1. The decision tree for the return on average assets (ROAA)

Variables used for splitting nodes are: ratio of non-income assets and total assets, ratio of capital and assets, market share and loan to total assets ratio. The banks are divided into the 6 groups shown in table 4.

Table 4. Leaf nodes of the decision tree for the return of average assets (ROAA)

Node number	Average ROAA	Split criteria
3	4,688000	Ratio of non-income assets and total assets higher than 4,54512
11	2,092586	1 - Ratio of non-income assets and total assets lower than 4,5412; 2 - ratio of loan to assets higher than 16,3300
10	1,162308	1- Ratio of non-income assets and total assets lower than 4,5412; 2 - ratio of loan to assets lower than od 16,3300
6	0,641538	1- Ratio of non-income assets and total assets lower than 4,5412; 2 – market share lower than 0,7215

Node number	Average ROAA	Split criteria
8	1,162632	1- Ratio of non-income assets and total assets lower than 4,5412; 2 – market share higher than 0,7215; 3 – Ratio of non-income assets and total assets lower than 1,5485
9	1,641613	1- Ratio of non-income assets and total assets lower than lower than 4,5412; 2 – market share higher than 0,7215; 3 – Ratio of non-income assets and total assets higher than 1,5485

On the basis of the results of the decision tree, the following results can be made: Profitability of the banks is positively influenced by ratio of capital and assets of the banks, market share and loan to assets ratio.

Detailed analysis has shown that banks with the ratio of non-income assets and total assets lower than 4,54 and ratio of capital and assets higher than 17,54 will have higher profitability than the banks with the similar ratio of non-income assets and total assets and lower ratio of capital and assets.

Banks with the ratio of non-income assets and total assets higher than 4,54 have higher profitability compared to the banks with the ratio lower than 4,54.

Banks with the market share higher than 0,72 with the ratio of non-income assets and total assets lower than 4,54 and with the ratio of capital and assets lower than 17,54 will be more profitable compared to the banks with lower market share and similar values of all other mentioned ratios. This confirms the previous researches which state that banks profitability is highly influenced by market share (Berger 2005, Frame and Kamerschen 2007).

Banks that belong to the same group according to the ratio of non-income assets and total assets and ratio of capital and assets will have different profitability due to the loan to assets ratio in a way that higher profitability will be accomplished by the banks with the higher value of loan to assets ratio (Bourke, 1989).

5. Conclusion

The aim of the paper is to design a model which would forecast the profitability of banks by their characteristics and the environment factors in order to maintain the stability of the banking sector. In our research profitability is measured by the return on average assets (ROAA) as the ratio of net income and average total assets. Data for this research consisted of data about the banks in Croatia over the period from 1999 to 2003. Also, methodology of decision tree is given with the results of the decision tree model (C&RT) for the banks in Croatia. Results have shown that profitability of the banks is positively influenced by ratio of capital and assets, market share and loan to assets ratio. Particularly, of the banks with the similar ratio of non-income assets and total assets higher profitability is accomplished by the banks with higher ratio of capital and assets. Further, if the banks belong to the group of those with similar values of ratio of non-income assets to total assets and ratio of capital and assets, profitability will be increased by higher value of market share as well as higher loan to assets ratio. Although it was expected that lower value of non-income assets to total assets ratio would increase the profitability, the case of Croatian banks has shown opposite influence. An explanation could be found in the fact that banks in Croatia realized their profitability on income from services and less on income stated in assets. In order to investigate this phenomenon it would be interesting to analyze income structure of the banks as well as non-income assets which we suggest as guidelines for further research.

6. References

1. Barr, R., L. M. Seiford and F.Thomas., 1994. „Forecasting Bank Failure: a non-parametric frontier estimation approach". *Recherches Economiques de Louvain* 60(4), 417-429.
2. Berger, A., 1995. „The relationship between capital and earnings in banking". *Journal of Money, Credit and Banking*, 27, 404-431.
3. Bourke, P., 1989., "Concentration and other determinants of bank profitability in Europe, North America and Australia." *Journal of Banking and Finance* 13, 65-79.
4. Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., 1984. *Classification and Regression Trees.* Belmont: Wadsworth.
5. Demirguc-Kunt, A. and Levine, R., 2001. "Financial Structure and Bank Profitability" in *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks, Markets, and Development*, Eds. Cambridge, MA: MIT Press.
6. Frame, W. S., and D. R. Kamerschen. 1997. "The Profit-Structure Relationship in Legally Protected Banking Markets Using Efficiency Measures". *Review of Industrial Organization,* 12, 9-22.
7. Lane, W. R., S. W. Looney and J. W. Wansley., 1986. „An Application of The Cox Proportional Hazards Model to Bank Failure". *Journal of Banking and Finance*. 10, 511-531.
8. Loh W. Y. and Shih Y. S., 1997. „Split Selection Methods for Classification trees". *Statistica Sinica* 7, 815-840.
9. Han, J., and Kamber, K., 2000. *Data Mining: Concepts and Techniques.* San Francisco: Morgan Kaufman.
10. McLachlan, G. J., 1992. *Discriminant Analysis and Statistical Pattern Recognition.* New York: Wiley Interscience.
11. Mitchell, T., 1997. „Decision Trees", in T. Mitchell. *Machine Learning*, London: McGraw-Hill.
12. Nuxoll, D.A., O'Keefe, J., and Samolyk, K., 2003. „Do Local Economic data improve off-site bank-monitoring model?". *FDIC Banking Review*, 15(2), 39-53.
13. Pavković, A., 2004. „Instrumenti vrednovanja uspješnosti poslovnih banaka". *Zbornik radova Ekonomskog fakulteta u Zagrebu*, 2(1), 179-191.
14. Salchenberger, L. M.; Cinar, E. M.; and Lash, N. A., 1992. „Neural networks: A new tool for predicting thrift failures". *Decision Sciences*, 23(4), 899-916.
15. Smirlock, M., 1985. „Evidence on the (Non) Relationship between Concentration and profitability in banking". *Journal of Money, Credit and Banking*, 17(1), 69-83.
16. Tam, K.Y. and Kiang, M.Y., 1992. "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Management Science*, 38, 926-947.

SQUEEZING-OUT PRINCIPLE IN FINANCIAL MANAGEMENT

Viljem Rupnik
INTERACTA, Ltd, Business Information Processing, Ljubljana, Parmova 53
e-mail: Viljem.Rupnik@siol.net

Abstract: When a firm is running its business at rock bottom level, the traditional book keeping categories may not be sufficient for various revitalization ideas. Furthermore, components of standard balance sheet and income statement are not apt to be put in functional relationships so as to carry out relevant optimization procedures; in addition, they are even insufficient and/or relevant to optimization criteria. Hereby, we try to suggest some way to get out of such a trouble.

Keywords: multidimensional and multi-criteria simulation based optimization, non-formal modelling, extended financial management.

1. AN INTENTION OF THE IDEA

Suppose we want to improve the firm's operations in terms of financial categories usually embraced in periodic financial statements as we practice every day. Moreover, a set of financial categories contained in financial statements, as a rule, do not contain various financial indicators which are usually expected to be used as optimization criteria. We are not even sure that a) financial categories from financial statements are sufficient and relevant to what we might wish to improve and b) a set of financial indicators may vary over firms. Thus, input and output variables in the course of financial managerial operations are not fixed.

A task of optimization requires some mapping to exist between input and output variables. Whatever output variable(s) is (are) chosen, we are not even sure about the relevant (or most responsible) input(s) to enter the game. True, to this part of a problem, an interaction analysis (IA) could be called for help, although the corresponding time series within the context of a given firm might not be sufficiently long for the results to be accepted for extrapolation/forecasting/prediction (see /1/).

Let us assume that for any subset of financial output categories we succeeded in finding a corresponding input subset satisfying our needs and conviction. The crucial question then accrues whether or not we are able to find some function/functional/operator as a means of mapping input into output subset. All we have at our hands is a "feeling" that such causalities may exist. Consequently, a decision modelling appears to be a formidable task to overcome for financial managers.

In the paper presented, we try to overcome this problem by using non-formal modelling (see /2/) to the extent which allows us to make a step beyond financial sphere; it may well happen that either input variables are not financial categories or output variables are quite different from financial variables. Thus, a bridge between financial management and system management of the firm might be enabled.

2. THE RELEVANT INITIAL DATA BASE

To illustrate the procedure of squeezing-out principle in financial management we propose to start with the following set of notions usually used by financial managers:

Basic categories of relevant data base

Variable	Formula (when exists)	financial category
Q		quantity of product sold
C		selling price
R	R=c*q	gross revenue
F		fixed cost
V		variable cost
S	S=F+V	total cost
EBIT	EBIT=R-S	gross profit before taxes and interest
P_a		total liability and equity
Z		financial leverage
End		common stock capital
Epd		preferred stock capital
E	E=(1-z)*P_a =End + Epd	total stock capital
D	D=Pa*z	debt capital
P		interest rate
O	O=p*D	Interest volume
EBT	EBT=EBIT-O	gross profit
D		tax rate
T	T=d*EBT	tax
R	(1-d)*EBT	net profit
N		total number of stocks
EPS	EPS=(r/n)	net profit per share (issued)
Ee	ee=E/n	expected nominal value of stock
W	w=EPS/ee	relative return per stock issued
W	W=EBIT/P_a	gross return on total liability and equity
Fnd		coefficient of net profit to common stock holders
Fpd		coefficient of net profit to preferred stock holders
Fre	fre=1-(fnd+fpd)	net profit to reinvestment coefficient
ROA	ROA=(fnd+fpd)*r/P_a	net profit to total capital/total equity and liabilities
ROE	ROE=(fnd+fpd)*r/(P_a-D)	net profit to own capital/total equity and liabilities
Beta		volatility
Krf		non-risk rate of return on own capital
Kr		risk rate of return on own capital
Rm		return of money market
Ked	ked=krf+(rm-krf)*Beta	expected return on common own capital
Kpd	kpd = [r*fpd/P_a]/se	expected return on preferred own capital
WACC	WACC=D*p/Pa+ + End* ked/P_a + Epd* kpd/P_a	weighted average cost of capital (based on debt, common and preferred capital)
Div	div=r*fnd/n	dividend paid
P	P=div/ke	estimated market price of stock
PEPS	PEPS=P/EPS	ratio between estimated market price of stock and its return
CI	CI=P*n/E	market price through book price of stock ratio
Reldiv	reldiv=div/P	estimated stock market price through dividend of stock ratio

Remark: when computing WACC we combined common own capital and retained profit, since we assumed both rates of return to be equal.

As we see, the above set of notions relies on balance sheet, income statement and share holding policy. Evidently, each item from the list above could be furthermore indented to receive finer results of the procedure. A list of categories is under our discretion of their constant or variable role within the procedure. Furthermore, within variable categories it is also possible to decide upon which of them are depending on the other. It is important to do that when we want to simulate the future financial behaviour of a firm. However, as we shall see, such relationships are very rare to be established. In both cases, e. i. input/output relations being capable to be expressed formally as well as those relations which can not be put in a formalized relationship will be considered as time variable mappings; the squeezing-out procedure thus becomes a task of multi-dimensional and multi-criteria optimization over time. Consequently, the same procedure may be specialised to the analysis of the past, present and future horizon (e. i. analysis, operation and planning).

Most of our interest will be spent on the future in terms of categories being essential to market economy. For example, one of the most frequent targets is EPS; at the moment, we ignore whether or not it can be expressed analytically. Moreover, we shall allow for each variable, either input or output one, to be discretized arbitrarily. The variations of input variables, determining EPS, we want to discover its maximum value, both over time and input variables domain. Since we allowed the dychotomy of all variables (e. i. their formal as well non formal presentation), the optimization procedure is therefore carried out through discrete simulation.

Simulation target could be a single category; thus we reveal its conditional minimum and maximum being dependent on input categories chosen as well as categories chosen as fixed parameters. Here, freedom of choosing fixed parameters is large.

3. SOFT SQUEEZING-OUT PRINCIPLE

3.1. An example on 1-dimensional 1-criterion optimization

Following the outline given in Ch.2 we see that from a set W of all financial categories quoted in a table of Ch. 2 we arrive to 1-dimensional and 1-criterion optimization problem with the following structure: **OIV=optimizing input variable: z**; **SIV=steering input variables**: **Pa, n, F, V, O, ee, EBIT/Pa, R; SP=steering parameters: p, krf, kr, beta, ked+kpd, fre, fnd, fpd, d; OOV=optimized ouput variable: EPS; IOV=induced output variables : ROA, ROE, div, P, PEPS, CI, WACC, D, E-D.**

Thus, formalizing it via partition W= **OIV** $\cup$ **SIV** $\cup$ **SP** $\cup$ **OOV** $\cup$ **IOV**, we have

$$\max_z EPS = \max_z (l-d)*[c*q(z)-V(q(z))-z*p*Pa/100-F]/n = EPS_o \qquad (1)$$

leading to simple sensitivity information. An important warning: 1-dimensional 1-criterion case like (1) in general does not exist for any 1-member OIV and OOV; we are lucky to establish an analogy to (1) for other outputs.

3.2. An example on 6-dimensional 1-criterion optimization

If we want to improve EPS, we can additionally activate some steering parameters by turning them into optimizing variables. The corresponding partition is: **OIV=optimizing input variables: z, P_a, n, F, V, R; SIV=steering input variables: O, ee, EBIT/Pa, R; SP= steering parameters: p, krf, kr, beta, ked, fre, fnd, fpd, d; OOV=optimized output variable: EPS; IOV= induced output variables: ROA, ROE, div, P, PEPS, CI, WACC, D.**

The corresponding search problem is stretched over OIV = (z, P_a, n, F, V, R) as an enriched set of optimizing variables in the sense of

$$maxEPS = \max (l-d)*[c*q(z)-V(q(z))-z*p*Pa/100-F]/n = EPS_o \qquad (2)$$

where a subset SP could be additionally activated through the simulation procedure in order to increase the EPS category. The largest degree of freedom is being found for the coefficients which are responsible for a splitting a profit into reinvestments, common and preferred stocks; thus we proved that a dividend policy is most frequent device to optimize EPS.

For an efficient share-holding policy we need to know: 1) what are the impacts of individual variables on EPS; 2) whether or not it is worthwhile to increase the number of variables involved; 3) the constraints on input optimizing variables; 4) how much steering parameters are stable.

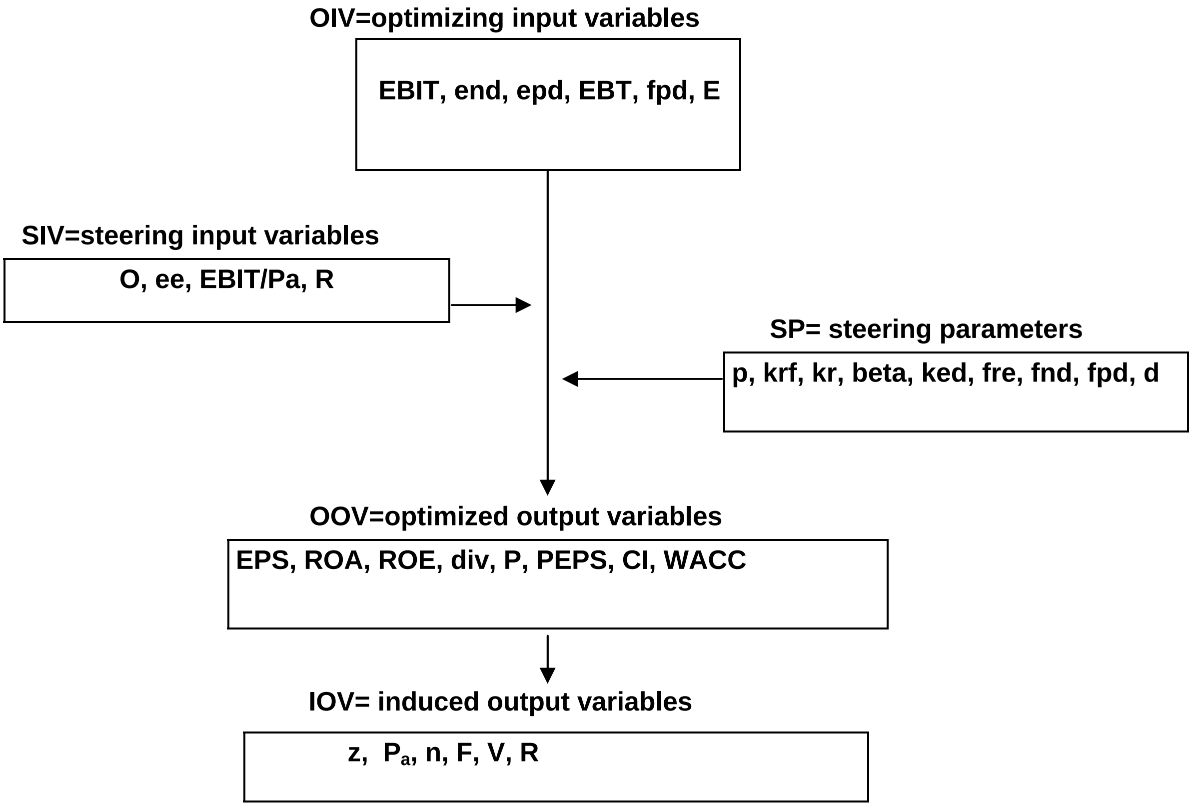
However, an important warning again: 6-dimensional (or any other finite-dimensional) 1-criterion case like (2) in general may not exist for any 1-member OOV subset; we are lucky to establish an analogy to (2).

Examples of (1) and especially (2) go far beyond the existing financial practice. They represent our endevour to “squize out” some scalar financial category, regardless of the question whether its analytic expression exists or not. The procedure has been named as “squeezing one” since it calls for optimization, whereas it is named as ”soft” due to the fact that we assumed the existence of formal presentation of 1-member subset OOV. To meet the demand of real life financial management, we need to enlarge the OIV subspace and at the same time to cover the case of soft squeezing-out procedure by an approach, subsequently described as “hard squeezing” procedure.

4. HARD SQUEEZING-OUT PRINCIPLE IN FINANCIAL MANAGEMENT

4.1. An example on 6-dimensional 8-criteria optimization

Let a partition of W be as shown by the block diagram below:



The corresponding formal presentation consists of six optimizing variables, ten optimized variables, four steering input variables and nine steering parameters (it is 6/10/4/9–optimization problem. In general, it is 6-dimensional and 8-criteria optimization problem for which a formal presentation can not be established. In addition, some of the variables are continuous, others are not. We are therefore compelled to switch to non-formal modelling (see /2/, /3/) in order to find (an approximate) solution.

A discussion of OIV:

a) interdependency of OIV variables is formalised.

The OIV set is composed of 6 not fully independent input variables. On the other case, when the role of OIV and IOV had been mutually interchanged, variables **z, P_a, n, F, V, R** were independent and therefore no dimension reduction would be possible. For the illustration, we shall deal with OIV as in the scheme above. Although, **fpd** is being dependent on **EBT** and **end** on one hand and **EBT** being dependent on **EBIT,** we can reduce the OIV, since optimal value of **EBIT** determines optimal value of **EBT**: we do not need to optimize over **EBT,** but only over **end** thus bringing optimal **fpd.** Thus, OIV subset is reduced to **EBIT, end, epd** and **E** which are all independent. The existence of formalised relationships among OIV variables helps us to reduce the scope of OIV. Thus, a) reduces to c) below.

b) interdependency of OIV variables is not formalised.

What to do, if there is no formalised relationship between the interdependent variables of OIV ? Can we reduce this set into a set of independent variables only?

c) independency of OIV variables.

This case is resistable to the condition of formalism; the reduced set OIV under a) exhibits such a case. In practice, we have to obey usual linear constraints imposed on variational spaces of OIV variables.

A discussion on OOV:

a) interdependency of OOV variables is formalised.

If there are formal relationships between them, the procedure is the same as in OIV case: a reduction of number of criteria is possible. Thus, a) reduces to c) below.

b) interdependency of OOV variables is not formalised.

What to do, if there is non- formal relationship between them? Can we reduce this set into a set of independent variables only?

c) independency of OIV variables.

This case is not effected by the existence of formalism; each criterion stands for its own role shaping the corresponding multicriteria optimization problem.

A discussion on OIV-OOV mapping:

a) A case of formalised mapping.

It is a typical multicriteria multidimensional optimization problem, where each criterion achieves its optimal value at different values of OIV variables. For example, under c) case of OIV and c) case of OOV

	EBIT	**end**	**epd**	**E**
EPS^0	0,15	0,262	0,375	1,062
ROA^0	0,015	0,035	0,775	0,705
ROE^0	0,015	0,035	0,775	0,705
Div^0	9,4%	9,6%	9,7%	9,9%
P^0	9,4%	9,6%	9,7%	9,9%
$PEPS^0$	1,59	2,79	3,98	8,53
CI^0	6,638	8,638	8,638	11,638
$WACC^0$	0,159	0,372	0,797	3,453

Each row shows optimal value of OOV mutually independent variables and corresponding optimal values of OIV variables. But, financial manager's decision has rest upon all 8 criteria (a joint decision). Thus, there are 8 criteria values **EPS^0 ,... $WACC^0$ t**o be simultaneously taken into account; the corresponding decision possibilities having no formal relationships available, require a non-formal modelling to find its »compromised« solution (see /2/, /3/).

EPS^0	0	0	0	0	0	0	0	0,15	0,262	0,375	1,062
0	**ROA^0**	0	0	0	0	0	0	0,015	0,035	0,775	0,705
0	0	**ROE^0**	0	0	0	0	0	0,015	0,035	0,775	0,705
0	0	0	**Div^0**	0	0	0	0	9,4%	9,6%	9,7%	9,9%
0	0	0	0	**P^0**	0	0	0	9,4%	9,6%	9,7%	9,9%
0	0	0	0	0	**$PEPS^0$**	0	0	1,59	2,79	3,98	8,53
0	0	0	0	0	0	**CI^0**	0	6,638	8,638	8,638	11,638
0	0	0	0	0	0	0	**$WACC^0$**	0,159	0,372	0,797	3,4537

b) A case of non-formalised mapping.
What to do, if there is non- formal relationship between them? Apparently, the two cases b) of OIV and OOV, mathematically speaking, refer to the same problem b) of non-formalised mapping a set OIV onto a set of OOV. This common case is discussed in Ch. 4.2.

4.2. Generalization of hard squizeeing-out procedure in financial management

As a theory of nonformal modelling allows to deal with attributes of different dimensions, magnitude and sign of correlation , we can extend the concept of hard squeezing-out principle to all subsets of W, provided constraints to feasible ranges and non-conflict regions have been introduced; they can be called **primary constraints**, entirely in hands of experienced financial analyst. In case of interdependent OIV and OOV variables it may happen that some additional, let us say, **secondary constraints**, are needed.

As it has been seen, the main difficulty in practicing hard squeezing-out principle is that of non-formality either of OIV, OOV or of mapping case. To generalise sufficiently the whole procedure is to respect fundamental assumption that output variables are not being capable to be formalised with respect to input variables

Before we start applying RKLR algorithm as a tool of non-formal modelling we have to establish causalities between input and output variables so as to assure their feasibility; in addition, pairwise, triple, quadruple, ect. infeasibilities are also to be foreseen (e.g. by means of interaction analysis, see /1/).However, it only reveals whether or not they are interconnected, but not their functional relationships.

Let us first have the simulation procedure being carried out upon OOV by using OIV under arbitrary partition of W, namely $W = \textbf{OIV} \cup \textbf{SIV} \cup \textbf{SP} \cup \textbf{OOV} \cup \textbf{IOV}$. It is important to notice that 5-pieces partition may bring some financial items which are not quite akin with respect to their dimensions; we shall offset this pecularity by a proposal of a generalization given below. Under any grid of simulation arbitrarily chosen we want to reveal a mapping (operator)

$$\Psi : \textbf{OIV} \rightarrow \textbf{OOV} \tag{3}$$

which is to bring the best (compromised) approximation to the optimal solution of m(OIV)/n(OOV) financial management decision under all primary constraints. Under an assumption, operator Ψ has no formal expression. If the cases a) and c) of both OIV and OOV sets were preferred in our context, we could simply use multidimensional constrained or unconstrained continuous or discrete optimization. Due to the absense of any formal description of their mutual relationships, we turn to simulation procedure.

After imposing primary constraints on OIV and OOV, we get their (in general, arbitrarily) discretized subsets $O\hat{IV}$ and $O\hat{OV}$, for which we want to find a shrinked operator $\hat{\Psi}$. Since it should bring the best (compromised) approximation to optimal decision, we have to check the Cartesian product Ω = $O\hat{IV}$ x $O\hat{OV}$ which is nothing but the input matrix to RKLR procedure as a non-formal modelling of decision algorithm. Here, an important note: if the highest rank (as a measure of quality of the solution) is to far from 1, we can repeat the same procedure carried out on Ω, but over some finer simulation grid streched over some sufficiently small vicinity of optimizing point of OIV.

5. DISCUSION ON EXPERIENCES AND CRITICISM

5.1. Some applied benefits

Practical experiences, based on soft squeezing-out principle, stem from application in SLO banking, trade and manufacturing companies. Benefits riped from this principle refer to questions like: How additionally to shift e.g. the maximal EPS upwards after we computed it, but not at satisfactory level? Are the environmental conditions convenient to assure the computed extrema underlying (3)? Have we remained/become competitive along the use of extrema obtained? Is the extremal financial policy endangered by some factors and how? What is the minimal average cost in your firm after extremal solution found? How far is your firm from the minimal average cost? And what about your marginal cost at extremal solution?

Furthermore: Can we apply the proposed model in case of producing more products? What is the position of your firm on the surface OOV, its domain OIV being fixed? What are the reasons of having a gap between extremal point and given operational point of that surface? If this difference is (component-wise) negative and absolutely increasing over time, what to do (activating SIV or SP or both?)? Shall we, in such a situation, augment the financial leverage and how far? If selling market is pushed to “the ceiling”, can we break it through? If answer is affirmative, what are their reflections on financial leverage and EPS? Does hiring cheaper credits slow down or accelerate the upward sloping of financial leverage curve? In the latter case, what other financial categories are influenced by such policy? If your firm merges with some other one, what would be an effect on slopping upwards of financial leverage curve? What about a case of its slopping down? What is IOV in case of your firm? Does your actual financial policy allow to choose variables which help to maximize EPS? And endless list of other similar questions.

5.2. Criticism

Most of applications refer to 1-dimensional criteria space, where the needs of practical management called for numerous versions of software (see /4/). There is a family of models, derived from (3) being governed by various special conditions and imposed on financial management environments. The reader may get a deeper insight from **http://www.atnet.si/interacta/ssop.html**., from where the role and functionality could be read off. For this purpose, we developed a powerful software, called ***Simulated book-accounting optimisation of profit (SKOP*** in Slovenian), which is a part of software family

SSOP (System simulation and optimisation of a firm) aiming at the corresponding broader class of financial management problems. SKOP family represents softverization of soft squizeeing-out principle used in financial management.

In the report above we spent a separate effort on independent variables in case of OIV as well in a case of OOV from a case of their interdependencies. Intuitively, we can deal with a mixed problem by, say, first solving the problem on independent variables in both cases and, then, switch to interdependent issues, provided in both cases that a formality is assured. However, a mixed case under the non-formality still requires some study.

Apparently, an optimal solution of 1-criterion optimization does not always coincide with optimal solution of more criteria optimization. Also, the quality of any suboptimal solution can be compared with optimal one inside $OI\hat{V}$ x $OO\hat{V}$, but inside (3).

6. REFERENCES

/1/ Jakulin A. and Bratko I., Quantifying and Visualizing Attribute Interactions, Faculty of Computer and Information Sciences, University of Ljubljana, Ljubljana, 2003.

/2/ Rupnik V., An Attempt to Non-formal Modelling, Proceedings of the 4th International Symposium on Operations Research, Preddvor, Slovenia, 1997.

/3/ Rupnik V., The shadowed decisions, Proceedings of the 4th International Symposium on Operations Research, Preddvor, Slovenia, 1997.

/4/ http://www.atnet.si/interacta/ssop.html.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 10

Production and Inventory

AHP METHOD AND LINEAR PROGRAMMING FOR DETERMINING THE OPTIMAL FURNITURE PRODUCTION AND SALES

Peter Bajt, Unec 82, 1381 Rakek, Slovenia
Lidija Zadnik Stirn, Biotechnical Faculty, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
bajt.peter@volja.net, lidija.zadnik@bf.uni-lj.si

Abstract: Producers and sellers have to be familiar with the consumers' habits and their criteria if they want to be competitive on the market. This is also true for furniture production and selling. Thus, in the paper we first deal with the problem of determining the optimal assortment of windows which are to be produced in the selected wood manufacturing company in Slovenia according to the consumers' needs and demands. This problem is assigned as a multicriteria problem, and AHP method is used to establish the most suitable material and style for windows produced and sold on the Slovene market. Further, linear programming was applied to optimize the selected windows' production according to the economic, technical and human resource constraints. Results of the research will be used as guidelines in reengineering of the selected Slovenian wood manufacturing company, i.e. for improving production and investment policy.

Key words: building furniture, windows, AHP method, linear programming

1. INTRODUCTION

In Slovenia there are many competitive producers of building furniture. They produce windows, outer doors, shutters and other products from different kinds of material (wood, wood/aluminium, artificial material, artificial material/aluminium and aluminium). Buyers take into account different criteria when choosing and consequently buying building furniture. The most frequently applied consumers' criteria are: price, easiness of maintenance, durability, environmental acceptability, thermal isolation and shape. Here we present only the problem of selection and production of the most suitable assortment of windows made in a selected Slovene wood manufacturing company. AHP method was used for solving a multicriteria decision making problem for the election of the most suitable type and material of windows according to the consumers' criteria (Saaty, 1994, Winston, 1994). Windows produced from artificial material were determined as the most appropriate. Further, taking into account that the technological process of windows made from artificial material consists of various working operations, we established for the next planning period the optimal plan of production with linear programming (Caine and Parker, 1996).

2. METHODS AND DATA

In the research we took into account a Slovene wood manufacturing company which produces windows (the name is not given here for the sake of data security). In order to find out which material and type of windows could be of the greatest interest to the consumers, we generated a survey and made a random selection of co-workers (they are regarded also as potential consumers) from the company's sale and production department. In the survey three types of windows were given as a decision option (windows made from artificial material, windows made from wood, windows made from wood and aluminium), and the following three criteria were considered: price, durability, maintenance and environmental acceptability of a window (Figure 2). Each co-worker was first informed about the 1-9 marking scale (example is given in Figure 1) and then asked to give the assessment for every pairwise comparison. The median of all assessments (Saaty, Aczel, 1983) was calculated for every comparison. Using these data, the most suitable material for windows production and sale in the department of the chosen wood manufacturing company was determined by AHP method and computer program Expert Choice 2000.

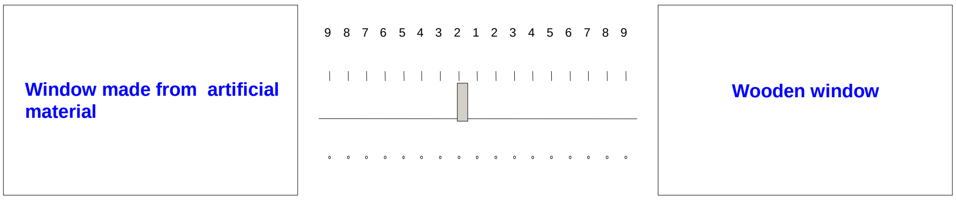


Figure 1: An example of marking scale for pairwise comparison

The windows from artificial material, which were selected by AHP as the most suitable regarding consumers' needs and wishes, could be produced in the company under consideration in various widths and heights. The following data were chosen in our research (Table 1):

- representatives of windows produced: P_1, P_2, P_3, P_4, P_5, P_6 in P_7
- the net profit per window produced in MU (monetary units)
- working operations (W.OP. 1 to W.OP. 7) with manufacturing times for each product
- available working hours in the following planned period.

Table 1: Production problem of windows, made from artificial material

	Product (piece)							
Working operation	P_1	P_2	P_3	P_4	P_5	P_6	P_7	Constraint (hours)
W.OP. 1	0,25	0,3	0,35	0,6	0,6	0,65	0,7	1440
W.OP. 2	0,2	0,25	0,25	0,45	0,45	0,5	0,5	1080
W.OP. 3	0,3	0,35	0,4	0,65	0,7	0,7	0,75	1620
W.OP. 4	0,2	0,2	0,2	0,35	0,4	0,4	0,45	900
W.OP. 5	0,2	0,2	0,25	0,45	0,45	0,5	0,5	1080
W.OP. 6	0,35	0,35	0,4	0,7	0,7	0,75	0,8	1710
W.OP. 7	0,2	0,25	0,25	0,45	0,45	0,5	0,55	1170
Profit (MU)	26	26	26	22	20	24	18	
Quantity (pieces)	x_1	x_2	x_3	x_4	x_5	x_6	x_7	

Excel 2003 with a sensitivity report was used for solving this linear production problem in which we were interested in:

- optimal production program of windows at which the net profit is maximal;
- which are still maximum allowed costs per hour if an additional working hour for operation W.OP. 4 is foreseen; and how many additional working hours are reasonable to initiate so that we are still able to increase the net profit,
- which products can increase the profit (value) and for how much so that the product would come in the optimal program.

3. RESULTS

From the decision tree (Figure 2), it is evident that the buyers assign environmental acceptability of the product (windows) as at least important. The most important of all criteria for the consumers is the price of the windows. The synthesis of the final assessment (Figure 3) reveals that the most suitable window for sale on the Slovene market produced in the selected company is to be made of artificial material (the final value of this alternative is

0,601, i.e., it covers 60,1 % of the total windows' production in the selected company). The results also show a small difference (only 0,017) in values between wooden and wood/aluminium windows. The calculated overall inconsistency ratio was 0,03, which is smaller than 0,1.

The production problem of the windows made of artificial material is according to Table 1 presented bellow, where x_1 to x_7 present the quantity (pieces) of products P_1 to P_7:

$$26x_1 + 26x_2 + 26x_3 + 22x_4 + 20x_5 + 24x_6 + 18x_7 \quad \text{maximum}$$

$$\begin{array}{rcl}
0{,}25x_1 + 0{,}3x_2 + 0{,}35x_3 + 0{,}6x_4 + 0{,}6x_5 + 0{,}65x_6 + 0{,}7x_7 & \leq & 1440 \\
0{,}2x_1 + 0{,}25x_2 + 0{,}25x_3 + 0{,}45x_4 + 0{,}45x_5 + 0{,}5x_6 + 0{,}5x_7 & \leq & 1080 \\
0{,}3x_1 + 0{,}35x_2 + 0{,}4x_3 + 0{,}65x_4 + 0{,}7x_5 + 0{,}7x_6 + 0{,}75x_7 & \leq & 1620 \\
0{,}2x_1 + 0{,}2x_2 + 0{,}2x_3 + 0{,}35x_4 + 0{,}4x_5 + 0{,}4x_6 + 0{,}45x_7 & \leq & 900 \\
0{,}2x_1 + 0{,}2x_2 + 0{,}25x_3 + 0{,}45x_4 + 0{,}45x_5 + 0{,}5x_6 + 0{,}5x_7 & \leq & 1080 \\
0{,}35x_1 + 0{,}35x_2 + 0{,}4x_3 + 0{,}7x_4 + 0{,}7x_5 + 0{,}75x_6 + 0{,}8x_7 & \leq & 1710 \\
0{,}2x_1 + 0{,}25x_2 + 0{,}25x_3 + 0{,}45x_4 + 0{,}45x_5 + 0{,}5x_6 + 0{,}55x_7 & \leq & 1170
\end{array}$$

$$x_i \geq 0 \qquad i = 1, 2, \ldots, 7$$

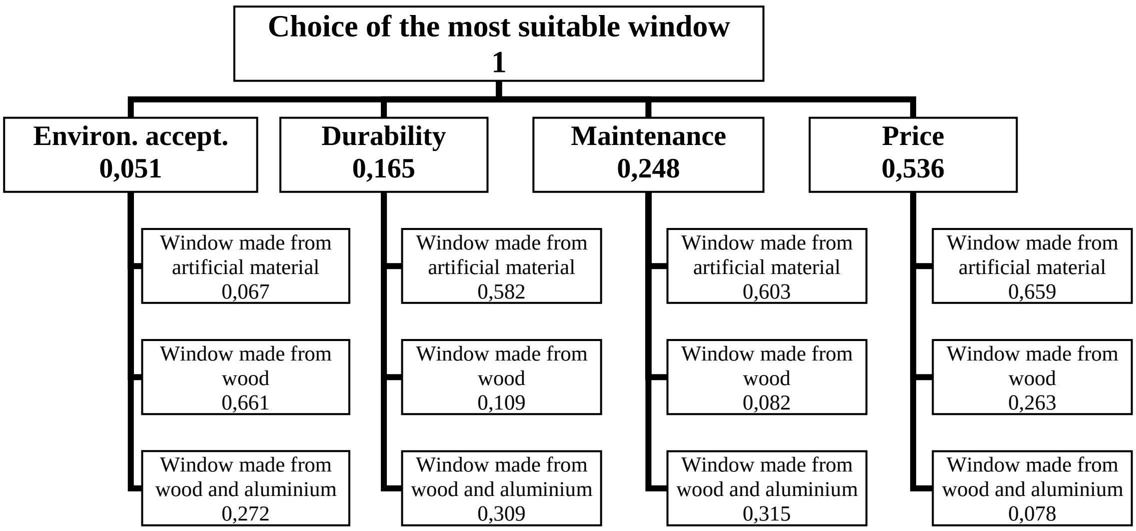


Figure 2: Decision making tree for choosing the most suitable window regarding the material

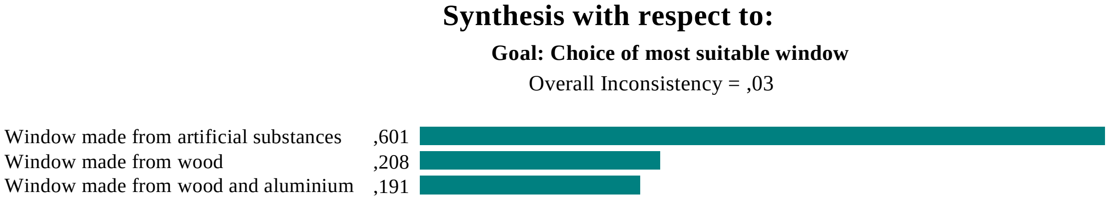


Figure 3: The assessment of suitable windows for production and sale

Using the Excel 3000 program we received the following optimal LP solution: producing only product P_1 (4500 pieces) the company's net profit amounts to 117000 MU. The results in Table 2 tell us that just W.OP 4 is fully utilized. Slack value is 0.

Table 2: Results concerning constraints at windows production problem

Name	Status	Slack
W.OP.1	Not binding	315
W.OP.2	Not binding	180
W.OP.3	Not binding	270
W.OP.4	Binding	0
W.OP.5	Not binding	180
W.OP.6	Not binding	135
W.OP.7	Not binding	270

The results given by sensitivity report (Table 3) show that besides the product P_1 also products P_2 and P_3 can be produced, each in the amount of 1500 pieces, to achieve equal maximum profit (they have value 0 in column "reduced costs"). It is evident from Table 4 (under column "allowable increase") that it is reasonable to increase capacity of W.OP.4 for at most 77,143 working hours. Hours exceeding this value will stay unused and the net profit will stay the same. Any hour of increase of W.OP 4 working capacity will increase the net profit by 130 MU (if the increase of the working capacity is "small enough" and there is no degeneration). Products P_4, P_5, P_6 in P_7 can come to the optimal program only if we increase their profit (values per piece): P_4 by 23,5 MU per piece, P_5 by 32 MU per piece, P_6 by 28 MU per piece and P_7 by 40,5 MU per piece (Table 3).

Table 3: Sensitivity report concerning adjustable cells (products)

Name	Final value	Reduced cost	Objective coefficient	Allowable increase	Allowable decrease
P1	4500	0	26	1E+30	0
P2	0	0	26	0	1E+30
P3	0	0	26	0	1E+30
P4	0	-23,5	22	23,25	1E+30
P5	0	-32	20	32	1E+30
P6	0	-28	24	28	1E+30
P7	0	-40,5	18	40,5	1E+30

Table 4: Results of constraints for the production problem

Name	Final value	Shadow price	Constraints R.H. side	Allowable increase	Allowable decrease
W.OP.1	1125	0	1440	1E+30	315
W.OP.2	900	0	1080	1E+30	180
W.OP.3	1350	0	1620	1E+30	270
W.OP.4	900	130	900	77,143	900
W.OP.5	900	0	1080	1E+30	180
W.OP.6	1575	0	1710	1E+30	135
W.OP.7	900	0	1170	1E+30	270

4. CONCLUSION

AHP method was used to determine the most acceptable windows for sale on the Slovene market from the buyers' point of view. In this sense the windows produced from artificial materials were chosen. These results also show that the price of the windows is the most important criterium for consumers. However, we noticed in our research that consumers consider at least the criterion of the environmental acceptability of products. It will also be very interesting to follow consumers' habits in the future because their awareness of the importance of environmental friendly products is increasing.

In the second part of the research we found that the company can gain the largest profit if it produces only product P_1 or products P_1, P_2 and P_3 in equal shares. Other products could come in optimal program only if their profits (values per piece) are increased. We can do this with the increase of sale prices or if we use more efficient technology. As we are not able to increase sale prices (in this case we will not be competitive on the Slovene market), we have to invest more knowledge and research in a new technology, and in such a way reach shorter production times of individual working operations.

5. REFERENCES

- Caine D.J, Parker B.J., 1996. Linear programming comes of age: a decision–support tool for every manager. Management Decision 34/4: 46-53.
- Excel 2003. Microsoft (computer program)
- Expert Choice 2000 (computer program)
- Saaty, T.L., Aczel, J., 1983. Procedures for Synthesizing Ratio Judgements. Journal of Mathematical Psychology, 1983/27/1, pp. 93-102.
- Saaty. T. L., 1994. Fundamentals of decision making and priority theory. RWS Publications, Pittsburgh.
- Winston, W. L., 1994. Operations Research; Applications and algorithms. Duxbury Press, Bellmont, CA.

SEMANTIC GRID BASED PLATFORM FOR ENGINEERING COLLABORATION

Matevž Dolenc, Robert Klinc and Žiga Turk
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
{mdolenc, rklinc, zturk}@itc.fgg.uni-lj.si

Abstract: The integration and interoperability of engineering software applications have been providing one of the most challenging environments for the application of information and communication technologies. The InteliGrid project combined and extended the state-of-the-art research and technologies in the areas of semantic interoperability, virtual organisations and grid technology to deliver a new semantic grid platform prototype enabling access to information, communication and processing infrastructure. The paper provides an overview of the developed semantic grid platform.

Keywords: grid technology, semantic grid, SOA, engineering, collaboration, ICT, InteliGrid

1. Introduction

The integration and interoperability of hundreds of engineering software applications supporting the design and construction of the built environment have been providing one of the most challenging environments for the application of information and communication technologies. The "islands of automation" problem [1] has been identified by the AEC community in the late 1980s, and several national and EU projects have been tackling the problem since. Conceptually, the integration solutions have been betting on the agreement on commonly accepted and standardized data structures, such as ISO-STEP or IAI-IFC standards. Projects such as COMBI [2], ATLAS [3], ToCEE [4], ISTforCE [5] and others proved both theoretically and with prototypes they developed that interoperability based on product data technology is achievable and the industry can benefit from it. But despite all research and development efforts such solutions are still rare in the industry. Since the focus of the above projects has been primarily the data structures describing the problem domain, the actual research communication platform prototypes used whatever was the information communication technology state-of-the-art at the time.

The statement by I. Foster [6] captures the essential requirements of collaboration inside the civil engineering sector: "the problem is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations ... not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving in industry. This sharing is highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs". This statement became one of the definitions of grid computing, particularly for the evolution of grid technology towards semantic grid. It gave ground to the InteliGrid [7] hypothesis that semantic grid technology could provide the solution to the above interoperability and information access problem.

The main goal of the InteliGrid project was to provide the engineering industries with challenging integration and interoperability needs a flexible, secure, robust, ambient accessible, interoperable, pay-per-demand access to (1) information, (2) communication and (3) processing infrastructure. The project addressed the challenge by successfully combined and extended state-of-the-art research and technologies in three key areas: (a) semantic interoperability, (b) virtual organisations, and (c) grid technology (see Figure 1) to provide standards-based collection of ontology based services and grid middleware in support of dynamic virtual organisations as well as grid enabled engineering applications. It was recognized that if a grid technology is to ensure the underlying engineering interoperability and collaboration infrastructure for a complex engineering virtual organisation the grid technology needs to support shared semantics. This is the area where major innovations and extensions of current grid middleware technologies are required.

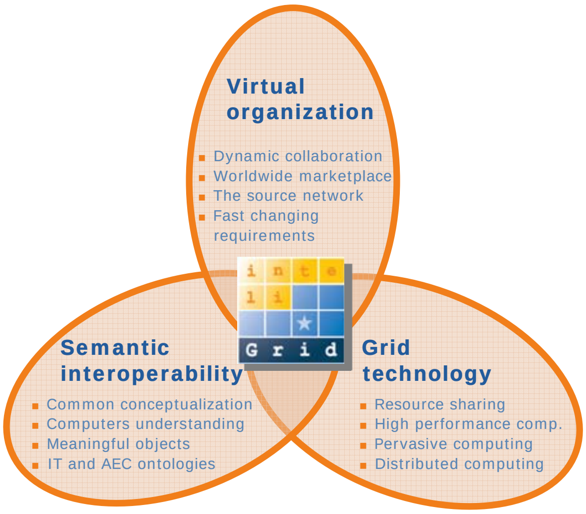


Figure 1: The InteliGrid project addressed three key technology areas.

2. Semantic grid architecture

The designed InteliGrid architectural framework draws together experiences from projects such as ToCEE and ISTforCE, the Service-Oriented Architecture (SOA) [8] and Model Driven Architectures [9]. InteliGrid's basic assumption is that software not only has to model the real world, it also has to model the technical resources that this software is using because these resources are becoming increasingly complex in a networked or grid environment. The InteliGrid framework architecture includes four layers shown in Figure 2: (a) the problem domain layer, (b) various conceptual models and ontologies, (c) the software layer which includes applications and services, (d) the layer of basic hardware and software resources, whereby both (c) and (d) are to some extent modelled also in (b). The software architecture (layer c) distinguishes between business applications, interoperability services, business services and grid middleware services. The concepts in the layer (b) are organised in the following ontologies: business ontology, organisation ontology, service ontology and meta-ontology. Services are loosely coupled and follow one of the most important SOA principles - they may be individually useful, or they can be composed to offer specific higher-level functionality. The following common characteristics can be defined for all InteliGrid services as main components:

- services are modular components that can be semantically described, registered, discovered and finally used by clients,
- services may be completely self-contained or depend on availability of other services,
- services are able to advertise details such as their capabilities, interfaces and supported communication protocols according to pre-defined concepts and ontologies, and
- all capabilities provided by services as well as communication and data channels among them and clients are protected by security and message level security mechanisms.

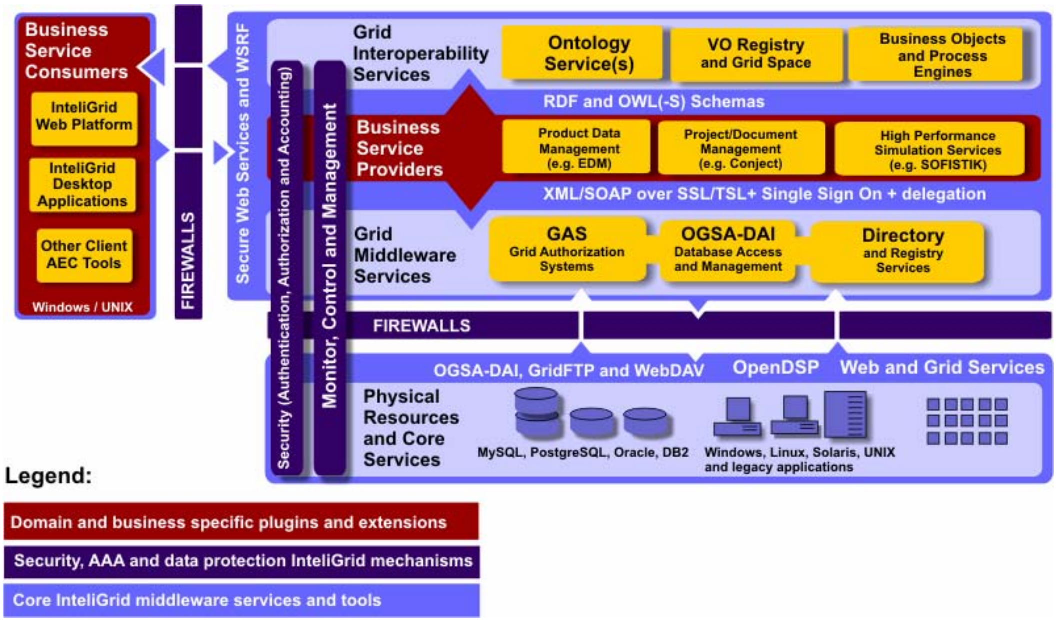


Figure 2 : InteliGrid high-level architecture components: end-user applications (left), services (right), and basic resources (bellow). Services are logically grouped into: business services (central), interoperability services (top), and grid midleware services (bottom).

Technically speaking, components are deployed either at some workstation or at a remote node on the grid. If on the grid, it is not important where they are deployed physically; the resource where they run will be very likely allocated dynamically. The grouping of the various services in Figure 2 is presented according to the logic of the service and does not necessarily imply who uses which. There are three main types of components in the InteliGrid platform:

- Domain and business specific applications. These applications are consumers of other services and are usually accessed through a web based portal interface although desktop applications can also make use of different available services.
- Secure Web Services and WSRF [10] compliant services. They can be further divided into: (1) interoperability services (top tier) that simplify the interoperability among all services, (2) domain and business specific services that perform some value added work. There are two kinds of business services: (a) collaboration services provide file and structured data sharing and collaboration infrastructure, and (b) vertical business services that create new design or plan information.
- Middleware services. These services offer traditional grid middleware functionality but extended with the particular needs of the InteliGrid platform. The services are based on mature grid technologies and their open source reference implementations – the underling service framework is based on the Globus Toolkit [10].
- Other resources. The bottom layer of the architecture consists of various physical infrastructure resources offered to the platform by suppliers. All these resources are available and can be accessed remotely through well defined interfaces and secure communication protocols. These services among others include: (1) services for remote data access, (2) remote application submission and control, etc.

When developing the described platform, the general design principle was that the use of new and advanced technologies such as grid technology, semantic web and grid services, etc. should not redefine the way end-users use the provided platform. Thus the InteliGrid platform appears to the user just as any other collaboration environment. It is the functionality and features of the shared environment that are making the difference. The feeling that there is in fact such a thing as the "InteliGrid Platform" is apparent only with specific activities, for example: getting or storing data, finding and running services and applications, etc. In actions like that the user will feel that the application that he is using is communicating with something - some services which are somewhere on the network, on the grid. An end user will have hands-on experience with domain and business specific components in the architecture. Only specific identified user types (e.g. grid administrator, virtual organisation CIO) will need to care about what is in the lower layers of the architecture.

3. Demonstration

An integrated demonstration from an architecture, engineering and construction (AEC) sector has been the basis for requirements and validation of the results. Although the developed engineering collaboration platform is designed to be used in different engineering domains the AEC sector has been identified as the most challenging environment for the application of the developed platform. Extensive description of the integrated demonstration [11] is out of scope of this paper so only a brief summary of the demonstration steps is presented bellow (Figure 3) together with representative screenshots demonstrating a selection of developed applications and services (Figure 4).

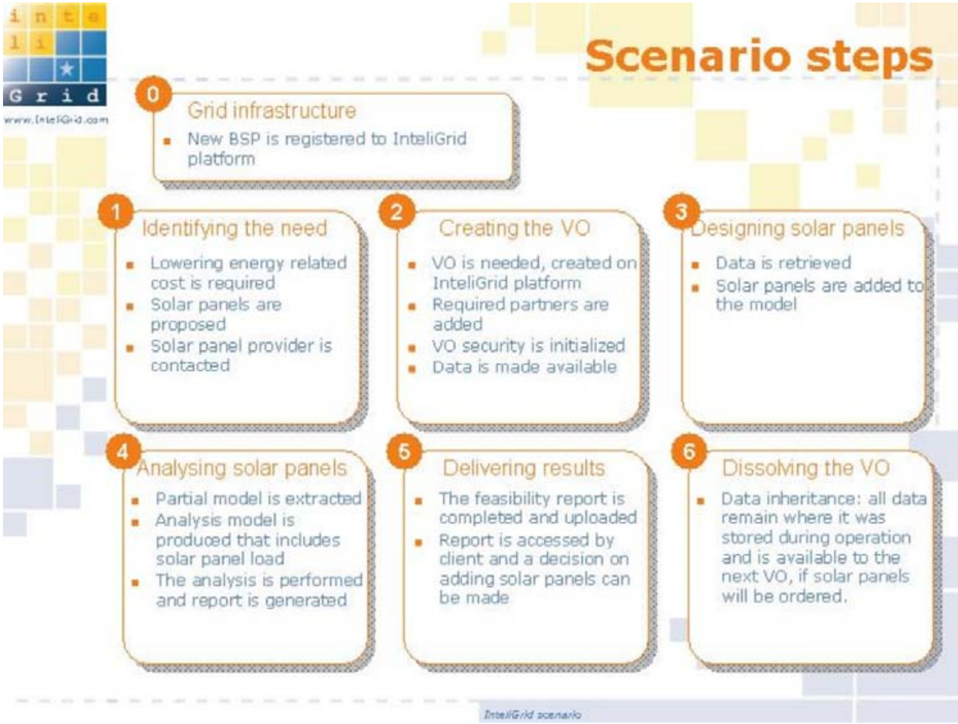
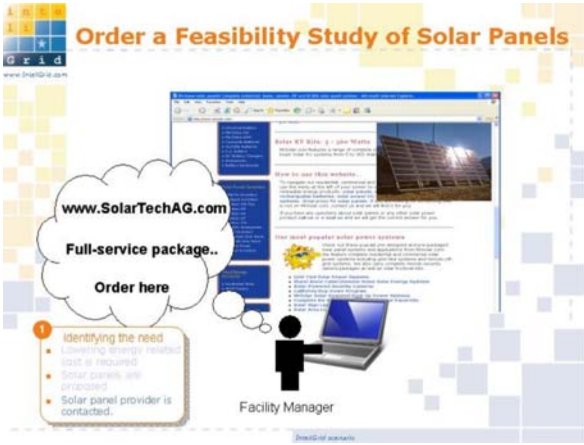


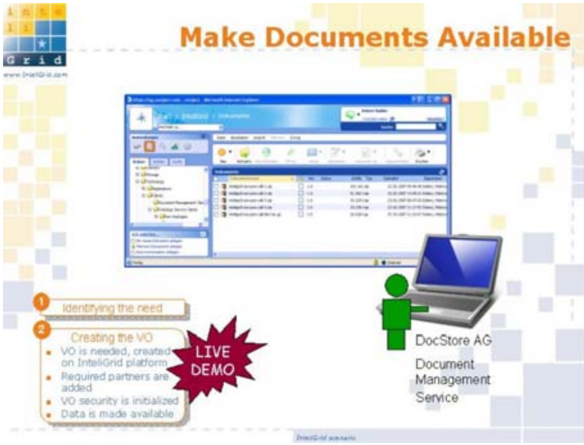
Figure 3 : The integrated demonstration includes 6 basic steps (parts) each demonstrating different aspect of the proposed InteliGrid solution.

4. Conclusions and lessons learned

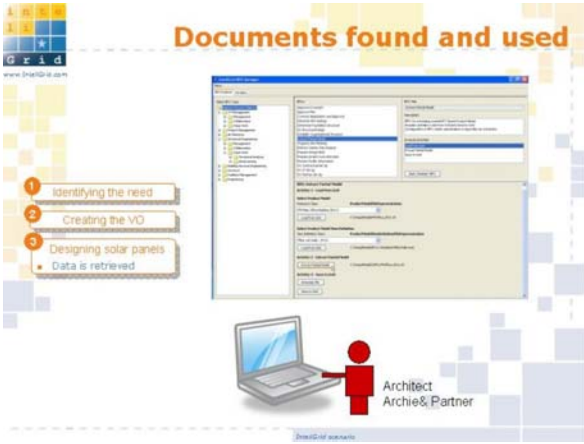
The presented semantic grid platform for engineering collaboration addresses the long standing problem of integration and interoperability in many engineering sectors. Although the described platform is not yet feature complete it has been successfully demonstrated several times that the approach taken by the InteliGrid project can provide solutions to various problems related to integration, interoperability, access to heterogeneous information, sharing of network resources, etc.



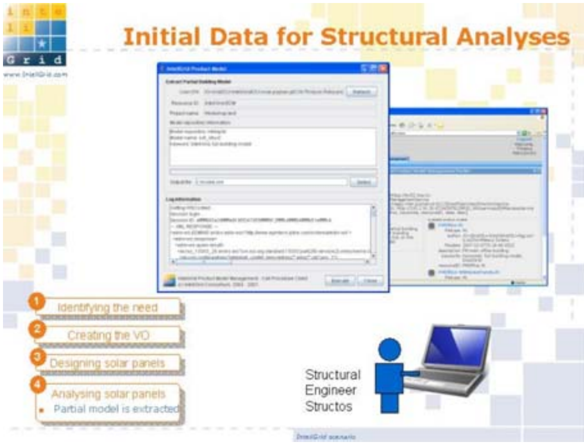
Step 1 – The InteliGrid platform is used to semantically search for relevant main contractor who takes the role of virtual organisation manager.



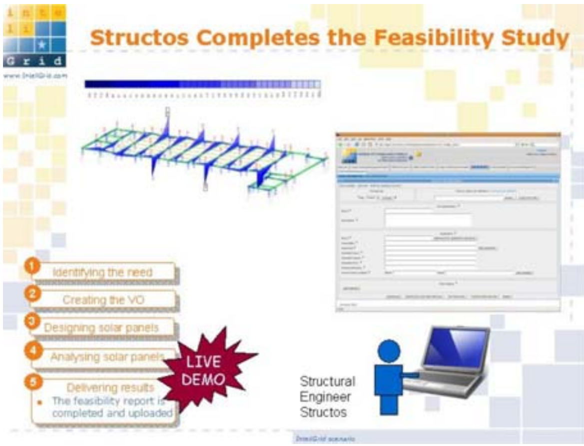
Step 2 – Initial data (document, specifications, design plans, etc.) is annotated and made available to the established virtual organisation.



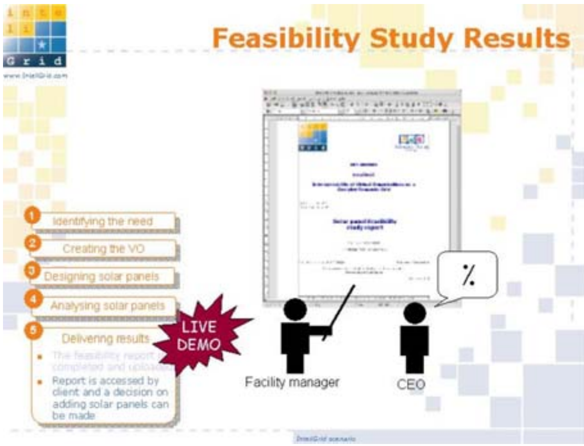
Step 3 – An end-user responsible for the designs finds the relevant documentation and delivers the modified architectural designs.



Step 4 – Structural engineer accesses partial model of the modified design.



Step 5 – Structural engineer utilises high-performance components to perform structural analysis.



Step 6 – The final report of the study is delivered to the client who makes a final decision about further investments

Figure 4 : Representative steps in the integrated demonstration showing the use of the platform as well as some of the developed applications and services.

Several important lessons learned during the platform design and development can be summarized as following:

- Combining multiple cutting edge technologies brings about a number of benefits. However, merging these technologies can also provide various problems on development level: (1) the difficulties in communicating concepts between sub-domains, (2) interface problems, (3) lack of the expected flexibility, etc.
- Basic technologies in the addressed ICT sub-domains are not yet stable to the extent that is required for rapid achievement of industry-relevant results.
- Stability of basic standards and tools for at least 2-3 years is needed to enable practical results in non-ICT industries.
- Semantic web technology, if appropriately applied, can considerably enhance available grid middleware and strengthen user orientation and user acceptance of grid solutions.
- The use of ontologies makes a lot of sense on the business layer. However, it is an open question whether there is a benefit to annotate lower layers and how “deep” such annotation should take place.
- The major gap in VO collaboration environments remains the lack of efficient interoperability on the data level.

References

[1] Hannus M. & Silen P. (2002). Islands of Automation, http://cic.vtt.fi/hannus/islands/

[2] Scherer R.J. (1995). EU-project COMBI - Objectives and overview. ECPPM, Proceeding: Product and Process Modelling in the Building Industry, Scherer (ed.), Balkema.

[3] Greening R. & Edwards M. (1995). ATLAS implementation scenario. ECPPM, Proceeding: Product and Process Modelling in the Building Industry, Balkema.

[4] ToCEE - Towards a Concurrent Engineering Environment in Building and Engineering Structures Industry. (1996). http://cic.vtt.fi/projects/tocee/index.html.

[5] Katranuschkov P., Scherer R.J. and Turk Z., (2001). Intelligent services and tools for concurrent engineering? An approach towards the next generation of collaboration platforms. ITcon Vol. 6, Special Issue Information and Communication Technology Advances in the European Construction Industry, pg. 111-128, http://www.itcon.org/2001/9

[6] Foster I., Kesselman C., Nick J., Tuecke S. (2002). The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration. Open Grid Service Infrastructure WG, http://www.globus.org/alliance/publications/papers/ogsa.pdf

[7] InteliGrid - Interoperability of Virtual Organizations on a Complex Semantic Grid, http://www.InteliGrid.com

[8] Erl T. (2005). Service-Oriented Architecture (SOA): Concepts, Technology, and Design. Prentice Hall PTR

[9] Kleppe A., Warmer J., Bast W. (2003). MDA Explained: The Model Driven Architecture--Practice and Promise. Addison-Wesley Professional, 1st edition

[10] Foster I., (2006). Globus Toolkit Version 4: Software for Service-Oriented Systems. IFIP International Conference on Network and Parallel Computing, Springer-Verlag LNCS 3779, p. 2-13.

[11] Dolenc M., Turk Z., Katranuschkov P., Krzysztof K., (2007). D93.2 Final report, The InteliGrid Consortium c/o University of Ljubljana, www.inteliGrid.com., http://www.inteligrid.com/data/works/att/d92_2.content.00832.pdf

AN EXTENDED APPROACH FOR PROJECT RISK MANAGEMENT

Janez Kušar[1], Lidija Bradeško[1], Lado Lenart[2] and Marko Starbek[1]
[1]Faculty of Mechanical Engineering, Aškerčeva 6, Ljubljana, Slovenia
[2]"Jožef Stefan" Institute, Jamova 39, Ljubljana, Slovenia
janez.kusar@fs.uni-lj.si

Abstract: In this paper an extended approach for risk-analysis method on product projects is presented. The emphasis is given to the solution, developed in the Faculty of Mechanical Engineering, supported by the MS Project software. In our solution a special attention is paid to the connection of individual activity risk analysis and the so-called status indicators. An important advantage of this solution is that the project manager and his team members are timely warned on a risk event and thus are ready for activation of the foreseen preventive and corrective measures.

Keywords: project management of orders, project risk management, status indicators

1. INTRODUCTION

Mass production was a prevailing production concept till the end of the 20th century, while today's companies favour a transition to the project type of production [1]. This is not only the case in companies which manufacture special equipment for new investments – this transition can also be seen in companies which have used mass production traditionally, e.g. in automotive industry [2]. The companies nowadays have to deal simultaneously with continuous and project processes.

Continuous processes are carried out for an "indefinite period of time"; they are used (according to the market demand) for providing new quantities of previously developed products.

Project processes are carried out once or in standard (modified) repetitions; they are aimed at achieving precisely defined objectives, for a known customer, and their duration is limited to a "definite period". Project processes can be either internal or market-oriented.

In spite of the fact that project processes are recurring, project risk management is very important, because these projects are very precisely defined in terms of deadlines and costs. Any discrepancy from the project plan can thus lead to business and competitive losses. Additionally, at the start of the project, the customer and the company jointly take a risk for successful project implementation and for good market sales of the product (e.g. automotive industry).

In continuation of this paper, emphasis will be given to practical aspects of risk management in product order projects, based on experience in implementing project management in Slovene companies.

2. PROJECT RISK MANAGEMENT

Project management consists of several processes; in [3] five key groups of project management processes are defined:

- initiating processes,
- planning processes,
- executing processes,
- controlling processes,
- closing processes.

Workgroups are responsible for implementation of individual project process, and they also assume responsibility for project risk management. Rojer [4] complemented project management processes with risk management processes, as presented in Figure 1.

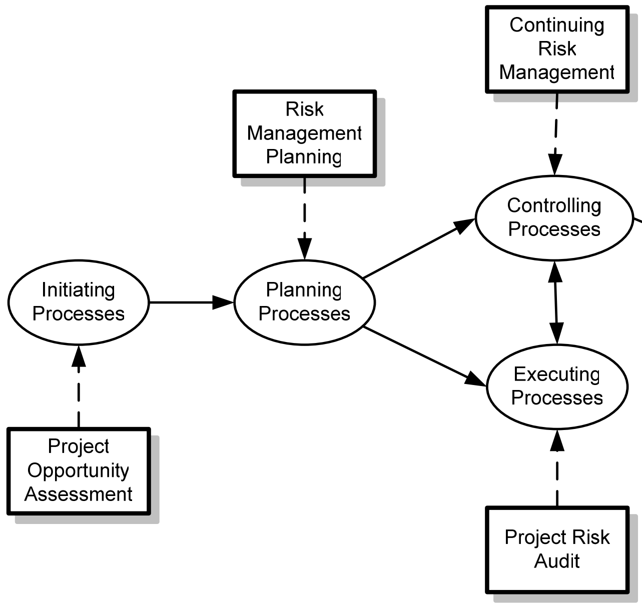


Figure 1: Risk management processes

Project risks are possible events or circumstances which can threaten the planned project implementation. Risk analysis is the most important tool, used by project managers for project processes risk management. Several methods are available for project risk analysis, especially for analysis of its activities [5], [6]. An analysis of available methods has revealed that the most suitable tool for project management of products is the *Critical success factors table*, as it represents an analytical aid for finding, evaluating, reducing and removing risks. It is elaborated by the project team, which is responsible for project planning and management. Design of the critical success factors table consists of risk analysis and risk management.

2.1. Risk analysis

Risk analysis consists of identification of problems or risk events, definition of the probability of their occurrence, evaluation of their consequences and incidences, and risk-level calculation [3].

During problem identification, the project team sequentially analyses all activities, defined in the project WBS. Possible problems of individual activities are entered into the critical success factors table (Table 1). If it is not possible to identify problems related to a particular activity, it is omitted.

Table 1: Critical success factors table

Risk analysis						**Risk management**		
No.	Activity/WBS code/ problem	Event probability EP	Estimate of consequences CE	Incidence estimate IE	Risk level RL	Measures P – preventive K - corrective	Responsibility	Indicator
1.	Activity 1							
2.	Activity 2							
n.	Activity n							

Quantitative risk analysis is defined by activity risk level, which is calculated on the basis of the following estimates:

- estimate of probability that a problem or risk event will occur,
- estimate of consequences of a problem or risk event,
- estimate of incidence of a problem or risk event.

During estimating, an interval scale from 1 to 5 [7] or a scale with estimated probability values is used [3]. The authors of this paper used the first option for the sake of its simplicity of use.

Probability that a problem or risk event will occur is estimated by using Table 2.

Table 2: Probability that a risk event will occur

Estimate	Event probability - EP
1	very small
2	small
3	medium
4	high
5	very high

In order to estimate the consequences of a problem or risk event, Table 3 is used.

Table 3: Estimate of consequences of an event

Estimate	Estimate of consequences - CE
1	very small
2	small
3	medium
4	high
5	very high

In [3] [7], the risk is defined only by estimating the probability that a risk event will occur and the estimated consequences. This article deals with project management of cyclically recurrent projects, so experience derived from similar past projects can be used for estimating the incidence probability of a risk event. Estimating the incidence of a problem occurring may seem unnecessary; however, the practice has proven that some problems, that affect the risk, are "chronically" recurring, although the company managements try to eliminate them.

Table 4 is used for estimating the incidence of a problem or risk event.

Table 4: Event incidence estimate

Estimate	Event incidence estimate - IE
1	never
2	very rarely
3	rarely
4	often
5	very often

Risk level (RL) of the activity is calculated by:

$$RL = EP \ x \, CE \ x \ IE$$

2.2. Risk management

If the risk analysis is done only on the basis of the estimated probability that an event will occur and on the estimate of its consequences, decision matrix can be chosen [3], [7] and on its basis it can be decided whether the risk is small, medium or high. The decision matrix is two-dimensional.

After addition of the risk-event-incidence factor, the decision problem becomes three-dimensional, so decisions cannot be made by using a two-dimensional matrix. We solved this problem by defining risk-level threshold values on the basis of experience:

- If $RL \leq 24$, the risk is small.
- If $25 \leq RL \leq 60$, the risk is medium.
- If $RL \geq 61$, the risk is high.

If the risk is small (normal) the project team does not specify any measures in advance. If the risk is medium, the project team prepares preventive measures, which are focused on elimination of sources for risk event occurring. If the risk event occurs nevertheless, the project team has to immediately create a corrective measure. If the risk is high, the project team prepares both preventive measures (to prevent that the risk event would occur) and corrective measures, which start processes for alleviation of risk-event consequences.

The project team enters the measures, together with bearers of responsibility, into Table 1, and defines indicators which warn project participants that project development requires starting an action. Project manager, project team, customer and operators of activities are responsible for project-risk monitoring and for the implementation of measures.

In practice, MS Project is often used as a tool for project management IT support, so the employees of the Centre of Excellence for Modern Automation Technologies on Faculty of Mechanical Engineering, Ljubljana, Slovenia, together with our partners in companies, decided that the above-presented extended risk-analysis methodology would be built into templates. Although in the server version of MS Project it is possible to use a risk-analysis tool, we estimate that from the user's perspective the proposed solution is simple, yet very effective.

3. CASE STUDY OF A PROJECT RISK ANALYSIS

As a case study of using the proposed method for project risk analysis in MS Project environment, we have chosen a simplified case of an order execution project.

For the purpose of the project risk analysis, the company management organised a creativity workshop, whose goal was to analyse all kinds of risks that may occur in projects in their company, to incorporate them (together with possible measures) into the project management rules of the company, and to make the critical success factors table, which would be used to extend the standard MS Project template. The table (which is a result of the creativity workshop) is presented in Figure 3.

Risk analysis

	Name	Risk description	Event probability	Estimate of onsequence	Incidence estimate	Risk level	Indicator	Measures	Responsibility	Hyperlink
0	⊟ PROJECT: ORDER		2,33	2,88	2,22	27,04	green			
1	⊟ PRODUCT DEVELOPMENT		2,67	3,33	3	32,33	yellow			
2	Product definition	Lack information	3	3	3	27	yellow	P:Weekly review	Project manager	
3	Prototype design	Repetition of part	1	3	2	6	green			
4	Test of prototype	Mistake on test equipmenti	4	4	4	64	red	P:Permanent control, K:Extra directive	Design engineer	Risk description.doc
5	⊟ PROCESS DEVELOPMENT		2	2	1,33	21,33	green			
6	Concept of technological procedures		0	0	0	0	green			
7	Tools and equipment	Supply delay	4	4	3	48	yellow	P: Weekly contact with supplier	Purchase	
8	Assembly and control equipment	Supply delay	4	4	3	48	yellow	P: Weekly contact with supplier	Purchase	
9	Logistics plan		0	0	0	0	green			
10	Process preparing		0	0	0	0	green			
11	Test production	Mistake on procedures	4	4	2	32	yellow		Technologist	
12	⊞ VALIDACITION OF PRODUCT AND PROCESS		2,4	2,2	1,8	26	green			
18	⊟ REGULAR PRODUCTION		2,25	4	2,75	28,5	green			
19	Start of regular production	Delay start	3	5	4	60	yellow	P:Control of conditional 5 days before start	Project manager	
20	Confirmation of validation	Mistake on procedures	3	4	3	36	yellow	P:Permanent control	Project manager	
21	Finish of FMEA	Term delay	2	2	2	8	green			
22	Regular (serial) production	Term delay	1	5	2	10	green			

Figure 3: Risk analysis and management table in MS Project

Project manager, team members and operators of activities can get the following data from Table 3:

- short definition of risks,
- event occurrence probability estimate,
- estimate of event consequences
- event-incidence estimate,
- risk level and risk indicator (in colours),
- responsibility for risk management,
- hyperlink to a document, where risks and measures are described in detail.

Risk indicators are coloured: green colour indicates low-risk-level activities, yellow colour indicates medium-risk-level activities and red colour indicates high-risk-level activities. Risk indicator colour also visually warns the project manager and team members on the risk levels of individual activities and on the expected preventive and corrective measures.

For a comparison of individual project risk with other projects, the risk level of the whole project is used. On the basis of [4] we decided that the risk level of tasks (groups of activities) and of the whole project would be calculated as an average risk level of activities (the lowest WBS project level). Naturally, the average project-risk-level can be just a statistical data, so it can be misleading if used uncritically. It can happen that a project has a low average risk level, although it contains high risk level activities. If risk event occurs in these activities, it can severely threaten the implementation of the project as to the expected scope, time and costs.

In addition to the risk indicator, other indicators can be added to Table 3; they warn us on other project-risk related dangers.

4. CONCLUSION

This article presents risk management in market-oriented projects, i.e. in product- and service projects. We have found that in such cyclically recurring projects, the causes of risk in the implementation of its activities are often similar and recurring.

To the well-known risk analysis method we have thus added the third parameter – the problem incidence. This data can be estimated on the basis of already completed project evaluation. The addition of this parameter has proven necessary in practical use, being required by both the customers of project products and by project management system auditors.

If the estimated problem incidence is high and it does not get lower in future similar projects, it is obvious that the company does not effectively eliminate the recurring problems. This is important data for the company management which should urgently undertake appropriate measures. Another goal of this method is therefore to gradually reduce the estimated problem incidences (target value is 1), and to make a (gradual) transition to a two-dimensional risk analysis.

In companies, MS Project is often used for project management support, so the employees of the Faculty of Mechanical Engineering, Ljubljana, Slovenia, together with our partners in companies, made an additional table to be added to the standard template used for the risk analysis. This template has proved very useful in practice, because in this way the project managers can use the same software for planning and for risk-management actions.

5. REFERENCES

1. Kendall I. G., Rollins C. S. (2003): Advanced Project Portfolio Management and the PMO, J. Ross Publishing, Inc.
2. Fleischer M., Liker K. J. (1997): Concurrent Engineering Effectiveness: Integrating Product Development Across Organisations, Hanser Garden Publications, Cincinnati
3. PMBOK Guide (2004), A guide to the project management body of knowledge, 3rd ed., Newtown Square: Project Management Institute.
4. Royer S. Paul (2002): Project Risk management – A Proactive approach, Management Concepts, Viena, Virginia
5. Cappels M. T. (2004), Financially Focused Project Management, J. Ross Publishing, Inc.
6. Goodpasture C. John (2004), Quantitative methods in project management, J. Ross Publishing, Inc.
7. Risk management guide for DOD acquisition (2006), sixth edition, Department of defence, USA

AN APPLICATION OF THE INTERACTIVE TECHNIQUE INSDECM-II IN PRODUCTION PROCESS CONTROL

Maciej Nowak
The Karol Adamiecki University of Economics in Katowice, Department of Operations Research
ul. 1 Maja 50, 40-287 Katowice, Poland
E-mail: nomac@ae.katowice.pl

Abstract
In the paper, a job-shop production system controlled by kanban discipline is considered. The decision problem consists in deciding what scheduling rule should be used, how many kanbans should be allocated to each operation, and what lot size should be applied. Three criteria are used for evaluating performance of each alternative: makespan, average work-in-progress level, and number of set-ups. Interactive multicriteria procedure for discrete decision making problems under risk INSDECM-II is employed for generating the final solution.

Keywords: Production process control; Kanban system; Multiple criteria analysis; Interactive approach; Uncertainty modeling;

1. Introduction

Each modern production facility aims at maximizing its productivity. Various activities are usually undertaken to achieve this goal. Implementation of a scheduling system suitable for the facility is undoubtedly of primary importance, since it results in increasing facility's capacity and improving service level. In practice it is not easy to evaluate production facility's productivity, as various issues have to be considered. On one hand, minimization of completion time is recognized to be very important. On the other, however, additional objectives, including work-in-progress level, tardiness, set-up times or machine utilization, are also considered. Since these objectives are in conflict, the decision maker faces (DM) a multicriteria problem.

The main characteristics of the production process considered in this paper are as follows:

- there are M work centres in the shop,
- each centre contains K_m identical machines of a given type,
- each machine can execute various operations, one operation can be executed at a time,
- once an operation is started on a machine, it must be finished on that machine,
- a set of N orders awaits processing in the shop,
- each order is composed of a list of operations,
- each operation requires a machine of a particular type, probability distributions of operations' completion times are known,
- different orders use machines in different sequences.

This study assumes that Just-in-Time (JIT) approach is used for scheduling production system. Production orders are broken into split-lots. The work flow is controlled by kanban cards. Different kanbans represent different operations that can be performed on a station. The job can be processed if corresponding kanban is available.

The problem that arises consists in deciding which rule should be used, how many kanbans should be allocated for each operation, and what lot size should be applied. In general, smaller lot-sizes reduce work-in-progress, but also increase the number of machine set-ups. Increasing the number of allocated kanbans improves machine utilisation, but may also increase average work-in-progress level. Finally, the performance of a scheduling rule depends on the performance measure that is used. Thus, the choice of the best triplet involving the Kanban lot size, the decision rule, and the number of kanbans constitutes a multicriteria problem.

Gravel et al. (1992) considered a similar problem and used ELECTRE method (Roy, 1985) to model outranking relations. Nowak et al. (2002) proposed a modified approach for this problem. They assumed that the DM is risk-prone and in a job-shop several products are usually processed simultaneously. In this paper interactive procedure in used for solving multicriteria problem.

2. Stochastic dominance rules

The methodology used in this paper uses Stochastic Dominance rules for comparing uncertain outcomes. Two groups of stochastic dominance relations are considered. The first one includes FSD, SSD, and TSD, which means first, second, and third degree stochastic dominance respectively. These rules can be applied for modeling risk averse preferences. Let $F(x)$ and $G(x)$ be cumulative distribution functions: $F(x) = \Pr(X_F \le x)$, $G(x) = \Pr(X_G \le x)$. Definitions of FSD, SSD and TSD are as follows:

$F(x) \succ_{\text{FSD}} G(x)$ if and only if $F(x) \ne G(x)$ and $H_1(x) = F(x) - G(x) \le 0$ for all $x \in [a,b]$

$F(x) \succ_{\text{SSD}} G(x)$ if and only if $F(x) \ne G(x)$ and $H_2(x) = \int_a^x H_1(y)dy \le 0$ for all $x \in [a,b]$

$F(x) \succ_{\text{TSD}} G(x)$ if and only if $F(x) \ne G(x)$ and $H_3(x) = \int_a^x H_2(y)dy \le 0$ for all $x \in [a,b]$

The second group of SD rules includes FSD and three types of inverse stochastic dominance: SISD, TISD1, TISD2 – second degree inverse stochastic dominance and third degree inverse stochastic dominance of the first and the second type. These rules can be applied for modeling risk seeking preferences. Let $\overline{F}(x)$ and $\overline{G}(x)$ be decumulative distribution functions defined as follows: $\overline{F}(x) = \Pr(X_F \ge x)$, $\overline{G}(x) = \Pr(X_G \ge x)$. Definitions of SISD, TISD1 and TISD2 are as follows:

$\overline{F}(x) \succ_{\text{SISD}} \overline{G}(x)$ if and only if $\overline{F}(x) \ne \overline{G}(x)$ and $\overline{H}_2(x) = \int_x^b \overline{H}_1(y)dy \ge 0$ for all $x \in [a,b]$

where: $\overline{H}_1 = \overline{F}(x) - \overline{G}(x)$

$\overline{F}(x) \succ_{\text{TISD1}} \overline{G}(x)$ if and only if $\overline{F}(x) \ne \overline{G}(x)$ and $\overline{H}_3(x) = \int_x^b \overline{H}_2(y)dy \ge 0$ for all $x \in [a,b]$

$\overline{F}(x) \succ_{\text{TISD2}} \overline{G}(x)$ if and only if $\overline{F}(x) \ne \overline{G}(x)$ and $\widetilde{H}_3(x) = \int_a^x \overline{H}_2(y)dy \ge 0$ for all $x \in [a,b]$

3. Interactive procedure INSDECM-II

The procedure presented in this study is a modified version of INSDECM technique proposed in Nowak (2006). It also exploits some ideas used in the approach proposed in Nowak (2004). The first procedure is based on the interactive multiple criteria goal programming approach (Spronk, 1981), the latter exploits the main ideas of the STEM technique (Benayoun et al., 1971).

INSDECM-II combines concepts that are used in multiple criteria goal programming and STEM method. In each iteration the ideal solution is generated. Next, a candidate alternative is generated. It is the one that is closest to the ideal solution according to the minimax rule. Additionally potency matrix, composed of the best and the worst values of average evaluations with respect to all criteria, is generated. The candidate alternative and potency matrix are presented. If the DM is satisfied with the proposal, the procedure ends, otherwise the DM is asked for defining restrictions on the values of distribution parameters. The consistency of such restrictions with stochastic dominance rules is analyzed. It is assumed that the restriction is not consistent with stochastic dominance rules if following conditions are simultaneously fulfilled:

– the evaluation of a_i with respect to criterion X_k does not satisfy the restriction,
– the evaluation of a_j with respect to criterion X_k satisfies the restriction,
– the evaluation of a_i with respect to X_k dominates corresponding evaluation of a_j under stochastic dominance rules.

The pair for which inconsistency takes place is presented and the DM is asked to confirm or relax the restriction. If the restriction is confirmed, the assumptions on the stochastic dominance rules that should be fulfilled are updated.

Let us assume the following notation:

K_1 – the set of indices of criteria, that are defined in such a way, that the larger values are preferred to smaller ones,

K_2 – the set of indices of criteria, that are defined in such a way, that the smaller values are preferred to larger ones,

A^l – set of alternatives considered in iteration l,

I^l – set of indexes i, such that $a_i \in A^l$,

μ_{ik} – average evaluation of alternative a_i in relation to attribute k,

$\mathbf{P}_1^l$ – potency matrix: $\mathbf{P}_1^l = \begin{bmatrix} \underline{\mu}_1^l \cdots & \underline{\mu}_k^l & \cdots & \underline{\mu}_m^l \\ \overline{\mu}_1^l \cdots & \overline{\mu}_k^l & \cdots & \overline{\mu}_m^l \end{bmatrix}$

where: $\overline{\mu}_k^l = \begin{cases} \max_{i\in \mathbf{I}^l}\{\mu_{ik}\} & \text{for } k \in \mathbf{K}_1 \\ \min_{i\in \mathbf{I}^l}\{\mu_{ik}\} & \text{for } k \in \mathbf{K}_2 \end{cases}$ $\underline{\mu}_k^l = \begin{cases} \min_{i\in \mathbf{I}^l}\{\mu_{ik}\} & \text{for } k \in \mathbf{K}_1 \\ \max_{i\in \mathbf{I}^l}\{\mu_{ik}\} & \text{for } k \in \mathbf{K}_2 \end{cases}$

Q – number of distribution parameters chosen by the DM for presentation in conversational phase of the procedure,

Q_1 – the set of indices of parameters, that are defined in such a way, that the larger values are preferred to smaller ones,

Q_2 – the set of indices of parameters, that are defined in such a way, that the smaller values are preferred to larger ones,

v_{ip} – value of p-th parameter for alternative a_i, $i = 1, \ldots, I^l$, $p = 1, \ldots, Q$,

$\mathbf{P}_2^l$ – additional potency matrix for attribute k in iteration l: $\mathbf{P}_2^l = \begin{bmatrix} \underline{v}_1^l \cdots & \underline{v}_q^l & \cdots & \underline{v}_Q^l \\ \overline{v}_1^l \cdots & \overline{v}_q^l & \cdots & \overline{v}_Q^l \end{bmatrix}$

where: $\overline{v}_k^l = \begin{cases} \max_{i\in \mathbf{I}^l}\{v_{iq}\} & \text{for } q \in \mathbf{Q}_1 \\ \min_{i\in \mathbf{I}^l}\{\mu_{iq}\} & \text{for } q \in \mathbf{Q}_2 \end{cases}$ $\underline{v}_k^l = \begin{cases} \min_{i\in \mathbf{I}^l}\{v_{iq}\} & \text{for } q \in \mathbf{Q}_1 \\ \max_{i\in \mathbf{I}^l}\{v_{iq}\} & \text{for } q \in \mathbf{Q}_2 \end{cases}$

$\eta_k^{\mathrm{FSD}}, \eta_k^{\mathrm{SSD}}, \eta_k^{\mathrm{TSD}}, \eta_k^{\mathrm{SISD}}, \eta_k^{\mathrm{TISD1}}, \eta_k^{\mathrm{TISD2}}$ – binary variables describing whether FSD, SSD, TSD, SISD, TISD1, TISD2 rule should be considered when comparing distributional evaluations of alternatives with respect to criterion X_k.

In INSDECM-II Generalized Stochastic Dominance (GSD) relation is used. This relation is defined as follows:

$$X_{jk} \succ_{\text{GSD}} X_{ik} \Leftrightarrow \left(X_{jk} \succ_{\text{FSD}} X_{ik} \wedge \eta_k^{\text{FSD}} = 1\right) \vee \left(X_{jk} \succ_{\text{SSD}} X_{ik} \wedge \eta_k^{\text{SSD}} = 1\right) \vee$$
$$\left(X_{jk} \succ_{\text{TSD}} X_{ik} \wedge \eta_k^{\text{TSD}} = 1\right) \vee \left(X_{jk} \succ_{\text{SISD}} X_{ik} \wedge \eta_k^{\text{SISD}} = 1\right) \vee$$
$$\left(X_{jk} \succ_{\text{TISD1}} X_{ik} \wedge \eta_k^{\text{TISD1}} = 1\right) \vee \left(X_{jk} \succ_{\text{TISD2}} X_{ik} \wedge \eta_k^{\text{TISD2}} = 1\right)$$

The operation of the procedure is as follows:

Initial phase:

1. Calculate average evaluations of alternatives with respect to attributes $\mu_{i\,k,}$ $i = 1, ..., n$, $k = 1, ..., m$.
2. Set: $\eta_k^{\text{FSD}} = 1, \eta_k^{\text{SSD}} = 1, \eta_k^{\text{TSD}} = 1, \eta_k^{\text{SISD}} = 1, \eta_k^{\text{TISD1}} = 1, \eta_k^{\text{TISD2}} = 1$, $l = 1$, $A^l = A$.

Iteration l:

3. Identify candidate alternative: $a_i := \arg \min_{j \in \mathbf{I}^l} \left\{ d^{\,l}_{jk} \right\}$, where $d^{\,l}_{jk}$ is calculated as follows:

   $$d^{\,l}_{jk} = \max_{k=1,\ldots m} \left\{ w_k^l \left| \overline{\mu}_k^{\,l} - \mu_{jk} \right| \right\} \qquad w_k^l = \frac{1}{r_k^l} \left[\sum_{i=1}^{m} \frac{1}{r_i^l} \right]^{-1} \qquad r_k^l = \left| \overline{\mu}_k^{\,l} - \underline{\mu}_k^{\,l} \right|$$

   In the case of a tie choose any a_i minimising the value of $d^{\,l}_{jk}$.
4. Present the data to the DM: average evaluations of the candidate alternative a_i with respect to attributes μ_{ik}, $k = 1, ..., m$, potency matrix $\mathbf{P}_1^l$.
5. Ask the DM whether he/she is satisfied with the data that are presented. If the answer is YES – go to 7.
6. Ask the decision maker to specify parameters of distributional evaluations to be presented; calculate distribution parameters v_{ip} for i such that $a_i \in A^l$, $p = 1, \ldots, Q$; calculate additional potency matrix $\mathbf{P}_2^l$; present additional potency matrix to the DM.
7. Ask the DM whether he/she is satisfied with the candidate alternative. If the answer is YES – the final solution is alternative a_i – go to 17, else – go to 8.
8. Ask the DM to specify an additional restriction.
9. Generate A^{l+1} the set of alternatives satisfying the restriction specified by the DM.
10. Calculate potency matrices $\mathbf{P}_1^{l+1}$ and $\mathbf{P}_2^{l+1}$; present matrices $\mathbf{P}_1^l$, $\mathbf{P}_2^l$, $\mathbf{P}_1^{l+1}$ and $\mathbf{P}_2^{l+1}$ to the DM; ask the DM whether he/she accepts the move from $\mathbf{P}_1^l$ and $\mathbf{P}_2^l$ to $\mathbf{P}_1^{l+1}$ and $\mathbf{P}_2^{l+1}$. If the answer is NO, then go to 4, else go to 11.
11. For each pair (a_j, a_i) such that $a_j \in A^l \setminus A^{l+1}$ and $a_i \in A^{l+1}$ identify GSD relation between $X_{j\,k}$ and $X_{i\,k}$. Generate the set of inconsistencies:

    $$\mathbf{N}^l = \left\{ \left(a_j, a_i\right), a_j \in \mathbf{A}^l \setminus \mathbf{A}^{l+1}, a_i \in \mathbf{A}^{l+1}, X_{jk} \succ_{\text{GSD}} X_{ik} \right\}$$
12. If $N^l = \varnothing$, then assume $l = l + 1$; go to 3, else go to 13.
13. Choose the first pair $\left(a_j, a_i\right) \in \mathbf{N}^l$; calculate: $\Pr(X_{ik} \le s_r)$, $\Pr(X_{jk} \le s_r)$, where:

    $$s_r = \min\left(\alpha_i, \alpha_j\right) + r \frac{\max\left(\beta_i, \beta_j\right) - \min\left(\alpha_i, \alpha_j\right)}{R} \quad \text{for } r = 0, 1, \ldots, R$$

    α_i, β_i – lower and upper bound for evaluations of X_{ik}

    α_j, β_j – lower and upper bound for evaluations of X_{jk}

    R – number of observations. Initially R can be set to 10, the DM can increase (decrease) the value of R if he/she finds the data to be not enough detailed (too detailed).

    Present the data to the DM pointing that a_j is to be rejected, while a_i is to be accepted. Ask the DM what is his/her decision – propose the decision maker:

    (a) accept a_i and reject a_j,

    (b) accept both a_j and a_i,

(c) reject both a_j and a_i.

If the DM's decision is (a), go to 14, if the decision is (b), go to 15, otherwise go to 16.

14. Update assumptions on DM's utility function:

$X_{ik} \succ_{\text{TSD}} X_{jk} \Rightarrow \eta_k^{\text{TSD}} = 0$ $\quad$ $X_{ik} \succ_{\text{SSD}} X_{jk} \Rightarrow \eta_k^{\text{SSD}} = 0, \eta_k^{\text{TSD}} = 0$

$X_{ik} \succ_{\text{TISD1}} X_{jk} \Rightarrow \eta_k^{\text{TISD1}} = 0$ $\quad$ $X_{ik} \succ_{\text{TISD2}} X_{jk} \Rightarrow \eta_k^{\text{TISD2}} = 0$

$X_{ik} \succ_{\text{SISD}} X_{jk} \Rightarrow \eta_k^{\text{SISD}} = 0, \eta_k^{\text{TISD1}} = 0, \eta_k^{\text{TISD2}} = 0$

go to 11.

15. Set: $\mathbf{A}^{l+1} = \mathbf{A}^{l+1} \cup \{a_j\}, \mathbf{N}^l = \mathbf{N}^l \setminus \{(a_j, a_i)\}$; go to 12.
16. Set: $\mathbf{A}^{l+1} = \mathbf{A}^{l+1} \setminus \{a_i\}, \mathbf{N}^l = \mathbf{N}^l \setminus \{(a_j, a_i)\}$; go to 12.
17. End of the procedure.

4. Illustrative example

To illustrate the procedure let us consider a shop with six machine centers. Four scheduling rules are considered: the first come – first served (FCFS) rule, the shortest processing time (SPT) rule, the same job as previously (SJP) rule, the shortest next queue (SNQ) rule. Four values of lot size are considered: 5, 10, 15, and 20, while the number of kanbans is assumed to be between 2 and 5. Thus, 64 triplets of parameters are considered. First, an exemplary production plan is analyzed. Series of simulation experiments are done for each alternate parameter triplets. Next, distributional evaluations with respect to three criteria are constructed. The final solution is generated as follows:

1. Calculation of average evaluations of alternatives with respect to attributes.
2. $\eta_k^{\text{FSD}} = 1, \eta_k^{\text{SSD}} = 1, \eta_k^{\text{TSD}} = 1, \eta_k^{\text{SISD}} = 1, \eta_k^{\text{TISD1}} = 1, \eta_k^{\text{TISD2}} = 1$, $l = 1$, $A^1 = A$.

Iteration 1:

3. Candidate alternative is identified: a_{25}
4. Presentation of the data to the DM: average evaluations of the candidate alternative: $\mu_{25\,1} = 275995$, $\mu_{25\,2} = 1409$, $\mu_{25\,3} = 2758$, potency matrix $\mathbf{P}_1^1$.

Potency matrix $\mathbf{P}_1^1$

	X_1	X_2	X_3
$\underline{\mu}_k^1$	323009	4552	4680
$\overline{\mu}_k^1$	261996	470	1513

5. The DM is satisfied with the data presented.
7. The DM is not satisfied with the candidate alternative.
8. The DM specifies additional restriction: $\Pr(X_{i\,1} \le 277250) \ge 0{,}98$
9. Set of alternatives satisfying the restriction specified by the DM is generated: $A^2 = \{a_1, a_2, a_3, a_5, a_6, a_7, a_8, a_9, a_{13}, a_{14}, a_{15}, a_{16}, a_{21}, a_{22}, a_{23}, a_{29}, a_{30}, a_{37}, a_{45}\}$
10. Potency matrix $\mathbf{P}_1^2$ is generated; matrices $\mathbf{P}_1^1$ and $\mathbf{P}_1^2$ are presented to the DM; the DM accepts the move from $\mathbf{P}_1^1$ to $\mathbf{P}_1^2$.

Potency matrix $\mathbf{P}_1^2$

	X_1	X_2	X_3
$\underline{\mu}_k^2$	274465	4552	3592
$\overline{\mu}_k^2$	261996	1407	1513

11. The set of inconsistencies is generated:

$N^1 = \{(a_{10}, a_3), (a_{10}, a_7), (a_{10}, a_8), (a_{10}, a_{21}), (a_{10}, a_{22}), (a_{10}, a_{23}), (a_{10}, a_{30}), (a_{10}, a_{37}), (a_{10}, a_{45}), (a_{11}, a_3), (a_{11}, a_{23}), (a_{11}, a_{45})\}$

For example for pair (a_{10}, a_3) following relations are identified:

$X_{10\,1} \succ_{\mathrm{SISD}} X_{3\,1}$ $\Pr(X_{10\,1} \le 277250) = 0{,}96$ $\Pr(X_{3\,1} \le 277250) = 1{,}00$

12. $N^1 \neq \varnothing$.
13. The pair $(a_{10}, a_3) \in \mathbf{N}^1$. The data are presented to the DM. The decision maker confirms the decision to accept a_3 and reject a_{10}.
14. Assumptions on DM's utility function are revised:

$$X_{10\,1} \succ_{\mathrm{SISD}} X_{3\,1} \Rightarrow \eta_1^{\mathrm{SISD}} = 0,\ \eta_1^{\mathrm{TISD1}} = 0,\ \eta_1^{\mathrm{TISD2}} = 0$$

11. The set of inconsistencies is generated.
12. $N^1 = \varnothing$, so $l = 2$.

The procedure continues until the decision maker accepts the candidate alternative.

5. Conclusions

Various objectives are taken into account when a scheduling problem is considered. Minimizing makespan, optimizing the use of machines, minimizing work is progress and minimizing the number of set-ups are usually considered to be important. As these criteria are in conflict, so a problem has a multicriteria nature.

The main purpose of this paper was to present comprehensive, yet simple methodology for decision problems in production process control. A new methodology for selecting values of parameters influencing the performance of a production facility was presented. Although this approach was applied in a job-shop environment, it could be easily adapted to other production systems.

The procedure uses two approaches: stochastic dominance and interactive methodology. The first is widely used for comparing uncertain prospects, the latter is a multiple criteria technique that is probably most often used in real-world applications. These two concepts has been combined in INSDECM-II procedure.

Acknowledgements

This research was supported by State Committee for Scientific Research (KBN) grant no 1 H02B 031 29.

References:

Benayoun, R., de Montgolfier, J., Tergny, J. and Larichev, C., 1971. Linear Programming with Multiple Objective Functions: Step Method (STEM). Mathematical Programming, 8, 366-375.

Gravel, M., Martel, J.M., Nadeau, R., Price, W. and Tremblay, R. (1992). A multicriterion view of optimal resource allocation in job-shop production. European Journal of Operational Research, 61, 230-244.

Nowak, M., 2004. Interactive approach in multicriteria analysis based on stochastic dominance. Control and Cybernetics, 33, 463-476.

Nowak, M., 2006. INSDECM – an interactive procedure for stochastic multicriteria decision problems. European Journal of Operational Research, 175, 1413-1430.

Nowak, M., Trzaskalik, T., Trzpiot, G. and Zaras, K., 2002. Inverse stochastic dominance and its application in production process control. In: Trzaskalik, T., Michnik, J. (Eds.), Multiple Objective and Goal Programming. Recent Developments. Physica-Verlag, Heidelberg, 362-376.

Spronk, J., 1981. Interactive Multiple Goal Programming. Martinus Nijhoff, The Hague.

MODIFICATION OF PRODUCTION-INVENTORY CONTROL MODEL WITH QUADRATIC AND LINEAR COSTS

Mirjana Rakamarić Šegić [1],Marija Marinović[2]
and Marko Potokar [3]
[1]Politechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia
email: mrakams@veleri.hr
Faculty of Arts and Sciences, Omladinska 14, Rijeka, Croatia
Tel.: 051/345-034 E-mail: marinm@ffri.hr
[3] Bankart d.o.o., Celovška 150, 1000 Ljubljana, Slovenia,
email: marko.potokar@bankart.si

Abstract: The objective of this paper is to modify the production-inventory model with quadratic and linear costs developed in /15/, for the case of infinite planning horizon taking into account discounting. We introduce constraints on the control variable for the case with constant positive demand. Finally, we perform analyses as to how the solution depends on the initial conditions for inventory and illustrate it with several examples.

Keywords: Production-inventory, optimal control.

1. Introduction

In the former paper /15/ we developed a production-inventory model for a firm that considers two types of costs: costs of producing and keeping the unit of product (linear costs) and extra costs resulting from the deviations of production and inventory levels from the desired levels (quadratic costs).

The idea emerged from the production-inventory control model named HMMS described in /8/ using calculus of variation techniques. In the following years, this model inspired many control theory formulations of the production planning problem, starting with Hwang, Fan and Ericson /9/, who introduced the maximum principle in their model. The advantage of the optimal control theory formulation of HMMS model lies in the simple implementation of constraints on the production rate. Another advantage is a simpler extension to the multi-item production, which Bergstrom and Smith /5/ implemented in 1970.

In 1972, Bensoussan /4/ presented the generalized optimal control theory formulation of continuous type in which he tried to encapsulate several types of HMMS models.

Similar model type, also implementing the methodology of the optimal control theory, was presented in detail by Thompson and Sethy /11/ in their book dated 1999.

All HMMS type models' goal functionals minimize only costs of inventory level and production rate deviations from the respective target values. We modified this in such a way that along with these types of costs, we also introduced linear costs for producing a unit of product and for keeping it in inventory stock and the discount rate. We did so because we consider that the costs of producing a unit of product or keeping it in the inventory stock is basically very different from the costs of production or inventory deviation from the desired level since they result from different causes.

We applied the control theory, particularly Pontriagyn`s maximum principle, to find the optimal paths for inventory and production decisions of a firm but we only considered the finite and not very long time horizon, so there was no need for discounting (the discount rate was assumed to be zero).

The objective in this paper is to adjust that model by taking into account the constant positive discount rate and we find the optimal paths for inventory and production decisions for infinite planning horizon. Than we make a special case for the constant positive demand and after that we add the constraint on the control variable (production), which gives different optimal decision rule for it. In the last part we perform analysis to show how the behavior of the model and optimal paths of both production and inventory depend on the level of initial inventory and we demonstrate it in several examples.

2. The model

A firm is producing some homogenous good and has a warehouse for inventory. The following data are needed to define the model:

P(t) - production rate at time t (control variable)

I(t) - inventory level at time t (state variable)

ρ - constant, nonnegative discount rate

$\hat{P}$ - constant, nonnegative, desired level of production

$\hat{I}$ - constant, nonnegative, desired level of inventory

a - constant, positive, extra inventory holding costs coefficient, resulting from the inventory deviation from the desired level (for example opportunity costs if inventory is higher than desired, or costs for maintaining empty warehouse space and still not having enough inventory to fulfill the order from buyers or the production needs)

b - constant, positive, extra production costs coefficient, resulting from the production deviation from the desired level (for example paying underemployed manpower due to lower production, or paying the overtime work, which is normally more expensive than the regular man-hours)

h - constant, positive, linear inventory holding cost coefficient for keeping unit of inventory

p - constant, positive, linear production cost coefficient for unit of product

S(t) - exogenous demand rate at time t, positive and continuously differentiable

T - length of planning period

I_0 - initial inventory level

Change of inventory level follows the usual stock-flow differential equation:

$$\dot{I}(t) = P(t) - S(t) \tag{2.1}$$

And the initial condition is $I(0) = I_0$

In the first paper, we minimized costs, which was expressed by the objective function of the model

$$\max\ J = -\int_0^T e^{-\rho t}[a(I-\hat{I})^2 + hI + b(P-\hat{P})^2 + pP]dt \tag{2.2}$$

We assumed that $\hat{P}$ is large enough and I_0 is small enough so that P will not become zero and we did not impose any constraints on P and on I.

We used a current value Hamiltonian

$$H = -a(I-\hat{I})^2 - hI - b(P-\hat{P})^2 - pP + \lambda(P-S)$$

concave in P, and applying maximum principle we deduced the decision rule for the optimal path of the control variable as

$$P^* = \hat{P} + \frac{1}{2b}(\lambda - p) \tag{2.3}$$

and we created TPBVP, which in the matrix form was given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \dot{I} \\ \dot{\lambda} \end{bmatrix} + \begin{bmatrix} 0 & \frac{-1}{2b} \\ -2a & -\rho \end{bmatrix}\begin{bmatrix} I \\ \lambda \end{bmatrix} = \begin{bmatrix} \hat{P} - \frac{p}{2b} - S \\ h - 2a\hat{I} \end{bmatrix} \tag{2.4}$$

Solving it as a simultaneous system of the first order differential equations, we obtained the expressions for the optimal paths of the state variable I, control variable P and adjoint variable λ

$$I^* = A_1 e^{r_1 t} + A_2 e^{r_2 t} + D(t)$$
$$P^* = r_1 A_1 e^{r_1 t} + r_2 A_2 e^{r_2 t} + S(t) + \dot{D}(t) \qquad (2.5)$$
$$\lambda^* = 2b(r_1 A_1 e^{r_1 t} + r_2 A_2 e^{r_2 t} - \hat{P} + S(t) + \dot{D}(t)) + p$$

where the constant A_1 and A_2 are

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} \dfrac{d_1 r_2 e^{r_2 T} - d_2}{r_2 e^{r_2 T} - r_1 e^{r_1 T}} \\ \dfrac{d_2 - d_1 r_1 e^{r_1 T}}{r_2 e^{r_2 T} - r_1 e^{r_1 T}} \end{bmatrix} \qquad (2.6)$$

D(t) is the name of the function representing particular solution for inventory ($D(t) = \bar{I}(t)$), from which we deduced particular solutions for adjoint variable too, as

$$\bar{\lambda} = p + 2b(S(t) + \dot{D}(t) - \hat{P}) \qquad (2.7)$$

the constants d_1 and d_2 are

$$d_1 = I_0 - D(0)$$
$$d_2 = \hat{P} - S(T) - \dot{D}(T) - \frac{p}{2b} \qquad (2.8)$$

and characteristic roots r_1 and r_2 are

$$r_{2,1} = \frac{\rho \pm \sqrt{\rho^2 + \frac{4a}{b}}}{2} \qquad (2.9)$$

2.1 Adjustment of the model for the case of infinite planning horizon

When the time horizon is long or even infinite, it is important to discount because otherwise all the solutions would be unbounded and would give an infinite values (which, in our case, is the worst, since we consider costs and aim to minimize them). So the continuous discount rate ρ in this case is assumed positive (ρ>0). Now we solve the model for the situation where T→∞. Since $r_1<0$, $r_2>0$, which also implies $r_1 - r_2 < 0$, it follows that dividing both numerator and denominator in the following expressions for constants A_1 and A_2 with continuous function $e^{r_2 T}$, gives:

$$\lim_{T\to\infty} A_1 = \lim_{T\to\infty} \frac{d_1 r_2 e^{r_2 T} - d_2}{r_2 e^{r_2 T} - r_1 e^{r_1 T}} = d_1 \qquad (2.1.8)$$

and

$$\lim_{T\to\infty} A_2 = \lim_{T\to\infty} \frac{d_2 - d_1 r_1 e^{r_1 T}}{r_2 e^{r_2 T} - r_1 e^{r_1 T}} = 0 \qquad (2.1.9)$$

So, in this case from (2.5), (2.1.8) and (2.1.9) the optimal paths become:

$$I^*(t) = d_1 e^{r_1 t} + D(t)$$
$$P^*(t) = r_1 d_1 e^{r_1 t} + S(t) + \dot{D}(t) \qquad (2.1.10)$$
$$\lambda^*(t) = 2b\left[r_1 d_1 e^{r_1 t} - \hat{P} + S(t) + \dot{D}(t) \right] + p$$

Since $r_1<0$ also implies that $d_1 e^{r_1 t}$ converges to zero when t tends to infinity, it can be easily seen that $I^*(t)$ converges to its particular solution D(t), which is actually an intermediate equilibrium level. It means that optimal "time path" converges and fulfills condition for dynamic stability of equilibrium.

2.2 Specialization of the model for the case of constant positive demand

For a constant S, the particular solution for $I^*(t)$ and the particular solution for λ, given by (2.7) becomes constant given by: $\bar{I}(t) = D$ (2.2.1)

$$\bar{\lambda} = p + 2b(S - \hat{P}) \qquad (2.2.2)$$

so $\dot{\bar{I}} = 0$ and $\dot{\bar{\lambda}} = 0$. When these are introduced into the system of differential equations (2.4), it changes into the following matrix equation:

$$\begin{bmatrix} 0 & -\frac{1}{2b} \\ -2a & -\rho \end{bmatrix} \begin{bmatrix} \bar{I} \\ \bar{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{P} - \frac{p}{2b} - S \\ h - 2a\hat{I} \end{bmatrix}$$

The determinant of the matrix on the left side is det=-a/b and the solutions for $\bar{I}$ and $\bar{\lambda}$ are obtained as:

$$\begin{bmatrix} \bar{I} \\ \bar{\lambda} \end{bmatrix} = \begin{bmatrix} \hat{I} + \frac{\rho b}{a}(\hat{P} - S) - \frac{\rho p + h}{2a} \\ 2b(S - \hat{P}) + p \end{bmatrix} \qquad (2.2.3)$$

So

$$D = \hat{I} + \frac{\rho b}{a}(\hat{P} - S) - \frac{\rho p + h}{2a} \qquad (2.2.4)$$

From the definition of constants d_1, d_2 in (2.8), since D is constant and $\dot{D} = 0$, they became:

$$d_1 = I_0 - D = I_0 - \hat{I} - \frac{\rho b}{a}(\hat{P} - S) + \frac{\rho p + h}{2a}$$

$$d_2 = \hat{P} - S - \frac{p}{2b} \qquad (2.2.5)$$

The optimal paths from (2.1.10) and with constant S (and constant D) are

$$\begin{aligned} I^* &= d_1 e^{r_1 t} + D \\ P^* &= r_1 d_1 e^{r_1 t} + S \\ \lambda^* &= 2b(r_1 d_1 e^{r_1 t} - \hat{P} + S) + p \end{aligned} \qquad (2.2.6)$$

2.3 Extension of the previous model by introducing constraint on the control variable

Until now, we assumed that $\hat{P}$ was large enough and I_0 small enough so that P will never become zero. It means that we included the interior solution implicitly, which hypothesis inadequately reflects the reality.

Now, we shall consider the case where there is constraint requiring the control variable P it to be nonnegative. ($P(t) \geq 0$). Again, we will assume that the demand S is a positive constant and the continuous discount rate ρ is positive. Since the solution now can be boundary, a different optimal decision rule for production, given by following equation, will be used

$$P^* = \max\{r_1 d_1 e^{r_1 t} + S, 0\} \qquad (2.3.1)$$

The first possibility is for P interior and the second is for P on its boundary. In the first possibility, as we have shown before, the optimal paths for interior solution for all three variables are given with (2.2.6).

2.4 Analysis of the solutions depending on the initial condition for inventory and examples

In this chapter, we shall perform analysis as to how the behavior of the model and optimal paths of both production and inventory depend on the initial inventory level.

Case 1

If I_0=D (note $\bar{I}(t) = D$) it follows from (2.2.5) that d_1=0 and from (2.2.6) that P^*=S, which is positive, so the solution is interior and $I^* = D$

From (2.3) follows
$$\bar{P} = \hat{P} + \frac{1}{2b}(\bar{\lambda} - p) \quad (2.4.1)$$

and from (2.7)
$$\bar{\lambda} = p + 2b(S - \hat{P}) \quad (2.4.2)$$

It can be deduced from these equations that $\bar{P} = S$. It means that in this case, since $I_0 = \bar{I}$ (or D), the optimal production path is $P^* = \bar{P}$ for every t.

The conclusion is:

If the initial inventory equals the particular solution for inventory, then the solution of the optimal production equals to its particular solution and they both equal to demand.

$$P^* = \bar{P} = S \quad (2.4.3)$$

It is interesting to notice that in this situation the optimal path for production depends only on demand and not on the parameters of the model.

Of course, if any parameter in the model changes, then the particular solution D for the inventory changes as well, and if the manager wants to keep production equal to demand, he must change initial inventory by setting it to the value of the particular solution D.

Case 2

For $I_0 \neq D$, from (2.2.5), (2.2.6), (2.3.1) and (2.3) the optimal solution is given by

$$P^*(t) = \max\{r_1(I_0 - D)e^{r_1 t} + S, 0\} = \max\{\hat{P} + \frac{1}{2b}(\lambda - p), 0\} \quad (2.4.4)$$

Case 2.1

For $I_0 \leq D$ (since $r_1<0$ and $(I_0-D)\leq 0$) it follows that the optimal production is always nonnegative, meaning that the solution is interior given with (2.2.6).

1.) Example (see Figure 1)

$\hat{P}$ =25 $\hat{I}$=15 ρ=0,5 S=20 a=0,5 b=0,5 h=8 p=10 ⇒ D=4,5 r_1= -0,780776 I_0=3<D=4,5

This example illustrates how the paths are moving when the initial inventory level is lower than the value of the particular optimal solution for inventory. At the beginning, its level is below the particular inventory line and consequently the production path is higher than demand. But as the production path approaches the demand path, the inventory level tends to its particular solution (meaning that it has the property of dynamic stability).

Case 2.2

For I_0>D, since $r_1(I_0$-D) is negative, $e^{r_1 t}$ is decreasing, S is assumed to be constant, it follows from (2.4.4) that P^* is increasing. So if the value for the zero moment P(0) is positive, the optimal production solution $P^*(t)$ will be positive at all times. We shall now find the initial conditions for which this is true.

The value of the initial production is P(0)=$r_1(I_0$-D)+S and if it must be positive:

$$r_1(I_0-D)+S>0 \quad /:r_1 \quad (r_1<0)$$

it follows that if the inventory level I_0 is lower then the value of $D - \frac{S}{r_1}$,$(I_0 < D - \frac{S}{r_1})$ the value of P(0) is positive and consequently $P^*(t)$ is positive and interior.

2) Example (see Figure 2)

All parameters remain the same except I_0=10, $\Rightarrow D-\frac{S}{r_1}=30{,}115528$, (I_0<30,115528; I_0< $\hat{I}$ =15)

Figure 1

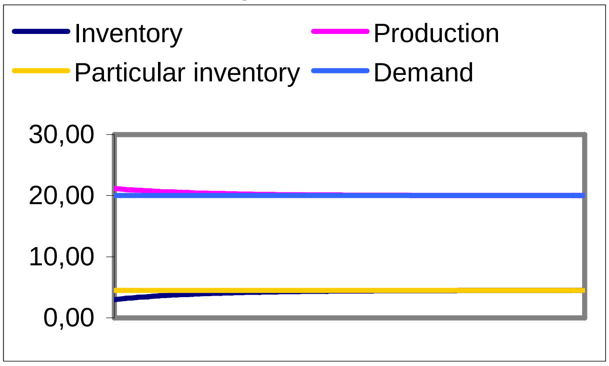


Figure 2

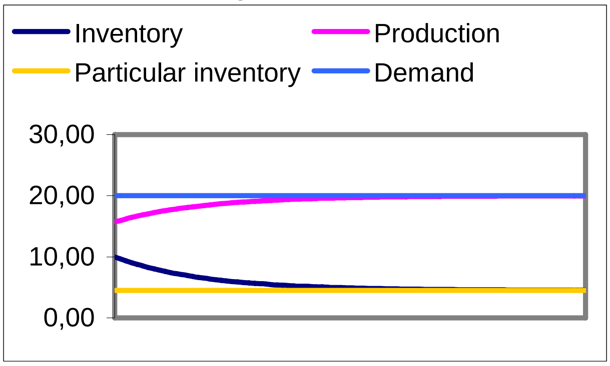


Figure 3

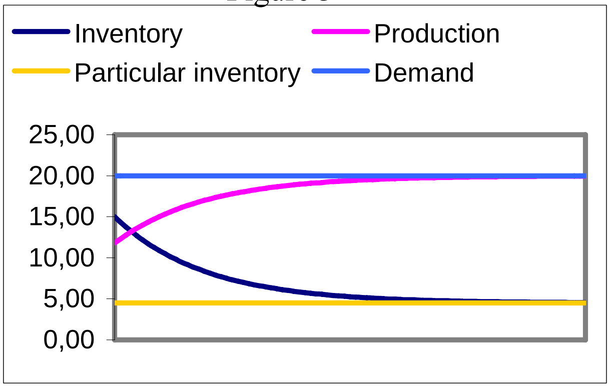
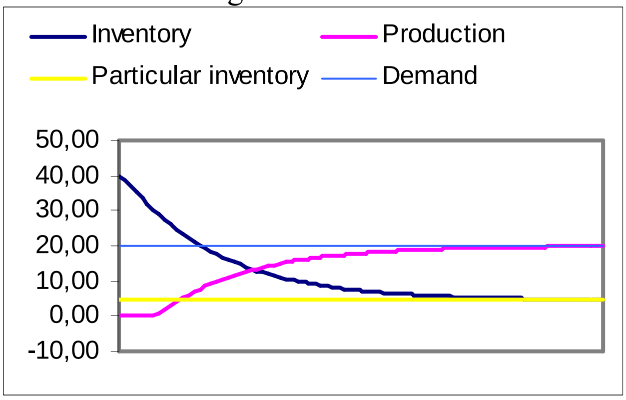


Figure 4



3) Example (see Figure 3)

D=4,5<I_0=15<30,115528 (I_0= $\hat{I}$) and the other parameters of the model remain the same

Examples 2 and 3 show the cases where the initial inventory value is higher than the value of the particular optimal solution for inventory. In such situations, the production is lower in the beginning of planning horizon because there is no need for such a large inventory. During the time, though, the production slightly increases in order to approach the demand and consequently the inventory level decreases and tends to its particular solution.

In examples analyzed till now, it can be noticed that they had an interior solution for production (control variable), while as T tends to infinity the optimal paths for production and inventory converge to their particular optimal solutions respectively.

Case 2.3
$$I_0 > D-\frac{S}{r_1} \tag{2.4.5}$$

When (2.4.5) is valid, the value of P(0) would be negative and the optimal production P^* given in (2.4.4) is zero until the moment t_1, where

$$P(t_1)=r_1(I_0-D)e^{r_1t_1}+S=0$$

which implies
$$e^{r_1t_1}=\frac{S}{r_1(D-I_0)} \tag{2.4.6}$$

What is the value of the moment t_1? From (2.2.6), (2.2.5) and (2.4.6) it can be deduced that the optimal inventory in the moment t_1 would be

$$I^*(t_1)=(I_0-D)e^{r_1t_1}+D=(I_0-D)\frac{S}{r_1(D-I_0)}+D$$
$$I^*(t_1)=D-\frac{S}{r_1} \tag{2.4.7}$$

Also, for $t \le t_1$ (since P^* =0) the equation of motion for inventory is different and it is expressed as follows: $\dot{I} = -S$

Its solution gives: $I(t)=I_0-St$ (2.4.8)

The expression (2.4.8) means that with no production, the inventory is decreasing from the initial inventory I_0, as the demand consumes it over time.

Since it is valid for $t \le t_1$, the inventory for the moment t_1 is given by

$$I(t_1)=I_0-St_1 \quad (2.4.9)$$

Equating (2.4.9) with (2.4.7) gives:

$$t_1 = \frac{I_0 - D}{S} + \frac{1}{r_1} \quad (2.4.10)$$

It can be proved that t_1 is positive because this situation exists only under the condition (2.4.5) assumed in this case

Proof: From (2.4.5) it follows: $I_0 - D + \frac{S}{r_1} > 0$ /:S

$$\frac{I_0 - D}{S} + \frac{1}{r_1} > 0$$

$$t_1 > 0$$

Until the moment t_1, the optimal inventory is given with the expression (2.4.9). From that moment onward, the problem can be considered as a new one, beginning in the moment t_1 and it has the new initial inventory given by

$$I^*(t_1) = D - \frac{S}{r_1} \quad (2.4.11)$$

From that moment onward, since the initial inventory satisfies the condition (2.4.5), the solution will be interior. It is important to notice that, because the initial moment for the second part of the problem is no longer zero but t_1, the time translation $t-t_1$ must be introduced. Finally, it gives the expression for the optimal inventory in this part of the problem as follows: $I^*{}' = (I^*{}'(0) - D)e^{r_1(t-t_1)} + D = (I^*(t_1) - D)e^{r_1(t-t_1)} + D$

When the (2.4.11) is introduced it gives

$$I^*{}' = -\frac{S}{r_1}e^{r_1(t-t_1)} + D \quad (2.4.12)$$

The optimal path for production can be deduced in a similar way:

$$P^*{}' = r_1(I'(0) - D)e^{r_1(t-t_1)} + S = r_1(I^*(t_1) - D)e^{r_1(t-t_1)} + S = r_1(-\frac{S}{r_1})e^{r_1(t-t_1)} + S$$

$$P^*{}' = S[1 - e^{r_1(t-t_1)}] \quad (2.4.13)$$

Finally, both optimal paths for inventory and production are given by following equations:

$$I^* = \begin{cases} I_o - St & 0 \le t \le \frac{I_0 - D}{S} + \frac{1}{r_1} \\ -\frac{S}{r_1}e^{r_1(t-t_1)} + D & t > \frac{I_0 - D}{S} + \frac{1}{r_1} \end{cases}$$

$$P^* = \begin{cases} 0 & 0 \le t \le \frac{I_0 - D}{S} + \frac{1}{r_1} \\ S[1 - e^{r_1(t-t_1)}] & t > \frac{I_0 - D}{S} + \frac{1}{r_1} \end{cases} \quad (2.4.14)$$

4) Example (see Figure 4)

$I_0=40>30,115528$ and the other parameters of the model remain the same

This example illustrates the situation where the initial inventory is so high that it exceeds the critical value D-S/r_1 which, as it was shown in case 2.3, causes boundary solution for optimal production. So, the optimal decision rule (2.4.4), as applied in the beginning of planning horizon, gives production equal to zero (boundary solution) and the decrease of inventory. It proceeds so until the moment t_1, given by (2.4.10), when inventory level reaches value of D-S/r_1=30,115528. From that moment onward, the optimal solution for production is again interior, and it follows the equation (2.4.14).

The purpose of outlined analysis is to show the dependence of the optimal solution of the model on the initial inventory level as to whether it has the boundary solution or not, which facilitates the management decision criteria setting.

3. Conclusion

In this paper we used the production-inventory model of a firm that produces a homogenous goods with linear and quadratic costs developed in our former paper and we extended the solution of the optimal paths for production (control variable) and inventory (state variable) to the case of infinite planning horizon. Than we introduced constraint on the control variable for the constant positive demand and solved it again using the optimal control theory. Finally we performed analyses as to how the solution depends on the initial conditions for inventory and presented a few examples.

References:

1. Axsäter S., Inventory Control, Kluwer's International Series, 2000.
2. A. C. Chiang, Elements of Dynamic Optimization,McGraw-Hill Inc. Singapore 1992.
3. J. A. Čibej, L. Bogataj, Sensitivity of quadratic cost functionals under stohastically perturbed controls in inventory systems with delays, IJPE 35 (1994) 265-270.
4. A. Bensoussan, A control theory approach to production models of the HMMS type, Working Paper 72 –19, E Institute for Advanced Studies in Management, Brussels (1972).
5. G. L. Bergstorm and B. E. Smith, Multi-item production planning: An extension of the HMMS rules, Management Science 16 (1970) B614-B629.
6. L. Bogataj, M. Bogataj, Dynamic Version of an Elementary Inventory Model, Proceedings of Second International Symp. on Inventories,Budapest,Hungary 1982.
7. L. Bogataj, Sensitivity of linear-quadratic systems with delay in the state and in control for perturbation of the system matrices, Glasnik matematički, Vol 24 (44) (1989),355-360.
8. C. C. Holt, F, Modigliani, J. F. Muth, H. A. Simon, Planning Production, Inventories, and Work Forces, Prentice-Hall, Englewood Cliffs, NJ, 1960.
9. C. L. Hwang, L. T. Fan and L. E. Ericson, Optimum production planning by the maximum principle, Management Science 13 (1967) 750-755.
10. L. S. Pontryagin, V. G. Boltianskii, R. V. Gamkrelidze, E. F. Mishchenko, The Mathematical Theory of Optimal Processes, Interscience Publishers, a division of John Wiley and sons, Inc. New York London Sydney (1965).
11. S. P. Sethi, G. L. Thompson, Optimal Control Theory, Application to Management Science and Economics, Kluwer Academic Publishers, Boston/Dordrecth/London, 1999.
12. H. M. Wagner, T. M. Whitin, Dynamic version of the economic lot size model, Management Science 5 (1958) 89-96.
13. Wallace J. Hopp, Mark L. Spearman, Factory Physiscs, sec ed, Irwin McGraw-Hill 2000.
14. W. L. Winston, Operations Research, Duxbury press, (1994.)
15. M. Rakamarić Šegić, J. Perić, L. Bogataj, Analysis of production-inventory control model with quadratic and linear costs, Proceedings of the 9th International Conference on Operational Research KOI 2002, Trogir, October, 2-4, str. 343-352.

FUNCTIONAL SEPARABILITY AND THE OPTIMAL DISTRIBUTION OF GOODS

Ilko Vrankić, Zrinka Lukač
Faculy of Economics Zagreb, Trg J. F. Kennedya 6, 10000 Zagreb, Croatia
{ivrankic,zlukac}@efzg.hr

Abstract: The Cobb-Douglas utility function plays a very important role in both consumer and production theory. It is the ordinal version of the utility function resulting from the Marshall's assumption of constant marginal utility of income. Based on the economic interpretation of this function's exponents we derive the two-phase algorithm for finding the optimal distribution of goods. The first phase consists of determining the optimal expenditure on the group of goods, while the second phase consists of determining the optimal expenditure on each good within the same group. By combining the two-phase programming and the consumer theory we develop the geometrical interpretation of the link between the price index and optimal quantity index through the original interpretation of the income expansion path.

Key words: efficient distribution of income, weakly separable utility function, income expansion path, two-phase programming

1. Introduction

The founders of subjective value theory, the well known economists Jevons, Menger and Walras, have developed their theory from the standpoint that utility is an additive and cardinally measurable quality embodied in a commodity whose consumption provides consumer with satisfaction. By Pareto's revolutionary act the cardinal consumer behavior theory has stepped aside to ordinal theory which uses weak preference relations as a way to describe consumer's taste. The ordinal theory assumes the axioms of completeness, transitivity, continuity, differentiability, nonsatiation and strict convexity. These axioms are usually substituted by requirement that the function used to describe consumer's preference or the preference function or the utility function is differentiable strongly increasing and strictly quasiconcave. In this way the consumer's choice problem of the most preferred bundle from the set of affordable bundles transforms into the problem of maximizing the utility subject to the budget constraint described by equality.

Associated to this optimization problem is the dual problem of minimizing the expenditure for a given consumer's welfare. However, if we divide commodities into the groups like food, shelter, entertainment and others, we face the problem of how to determine optimal expenditure on each group of commodities and then how to determine the optimal expenditure on each commodity within the same group. In this way we obtain the two-phase efficient income distribution problem. This problem, which involves the aggregating across the commodities as well as the separable decision making, is one of the most important problems in both theory and practice. The functional separability, which is different from Hicksian separability, has attracted the attention of many economists, among them of a well known economist William Moore Gorman who was one of the pioneers in the field of separability. The economist Charles Blackorby has in the same line of work devoted much attention to the interrelationship between the separability a multi-phase programming.

The purpose of this paper is not to explore all possible links among different kinds of separability and multi-phase programming but to derive the necessary and the sufficient conditions for two-phase programming by combining optimization and consumer behavior theory. The fact that both the one-phase and two-phase programming models give the same solution will be illustrated by a numerical example, thus making it easier to comprehend the interrelationship between the separability and multi-phase efficient distribution of the limited consumer's income. We hope that this paper will ease the difficulties resulting from the very demanding literature. Moreover, we hope to link the general and methodological knowledge of both optimization and economic theory regarding the consumer's choice into one unique, indivisible and interesting entity in an original and unusual way.

In Section 2 we start from the additively separable cardinal utility function and then replace it with its ordinal version. In process of doing this we naturally obtain the two-phase algorithm for efficient distribution of goods, where special attention is given to the necessary and the sufficient conditions for the two-phase programming. In Section 3 we supplement the algorithm by theorem and proof showing that the solutions of the one-phase and the corresponding two-phase programming models concide. These findings are illustrated by the example of generalized Cobb-Douglas utility function in Section 4. The last section summarizes the discussion presented in previous sections as well as clearly indicates the direction for further research.

2. From additive commodity quantity to two-phase programming

From the very beginning the separability has played an important role within the subjective value theory. Namely, the very well known economists Jevons, Menger and Walras, the founders of the subjective value theory, have started from the standpoint that the commodity utilities are independent and that the overall utility is separable, i.e.

$$u(x_1, x_2, ..., x_n) = u^1(x_1) + u^2(x_2) + ... + u^n(x_n). \tag{1}$$

Therefore the marginal utilities of individual goods are independent of consumption of other goods, i.e.

$$u_{jk} = 0 \qquad k \neq j\ . \tag{2}$$

The three founders of the mechanics of self-interest and utility have also used The Law of Diminishing Marginal Utility, which is the central law of cardinal theory, i.e.

$$u_{jj} < 0, \tag{3}$$

They were followed by Marshall who use this law under the name The Law of Satiable Wants. He has also used money as an invariable or at least nearly invariable measure of subjective satisfaction caused by consumption of goods. According to Samuelson, the Marshall's assumption of constant marginal utility of income with respect to the change of the price vector has very important implications. From this assumption it follows that the cardinal utility function has the form of

$$u(x_1, x_2, ..., x_n) = b + a\sum_{i=1}^{n} \alpha_i \ln x_i \qquad a, b, \alpha_i \in R, \quad a, \alpha_i > 0. \tag{4}$$

From the model of maximizing utility subject to the budget constraint

$$\begin{aligned} &\max_{\mathbf{x} \geq 0}\ u(\mathbf{x}) \\ &\text{s.t.}\ \ \mathbf{px} = M \end{aligned} \tag{5}$$

where $\mathbf{p} = (p_1, p_2, ..., p_n)$ is the price vector and M is the income, we derive Marshallian demand functions. In this case they take the form of

$$x_i^M(\mathbf{p}, M) = \frac{\alpha_i}{\sum_{j=1}^{n} \alpha_j} \cdot \frac{M}{p_i} \tag{6}$$

where coefficients α_i describe how the consumer distributes his limited income among the goods.

Example 1. *Let us now consider the following simple example of consumer whose utility function takes the form of the*

$$u(x_1, x_2, x_3) = \frac{1}{2}\ln x_1 + \frac{1}{3}\ln x_2 + \frac{1}{6}\ln x_3 . \tag{7}$$

It describes the consumer who spends half of its income on the first commodity. However, the consumer can also obtain this result in two steps: in first step he sets apart $\frac{1}{2}+\frac{1}{3}=\frac{5}{6}$ *of his income for purchasing the group of first two commodities. Then, he spends*

$$\frac{\frac{1}{2}}{\frac{1}{2}+\frac{1}{3}} = \frac{3}{5} \tag{8}$$

of that income on purchasing the first commodity. The second step is the result of maximizing the expenditure utility for the group of the first two commodities subject to the budget constraint. It reflects the equilibrium between the overall expenditure on the group of first two commodities and the part of the income that is spent on this group of commodities:

$$\begin{aligned} &\max_{x_1, x_2 \geq 0} \; v(x_1, x_2) = \frac{1}{2}\ln x_1 + \frac{1}{3}\ln x_2 \\ &\text{s.t.} \qquad p_1 x_1 + p_2 x_2 = \frac{5}{6}M \end{aligned} \tag{9}$$

The natural generalization of this discussion leads us towards two-phase programming. First we divide commodities into two groups:

$$\mathbf{x} = (\mathbf{y}, \mathbf{z}) \quad , \mathbf{y} \in \mathbf{R}_+^m, \mathbf{z} \in \mathbf{R}_+^{n-m} . \tag{10}$$

We do the same with the price vector:

$$\mathbf{p} = (\mathbf{q}, \mathbf{r}) \quad , \mathbf{q} \in \mathbf{R}_{++}^m, \mathbf{r} \in \mathbf{R}_{++}^{n-m} . \tag{11}$$

Now, the first step consists of determining how much to spend on each group of commodities. The second step consists of determining how much to spend on each commodity within the same group.

Since according to the ordinal consumer theory the weak preference relations are one of the basic means of describing consumer's behavior, it is therefore natural to explore how the preference structure resulting from the two-phase programming fits in the general framework of the consumer behavior theory.

First of all, induced preferences determined by the consumption of commodities within the same group are independent of the consumption of commodities from outside the group. Such preferences are called weakly separable and they can be replaced by the ordinal utility function of the form

$$u(\mathbf{y}, \mathbf{z}) = U[v(\mathbf{y}), \mathbf{z}] \tag{12}$$

where $v(\mathbf{y})$ is the utility function replacing the induced preferences. Furthermore, $U[v(\mathbf{y}), \mathbf{z}]$ is strictly increasing with respect to the first variable, i.e. quantity index for the commodities within the same group.

By using the price vector for the commodities from the same group and the expenditure for this group of goods we can determine conditional demand functions for the commodities belonging to that group

$$\mathbf{y}^M(\mathbf{q}, m). \tag{13}$$

This is achieved by solving the model

$$\max_{\mathbf{y}\geq 0} \quad v(\mathbf{y}) \qquad \text{s.t.} \quad \mathbf{qy} = m \tag{14}$$

It remains to to determine the optimal expenditure on this group of commodities,

$$\mathbf{qy}^M(\mathbf{q}, \mathbf{r}, M), \tag{15}$$

i.e. the expenditure for which the demand function and conditional demand function coincide,

$$\mathbf{y}^M(\mathbf{q}, \mathbf{r}, M) = \mathbf{y}^M\left[\mathbf{q}, \mathbf{qy}^M(\mathbf{q}, \mathbf{r}, M)\right] \tag{16}$$

Once we know the optimal quantity index for the commodities within the same group, we can obtain the optimal expenditure by solving the model (17) of minimizing the expenditure for a given level of utility

$$e(\mathbf{q}, v) = \min_{\mathbf{y}\geq 0} \quad \mathbf{qy} \qquad \text{s.t.} \quad v(\mathbf{y}) = v \tag{17}$$

where $e(\mathbf{q}, v)$ is the induced expenditure function.

Therefore, one should solve the problem

$$\max_{v\geq \mathbf{0}, \mathbf{z}\geq \mathbf{0}} \quad U(v, \mathbf{z}) \qquad \text{s.t.} \quad e(\mathbf{q}, v) + \mathbf{rz} = M \tag{18}$$

In general case, (18) is a non-linear programming problem, since the overall expenditure function $e(\mathbf{q}, \mathbf{v})$ is generally non-linear.

However, in case when the induced preference utility function is linearly homogenous, we know that the expenditure expansion path is a ray coming from the origin. As we move along this curve, the marginal rate of substitution between the goods remains the same. Furthermore, the minimal expenditure on commodities within the same group is proportional to consumption utility of the commodities from the same group.

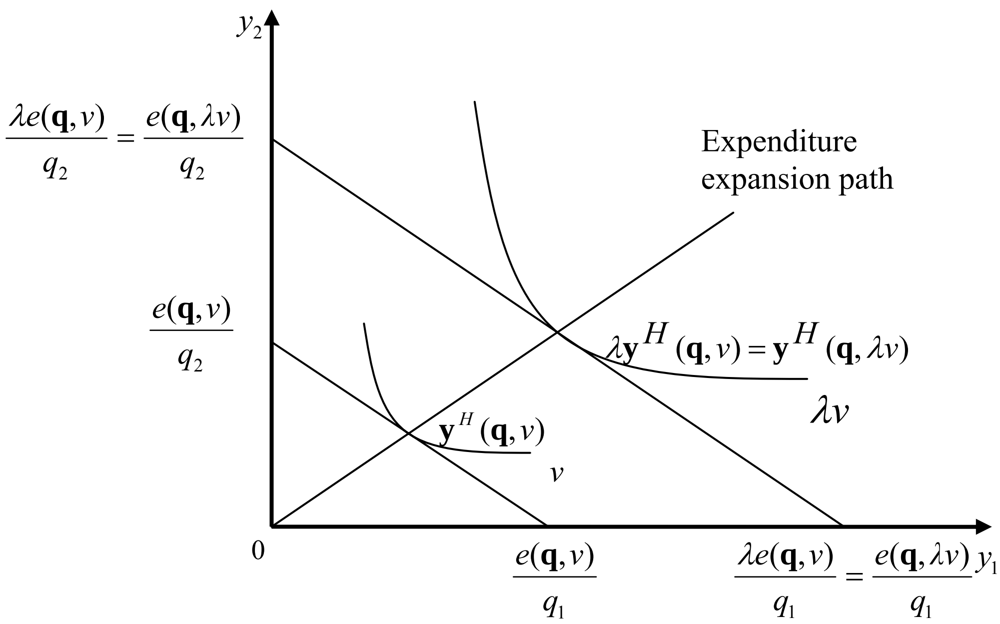


Figure 1. Linearly homogenous preferences and multiplicatively separable minimal expenditure function

Figure 1 depicts such a situation of linearly homogenous preferences and multiplicatively separable minimal expenditure function. By index H we denote the Hicksian demand function which determines the quantity of goods for which the expenditure for a given level of utility is minimal. The proportional change of these quantities causes the proportional change of utilities as well as the proportional change of expenditures.

Therefore, in case when the induced preference utility function is linearly homogenous, the minimal expenditure function is multiplicatively separable with respect to the price of commodities as well as to expenditure utility of the commodities within the same group

$$e(\mathbf{q}, v) = e(\mathbf{q}, v \cdot 1) = ve(\mathbf{q}, 1) = e(\mathbf{q}) \cdot v \tag{19}$$

Let us now illustrate this discussion on the simple utility function presented in Example 1. Since the strongly increasing transformations do not change the sequence of the utility indexes, without the loss of generality we can use them to describe the consumer preferences. Therefore, instead of (7) we can consider the following ordinal utility function:

$$e^{\frac{6}{5}u(\mathbf{x})} = x_1^{\frac{3}{5}} x_2^{\frac{2}{5}} x_3^{\frac{1}{5}} \tag{20}$$

The corresponding induced utility function for the consumption of the first two goods is now equal to

$$v(x_1, x_2) = x_1^{\frac{3}{5}} x_2^{\frac{2}{5}}. \tag{21}$$

Now the price index is equal to

$$e(\mathbf{q}) = e(p_1, p_2) = \frac{5}{2^{\frac{2}{5}} \cdot 3^{\frac{3}{5}}} \cdot p_1^{\frac{3}{5}} \cdot p_2^{\frac{2}{5}} \tag{22}$$

From the maximization model

$$\begin{aligned} &\max_{v, x_3 \geq 0} \quad v x_3^{\frac{1}{5}} \\ &\text{s.t.} \quad \frac{5}{2^{\frac{2}{5}} \cdot 3^{\frac{3}{5}}} \cdot p_1^{\frac{3}{5}} \cdot p_2^{\frac{2}{5}} \cdot v + p_3 x_3 = M \end{aligned} \tag{23}$$

we obtain the optimal consumption quantity index for the first two goods, equal to

$$v = \frac{M}{2^{\frac{3}{5}} \cdot 3^{\frac{2}{5}} \cdot p_1^{\frac{3}{5}} \cdot p_2^{\frac{2}{5}}} \tag{24}$$

as well as the optimal expenditures for this group of commodities:

$$e(p_1, p_2, p_3) = e(p_1, p_2, 1) \cdot v = \frac{5M}{6}. \tag{25}$$

In order to obtain the consumption of the first commodity, we solve the problem

$$\begin{aligned} &\max_{x_1, x_2 \geq 0} \quad x_1^{\frac{3}{5}} x_2^{\frac{2}{5}} \\ &\text{s.t.} \quad p_1 x_1 + p_2 x_2 = \frac{5M}{6} \end{aligned} \tag{26}$$

As expected, by solving the problem (26) we obtain the result that the consumer spends half of its income on the first commodity, i.e.

$$x_1^M(p_1,p_2,\frac{5M}{6}) = \frac{\frac{3}{5}\cdot\frac{5M}{6}}{p_1} = \frac{M}{2p_1} = x_1^M(p_1,p_2,p_3,M). \quad (27)$$

3. Problem description and solution algorithm

In previous section we have shown how to find the efficient distribution of limited income by using two phases. In the first phase we determine the commodity quantity index for the commodities within the same group. This index corresponds to the utility index of the bundle consisting of the commodities belonging to the same group.

The weakly separable preferences, which make the first phase possible, can now be replaced by the preference function of the form

$$u(\mathbf{y},\mathbf{z}) = U[v(\mathbf{y}),\mathbf{z}]. \quad (28)$$

Using this function we obtain the optimal commodity quantity index for the commodities within the same group by solving the following optimization problem:

$$\begin{aligned} &\max_{\mathrm{v},\mathbf{z}\geq\mathbf{0}} \ U(\mathrm{v},\mathbf{z}) \\ &\text{s.t.} \quad e(\mathbf{q},\mathrm{v}) + \mathbf{rz} = M \end{aligned} \quad (29)$$

This problem involves a non-linear constraint. However, in case of linearly homogenous induced preference function $v(\mathbf{y})$ we obtain the simplified problem (30):

$$\begin{aligned} &\max_{\mathrm{v}\geq 0,\mathbf{z}\geq\mathbf{0}} \ U(\mathrm{v},\mathbf{z}) \\ &\text{s.t.} \quad e(\mathbf{q})\cdot\mathrm{v} + \mathbf{rz} = M \end{aligned} \quad (30)$$

Thereby the price index $e(\mathbf{q})$ represents the minimal expenditure on this group of commodities corresponding to the unit quantity index. The optimal expenditure on this group of goods are now obtained by multiplying the price index with optimal quantity index.

The second phase consists of determining the expenditure on each commodity within the group. In order to do so, we have to solve the following optimization problem:

$$\begin{aligned} &\max_{\mathbf{y}\geq 0} \ v(\mathbf{y}) \\ &\text{s.t.} \quad \mathbf{qy} = m \end{aligned} \quad (31)$$

The problem (31) is dual to the optimization problem

$$\begin{aligned} &\min_{\mathbf{y}\geq 0} \ \mathbf{qy} \\ &\text{s.t.} \quad v(\mathbf{y}) = v \end{aligned} \quad (32)$$

which has already been solved when determining the price index. Having known this solution, we obtain the overall problem solution by multiplying the optimal quantity index with the group commodity quantities giving the unit quantity index minimal expenditures.

It is obvious now that in this way we obtain the same solution as when solving the following utility maximization problem subject to the income constraints:

$$\begin{aligned} &\max_{\mathbf{y},\mathbf{z}\geq 0} \ u(\mathbf{y},\mathbf{z}) \\ &\text{s.t.} \quad \mathbf{qy} + \mathbf{rz} = M \end{aligned} \quad (33)$$

Therefore, given the assumption mentioned in Section 2, we have the following theorem:

Theorem 1.

$$\mathbf{y}^M(\mathbf{q},\mathbf{r},M) = \mathbf{y}^M\left[\mathbf{q},\mathbf{q}v^M(\mathbf{q},\mathbf{r},M)\right]. \quad (34)$$

Proof. *Vector*

$$\mathbf{y}^M(\mathbf{q},\mathbf{r},M) \tag{35}$$

belongs to the set of affordable bundles determined by the prices of commodities belonging to the same group and by optimal expenditures on this group of commodities. If (35) is different from the unique vector maximizing the utility index of this group of commodities minimal expenditure quantity index

$$\mathbf{y}^M\left[\mathbf{q},\mathbf{q}y^M(\mathbf{q},\mathbf{r},M)\right], \tag{36}$$

we would have the following inequality:

$$v\left\{\mathbf{y}^M\left[\mathbf{q},\mathbf{q}y^M(\mathbf{q},\mathbf{r},M)\right]\right\} > v\left[\mathbf{y}^M(\mathbf{q},\mathbf{r},M)\right] \tag{37}$$

However, since U *is strongly increasing with respect to the consumption quantity index for the commodities within the group and vector*

$$\left\{\mathbf{y}^M\left[\mathbf{q},\mathbf{q}\mathbf{y}^M(\mathbf{q},\mathbf{r},M)\right],\mathbf{z}^M(\mathbf{q},\mathbf{r},M)\right\} \tag{38}$$

is an element of the affordable bundle space, it follows that vector (38) is not maximizing the overall utility. This proofs the theorem. ■

4. Optimal Distribution of Goods for the Generalized Cobb-Douglas Utility Function

In this section we derive the optimal distribution of goods for the generalized Cobb-Douglas utility function

$$u(x_1,\dots,x_n) = A\cdot\prod_{i=1}^{n} x_i^{\alpha_i} \tag{39}$$

by using the two-phase programming algorithm described in previous sections.

First we divide the goods into two groups:

$$\mathbf{y} = (x_1,\dots,x_l) \text{ and } \mathbf{z} = (x_{l+1},\dots,x_l) \tag{40}$$

We do the same with the price vector, thus obtaining

$$\mathbf{q} = (p_1,\dots,p_l) \text{ and } \mathbf{r} = (p_{l+1},\dots,p_n) \tag{41}$$

In order to obtain a linearly homogenous induced preference utility function, we consider a strongly increasing transformation of the overall utility function

$$u^{\frac{1}{\alpha}}(x_1,\dots,x_n) = A^{\frac{1}{\alpha}}\cdot\prod_{i=1}^{n} x_i^{\frac{\alpha_i}{\alpha}} \tag{42}$$

where

$$\alpha = \alpha_1 + \dots + \alpha_l \tag{43}$$

The corresponding induced utility function for the consumption of the first group of goods is now equal to

$$v(x_1,\dots,x_l) = \prod_{i=1}^{l} x_i^{\frac{\alpha_i}{\alpha}}. \tag{44}$$

Now the price index is equal to

$$e(\mathbf{q}) = e(p_1,...,p_l) = \alpha \cdot \prod_{i=1}^{l}\left(\frac{p_i}{\alpha_i}\right)^{\frac{\alpha_i}{\alpha}} \tag{45}$$

From the maximization model

$$\begin{aligned} &\max_{v,x_{l+1},...,x_n \geq 0} \quad A^{\frac{1}{\alpha}} \cdot v \cdot \prod_{i=l+1}^{n} x_i^{\frac{\alpha_i}{\alpha}} \\ &\text{s.t.} \quad \alpha \cdot \prod_{i=1}^{l}\left(\frac{p_i}{\alpha_i}\right)^{\frac{\alpha_i}{\alpha}} \cdot v + \sum_{j=l+1}^{n} p_j x_j = M \end{aligned} \tag{46}$$

we obtain the optimal consumption quantity index for the first group of goods, equal to

$$v = \frac{M}{\alpha + \sum\limits_{j=l+1}^{n} \alpha_j} \cdot \prod_{i=1}^{l}\left(\frac{\alpha_i}{p_i}\right)^{\frac{\alpha_i}{\alpha}} \tag{47}$$

as well as the optimal expenditures for the first group of commodities:

$$e(p_1,...,p_n) = e(p_1,...,p_n,1) \cdot v = \frac{\alpha \cdot M}{\alpha + \sum\limits_{j=l+1}^{n} \alpha_j}. \tag{48}$$

In order to obtain the consumption of the first group commodities, we solve the problem

$$\begin{aligned} &\max_{x_1,...,x_l \geq 0} \prod_{i=1}^{l} x_i^{\frac{\alpha_i}{\alpha}} \\ &\text{s.t.} \quad p_1 x_1 + ... + p_l x_l = \frac{\alpha \cdot M}{\alpha + \sum\limits_{j=l+1}^{n} \alpha_j} \end{aligned} \tag{49}$$

By solving the problem (49) we obtain that for each good j from the first group of commodities the consumer spends

$$x_i^M\left(p_1,...,p_l, \frac{\alpha \cdot M}{\alpha + \sum\limits_{j=l+1}^{n} \alpha_j}\right) = \frac{M \cdot \alpha_i}{p_i \cdot \left[\alpha + \sum\limits_{j=l+1}^{n} \alpha_j\right]} \tag{50}$$

which corresponds to the solution of the constrained one-phase overall utility maximization problem.

5. Conclusion

The Marshall's assumption of constant marginal utility of income has played an important role in the cardinal consumer behavior theory. It has resulted in the utility function whose ordinal version, the Cobb-Douglas utility function, plays an important role in both consumer and production theory. Based on the economic interpretation of the exponents of the Cobb-Douglas function we determine how much of income to spend on each commodity. It also leads to the two-phase programming. The preferences which make the two-phase efficient distribution of goods possible are being replaced by the weakly separable preference function having the consumption quantity index of the goods within the same group as one of its arguments. Thereby we also find the optimal quantity index in both models of maximizing the utility and minimizing the expenditure which are mutually dual.

By using the property that for linearly homogenous induced preference utility function the linear consumption expansion curve is linear, we can convert the optimal quantity index into proportionality factor and thus find the optimal expenditure for the group of goods as well as the optimal consumption of commodities within the group. The theorem and the proof showing that the solutions of the one-phase and the corresponding two-phase programming model's coincide are illustrated by the historically important Cobb-Douglas utility function.

It is clear that the direction for the further research is determined by the consumption expansion curve which is generally not linear. Therefore, there's a very challenging task of exploring the problems resulting from the complex relationship between the quantity index and the price index for the goods within the group in front of us as well as the numerical application of these problems.

References

[1] Blackorby, C., Primont, D. and Russell, R.R. (1978b), *Duality, Separability and Functional Structure: Theory and Economic Applications*. New York: American Elsevier.

[2] Gorman, W. M. (1959) "Separable Utility and Aggregation", *Econometrica* 27: 469-81

[3] Gorman, W. M. (1968) "The Structure of Utility Functions", *Review of Economic Studies* 35: 369-90

[4] Gorman, W. M. (1976) "Trick with utility Functions", in *Essays in Economic Analysis*, edited by M. Artis and R. Nobey. Cambridge University Press, pp. 2111-43

[5] Jevons, W. S. (1871). *Theory of Political Economy.* Fifth edition, edited by Collison Black (1970). Pelican Books.

[6] Marshall, A. (1890). *Principles of Economics*. London, Macmillan

[7] Menger, C. (1871). *Principles of Economics.* Translated by J. Dingwal and B.F. Hoselitz, Glancoe, Illinois, The Free Press, 1950.

[8] Pareto, V. (1906). *Manual of political Economy*. First translation in english 1971. New York: Augustus M. Kelly Publishers.

[9] Samuelson, P.A. (1942). "Constancy of the Marginal Utility of Income", in Oscar Lange et al., *Studies in Mathematical Economics and Econometrics: In Memory of Henry Schultz.* Chicago

[10] Walras, L. (1874). *Elements d'economic politique pure*, Lausanne, L. Corbaz. English translation by William Jaffe (1954) *Elements of Pure Economics*, London: Allen and Unwin.

[11] Walras, L. (1892) "Geometrical Theory of the Determination of Prices", *Annals of the American Academy of Political and Social Science*, July, pp. 47-64

EXPECTED AVAILABLE INVENTORY AND STOCKOUTS IN CYCLICAL RENEWAL PROCESSES

Kangzhou Wang[ab], Marija Bogataj[b]
[a]Lanzhou Polytechnical College, Department of Basic Science, Lanzhou, China
[b]University of Ljubljana, Faculty of Economics, Kardeljeva ploščad 17, 1000 Ljubljana, Slovenia
kangzhou.wang@hotmail.com, marija.bogataj@ef.uni-lj.si

Abstract: In stochastic material requirements planning (MRP) systems external demand is often described as renewal process. In this paper we consider the case when demand is described as cyclical compound Poisson process, i.e. external demand is generated by individual events separated by independent stochastic time intervals being exponentially distributed and quantities of demand are considered as a sequence of independent cyclical random variables. As the main results we present the general expression for expected stockout and expected available inventory in MRP systems.

Key Words: Compound Poisson Process, Cyclical Demand, Laplace Transforms, MRP.

1. Introduction

MRP is a system that controls inventory levels, plans production, helps to supply management with important information and supports the manufacturing control system with respect to the production of parts and assembly of them. In modern literature the study of MRP systems has received higher attention also at academic world (starting with Grubbström and his Linköping School). Extensions have been made to connect these studies with other theories, especially to give theoretical background to the supply chain management (Bogataj M. and Bogataj L., 2004). This approach improves the studies how to reduce all kind of risks in total supply chain when they are interacting (Bogataj D. and Bogataj M., 2007).

With the objective of obtaining optimal solutions, when timing and quantity of production are decision variables, quantitative aspects of planning and inventory control have resulted in several articles in journals and other publications about MRP and similar multi - level production-inventory systems. One breakthrough in this direction is the application of transforms (Laplace, z-, $\cdots$) and input-output analysis to MRP. Already in 1967 Grubbström pointed out that input-output analysis and Laplace (or z-) transforms improve the approach to MRP studies. The intensity of studying MRP systems in frequency domain has increased after 1997, when Mini-Symposium in Storlien widely opened the door to this theory (see Bogataj and Grubbström, 1997), though important contribution to further theoretical study of stochastic properties of MRP systems has been given already a year before the Mini - Symposium by Grubbström, 1996 . The study here is based on important contributions of Grubbström and members of his Linköping School (1999, 2000, 2003).

In previous studies the demand process was assumed as a renewal process and the quantity of each demand was supposed to be equal to 1. Usually in real world the arrivals of customers to the market are Poisson distributed and very often the demand size is not only one product. The first two papers which pointed out the necessity to introduce compound Poisson distributed demand in stochastic MRP models were papers of Bogataj and Bogataj (1998a, 1998b). Thorough study of compound distributions of demand in MRP systems was later given by Grubbström and Tang (2006) and Tang and Grubbström (2006).

In many real world cases the size of demand in renewal process has a special characteristic of periodicity. On the market seasonal movements of demand are well known for many different products. In this paper we consider the case when demand is described as cyclical compound Poisson process, i.e. external demand is generated by individual events separated by independent stochastic time intervals being exponentially distributed and quantities of demand are considered as a sequence of independent cyclical random variables. As the main results we present the general expression for expected stockout and available inventory in such a system. In this paper fundamental equations still form the main structure of material requirements planning model as it is suggested in several papers by R.W. Grubbström.

2. Transform Theorem and Fundamental Equations

Let us state some useful theorems on Laplace transforms which will be used later. The most important properties are:

Filtering property: $\int_0^\infty f(t)\delta(t-a)\mathrm{d}t = f(a)$, if $f(t)$ is continuous function on $[0,\infty)$.

Time differentiation: $\pounds\{f'(t)\} = s\pounds\{f(t)\} - f(0)$.

Derivative of transform: $\mathrm{d}^n \pounds\{f(t)\}/\mathrm{d}s^n = (-1)^n \pounds\{t^n f(t)\}$.

First translation theorem(shift on s-axis): $\pounds\{e^{at} f(t)\} = \pounds\{f(t-a)\}$.

The inverse transform of $\tilde{f}(s)$ may be(where β is chosen such that the integral will converge) computed as $f(t) = \frac{1}{2\pi i}\int_{\beta-i\infty}^{\beta+i\infty} e^{st}\tilde{f}(s)\mathrm{d}s$

The cumulative property: $E[\pounds\{f(t,T)\}] = \pounds\{E[f(t,T)]\}$, $f(t,T)$ is any function of time and stochastic variable T.

Also we have $\pounds\{e^{at}\} = \frac{1}{s-a}$, $\pounds\{\delta(t-t_i)\} = e^{-st_i}$, $\pounds\{H(t-t_i)\} = \frac{e^{-st_i}}{s}$. Some inverse Laplace transforms formulae are as follows: $\pounds^{-1}\{\frac{1}{t^n}\} = [(n-1)!]^{-1}t^{n-1}$, $\pounds^{-1}\{\frac{1}{s^2+a^2}\} = a^{-1}\sin at$, $\pounds^{-1}\{\frac{s}{s^2+a^2}\} = \cos at$.

The fundamental equations of MRP theory are balance equations describing the time development of inventory, backlogs and allocations(for details see R.W. Gubbström, 1999). Let there be N items in the system altogether. Demand $\mathbf{D}$, backlog $\mathbf{B}$ and production $\mathbf{P}$ are represented by N-dimensional column vectors each being a function of time. These vectors are rates with the dimension units per time unit and they are turned into Laplace transforms denoted by tildes or by $\pounds\{\cdot\}$, cumulative values(time integrals) of functions are denoted by bars and inverse transforms by $\pounds^{-1}\{\cdot\}$. For the production of one unit of item j, there is a need in the amount of h_{kj} of item k, and there is a lead time τ_j ahead of the completion of the production at which the components are needed. The h_{kj} are arranged into the square input matrix $\mathbf{H}$ describing the product structures of all relevant products. The lead times $\tau_1, \tau_2, \cdots, \tau_N$, create internal demands and are represented by a diagonal matrix $\tilde{\tau}$, the lead time matrix, having $e^{s\tau_j}$ in its jth diagonal position, where s is the complex Laplace frequency. $\tilde{\mathbf{H}} = \mathbf{H}\tilde{\tau}$ is the generalized input matrix and it captures component requirement together with their requirement timing.

The available inventory $\tilde{\mathbf{R}}(s)$ is cumulative production $\tilde{\mathbf{P}}(s)/s$ less cumulative demand $\tilde{\mathbf{D}}(s)/s$ and cumulative internal demand $\mathbf{H}\tilde{\tau}(s)\tilde{\mathbf{P}}(s)/s$ and plus backlog $\tilde{\mathbf{B}}(s)$, we obtain

$$\tilde{\mathbf{R}}(s) = \frac{\tilde{\mathbf{R}}(0) - \tilde{\mathbf{B}}(0) + (\mathbf{I} - \mathbf{H}\tilde{\tau}(s))\tilde{\mathbf{P}}(\mathbf{s}) - \tilde{\mathbf{D}}(s)}{s} + \tilde{\mathbf{B}}(s).$$

For any item, its available inventory and its backlog cannot be positive at the same time, since a delivery takes place from available inventory as soon as there is an unsatisfied external demand. Hence, if for any component $R_j(t) > 0$ at time t, then $B_j(t) = 0$, and vice versa. Therefore, $\mathbf{R}(t)$ and $\mathbf{B}(t)$, both being nonnegative, may be written as

$$\tilde{\mathbf{R}}(s) = \left[\frac{\tilde{\mathbf{R}}(0) - \tilde{\mathbf{B}}(0) + (\mathbf{I} - \mathbf{H}\tilde{\tau}(s))\tilde{\mathbf{P}}(\mathbf{s}) - \tilde{\mathbf{D}}(s)}{s}\right]^+,$$

$$\tilde{\mathbf{B}}(s) = \left[\frac{\tilde{\mathbf{B}}(0) - \tilde{\mathbf{R}}(0) - (\mathbf{I} - \mathbf{H}\tilde{\tau}(s))\tilde{\mathbf{P}}(\mathbf{s}) + \tilde{\mathbf{D}}(s)}{s}\right]^+,$$

where $[\cdot]^+$ is the maximum operator Max$\{0,\cdot\}$ operating on a s function. The equations above defining the development of $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{B}}$ we call the fundamental equations.

3. Cyclical Demand

In this paper we just consider single end item system. We use P, B and D to denote the production, backlog and external demand of the end item. As we assumed, the external demand is generated by individual events separated by independent and identically distributed time intervals Y_i having exponential distribution, and the sizes of each demand X_i are independent and cyclical random variables, $i = 1, 2, \cdots$, i.e. for fixed constant number of periods m, we have $F_{X_j}(x) = F_{X_{m+j}}(x) = F_{X_{2m+j}}(x) = \cdots = F_{X_{nm+j}}(x) = \cdots$, $1 \leq j \leq m$, $n \in N$, $F_{X_{nm+j}}(x) = P\big(X_{nm+j} \leq x\big)$. As the well-known property of the Poisson process, the ith demand occurs in time $T_i = \Sigma_{l=0}^{i} Y_l$, and the distribution function of T_i is a Gamma distribution with parameters i and λ. The Laplace transform of the distribution of T_i can be written as

$$\pounds\{f_{T_i}(t)\} = \pounds\Big\{\frac{\lambda e^{-\lambda t}(\lambda t)^{i-1}}{(i-1)!}\Big\} = \frac{\lambda^i}{(i-1)!}\pounds\{e^{-\lambda t}t^{i-1}\}.$$

Using the first shift theorem $\pounds\{e^{at}f(t)\} = F(s-a)$, where $\pounds\{f(t)\} = F(s)$, and $\pounds\{t^{i-1}\} = \frac{(i-1)!}{s^i}$, we have

$$\pounds\{f_{T_i}(t)\} = \frac{\lambda^i}{(i-1)!}\frac{(i-1)!}{(s+\lambda)^i} = \big(\frac{\lambda}{s+\lambda}\big)^i.$$

Hence,

$$E[e^{-sT_i}] = \int_0^\infty e^{-st} f_{T_i}(t)\mathrm{d}t = \pounds\{f_{T_i}(t)\} = \big(\frac{\lambda}{s+\lambda}\big)^i. \tag{1}$$

We now assume that external demand follows a stochastic process $D(t)$ of the renewal type, i.e.

$$D(t) = \sum_{i=1}^{\infty} X_i \delta(t - T_i), \tag{2}$$

which is made up of sequence if unit impulses $\delta(\cdot)$, i.e. Dirac delta functions. Then, using Eq.(1), we obtain

$$\begin{aligned}\pounds\{E[D(t)]\} &= E\big[\pounds\{D(t)\}\big] = E\big[\sum_{i=1}^{\infty}\pounds\{X_i\delta(t-T_i)\}\big] = E\big[\sum_{i=1}^{\infty}X_i\pounds\{\delta(t-T_i)\}\big]\\ &= E\big[\sum_{i=1}^{\infty}X_i e^{-sT_i}\big] = \sum_{i=1}^{\infty}E\big[X_i e^{-sT_i}\big] = \sum_{i=1}^{\infty}E[X_i]E\big[e^{-sT_i}\big] = \sum_{i=1}^{\infty}E[X_i]\big(\frac{\lambda}{s+\lambda}\big)^i.\end{aligned}$$

Since the demand sizes X_i are cyclical we denote $\mu_i = E[X_i]$, that is, $\mu_j = \mu_{m+j} = \cdots = \mu_{nm+j} = \cdots$, where $1 \leq j \leq m$. Because the number of demand events goes to infinity, also the number of cycles goes to infinity. Then we have

$$\begin{aligned}\pounds\{E[D(t)]\} &= \sum_{i=1}^{\infty}E[X_i]\big(\frac{\lambda}{s+\lambda}\big)^i = \mu_1\big(\frac{\lambda}{s+\lambda}\big) + \mu_2\big(\frac{\lambda}{s+\lambda}\big)^2 + \cdots + \mu_m\big(\frac{\lambda}{s+\lambda}\big)^m\\ &\qquad + \mu_1\big(\frac{\lambda}{s+\lambda}\big)^{m+1} + \mu_2\big(\frac{\lambda}{s+\lambda}\big)^{m+2} + \cdots + \mu_m\big(\frac{\lambda}{s+\lambda}\big)^{2m} + \cdots\\ &= \mu_1\sum_{n=0}^{\infty}\big(\frac{\lambda}{s+\lambda}\big)^{nm+1} + \mu_2\sum_{n=0}^{\infty}\big(\frac{\lambda}{s+\lambda}\big)^{nm+2} + \cdots\\ &\qquad + \mu_{m-1}\sum_{n=0}^{\infty}\big(\frac{\lambda}{s+\lambda}\big)^{nm+m-1} + \mu_m\sum_{n=0}^{\infty}\big(\frac{\lambda}{s+\lambda}\big)^{nm+m}\\ &= \sum_{j=1}^{m}\mu_j\sum_{n=0}^{\infty}\big(\frac{\lambda}{s+\lambda}\big)^{nm+j}.\end{aligned}$$

Meanwhile, we can obtain

$$\sum_{n=0}^{\infty}(\frac{\lambda}{s+\lambda})^{nm+j} = (\frac{\lambda}{s+\lambda})^{j}\sum_{n=0}^{\infty}(\frac{\lambda}{s+\lambda})^{nm} = (\frac{\lambda}{s+\lambda})^{j}\sum_{n=0}^{\infty}\Big((\frac{\lambda}{s+\lambda})^{m}\Big)^{n}$$
$$= (\frac{\lambda}{s+\lambda})^{j}\Big(1-(\frac{\lambda}{s+\lambda})^{m}\Big)^{-1},$$

where the following condition should be fulfilled: $\left|\frac{\lambda}{s+\lambda}\right| < 1$. Finally the Laplace transform of expected external demand rate will be

$$\pounds\{E[D(t)]\} = \sum_{j=1}^{m}\mu_j(\frac{\lambda}{s+\lambda})^{j}\Big(1-(\frac{\lambda}{s+\lambda})^{m}\Big)^{-1} = \Big(1-(\frac{\lambda}{s+\lambda})^{m}\Big)^{-1}\sum_{j=1}^{m}\mu_j(\frac{\lambda}{s+\lambda})^{j}. \quad (3)$$

We also assume that external cumulative demand follows a compound Poisson process having cumulative demand $\bar{D}(t)$, i.e.

$$\bar{D}(t) = \sum_{i=1}^{N(t)} X_i = \sum_{i=1}^{\infty} X_i H(t-T_i), \quad (4)$$

where $H(\cdot)$ is a Heaviside function, $N(t)$ follows a Poisson process with rate λ, representing the number of demand events since time $t = 0$. The Laplace transform of expected accumulative demand is therefore

$$\begin{aligned}\pounds\{E[\bar{D}(t)]\} &= \frac{1}{s}\pounds\{E[D(t)]\} = \frac{1}{s}\Big(1-(\frac{\lambda}{s+\lambda})^{m}\Big)^{-1}\sum_{j=1}^{m}\mu_j(\frac{\lambda}{s+\lambda})^{j} \\ &= \sum_{j=1}^{m}\mu_j\lambda^j\frac{(s+\lambda)^{m-j}}{s((s+\lambda)^m-\lambda^m)} = \frac{\mu_1\lambda(s+\lambda)^{m-1}+\mu_2\lambda^2(s+\lambda)^{m-2}+\cdots+\mu_m\lambda^m}{s^2\left((s+\lambda)^{m-1}+\lambda(s+\lambda)^{m-2}+\cdots+\lambda^{m-1}\right)} \\ &= \frac{\sum_{j=1}^{m}\mu_j\lambda^j(s+\lambda)^{m-j}}{s^2\sum_{j=0}^{m-1}(s+\lambda)^{m-1-j}\lambda^j}.\end{aligned} \quad (5)$$

Obviously, $s_0 = 0$ is the second-order pole of $\pounds\{E[\bar{D}(t)]\}$, and

$$s_k = (e^{\frac{2k\pi}{m}i}-1)\lambda = (\cos\frac{2k\pi}{m}-1+i\sin\frac{2k\pi}{m})\lambda \quad (6)$$

are the simple poles, where $k = 1, 2, \cdots, m-1$. Here i is equal to $i = \sqrt{-1}$. If we denote $s_k = a_k + ib_k$, where a_k and b_k are all real numbers. Then the expression of $\pounds\{E[\bar{D}(t)]\}$ can be rewritten as

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{m}\frac{\mu_j\lambda^j(s+\lambda)^{m-j}}{s^2(s-s_1)(s-s_2)\cdots(s-s_{m-1})} = \sum_{j=1}^{m}\frac{\mu_j\lambda^j(s+\lambda)^{m-j}}{s^2\prod_{i=1}^{m-1}(s-s_i)}. \quad (7)$$

As the well-known property of complex roots, s_1 and s_{m-1}, s_2 and s_{m-2}, etc, $s_{\frac{m-1}{2}}$ and $s_{\frac{m+1}{2}}$ are conjugate each other when m is an odd number. When m is an even number, then s_1 and s_{m-1}, s_2 and s_{m-2}, etc, $s_{\frac{m}{2}-1}$ and $s_{\frac{m}{2}+1}$ are conjugate each other, respectively, and the $\frac{m}{2}th$ simple root $s_{\frac{m}{2}}$ equals to -2λ. Then we can obtain

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{m}\frac{\mu_j\lambda^j(s+\lambda)^{m-j}}{s^2\prod_{i=1}^{\frac{m-1}{2}}\left((s-a_i)^2+b_i^2\right)},$$

when m is an odd number. Because $1 \le i \le \frac{m-1}{2}$, we have $a_i = \lambda(\cos\frac{2i\pi}{m} - 1) < 0$ and $b_i = \lambda \sin\frac{2i\pi}{m} > 0$. Meanwhile, if m is an even number, we have

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{m} \frac{\mu_j \lambda^j (s+\lambda)^{m-j}}{s^2(s+2\lambda)\prod_{i=1}^{\frac{m}{2}-1}\left((s-a_i)^2+b_i^2\right)},$$

where $a_i = \lambda(\cos\frac{2i\pi}{m} - 1) < 0$ and $b_i = \lambda \sin\frac{2i\pi}{m} > 0$.

In algebra, we have following formula,

$$\begin{aligned}\frac{P(s)}{Q(s)} &= \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{(s-a_1)^{k_1}(s-a_2)^{k_2}\cdots(s^2+p_1 s+q_1)^{l_1}(s^2+p_2 s+q_2)^{l_2}\cdots} \\ &= \frac{A_1}{s-a_1} + \frac{A_2}{(s-a_1)^2} + \cdots + \frac{D_1 s + E_1}{s^2+p_1 s+q_1} + \frac{D_2 s+E_2}{(s^2+p_1 s+q_1)^2} + \cdots \\ &\quad + \frac{D_{l_1} s + E_{l_1}}{(s^2+p_1 s+q_1)^{l_1}} + \frac{F_1 s+G_1}{s^2+p_2 s+q_2} + \cdots + \frac{F_{l_2} s + G_{l_2}}{(s^2+p_2 s+q_2)^{l_2}} + \cdots.\end{aligned} \tag{8}$$

where $a_1, a_2, \cdots; p_1, p_2, \cdots; q_1, q_2, \cdots$ are real numbers with $p_i^2 - 4q_i \le 0$, and $k_1, k_2,\ l_1, l_2, ...$ are positive integers, the terms $(s - a_i)$ are the linear factors of $Q(s)$ which correspond to real roots of $Q(s)$, and the terms $(s^2 + p_i s + q_i)$ are the irreducible quadratic factors of $Q(s)$ which correspond to pairs of complex conjugate roots of $Q(s)$, and the degree of numerator $P(s)$ is strictly smaller than the degree of the denominator $Q(s)$. Following the above discussion, with the aid of formula (8), if m is an odd number, we get

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{m} \mu_j \lambda^j \Big(\frac{C_{j1}}{s} + \frac{C_{j2}}{s^2} + \sum_{i=1}^{\frac{m-1}{2}} \frac{A_{ji}s + B_{ji}}{(s-a_i)^2+b_i^2}\Big),$$

where A_{jk}, B_{jk} and C_{jk} are real constants, if we denote $\frac{P(s)}{Q(s)} = \frac{(s+\lambda)^{m-j}}{s^2\prod_{i=1}^{\frac{m-1}{2}}\left((s-a_i)^2+b_i^2\right)}$, then we have $Q(s) = s^2 H_1(s)$, where $H_1(s) = \prod_{i=1}^{\frac{m-1}{2}}\left((s-a_i)^2+b_i^2\right)$, and $H_1(0) \ne 0$. Using the formulae in (Zwillinger, 2003, pp.87-88), we have

$$\begin{aligned} C_{j1} &= \frac{\mathrm{d}}{\mathrm{d}s}\Big(\frac{P(s)}{H_1(s)}\Big)\Big|_{s=0} = \frac{(m-j)\lambda^{m-j-1}}{\prod_{i=1}^{\frac{m-1}{2}}(a_i^2+b_i^2)} + \frac{\lambda^{m-j}\sum_{k=1}^{\frac{m-1}{2}}\Big(2a_k\prod_{i\ne k}(a_i^2+b_i^2)\Big)}{\prod_{i=1}^{\frac{m-1}{2}}(a_i^2+b_i^2)^2}, \\ C_{j2} &= \frac{P(0)}{H_1(0)} = \frac{\lambda^{m-j}}{\prod_{i=1}^{\frac{m-1}{2}}(a_i^2+b_i^2)} > 0. \end{aligned}$$

For getting the coefficients of quadratic factor, we still denote $Q(s) = \big((s-a_k)^2+b_k^2\big)H_2(s)$, where $H_2(s) = s^2\prod_{i\ne k}\left((s-a_i)^2+b_i^2\right)$, then

$$\frac{P(s)}{Q(s)} = \frac{A_{jk}s + B_{jk}}{\left((s-a_k)^2+b_k^2\right)} + \frac{G(s)}{H_2(s)},$$

because A_{jk} and B_{jk} are both real, after multiplying the above equation by $Q(s)$, a root of $(s-a_k)^2+b_k^2$(i.e. s_k) is substituted for s, then the values of A_{jk} and B_{jk} can be inferred from this single complex equation by equating real and imaginary parts. (Since $(s-a_k)^2+b_k^2$ divides $Q(s)$, there are no zeros in the denominator.) That is,

$$A_{jk} = Re\Big(\frac{P(s_k)}{H_2(s_k)}\Big) = Re\Big(\frac{(s_k+\lambda)^{m-j}}{s_k^2\prod_{i\ne k}\left((s_k-a_i)^2+b_i^2\right)}\Big),$$

$$B_{jk} = Im\Big(\frac{P(s_k)}{H_2(s_k)}\Big) = Im\Big(\frac{(s_k+\lambda)^{m-j}}{s_k^2\prod_{i\ne k}\left((s_k-a_i)^2+b_i^2\right)}\Big).$$

where $i = 1, 2, \cdots, \frac{m-1}{2}$.

In case m is an even number, we get

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{m} \mu_j \lambda^j \Big(\frac{C_{j1}}{s} + \frac{C_{j2}}{s^2} + \frac{C_{j3}}{s+2\lambda} + \sum_{i=1}^{\frac{m}{2}-1} \frac{A_{ji}s + B_{ji}}{(s-a_i)^2 + b_i^2} \Big),$$

where A_{jk}, B_{jk} and C_{jk} are real constants. Using the same method for getting the all coefficients when m is an odd number, we obtain

$$\begin{aligned}
C_{j1} &= \frac{\sum_{k=1}^{\frac{m}{2}-1} \big(a_k \prod_{i\neq k}(a_i^2 + b_i^2)\big)}{\lambda \prod_{i=1}^{\frac{m}{2}-1}(a_i^2 + b_i^2)^2} + \frac{(2m-2j-1)\lambda^{m-j-2}}{4\prod_{i=1}^{\frac{m}{2}-1}(a_i^2 + b_i^2)}, \\
C_{j2} &= \frac{\lambda^{m-j}}{2\lambda \prod_{i=1}^{\frac{m}{2}-1}(a_i^2 + b_i^2)} > 0, C_{j3} = \frac{(-\lambda)^{m-j}}{4\lambda^2 \prod_{i=1}^{\frac{m}{2}-1}\big((2\lambda + a_i)^2 + b_i^2\big)}, \\
A_{jk} &= Re\Big(\frac{(s_k + \lambda)^{m-j}}{s_k^2 (s_k + 2\lambda) \prod_{i\neq k}\big((s_k - a_i)^2 + b_i^2\big)} \Big), \\
B_{jk} &= Im\Big(\frac{(s_k + \lambda)^{m-j}}{s_k^2 (s_k + 2\lambda) \prod_{i\neq k}(\big((s_k - a_i)^2 + b_i^2\big)} \Big),
\end{aligned}$$

where $i = 1, 2, \cdots, \frac{m}{2} - 1$.

Now, using some tables of inverse Laplace transforms, we can get $\pounds^{-1}\{\frac{C_{j1}}{s}\} = C_{j1}$, $\pounds^{-1}\{\frac{C_{j2}}{s^2}\} = C_{j2}t$, $\pounds^{-1}\{\frac{C_{j3}}{s+2\lambda}\} = C_{j3}e^{-2\lambda t}$, and

$$\begin{aligned}
\pounds^{-1}\Big\{ \frac{A_{jk}s + B_{jk}}{(s-a_k)^2 + b_k^2} \Big\} &= \pounds^{-1}\Big\{ \frac{A_{jk}(s-a_k) + B_{jk} + a_k A_{jk}}{(s-a_k)^2 + b_k^2} \Big\} \\
&= \pounds^{-1}\Big\{ \frac{A_{jk}(s-a_k)}{(s-a_k)^2 + b_k^2} \Big\} + \pounds^{-1}\Big\{ \frac{B_{jk} + a_k A_{jk}}{(s-a_k)^2 + b_k^2} \Big\} \\
&= e^{a_k t} A_{jk} \cos(b_k t) + e^{a_k t} b_k^{-1} (B_{jk} + a_k A_{jk}) \sin(b_k t).
\end{aligned}$$

Finally, we obtain the expected value of the cumulative demand for any time $t \in (0, T)$, where T would be also ∞, as following

$$\begin{aligned}
E[\bar{D}(t)] = \sum_{j=1}^{m} \mu_j \lambda^j \Big(C_{j1} + C_{j2}t + C_{j3}e^{-2\lambda t} + \sum_{k=1}^{l} \big(e^{a_k t} A_{jk} \cos(b_k t) \\
+ e^{a_k t} b_k^{-1} (B_{jk} + a_k A_{jk}) \sin(b_k t) \big) \Big).
\end{aligned} \tag{9}$$

where, if m is an odd number, $l = \frac{m-1}{2}$ and $C_{j3} \equiv 0$; if m is an even number, $l = \frac{m}{2} - 1$. A_{jk}, B_{jk} and C_{jk} are constants as have been stated previously.

In expression (9), since a_k is always smaller than zero and C_{j2} is always greater than zero, we have following conclusion: with the increase of time variable t, the values of the third and the fourth terms in expression (9) will rapidly decrease, the cumulative demand goes to an increasing linear function when time variable t goes to infinity. It shows that the greater is the value of time variable t, the smaller is the influence of cyclical demand on total demand.

For illustrating our method of obtaining the result of $E[\bar{D}(t)]$, we will give one numerical example in following. Assuming $m = 4$, that is, the case of even number of periodical units , and $\lambda = 1$. Then using the expression (6), we obtain $s_0 = 0$, $s_1 = -1 + i$, $s_2 = -2$ and $s_3 = -1 - i$. Also, assuming $\mu_1 = 2$, $\mu_2 = 4$, $\mu_3 = 3$, $\mu_4 = 1$, using the expression of $\pounds\{E[\bar{D}(t)]\}$ when the number of periodical units m is even, we have

$$\pounds\{E[\bar{D}(t)]\} = \sum_{j=1}^{4} \frac{\mu_j \lambda^j (s+\lambda)^{4-j}}{s^2\big((s-a_1)^2 + b_1^2\big)(s+2\lambda)} = \frac{2(s+1)^3}{s^2\big((s+1)^2+1\big)(s+2)}$$

$$+\frac{4(s+1)^2}{s^2\big((s+1)^2+1\big)(s+2)}+\frac{3(s+1)}{s^2\big((s+1)^2+1\big)(s+2)}+\frac{1}{s^2\big((s+1)^2+1\big)(s+2)}$$
$$=2\Big(\frac{\frac{3}{8}}{s}+\frac{\frac{1}{4}}{s^2}+\frac{-\frac{1}{4}s}{\big((s+1)^2+1\big)}+\frac{-\frac{1}{8}}{s+2}\Big)+4\Big(\frac{\frac{1}{8}}{s}+\frac{\frac{1}{4}}{s^2}+\frac{-\frac{1}{4}s-\frac{1}{2}}{\big((s+1)^2+1\big)}+\frac{\frac{1}{8}}{s+2}\Big)$$
$$+3\Big(\frac{-\frac{1}{8}}{s}+\frac{\frac{1}{4}}{s^2}+\frac{\frac{1}{4}s}{\big((s+1)^2+1\big)}+\frac{-\frac{1}{8}}{s+2}\Big)+\Big(\frac{-\frac{3}{8}}{s}+\frac{\frac{1}{4}}{s^2}+\frac{\frac{1}{4}s+\frac{1}{2}}{\big((s+1)^2+1\big)}+\frac{\frac{1}{8}}{s+2}\Big)$$
$$=\frac{1}{2s}+\frac{5}{2s^2}+\frac{-\frac{1}{2}s-\frac{3}{2}}{\big((s+1)^2+1\big)}.$$

Using formula (9), we can get the last result of $E[\bar{D}(t)]$ as

$$E[\bar{D}(t)]=\pounds^{-1}\Big\{\frac{1}{2s}\Big\}+\pounds^{-1}\Big\{\frac{5}{2s^2}\Big\}+\pounds^{-1}\Big\{\frac{-\frac{1}{2}(s+1)-1}{\big((s+1)^2+1\big)}\Big\}=\frac{1}{2}+\frac{5}{2}t-\frac{1}{2}e^{-t}\cos t-e^{-t}\sin t.$$

Especially, if $m=1$, that is, demand sizes are identically distributed random variables, if we denote $\mu=\mu_1=\mu_2=\cdots=\mu_n=\cdots$, we can easily get

$$E[D(t)]=E\big[\sum_{i=1}^{\infty}X_i\delta(t-T_i)\big]=\sum_{i=1}^{\infty}E[X_i]E\big[\delta(t-T_i)\big]=\sum_{i=1}^{\infty}\mu\int_0^{\infty}\delta(t-x)f_{T_i}(x)\mathrm{d}x$$
$$=\sum_{i=1}^{\infty}\mu\int_0^{\infty}\delta(x-t)f_{T_i}(x)\mathrm{d}x=\sum_{i=1}^{\infty}\mu f_{T_i}(t)=\sum_{i=1}^{\infty}\mu\lambda e^{-\lambda t}\frac{(\lambda t)^{i-1}}{(i-1)!}=\mu\lambda e^{-\lambda t}\sum_{i=0}^{\infty}\frac{(\lambda t)^i}{i!}$$
$$=\mu\lambda e^{-\lambda t}e^{\lambda t}=\lambda\mu,$$

Then, we have $E[\bar{D}(t)]=\int_0^t E[D(t)]\mathrm{d}t=\lambda\mu t$.

If production takes place in batches of the size P_i at time t_i, $i=1,2,\cdots$ and is assumed to be a deterministic time function $P(t)=\sum_{i=1}^{\infty}P_i\delta(t-t_i)$ having the cumulative $\bar{P}(t)$, $\bar{P}(t)=\sum_{i=1}^{\infty}P_iH(t-t_i)$, where $\bar{P}(t)$ is a staircase function with steps of height P_i and widths $t_{i+1}-t_i$ with a first step at $t=t_1$, as the definition of $\bar{P}_i$ and $\bar{P}(t)$, we also have $\bar{P}_i=\bar{P}(t)$, $t\in[t_i,t_{i+1})$, $i=0,1,2,\cdots$, $\bar{P}_i=\bar{P}_{i-1}+P_i$. During the ith interval, the probability of stockout $x\geq 0$ at time t(in the ith interval) will be

$$P\big(B(t)\leq x\big)=P\big(\bar{D}(t)\leq x+\bar{P}(t)\big).$$

So, the Laplace transform of expected stockouts can be written

$$\begin{aligned}\pounds\{E[B(t)]\}&=\pounds\{E[\bar{D}(t)]\}-\pounds\{\bar{P}(t)\}\\&=\frac{1}{s}\Big(1-\big(\frac{\lambda}{s+\lambda}\big)^m\Big)^{-1}\sum_{j=1}^{m}\mu_j\big(\frac{\lambda}{s+\lambda}\big)^j-\frac{\tilde{P}(s)}{s}.\end{aligned}\tag{10}$$

Hence, if we assume $B(0)=\bar{P}(0)=\bar{D}(0)=R(0)=0$, and applying Eq.(9), the expected stockouts will be

$$\begin{aligned}E[B(t)]=\Big[E[\bar{D}(t)]-\bar{P}(t)\Big]^+&=\Big[\sum_{j=1}^{m}\mu_j\lambda^j\Big(C_{j1}+C_{j2}t+C_{j3}e^{-2\lambda t}\\&+\sum_{k=1}^{l}\big(e^{a_kt}A_{jk}\cos(b_kt)+e^{a_kt}b_k^{-1}(B_{jk}+a_kA_{jk})\sin(b_kt)\big)\Big)-\bar{P}(t)\Big]^+.\end{aligned}\tag{11}$$

The expected available inventory can be expressed as

$$\begin{aligned}E[R(t)]=\bar{P}(t)-E[\bar{D}(t)]=\bar{P}(t)&-\sum_{j=1}^{m}\mu_j\lambda^j\Big(C_{j1}+C_{j2}t+C_{j3}e^{-2\lambda t}\\&+\sum_{k=1}^{l}\big(e^{a_kt}A_{jk}\cos(b_kt)+e^{a_kt}b_k^{-1}(B_{jk}+a_kA_{jk})\sin(b_kt)\big)\Big).\end{aligned}\tag{12}$$

4. Conclusion

In this work we developed some parameters of MRP model for a multi-stage, single end-product cases with cyclical demand in compound Poisson process i.e. external demand is generated by individual events separated by independent stochastic time intervals being exponentially distributed and quantities of demand are considered as a sequence of independent cyclical random variables. We found out that the useful expression for expected cumulative demand for any final time t can be derived which enables us to evaluate behavior of MRP very easy and straightforward for any final number of periodical units in the cycle and for any pattern of cycles even when time horizon is infinite.

References

[1] Bogataj, L., Grubbström, R. W. (1997), *Input-output Analysis and Laplace Transforms in Material Requirements Planning*, Storlien, Sweden.

[2] Bogataj, M., Bogataj, L. (1998a), Compound distribution of demand in location-inventory problems. V: PAPACHRISTOS, Sotirios (ur.), GANAS, Ioannis (ur.). Third ISIR summer school, Ioannina, 1998. *Inventory modeling in production and supply chains: research papers.* Ioannina: International Society for Inventory Research: University of Ioannina, pp15-46.

[3] Bogataj, L., Bogataj, M. (1998b), Input-Output analysis applied to MRP models with compound distribution of total demand, *Proceeding 12th International Conference on Input-Output Techniques*, (Erik Dietzenbacher, Ed.), International Input-Output Organization, New York.

[4] Bogataj, M., Bogataj, L. (2004), On the compact presentation of the lead times perturbations in distribution networks, *International Journal of Production Economics,* Vol.88(2), pp145-155.

[5] Bogataj, D., Bogataj, M. (2007), Measuring the supply chain risk and vulnerability in frequency space, *International Journal of Production Economics,* Vol.108, pp291-301.

[6] Churchill, R. V., Brown, J. W. (1984), *Complex Variables and Applications*, 4th ed, McGraw-Hill Inc., Tokyo.

[7] Grubbström, R. W. (1967), On the application of Laplace transform to certain economic problems, *Management Science,* Vol.13, pp558-567.

[8] Grubbström, R. W. (1996), Stochastic properties of a production-inventory process with planned production using transform methodology, *International Journal of Production Economics,* Vol.45,
pp407-419.

[9] Grubbström, R. W. (1999), A net present value approach to safety stocks in a multi-level MRP system, *International Journal of Production Economics,* Vol.59, pp361-375.

[10] Grubbström, R. W., Tang, O. (2000), An overview of input-output analysis applied to production-inventory system, *Economic Systems Research,* Vol.12, pp3-25.

[11] Grubbström, R. W. (2003), A stochastic model of multi-level/multi-stage capacity-constrained produced-inventory system, *International Journal of Production Economics,* Vol.81-82, pp483-494.

[12] Grubbström, R. W., Tang, O. (2006), The moments and central moments of a compound distribution, *European Journal of Operational Research*, Vol.170(1), pp106-119.

[13] Tang, O., Grubbström, R. W. (2006), On using higher-order moments for stochastic inventory systems, *International Journal of Production Economics*, Vol.104(2), pp454-461.

[14] Zwillinger, Daniel. (2003), *CRC standard mathematical tables and formulae*, 31st ed, Chapman and Hall/CRC, Boca Raton .

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 11

Education and Statistics

STOCK PRICES TEHNICAL ANALYSIS

Josip Arnerić, Elza Jurun, Snježana Pivac
University of Split, Faculty of Economics
Matice hrvatske 31, 21000 Split, Croatia
jarneric@efst.hr, elza@efst.hr, spivac@efst.hr

Abstract: This paper establishes technical analysis of stock prices based on average trading prices. Analysis procedure begins with defining average prices on daily basis which are involved in stock market investment decisions for the first time in financial theory as well as in practice. Namely, all theoretical statements are confirmed by movements of Podravka stocks, as component of CROBEX index on Zagreb Stock Exchange. Using exponential smoothing methodology difference between short-term and long-term investment strategy is defined according to bull and bear signals.

Keywords: average trading prices, technical analysis, exponential weighted moving average method, rolling standard deviation, Bollinger's range, bull and bear signals

1. INTRODUCTION

The approaches used to analyze stocks and make investment decisions are divided into two categories: fundamental analysis and technical analysis. Fundamental analysis involves analyzing the characteristics of a company in order to estimate its "value". Technical analysis takes a completely different approach; it doesn't care about the "value" of a company. Technicians, sometimes called chartists, are only interested in the price movements in the market.

Technical analysis studies supply and demand in a market in an attempt to determine what direction, or trend, will continue in the future. Technical analysis is a method of evaluating stocks by analyzing the statistics of the past prices movements and volume. Therefore, it uses charts and other tools such as indicators and oscillators to identify patterns that can suggest future movements. Technical analysis relays on three basic assumptions:

- at any given time, a stock's price reflects everything that has or could affect the company - including fundamental factors;
- the repetitive nature of price movements is attributed to market psychology, i.e. market participants tend to provide a consistent reaction to similar market situations. It means that history tends to repeat itself;
- price movements are believed to follow trends.

2. PRICE MOVEMENTS IN TRENDS

In technical analysis it has been shown that after a trend of price movements has been established, the future price movement is more likely to be in the same direction. Most technical trading strategies are based on this assumption.

Empirical researches discover two types of trend distinguishing according to:

- time structure and
- general direction.

According to time structure there are long-term trends, intermediate trends or short-term trends. These are connected with investment strategies. Namely, there is a significant difference between an investor and a trader. It means that an investor expects profit only in long-term period, while traders prefer to profit in short-term period. So, it can be defined that long-term investment strategy is associated within time frame of 50 trading days; intermediate strategy is adapted with 20 trading days, while short-term investment strategy is associated within 10 trading days. Furthermore, these ranges are suggested by John Bollinger. According to general direction prices could trend up, trend down, or trend sideways. In financial literature synonym for uptrend market is bull market. A bull market tends to be associated with increasing investor confidence, motivating investors to buy in anticipation of further capital earnings. Technical term for downtrend market is bear market. A bear market tends to be accompanied by widespread pessimism. Investors anticipating further losses are motivated to sell.

3. SMOOTHING TECHNIQUES AND ROLLING ESTIMATES

Smoothing techniques are very often used in financial literature in general. In this contest it will be used primarily as an indicator of "bullish" or "bearish" signs in the stock market. Precisely speaking, in this paper smoothing techniques will be used to decline stochastic variations. The simplest method of smoothing time series is simple moving average (SMA) method[1], which is given by:

$$SMA_t(k) = \frac{1}{k}\left(P_t + P_{t-1} + P_{t-2} + ... + P_{t-k+1}\right), \tag{1}$$

where *k* is number of previous periods for which prices are observed.

In other words, forecast value for one period ahead is the simple average of current stock price and previous $t-(k-1)$ prices. Choice of period *k* depends on the particular purpose of the research. Namely, it has to be noticed that the larger the choice of *k* , the smoother the series will be. The main disadvantage of SMA is all observations are equally weighted. To mitigate the effects of extreme observations on moving average estimates can be weighted differently. Therefore, a common procedure that puts more weight on the most recent observations is based on exponentially declining weights and the resulting weighted moving average is called exponential weighted moving average (EWMA). According to exponential smoothing method forecast values could be calculated recursively:

$$\hat{P}_t = (1-\lambda)\cdot P_t + \lambda\cdot\hat{P}_{t-1} \tag{2}$$

where $\hat{P}_t$ is present period forecast and $\hat{P}_{t-1}$ is previous period forecast.

By continuous substitution equation in (2) becomes:

$$\begin{aligned}\hat{P}_t &= (1-\lambda)\cdot P_t + \lambda(1-\lambda)P_{t-1} + \lambda^2(1-\lambda)P_{t-2} + ...\\ &+ \lambda^i(1-\lambda)P_{t-i} + ... \qquad\qquad i = 0,1,2,...,k-1\end{aligned}. \tag{3}$$

In equation (3) parameter lambda $0<\lambda<1$ is called smoothing constant. When k converges to infinity, relation (3) can be noticed as:

$$\hat{P}_t(\lambda) = (1-\lambda)\sum_{i=0}^{\infty}\lambda^i P_{t-i}\ . \tag{4}$$

From relation (4) follows that $w_i \to 0$ when $\lambda^i \to 0$, according to:

$$\hat{P}_t(k) = \sum_{i=0}^{k-1} w_i P_{t-i} \quad , \quad w_i = \frac{\lambda^{i-1}}{\sum_{i=0}^{k-1}\lambda^{i-1}}\ . \tag{5}$$

From equations (2), (3) and (4) it follows that the closer lambda is to unity the more weight is put on the previous period's estimate relative to the current period's observation. Therefore weights are decreasing in exponential manner. However, weights can decrease slowly or faster. Classical approach of forecasting time series suggests lambda with minimal sum of squared errors. Because the interest of this paper is to describe pattern of a time

[1] Enders, W., *Applied Econometric Time Series,* Second Edition, Alabama: Wiley, 2004., p 48.

series, this suggestion will not be strongly considered. Namely, for the smoothing purpose, parameter lambda can be estimated according to chosen period *k*. This is in a close connection with time structures of investment strategies.

That's why parameter lambda can be estimated in following way:

$$\hat{\lambda} = \frac{k-1}{k+1}. \tag{6}$$

According to formula (6) lambda is a real number from interval $\langle 0, 1 \rangle$. The largest is period *k* the closer lambda is to unity. Therefore, weights are decreasing slowly and time series is smoother.

4. BOLLINGER BANDS

Bollinger Bands were created by John Bollinger[2]. Bollinger Bands are plotted at above and below a moving average (EWMA), where the standard deviation is a measure of volatility. In this paper the rolling standard deviation of relevant prices is used:

$$SD_t = \hat{\sigma}_t \sqrt{\frac{k}{k-1}}, \quad t = k, \ldots, n, \tag{7}$$

where *k* is number of periods within the rolling standard deviations are computed. In equation (7) factor $\sqrt{\frac{k}{k-1}}$ ensures unbiased estimation.

During periods of extreme price changes (i.e., high volatility), the bands indicate to divergence. During periods of stagnant pricing (i.e., low volatility), the bands narrow to contain prices. The longer prices remain within the narrow bands the more likely a price breakout. They are one of the most powerful concepts available to the technically based investor, but they do not, as is commonly believed, give absolute buy and sell signals. What they do is the answer to the question of whether prices are high or low on a relative basis. Using this information, an investor can make buy and sell decisions, confirming price action. Closing prices are most often used to compute Bollinger Bands, while other variations can also be used. For example the typical price (TP) is:

$$TP = \frac{high + low + close}{3}. \tag{8}$$

The weighted price (WP) is defined:

$$WP = \frac{high + low + 2 \cdot close}{4}. \tag{9}$$

In this paper for the first time we suggest so called the *real weighted price* (RWP):

$$RWP_t = \frac{\sum_{i=1}^{m} p_i q_i}{\sum_{i=1}^{m} q_i} = \frac{turnover}{volume}, \quad \forall t, \tag{10}$$

where *m* is number of transactions in a current trading day *t*, while p_i is executive price of *i-th* transaction, and q_i is trading quantity. According to (8) and (9) equations the advantage

[2] Colby, R.W., *The Encyclopedia of Technical Indicators*, McGraw-Hill, New York, 2003., p 188.

of the real weighted price is in the fact that executive prices are weighted by trading quantity, i.e. executive price which is traded more has greater weight and *vice versa*.

Bollinger recommends using a 20-day simple moving average for the centre band and 2 standard deviations for the outer bands. The length of the moving average and number of deviations can be adjusted. In this paper the lengths of the moving average are 10 and 50 respectively and numbers of deviations are 1,5 and 2,5 according to comparison of the short-term and long-term investment strategies.

It can be concluded that Bollinger Bands serve two primary functions:

- to identify periods of high and low volatility, and
- to identify periods when prices are at extreme levels.

Even so, a security can become overbought or oversold for an extended period of time. Knowing whether or not prices are high or low on a relative basis can enhance the interpretation of other indicators and oscillators. Therefore, the *relative strength index* (RSI)[3] is used. The RSI is a oscillator showing price strength by comparing upward and downward movements.

$$RSI = 100 - \left(\frac{100}{1 + \frac{U}{D}} \right), \tag{11}$$

where U is an absolute value of the moving average of upward executive price change, and D is an absolute value of the moving average of downward executive price change. The RSI is oscillator that ranges between 0 and 100. In situation when RSI reaches the 70% a security is considered to be overbought, or oversold at the level of 30%. Generally, if the RSI rises above 30% it is considered bullish for the underlying stock. Conversely, if the RSI falls below 70%, it is a bearish signal. The centreline for RSI is 50%. Levels of 80% and 20% are also used.

5. TEHNICAL ANALYSIS IN CROATIA

The complete procedure of presented technical analysis is established using observations of Podravka stocks as the most frequently traded stock from CROBEX index at Zagreb Stock Exchange.

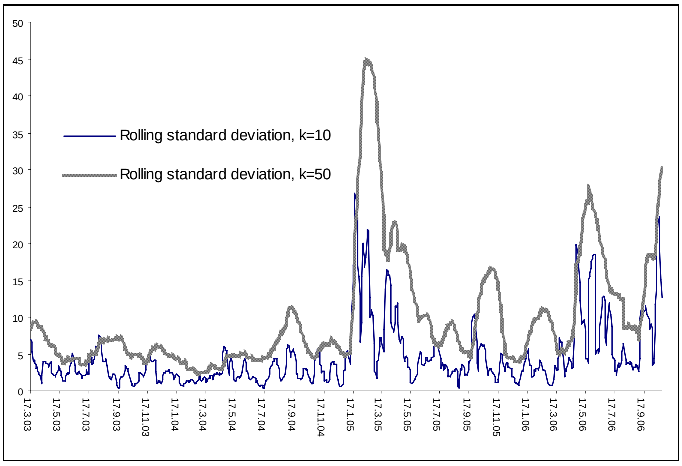
Figure 1 shows by one line rolling standard deviation movements from the point of investor's view (long-term trading periods) and the other line are the movements from the point of trader's view (short-term trading periods).

Representative for long term periods is rolling standard deviation for time frame of 50 trading days and representative for short term periods is rolling standard deviation of 10 trading days.

Estimated according to real weighted prices short term Bollinger Bands are presented in Figure 2.

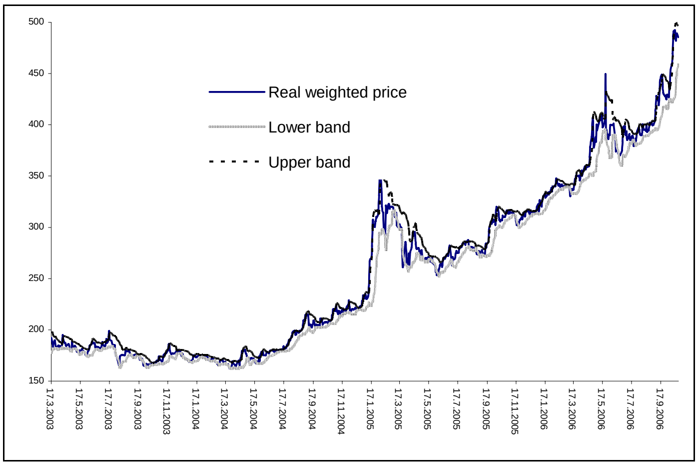
[3] Colby, R.W., *The Encyclopedia of Technical Indicators*, McGraw-Hill, New York, 2003., p 610.

Figure 1. Rolling standard deviation estimates for short-term and long-term trading periods



Source: According to data on www.zse.hr

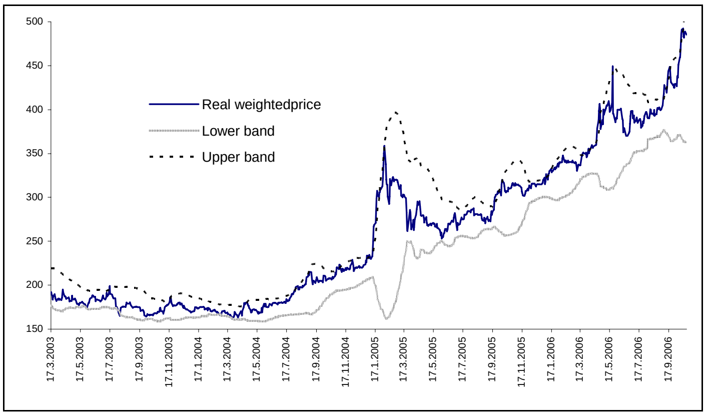
Figure 2. Short-term Bollinger Bands estimates according to real weighted prices



Source: According to data on www.zse.hr

Long-term Bollinger Bands estimated according to real weighted prices are illustrated in Figure 3.

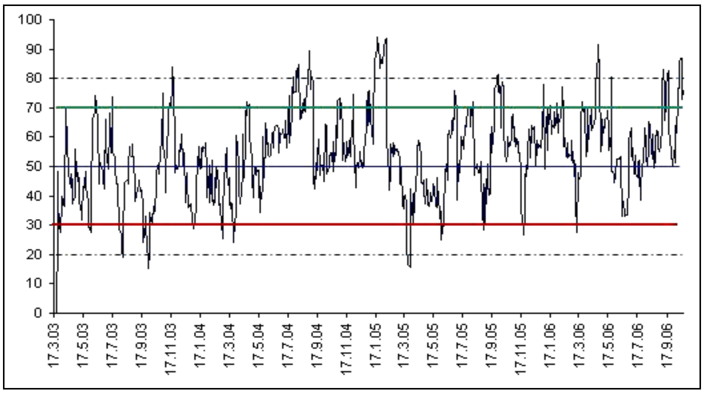
Figure 3. Long-term Boolinger Bands estimates according to real weighted prices



Source: According to data on www.zse.hr

Apart from various indicators technical analysis requires measurements using oscillators. An example of oscillator relative strength index is used in this analysis.

Figure 4. Relative strength index oscillator



Source: According to data on www.zse.hr

Financial theory and practice recommended a smoothing period of 14. This is by reckoning of EWMA smoothing i.e. $1-\lambda=1/14$ or $k=27$. This choice of k in Figure 4 is managed by the fact that it is reasonable to chose intermediate time period.

6. CONCLUSION REMARKS

One of the aims of this research is to discover reliable signals to buy or to sell on the stock market using technical analysis. Using exponential smoothing methodology difference between short-term and long-term investment strategy is defined according to bull and bear signals. According to concrete data about Podravka stocks at Zagreb Stock Exchange from March 17th 2003 to September 17th 2006 (903 trading days) the final technical analysis results in Table 1 give precise suggestions to buy or to sell on the Stock market from the point of view of short-term and long-term strategy.

Table 1. Investment decisions according to performed technical analysis

Short-term strategy	Long-term strategy			Trading days
	buy signal	no signal	sell signal	
Buy signal	3	85	-	88
no signal	-	622	21	643
sell signal	-	120	52	172
Trading days	3	827	73	903

Source: According to data on www.zse.hr

At the end the final decision to buy or to sell depends on the concrete investor (trader) preference to risk more in order to earn more. Apart from technical analysis results for such a decision additional capital market information will be used.

REFERENCES

1.Colby, R.W., *The Encyclopedia of Technical Indicators*, McGraw-Hill, New York, 2003.

2.Enders, W., *Applied Econometric Time Series,* Second Edition, Alabama: Wiley, 2004.

3.The Zagreb Stock Exchange: http:// www.zse.hr

TESTING FOR GRANGER CAUSALITY BETWEEN ECONOMIC SENTIMENT INDICATOR AND GROSS DOMESTIC PRODUCT FOR THE CROATIAN ECONOMY

Vlasta Bahovec
Mirjana Čižmešija
Nataša Kurnoga Živadinović
University of Zagreb - Faculty of Economics & Business
Trg J. F. Kennedya 6
10000 Zagreb, Croatia
bahovec@efzg.hr
mcizmesija@efzg.hr
nkurnoga@efzg.hr

Abstract: Granger causality test is useful to determining whether one time series is useful in forecasting another. In this paper we examine wheter data from Business Tendency Surveys are useful for forecasting some selected referent macro economy variables in the short run. We compare the Economic Sentiment Indicator – ESI as a composite indicator for the whole Business survay with the GDP for Croatia. It is evident that the ESI in Croatia correctly predict changes in a national economy (expresed with GDP growht) two quarters in advanse.

Keywords: Granger Causality**,** Augmented Dickey-Fuller test, ESI – Economic Sentiment Indicator, GDP – Gross Domestic Product

Introduction

A Business Survey is a method of gathering information about the economic agents´ perception of their environment. This qualitative survey is based on observing, following, explaining and forecasting changes in the business climate [7]. Qualitative data on businessmen's perceptions of their economic environment are translated into quantitative indicators. These are: Industrial Confidence Indicator (ICI), Construction Confidence Indicator (BCI), Retail Trade Confidence Indicator (RTCI), Consumer Confidence Indicator (CCI) and Services Confidence Indicator (SCI). The Economic Sentiment Indicator (ESI) is a composite indicator, deriving from Business and Consumer Surveys (BCS)[1] and includes all variables components of composite indicators which are mentioned above [4], and it should be compared with the referent series which reflects movements in the economy as a whole. This referent series for ESI is GDP. This series can be used to test the explanatory performance of the ESI.

There are many research results which have shown that there are causalities between real economical movements and business survey results [2], [3], [6]. The Business Surveys provide information on a wide range of variables that are useful in monitoring cyclical developments. Based on this information movements in national economic activity can be predicted. In some countries, and according to conducted researches, it was displayed, that even when the results of the Business Survey have the strongest correlation with the real economic movements in the same period of time (t-0), i.e. in the same quarter or the same month, due to the nature of their origination and the publication time, which is always prior to the result publication of the Official Central Statistic; these results are a very good base for a short term business forecast. The aim of this paper is to test a potential predictability

[1] ESI for Croatia has got variables - components of ICI, RTCI and BCI because the time series of the Consumer Survey are to short and the Services Survey is not yet being conducted in Croatia.

power of ESI for the GDP. For this purpose we used some qualitative methods in comparing two time series, regression and correlation analysis and Granger causality test [5].

Granger causality tests

The relationship between two variables can be captured by a VAR model [1], [4]. It is possible that the variable Y causes the variable X, that the variable X causes Y, there is a bi-directional feedback (causality among the variables), and the two variables are independent. Granger [5] developed a test that defined causality: a variable Y is said to Granger – cause X, if X can be predicted with greater accuracy by using past values of the Y variable rather than not using such past values, all other terms remaining unchanged. The Granger causality test for the two variables Y and X involves the estimation of the following pair of regressions [1]:

$$y_t = a_1 + \sum_{i=1}^{n} \beta_i x_{t-i} + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_{1t} \qquad (1)$$

and for the ESI and GDP variables model is:

$$GDP_t = a_1 + \sum_{i=1}^{n} \beta_i ESI_{t-i} + \sum_{j=1}^{m} \gamma_j GDP_{t-j} + e_{1t}\,. \qquad (2)$$

$$x_t = a_2 + \sum_{i=1}^{n} \theta_i x_{t-i} + \sum_{j=1}^{m} \delta_j y_{t-j} + e_{2t} \qquad (3)$$

and for the ESI and GDP variables model is:

$$ESI_t = a_2 + \sum_{i=1}^{n} \theta_i ESI_{t-i} + \sum_{j=1}^{m} \delta_j GDP_{t-j} + e_{2t} \qquad (4)$$

Where it is assumed that both ε_{yt} and ε_{xt} are uncorrelated white-noise error terms. In this model it can be the following different cases[1]:

- The lagged *x* terms in (1) may be statistically different from zero as a group, and the lagged *y* terms in (3) not statistically different from zero. It is the case that x_t causes y_t.
- The lagged *y* terms in (3) may be statistically different from zero as a group, and the lagged *x* terms in (1) not statistically different from zero. It is the case that y_t causes x_t
- Both sets of *x* and *y* terms are statistically different from zero in (1) and in (3) so that we have bi-directional causality.
- Both sets of *x* and *y* terms are not statistically different from zero in (1) and in (3) so that x_t is independent of y_t.

The procedure in order to conduct the Granger causality test has the following steps [1]:

1) Regress y_t on lagged *y* terms as in the model:

$y_t = a_1 + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_{1t}$ or $GDP_t = a_1 + \sum_{j=1}^{m} \gamma_j GDP_{t-j} + e_{1t}$ and compute restricted residual sum of squares, RSS_R.

2) Regress y_t on lagged *y* terms plus lagged *x* terms as in the model:

$$y_t = a_1 + \sum_{i=1}^{n} \beta_i x_{t-i} + \sum_{j=1}^{m} \gamma_j y_{t-j} + e_{1t} \text{ or } GDP_t = a_1 + \sum_{i=1}^{n} \beta_i ESI_{t-i} + \sum_{j=1}^{m} \gamma_j GDP_{t-j} + e_{1t} \text{ and}$$

compute unrestricted residual sum of squares RSS_U.

3) The null hypothesis implies that x_t does not cause y_t and alternative hypothesis imply that x_t does cause y_t:

$$H_0 \ldots\ldots\ldots \sum_{i=1}^{n} \beta_i = 0$$

$$H_1 \ldots\ldots\ldots \sum_{i=1}^{n} \beta_i \neq 0$$

4) The Granger causality statistic is the F-statistic:

$$F = \frac{\frac{(RSS_R - RSS_U)}{m}}{\frac{RSS_U}{(n-k)}} \quad (5)$$

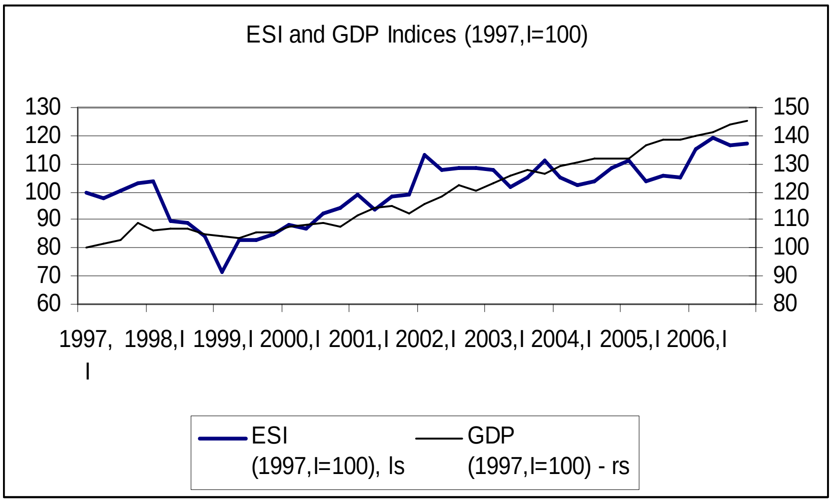
which follows $F_{m,n-k}$ distribution, where *k is n+m+1.*

Empirical results

The data that we are going to analyze were two time series of quarterly frequencies of ESI and GDP for Croatia under the period from the first quarter of 1997 to the fourth quarter of 2006.[2]

Based on the graphic illustration of the mentioned series (Figure 1) and the coefficients of determination (R-squared, for the different lags in quarters, table 1), a connection between ESI and GDP can be noted. The characteristic of ESI preceding GDP can be determined through the Granger causality test.

Figure 1



[2] ESI is published in periodical Privredni vjesnik. GDP published Croatian Central Bureau of Statistics.

Table 1

Dependent variable	Independent Variable	R-squared	t-Statistic (independent variable)	Probability (independent variable)[3]
GDP	ESI (- 0)	0.575105	7.171735	0.0000
	ESI (-1)	0.581464	7.213982	0.0000
	ESI (-2)	0.581081	7.066500	0.0000
	ESI (-3)	0.516615	6.116054	0.0000
	ESI (-4)	0.451526	5.290573	0.0000
	ESI (-6)	0.430967	4.922985	0.0000
	ESI (-8)	0.353105	4.046651	0.0003

The assumption for conducting Granger causality test is that two variables (in our case ESI and GDP) are stationary. The Augamented Dickey – Fuller test for unit roots has been calculated (The program support EViews is applied, [8]). For all of the series the null hypothesis of no stationarity can be rejected at 5% significance level. Than we made simple differences of ESI (D(ESI)) and GDP (D(GDP)). This time series are stationary (see Figure 2). The results of Augmented Dickey – Fuller test for unit roots are in table 2.

Figure 2
Simple differences of ESI and of GDP

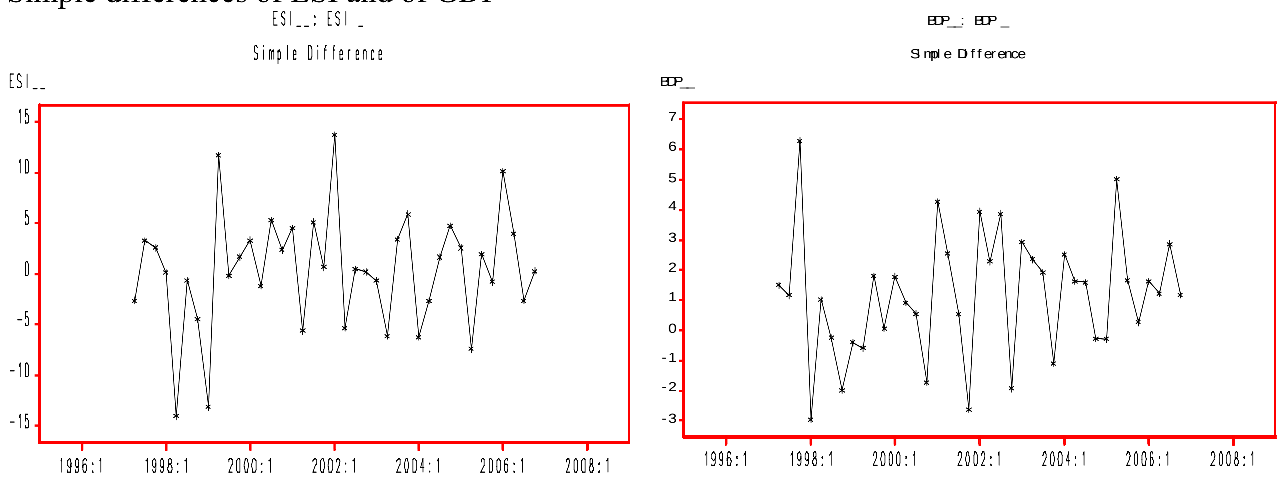


Table 2

Augmented Dickey-Fuller test statistic	t-Statistic	Prob.
Null Hypothesis: ESI has a unit root	-1.278719	0.6298
Null Hypothesis: D(ESI) has a unit root	-7.149265	0.0000
Null Hypothesis: GDP has a unit root	-2.185373	0.4830
Null Hypothesis: D(GDP) has a unit root	-6.662862	0.0000

The results of applied Granger causality test for ESI and GDP are presented in the next tables.[4]

[3] Independent variable is statistically significant at the level of significance of 5% for all regressions.
[4] For the monthly data, the reasonable lag terms can be range from 1 to 12 or 24; for the quarterly data the lag terms can be range from 1 to 4, 8, 12, etc. ; for the annual data it can be less.

Table 3
Granger Causality Tests

Direction of causality	Number of lags	F-value	Probability
D(ESI) does not Granger Cause D(GDP)	2	4.53340	*0.01847*
	3	6.02677	*0.00255*
	4	5.31018	*0.00291*
	6	3.46820	*0.01631*
	8	1.33005	0.30599

Table 4
Granger Causality Tests

Direction of causality	Number of lags	F-value	Probability
D(GDP) does not Granger Cause D(ESI)	2	0.02246	0.97780
	3	0.05547	0.98247
	4	0.08964	0.98489
	6	1.89151	0.13213
	8	1.52338	0.23452

For these results, for lags 2, 3 or 4 quarter we can not accept the null hypothesis that ESI does not Granger GDP (in terms of simple differences). It means that ESI does Granger cause GDP.[5] For the same lags, we can accept the null hypothesis that GDP does not Granger Cause ESI (at the level of significance of 5%). It appears that Granger causality runs one way from ESI to GDP and not the other way.

Conclusion

The relationship between ESI and GDP is very important in interpreting the characteristic of ESI preceding GDP. Based on the changes of ESI for Croatian Business Survey, we are able to predict the movement of the entire national economic activity two or three quarters in advance[6]. The graphic illustration, the coefficients of determination and the results of the Granger causality test confirm the mentioned ESI quality of Croatia. Its value as a prognostic indicator is even greater, if we consider the fact that the results of the Business Survey, which were earlier accessible to users than the data of the Central Bureau of Statistics. The results of the application of the Granger causality test also show that the ESI determines the level and the movement of GDP, whereby there is no invert connection, i.e. there is no Granger causality, according to which GDP determines ESI. The Granger causality runs one-way from ESI to GDP and not the other way around. It is expected that the ESI quality of Croatia, from its prognostic power point of view, will reinforce even more with the introduction of new component variables in ESI, i.e. with starting Service Surveys, which Croatia does not conduct by now, as well as with the inclusion of the results of the Customer Survey, whose time series will soon be long enough, to be included as components in ESI.

References

[1] Asteriou, D. (2006). Applied Econometrics – A modern Approach using EViews and Microfit. Palgrave Macmillan. New York

[5] For lag 1 ESI does not cause GDP (at the level of significance 5%).
[6] With the lead of two quarter, in around 70% of the cases, ESI predicts correctly the changes in the direction of GDP.

[2] Bukovšak, M. (2006). Anketa pouzdanja potrošača u Hrvatskoj. Hrvatska narodna banka – Istraživanja. Zagreb

[3] Gayer, C. (2004). Forecast Evaluation of European Commission Survey Indicators, 27th CIRET Conference. Warsaw, September 2004.

[4] Gujarati, D. N. (2003). Basic Econometrics, 4th Edition. McGraw-Hill. New York

[5] Granger, C. J. (1969.). Investigating Causal Relationships by Econometrics Models and Cross Spectral Methods, Econometrica, Vol. 37, pp. 425-435.

[6] Holickova, E. (2005). Business Survey and Short – Term Projection. OECD workshop on International Development of Business and Consumer Tendency Surveys. Brussels, 14-15. November 2005.

[7] *The joint harmonised EU programme of business and consumer surveys*, User guide (updated 07/06/2007). European Economy. European Commission, Directorate-General for Economic and financial affairs.
http://ec.europa.eu/economy_finance/indicators/business_consumer_surveys/userguide_en.pdf

[8] Program support EViews and SAS

STUDENT SATISFACTION WITH QUANTITATIVE SUBJECTS

Majda Bastič
Faculty of Economics and Business
majda.bastic@uni-mb.si

Abstract: The student satisfaction with knowledge and competences received by a subject influences both a status of the subject in a study programme as well as a status of the higher education institution in the competitive educational environment. Therefore, the most important determinants of student satisfaction were investigated on a sample of 239 students who estimated their satisfaction with two subjects, i.e. operations research and operations management. Perceived performance, learner empowerment, and learning process were found as important determinants of student satisfaction.

Keywords: Student Satisfaction, Universities, Quantitative Subjects, Slovenia

1. Introduction

Slovenian university institutions have faced an increasingly competitive environment and numerous challenges associated with the Bologna Process (BP). One of the most important goals of the BP is that of making Europe "the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and social cohesion" (Berlin Communique, 2003). Higher Education (HE) plays a key role in furthering the successful transition to this new economic model by preparing graduates to be capable to successfully face these new challenges. The trend toward a "learning society" has been widely accepted and consolidated for some time. Reflecting on different aspects which characterise this trend, the relevance of focusing on competences becomes apparent.

The great importance of competences as desired learning outcomes was one of the reasons that Tuning project was developed (Final Report, 2001). Thirty generic competences were selected from three categories: instrumental, interpersonal and systemic. Respondents from all Europe were asked to rate both the importance and the level of achievement in each competence by educational programme, and also to rank the five most important competences. One of the most striking conclusions of this study is the remarkable correlation (Spearman correlation is 0.973, $p < 0.01$) between the ratings given by employers and those given by graduates all over Europe. In their opinion, the most important competences to be developed are: capacity for analysis and synthesis, capacity to learn, problem solving, capacity for applying knowledge in practice, capacity to adapt to new situations, concern for quality, information management skills, ability to work autonomously, and teamwork. At the other end of the scale we find: understanding the cultures and customs of other countries, appreciation of diversity and multiculturality, ability to work in an international context, leadership, research skills, project design and management, and knowledge of a second language.

Numerous studies have shown that the long-term success of a firm is closely related to its ability to adapt to customer needs and changing preferences (Li et al., 2006). Therefore, consumer satisfaction has long been recognized in marketing thought and practice as a central concept as well as an important goal of all business activities (Yi, 1990). Far few studies have been performed to investigate student satisfaction in HE. Because these few studies were focused more on the university-level satisfaction, the relationship between the competences developed in learning process and the student satisfaction was not taken into account (Elliot and Healy, 2001). The objective of this study is to find the most important determinants which influence the subject-level satisfaction. These determinants will be found with data referring to two quantitative subjects, i.e. operations research (OR) and operations management (OM).

2. Student satisfaction model

Student satisfaction model applied in this study was partly based on the theoretical framework of the customer satisfaction model where a perceived performance and value of product or service are taken as antecedents of customer satisfaction. It has a positive impact on customer loyalty (Chan et al., 2003). Considering the role of service provider (learner) and service process (learning process) in achieving service quality it was assumed that student satisfaction depends on perceived performance, learner empowerment and learning process. Perceived value expressing the perceived level of product quality relative to its price paid by customer was not included in the model because the full-time students do not pay a tuition fee.

According to Yi (1990) product-level consumer **satisfaction** can be generally defined as the consumer's response to the evaluation of the perceived discrepancy between some comparisons (e.g. expectations) and the perceived performance of the product. Satisfaction should be measured with its antecedents and consequences in an equation system to estimate their relationships with their indicators as well as with each other. Although consensus has not been reached on how to measure consumer satisfaction, various studies revealed three important aspects: i) general or overall satisfaction; ii) confirmation of expectation, i.e. the degree to which performance exceeds or falls short of expectations; and iii) comparison to ideal, i.e. the performance relative to the consumer's hypothetical ideal product.

Elliot and Healy (2001) defined **student satisfaction** as a short-term attitude that results from the evaluation of their experience with the education service received. We applied three indicators to measure the subject-level satisfaction. These were first, the overall estimate of satisfaction with the subject measured on the five point scale from "very low estimate" to "very high estimate"; second, the probability that the subject will be recommended to other students was measured on the five point scale from "certainly not" to "certainly" and third, the extent to which the subject fulfilled their expectations measured on the five point scale from "not fulfilled" to "exceed the expectations".

Perceived performance is usually referred to perceived quality which stands for the consumer's global judgement of the overall excellence of a product (Anderson et al., 1994). In the literature, perceived performance is viewed as one of the antecedents of consumer satisfaction. Two primary components were revealed as components of perceived performance. They are customization or fitness for use, which relates to whether the product can meet various consumer needs, and reliability, which relates to whether the product can be free from deficiencies for a long period of time.

We assumed that the perceived performance of a subject consists of two components, i.e. knowledge and competences. They relate to whether the subject can meet the students' needs associated with their employability. 3 indicators regarded knowledge and 2 indicators associated with competences developed through learning process were included into a questionnaire. The indicators measured obtained knowledge are the extent of theoretical knowledge, the extent of knowledge usable in practice, and the extent of knowledge which is appreciated by the employers. The next two indicators refer to acquired competences. They are a student's capability to transform economic or business problem into an appropriate mathematical model, and a capability to apply knowledge obtained in making better and more reliable decisions.

The **learner empowerment** relates to the quality of the learner's interaction with both the use and the learning of subject. An empowered learner would thus be able to analyse the students' strengths and weaknesses with respect to specific situations or problems, to evaluate what they need to learn in order to meet their objectives, and to make informed decisions about how to go about achieving these goals. In other words, an empowered learner is one who has acquired transferable learning skills which go beyond the confines of a given level of competence in a subject (Tudor, 2005).

We used 6 indicators to measure the learner empowerment. They referred to how difficult was to follow the learner's lectures, did the learner illustrate the theory with the cases from practice, did the learner motivate the students for co-operation in learning process, did the learner motivate the students for deep study, did the learner respond to the students' questions, and how much did the learner help students when they faced up to study problems.

The **learning process** can be estimated with regard to knowledge and competences obtained by lectures, exercises, personal contacts with the learner, and e-learning. It was expected that there is a positive relationship between learning process and student satisfaction.

All indicators referring to the perceived performance, the learner empowerment, and the learning process were measured on a five point scale, where 1 means much less than a student expected, and 5 means much more than a student expected.

Considering the objective of the research the following hypotheses were tested.

H1. The student satisfaction depends on the perceived performance, the learner empowerment, and the learning process.

H2. The perceived performance is one of the most important factors influencing the student satisfaction with the subject.

H3. The student satisfaction depends mainly on knowledge usable in practice followed by capability to apply knowledge in making better and more reliable decisions, and capability to transform the economic or business problem into an appropriate mathematical model.

H4. The part-time students are satisfied with quantitative subjects more than the full-time students.

H5. The part-time students perceive higher extent of all kinds of knowledge and competences obtained by quantitative subjects than the full-time students.

3. Data collection and analyses

The survey of student satisfaction was centred on two subjects, i.e. OR and OM, on two faculties of University of Maribor, i.e. the Faculty of Economics and Business and the Faculty of Logistics, and their full- and part-time students.

Between March and May 2007, the students of the fourth semester were asked to assess the subjects (OR and OM) in respect of indicators listed in a questionnaire. A pre-test with 10 respondents was used to check that the text was drafted clearly. Once the actual main study has been carried out, a total of 239 usable questionnaires were available for a detailed analysis. The subject OR was estimated by 138 students, and 101 students estimated the subject OM. The sample consists of 169 full-time students and 70 part-time students.

In order to test hypothesis H1 the constructs the perceived performance, the learner empowerment, the learning process, and the student satisfaction were built and their reliability were assessed. Procedures such as Cronbach's alpha, the item to total correlation, and exploratory and confirmatory factor analysis can be applied to test their reliability. As a rule, the performance target of Nunnally (1978) was used as a guide and this requires an alpha value of 0.7. The calculation of Cronbach's alpha is followed by an explorative factor analysis, which provides an indication in respect of discriminant and convergence validity (Hair et al., 1998). It is suggested that those measurement items that have a low factor loading (< 0.4) are eliminated. For this reason, the item 'exercise' which was used to measure learning process was eliminated in our case. In Table 1, mean, standard deviation, Cronbach's alpha and the percent of variance explained are given.

Table 1. Description of factors

Factor	Mean	Standard Deviation	Variance explained	Cronbach's α
Learner empowerment	3.384	0.996	54.858	0.838
Performance	3.151	1.026	67.023	0.879
Learning process	3.093	1.173	58.824	0.650
Satisfaction	3.275	1.042	77.611	0.854

In order to investigate which factor constitutes the best predictor(s) of the student satisfaction the regression analysis was carried out. Factors the learner empowerment, the perceived performance, and the learning process were used as independent variables whereas the factor student satisfaction was chosen as a dependent variable. With ordinary least square method, the estimated standardized regression coefficients (b_j) and multiple coefficient of determination (R^2) for regression equation were estimated. All three variables attained statistical significance in the equation. A total of 56.2 percent of variance of the student satisfaction was explained. Thus, the hypothesis H1 is confirmed.

The perceived performance was found as the most important factor influencing the student satisfaction (b_1=0.352), followed by the learner empowerment (b_2=0.340), and the learning process (b_3=0.184). Thus, the hypothesis H2 is also confirmed.

To test hypothesis H3 the relationship between the factor student satisfaction and the items measured the perceived performance was analyzed. Pearson correlation coefficients were computed to test for these relationships. All correlations were positive and significant at the 0.01 level. They showed that all kinds of knowledge and competences obtained by OR and OM had important and positive impact on the student satisfaction. However, the highest correlation coefficient belonged to the knowledge usable in practice (r=0.572). It is followed by the capability to apply the knowledge of OR and OM in making better and more reliable decisions (r=0.549), and the capability to transform an economic or business problem into an appropriate mathematical model (r=0.543). The correlation between the student satisfaction and the knowledge required by the employers took the last place (r=0.519). Therefore, the hypothesis H3 is also confirmed.

Independent t-tests were conducted to see whether the working experience and better understanding of knowledge needed in the organisations for their further growth have any relationship to the perceived performance of subjects. The respondents were classified into two groups because it was assumed that the part-time students with working experience better know the knowledge needed in their organisations. 169 full-time students were classified in one group; 70 part-time students were in the other group. The student satisfaction was measured with three items scale including subject estimate, the probability that the subject will be recommended to other students, and the extent to which the subject fulfilled the student expectation. The part-time students estimated all three items higher than the full-time students (see Table 2). All three mean differences are significant, which confirms the hypothesis H4. These results allow us the interpretation that working experience and better understanding of knowledge needs in organizations help the part-time students to better understand optimization methods and especially possibilities for their use in practice, which increase their satisfaction with these two subjects.

The mean values referring to the extent of knowledge obtained and capabilities developed through learning process of OR and OM were compared by t-tests. Again, the students were classified into two groups. The results indicate that all mean differences were statistically significant (see Table 3). Remarkable differences belonged to the capability to transform economic or business problem into an appropriate mathematical model, and the extent of the theoretical knowledge obtained. Taking into account these results, the hypothesis H5 is also confirmed.

Table 2. Analysis of mean differences in satisfaction between two groups of students

Variable	Group	Mean	Std. Deviation	Sig. 1-tailed
Estimate	Full-time	3.31	0.972	0.000
	Part-time	3.89	0.843	
Recommendation	Full-time	2.92	1.147	0.000
	Part-time	3.91	0.981	
Fulfilment of expectation	Full-time	2.94	0.904	0.000
	Part-time	3.65	0.860	

Table 3. Analysis of mean differences in knowledge and competences between two groups of students

Variable	Group	Mean	Std. Deviation	Sig. 1-tailed
Theoretical knowledge	Full-time	3.24	0.934	0.018
	Part-time	3.51	0.913	
Knowledge usable in practice	Full-time	2.94	1.073	0.023
	Part-time	3.25	1.098	
Capability to transform problem into an appropriate mathematical model	Full-time	3.27	1.026	0.007
	Part-time	3.61	0.906	
Capability to apply knowledge for better and more reliable decisions	Full-time	2.94	1.067	0.047
	Part-time	3.20	1.145	
Knowledge required by employers	Full-time	2.92	0.963	0.043
	Part-time	3.16	1.081	

4. Conclusions

The results of the survey presented in this paper show that the student satisfaction with the subjects OR and OM depends mainly on the subject performance defined by knowledge obtained and competences developed as well as on the learner empowerment. The knowledge which can be used in practice and capability to apply knowledge in making better and more reliable decisions had the most important impact on the perceived performance of subjects and consequently on the student satisfaction. The part-time students assessed both subjects higher than the full-time students and perceived higher extent of knowledge and competences obtained by these two subjects. It is probably the consequence of working experience and every day life in the organizations which improve their understanding of market demands and the required knowledge and competences of employees to successfully meet the market requirements. We can not forget the support of the empowered learner in the learning process and their contribution to the perceived performance of subject. It is very important for them to have capability to analyse the students' strengths and weaknesses, to understand the students' needs and objectives and to find effective ways to meet their needs and achieve their goals.

The lack of working experience probably prevents the full-time students to see more possibilities for use of OR and OM knowledge in practice. These findings call for an improvement of the learning process with more cases where the way how to solve business problems will be presented. The co-operation with experts who will present the students relevant problems in practice and the ways how they were or should be solved could also improve the perceived performance of both subjects. The students' co-operation in research teams could be another way to improve their satisfaction.

The results of this study were obtained in Slovenia which is one of the post-transition countries. It will be interested to investigate the perceived performance of similar subjects in more developed countries to reveal whether the level of development and higher market demands influence the subject performance and the student satisfaction with quantitative subjects.

References

Anderson, E.W. and C. Fornell (2000). Foundation of the American customer satisfaction index. *Journal of Total Quality Management*, Vol. 11, No. 7: 869-82.

Berlin Communique 2003. *Communique of the Conference of Ministers responsible for Higher Education, Berlin 19 September 2003*, "Realising the European Higher Education Area", www.bologna-berlin2003.de/pdf/Communique1.pdf

Chan, L.K., Y.V. Hui, H.P. Lo, S.K. Tse, G.K.F. Tso and M.L. Wu (2003). Consumer satisfaction index: new practice and findings. *European Journal of Marketing*, Vol. 37 No. 5/6: 872-909.

Elliot, K.M. and M.A. Healy (2001). Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education*, Vol. 10 No. 4: 1-11.

Final Report of Tuning Educational Structures in Europe (2001-2002), Part One and Two. Available: http://odur.let.rug.nl./TuningProject/doc_tuning_plase1.asp

Hair, J.F., R.E. Anderson, R.L. Tatham and W.C. Black (1998). *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ.

Li, B., M.W. Riley, B. Lin and E. Qi (2006). A comparison study of customer satisfaction between the UPS and FedEx: an empirical study among university customers. *Industrial Management & Data Systems*, Vol. 106, No.2: 182-99.

Nunnaly, J. C. (1978.), P*sychometric Theory,* New York: McGraw-Hill Book Company.

Tudor, Ian (2005). The Challenge of the Bologna Process for Higher Education language teaching in Europe. ENLU website.

Yi, Y. (1990). A critical review of consumer satisfaction, in Zeithaml, V. A. (Ed.), *Review of Marketing,* American Marketing Association, Chicago, IL: 68-123.

CHI-SQUARE VERSUS PROPORTIONS TESTING - CASE STUDY ON TRADITION IN CROATIAN BRAND

Ivan Bodrožić
University of Split, Faculty of Theology
Zrinsko Frankopanska 19, 21000 Split, Croatia

Elza Jurun, Snježana Pivac
University of Split, Faculty of Economics
Matice hrvatske 31, 21000 Split, Croatia
elza@efst.hr, spivac@efst.hr

Abstract: By this paper authors try to establish the much more common procedure of proportions testing with the same conclusions in statistical sense versus more complex Chi-square of independence. Case study is very interested and useful in scientific sense by itself, not only as the practical example used to sustain mathematical-statistic analysis conclusions. In this quantitative research in social science more than 600 questionnaires had been analysed to perceive importance of tradition by purpose of involving it in modern Croatian brand.

Keywords: proportions testing, Chi-square of independence, expected count problem, survey research, tradition, Croatian brand

1. INTRODUCTION

The most common procedure of testing independence of variables with nominal measure in classical statistic sense is Chi-square. It is rather complex procedure and sometimes it does not lead to final conclusions, because of statistic-technical backsets[1]. There are two typical situations when statistic-technical backset appears: a great number of variations of nominal variable/s that on artificial way increases degrees of freedom; a great number of expected counts less than 5.

Both situations result with wrong, usually opposite conclusion about independence of variables with nominal measure. That is reason why a much more precise procedure based on proportions testing is established in this paper. Namely, authors have perceived that in such kind of testing valid results can be realized using hypothesis testing about difference of two proportions. During quantitative researches in social science authors have recognized it as the most appropriate approach to survey research analysis[2]. So, in this paper a survey research on tradition in Croatian brand is presented to verify those statements.

2. CASE STUDY

Zabiokovlje is a part of Split-Dalmatian County. It is situated in very picturesque nook of Croatian mainland. Till today it was pure, crag, agricultural and traffic isolated area. Nowadays a large part of Zabiokovlje is a big building site. New modern highway changes role and perspective of Zabiokovlje in Croatian economy. Namely, this area is natural corridor towards South-Adriatic coast in national setting and transport corridor for passengers and goods towards South-East Europe. Accordingly of its longtime segregation there are traditional customs, cultural heritage, traditional culinary, health food and

---

[1] Aron A., Aron E., Coups E., *Statistics for the Behavioural and Social Sciences,* 4th edition, Prentice Hall, Cambridge, 2007., (Chapter 4: Some Key Ingredients for Interferential Statistics.)

[2] Pivac S., Rozga A., *Statistika za sociološka istraživanja*, University of Split, Faculty of Philosophy Split, 2006. pp 9-71.

untouched nature. These are not only sociological or ethnic phenomenon but nowadays especially the great possibility to become a part of Croatian brand through ecological and rural tourism offers. Without wide explanation of all possibilities it is relevant to mention that more than ten years manifestation "Glumci u Zagvozdu" as a part of cultural heritage of Zabiokovlje attracts numerous domestic and foreign tourists during the whole summer time.

This research was initiated in order to discover possibilities of including Zabiokovlje in Croatian brand in sense of sustainable development. A survey has been carried out in this area and more than 600 questionnaires have been analysed. It has been found out that the interviewed (who were older than 18) form a sample which confirms to all the contemporary requests of statistics. So it is representative and random. The given answers which could at first be defined as easy-going, don't remain on the level of description of behaviour forms, but help us to reveal the essence of relations towards heritage as well as relations to modern social environment. Because of the length limitations imposed by such a paper it won't be possible to mention all the results and dimensions of this extensive survey. So, this paper is focused to analyse those parts of survey research that can be used also to confirm author's proposition that proportions testing gives answer on Chi-square of independence as well.

3. METHODOLOGY MENTIONS

3.1. Chi-square independence testing

Chi-square is common and the most applied procedure for independence testing. According to journal "Science", Chi-square is one of 20 the most important scientific finding of 20th century. It is frequency based statistic, doesn't assumed distribution form and belongs to nonparametric tests. Among numerous possibilities of applying the independence testing is one of the most popular. The essence of algorithm is statistical significant difference between empirical and theoretic frequencies. Sums of empirical frequencies and theoretical frequencies are identical and their arrangement in distribution leads to final conclusion of testing. Theoretical frequencies are calculated under assumption of null hypothesis, and if the difference between them and empirical frequencies is statistical significant null hypothesis is not valid.

At the beginning null (H_0) and alternative (H_1) hypotheses have to be defined.

$$\begin{aligned} &H_0 : ....P_{ij} = P_{i\bullet} \cdot P_{\bullet j}, \qquad \forall i, \forall j \\ &H_1 : ....\exists P_{ij} \neq P_{i\bullet} \cdot P_{\bullet j} \end{aligned} \qquad (1)$$

where P_{ij} are frequencies and $P_{i\bullet}$, $P_{\bullet j}$ are marginal frequencies of relevant variables. Null hypothesis in (1) assumed independence between two variables. It is necessary to note that this testing requests frequencies in contingence table (i.e. final table of Chi-square accounting) not to be to undersized. The general principle doesn't allow expected frequency to be less than 5. The practice notices divergence from this rule but very rarely. For example, in contingence tables larger than 2x2 (2 rows and 2 columns) the smallest expected frequency less than 1 can be allowed under condition that there are no more than 20% frequencies less than 5. When in such a table there are great number of expected counts less than 5 statistic theory proposes aggregation of relevant similar variables into the same group. Namely, when expected frequencies are very small, it can lead to wrong conclusion i.e. unrealistic rejection of the null hypothesis. Further more, a great number of variations of nominal variable/s on artificial way increases degrees of freedom, which can lead to unrealistic acceptation of the null hypothesis. In the case of the smallest contingence tables order 2x2, with expected frequencies less than 5, hypothesis testing by Chi-square independence can't be carried out. For these tables sample size more than 40 is necessary and using of some additional tests is required.

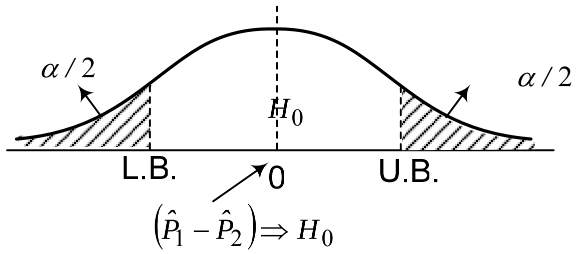
3.2. Proportions testing

At the beginning null (H_0) and alternative (H_1) hypotheses have to be defined.

$$\begin{aligned} H_0 &: ....P_1 - P_2 = 0 \\ H_1 &: ....P_1 - P_2 \neq 0 \end{aligned} \quad (2)$$

where P_1 and P_2 are proportions of relevant variables. Null hypothesis in (2) assumed that there is no difference between P_1 and P_2. That is coherent situation when Chi-square null hypothesis is acceptable. The conclusion about acceptance or rejection of null hypothesis is based on difference ($\hat{P}_1 - \hat{P}_2$) between relative frequencies from relevant samples (according to figure 1). If the difference between proportions (relative frequencies) from the sample ($\hat{P}_1 - \hat{P}_2$) is in the interval between lower bound (L.B.) and upper bound (U.B.) null hypothesis can be accepted as valid at chosen significant test level (α).

Figure 1: Null hypothesis acceptance for proportions testing



Source: Author's construction.

4. CASE STUDY

As it is mentioned the basic aim of this research is involving tradition in Croatian brand. Zabiokovlje as a part of Split-Dalmatian County, which successfully maintain its economy within sustainable development, is chosen as the survey research area[3]. For the purpose of this paper only parts of the extensive on-going research related to the tradition in Croatian brand have been prepared to confirm the basic statistic results of comparison between Chi-square and proportions testing. From the wide specter of related variables with nominal measure and modifications of offered answers in questionnaires three combinations have been selected for presentation.

Table 1: Crosstabulation analysis of the answers about traditional decoration across sex

Count		Traditional decoration on holidays				Total
		always with great importance for me	often	sometimes	never	
Sex	female	291	63	12	1	367
	male	197	41	22	7	267
Total		488	104	34	8	634

Source: Survey research results.

[3] Arnerić J., Jurun E., *Cross-Tabulation Analyses of the Survey Research on the Moral Values*, Proceedings of the 8th International Symposium on Operational Research, SOR'05, Nova Gorica, Slovenia, 2005, pp 147-152.

Table 1 presents results of crosstabulation analysis about relation between sex and traditional decoration on holidays. It is obvious that there is difference between women's and men's attitude towards traditional decoration especially in their ambiences during national and religious holidays.

Table 2: Chi-square test results about traditional decoration across sex

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	14,797[a]	3	,002
Likelihood Ratio	15,118	3	,002
Linear-by-Linear Association	8,964	1	,003
N of Valid Cases	634		

a. 2 cells (25,0%) have expected count less than 5. The minimum expected count is 3,37.

Source: According to survey research results.

Chi-square test results from Table 2 confirm rejecting of null hypothesis i.e. there is dependence between sex and traditional decoration at 0,2% significant level. Although 25% expected counts are less than 5, Chi-square test results are valid because the minimum expected count is larger than 1 i.e. it is 3,37.

In opposite case it would be impossible to make conclusion, because methodology requests aggregation of relevant similar variables in rows and/or columns. In this case it is not possible because there are only two rows. For the purpose of proportions testing for independence existence, it is necessary to compute results for all the combinations of variations of nominal variable/s, because only one statistically confirmed proportion difference may cause rejection of Chi-square null hypothesis.

For this case, proportions testing results, about difference between women's and men's who sometimes use traditional decoration on holiday, are presented.

Interval of null hypothesis acceptance (5% significance level) is: $[\pm 0{,}035517]$. Since, the proportions difference from the sample is $(\hat{p}_1 - \hat{p}_2) = -0{,}0497$ null hypothesis can be rejected. So, it can be concluded that there are statistically significant difference between female and male attitude towards traditional decoration.

Table 3 presents results of crosstabulation analysis about relation between financial status of interviewed and traditional culinary.

Table 3: Crosstabulation analysis of the answers about traditional culinary across financial status

Count		Traditional culinary				Total
		always with great importance for me	often	sometimes	never	
Financial status	very good	51	15	5	3	74
	good	167	70	29	4	270
	modest but sufficient	110	62	29	1	202
	low	43	14	11	3	71
	very low	9	3	5	0	17
Total		380	164	79	11	634

Source: Survey research results.

Chi-square test results from Table 4 confirm rejecting of null hypothesis i.e. there is dependence between financial status of interviewed and their attitude towards traditional culinary at 4,8% significant level..

For this case, proportion testing results, for the interviewed who enjoy traditional culinary always with great importance, are presented. From this group of interviewed answers of those who have very good and those who have modest but sufficient financial status are compared.

Interval of null hypothesis acceptance (5% significance level) is: $[\pm 0{,}066911]$. Since, the proportions difference from the sample is $(\hat{p}_1 - \hat{p}_2) = 0{,}144635$ null hypothesis can be rejected. So, it can be concluded that there are statistically significant difference between enjoying in traditional culinary among those who have very good and those who have modest but sufficient financial status. So, Chi-square results are once again confirmed by proportions testing.

Table 4: Chi-square test results about traditional culinary across financial status

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21,131	12	,048
Likelihood Ratio	20,144	12	,064
Linear-by-Linear Association	4,427	1	,035
N of Valid Cases	634		

Source: According to survey research results.

Table 5 presents results of crosstabulation analysis about relation between interviewed education and their way of holiday spending.

Table 5:Crosstabulation analysis of the answers about holiday spending across education

		For me holiday is the day					
Count		as any other day	when I can finish works I didn't manage in last week	when I enjoy the family atmosphere, a good meal and TV	possibility to visit the places of country heritage	possibility to meet parents, the sick and the old; possibility to read, Internet, and to be at peace with yourself	Total
Education	Without education or Incomplete elementary school	19	3	23	0	3	48
	Elementary school	20	10	30	3	6	69
	Skilled worker	21	22	30	6	9	88
	Secondary higher education	72	53	130	11	44	310
	University degree	10	13	24	4	11	62
	Higher degree	5	8	18	4	8	43
	MSc or PhD	4	0	5	2	3	14
Total		151	109	260	30	84	634

Source: Survey research results.

Taking into account that the original contingence table of Chi-square had great number of expected counts less than 5, according to statistic theory, counts of relevant similar variables are aggregated into the same group. Hence, in Table 6 there are Chi-square independence test results about holiday spending across education, where interviewed with incomplete elementary school and without education, as well as those with MSc and PhD, are grouped in the same subsets. For the same reason, aggregation across the columns has be done with the groups of those for whom holiday is the day of possibility to meet parents, the sick and the old, and those for whom holiday is the day of possibility to read, internet and to be at peace with themselves although they are completely different type of people.

Chi-square test results from Table 6 confirm rejecting of null hypothesis i.e. there is dependence between relevant variables at 4,3% significant level. The sample is larger than 600 and size of expected counts can be tolerated.

For this case, proportion testing results, for the interviewed for whom holiday is day as any other day, are presented. From this group of interviewed answers of those without education and/or with incomplete elementary school and those with higher degree are compared.

Interval of null hypothesis acceptance (5% significance level) is: $[\pm 0{,}181353]$. Since, the proportions difference from the sample is $(\hat{p}_1 - \hat{p}_2) = 0{,}279554$ null hypothesis can be rejected. So, it can be concluded that there are statistically significant difference between holiday spending among those who have higher degree and those without education and/or with incomplete elementary school.

Table 6: Chi-square test results about holiday spending across education

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	37,106(a)	24	,043
Likelihood Ratio	41,334	24	,015
Linear-by-Linear Association	13,057	1	,000
N of Valid Cases	634		

a 9 cells (25,7%) have expected count less than 5. The minimum expected count is ,66.

Source: According to survey research results.

5. CONCLUSION REMARK

This paper focused establishing the much more common procedure of proportions testing with the same conclusions in statistical sense in the situations when Chi-square of independence is not possible for technical reasons. It is only a part of an extensive on-going scientific research about tradition in Croatian brand. This part of the research by proportions testing confirms the same conclusions as Chi-square of independence.

REFERENCES

1. Arnerić J., Jurun E., *Cross-Tabulation Analyses of the Survey Research on the Moral Values*, Proceedings of the 8th International Symposium on Operational Research, SOR'05, Nova Gorica, Slovenia, 2005.
2. Aron A., Aron E., Coups E., *Statistics for the Behavioural and Social Sciences,* 4th edition, Prentice Hall, Cambridge, 2007.
3. Pivac S., Rozga A., *Statistika za sociološka istraživanja*, University of Split, Faculty of Philosophy Split, 2006.

MULTIRESOLUTION AND CORRELATION ANALYSES OF GDP IN EUROZONE VS. EU MEMBER COUNTRIES

Robert Volčjak
Economic Institute of the Law School
Prešernova 21, SI-1000 Ljubljana
Slovenia
robert.volcjak@eipf.si

Vesna Dizdarević
Promo + d.o.o.
Perčeva 4, SI-1000 Ljubljana
Slovenia
promoplus@siol.net

Abstract: In this paper the business cycles and especially their convergence in the Euro zone assumed as to satisfy the Optimal Currency Area (OCA) is being considered. Multiresolution decomposition of GDP growth signal is used and correlation coefficients are computed for decomposed signal to assess the numerical values of synchronicities of business cycles. The conclusion is that the Euro zone in many ways confirm OCA theory and that the most of the new members of the EU might experiences some difficullies if joining the Euro too early.

Keywords: wavelets, multiresolution analysis, business cycles, Euro zone

1. Introduction

The topic of business cycles and especially convergence has received a great deal of attention in recent years, mainly motivated by the economic and monetary union in Europe (EMU). According to optimal currency area (OCA) theory, developed roughly four decades ago, two countries or regions will benefit from a monetary union if they share similar business cycles, trade intensively, and rely on efficient adjustment mechanisms (e.g., labor mobility, price flexibility of production factors,...) to smooth out asymmetric shocks. Consequently, efforts have been made to quantify the synchronicity of business cycles among the core members of the European Union (EU) and the new ten EU members that joined as of May 2004 (Crowley, Lee, 2005). The reason for these efforts is that if business cycles of the Euro zone countries are asynchronous, then the monetary union may not be as beneficial. The new EU countries should not rush too early to adopt the Euro unless their economies meet the conditions set by the OCA theory. Even for the countries already in the Euro zone asynchronous business cycles may spell some trouble as the European Central Bank set its monetary policy (e.g. interest rate) for the whole Euro-zone. It is well known fact that business cycles can be statistically decomposed into components with different frequencies (trend, season, noise). Therefore it is a natural way to use multiresolution analysis as a tool to decompose business cycles, defined in this paper by the dynamics of gross domestic product (GDP), into the components with well defined frequencies that allow the comparison among them. The synchronicity of business cycle components can be measured in many ways. Here the usual correlaton coefficients between components are used for the easier interpretation of the results. Section 2 considers a brief overview of wavelets and multiresolution analysis, in section 3 obtaining the data and methods of calculations are described, main results and conclusions are given in section 4 and section 5 lists literature and sources used in the paper.

2. Wavelets and MRA

Wavelets, respectively multiresolution analysis (MRA) enable decomposing a signal (e.g., a time series of GDP, industrial production, inflation, stock returns) into high and low frequency components (Chui, 199; Percival, Walden, 2000). High frequency (irregular) components describe the short-run dynamics, whereas low-frequency components represent the long-term behavior of a signal. Identification of the business cycle involves retaining intermediate frequency components of a time series. That is, we disregard very high- and very low-frequency components. For instance, it is customary to associate a business cycle with cyclical components between 6 and 32 quarters (Burda, Wyplosz, 2005).

A function or signal can be viewed as composed of a smooth (or trend) background and fluctuations or details on top of it. The distinction between the smooth part and the details is determined by the resolution, that is, by the scale below which the details of a signal cannot be discerned. At a given resolution, a signal is approximated by ignoring all fluctuations below that scale. By progressively increasing the resolution, at each stage of the increase in resolution finer details are added to the coarser description, providing a successively better approximation to the signal. Eventually when the resolution goes to infinity, the exact signal is recovered. The intuitive description above can be written formal as follows. The resolution level is labelled by an integer j. The scale associated with the level j=0 is set to unity and that with the level j is set to $1/2^j$. Consider a function $f(t)$. At resolution level j the function is approximated by $f_j(t)$. At the next level of resolution $j+1$, the details at that level denoted by $d_j(t)$ are included and the approximation to $f(t)$ at the new resolution level is then $f_{j+1}(t)=f_j(t)+d_j(t)$. The original function $f(t)$ is fully recovered when the resolution tends to infinity:

$$f(t) = f_j(t) + \sum_{k=j}^{\infty} d_k(t).$$

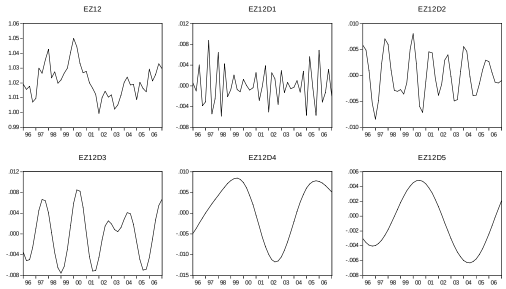
The word multiresolution (MR) refers to the simultaneous presence of different resolutions. The above equation represents one way of decomposing the function $f(t)$ into a smooth part plus details. By analogy, the space of square integrable functions, $L^2(\mathbf{R})$, may be viewed as composed of a sequence of subspaces $\{W_k\}$ and V_j, such that the approximation of $f(t)$ at resolution j, i.e. $f_j(t)$, is in V_j and the details $d_k(t)$ are in W_k. Functions which are used for these reasons are called wavelets, the practical procedures for applications of wavelet analysis commonly utilize a discrete wavelet transform (DWT). The most commonly used wavelets' families are the orthogonal ones.In this paper the quarterly data are analyzed. MR scales are such that scale (or detail) 1 (D1) is associated with 1-2 quarters dynamics, scale 2 (D2) with 2-4 quarters dynamics, scale 3 (D3) is with 4-8 quarters or 1-2 years dynamics, scale 4 (D4) is with 8-16 quarters or 2-4 years dynamics and scale 5 (D5) with 16-32 quarters or 4-8 years dynamics.

3. Data and methodology

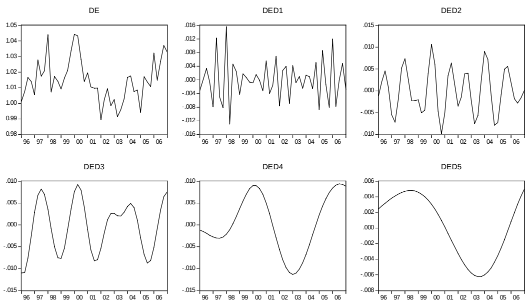
Quarterly data for GDP of EU countries, measured in millions of euros at constant 1995 prices and exchange rates, was obtained from the EUROSTAT. Most data ranges from 1995Q1 to 2007Q1 except for Romania (1999Q1-2007Q1), Ireland and Croatia (1997Q1-2007Q1). From these due to potential seasonality in the data the business cycle time series for a country i was computed as $GDP_{i,t}/GDP_{i,t-4}$. Thus obtained time series was then feeded to the MATLAB software package with the wavelets toolbox through which the MRD of every GDP growth series was decomposed into smooth level and five detail levels D1-D5, using, due to their indefinitely derivability, Meyer family of wavelet functions. In Figure 1, panels (a) – (e) the different levels can be seen together with the original signal for the Euro zone (EZ12), Germany (DE), the Netherland (NL), United Kingdom (UK) and Slovenia (SI) respectively. Various levels of synchronicity at different scales of details can be seen even better in Figure 2, where the components at scales D5, D4 and D3 are shown on panels (a)-(c) respectively for the above mentioned countries plus Estonia (EE) and Czech Republic (CZ).

Figure 1

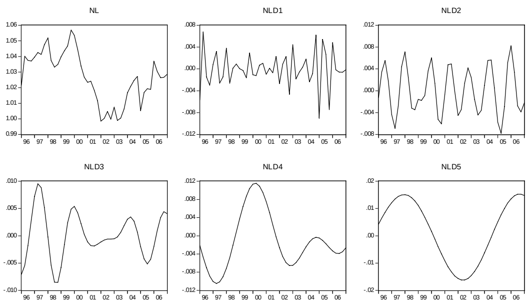
(a)



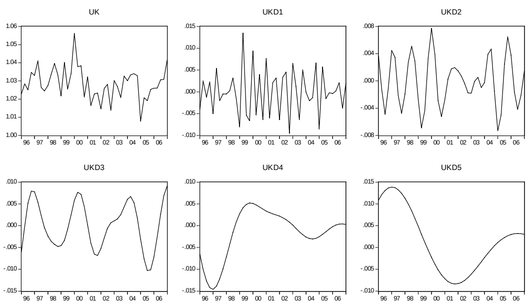
(b)



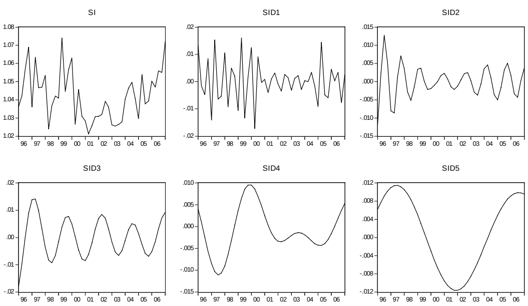
(c)



(d)



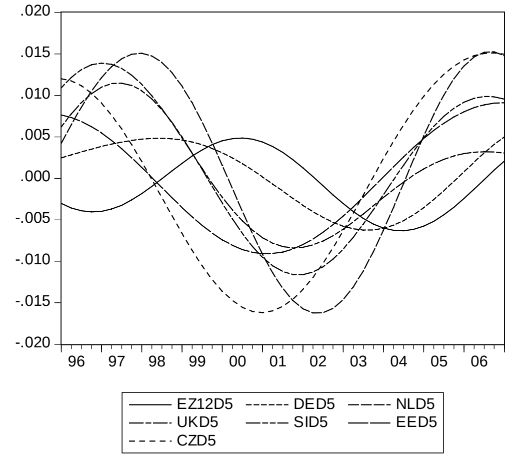
(e)



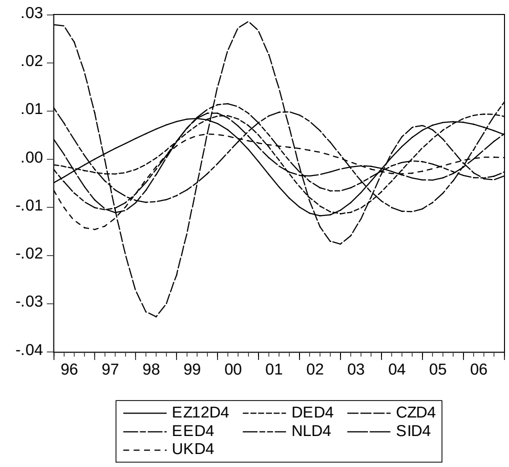
Sources: Eurostat, own calculations

Figure 2

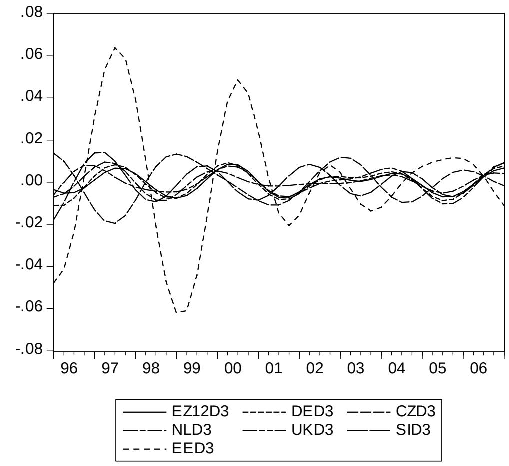
(a)



(b)



(c)



Sources: Eurostat, own calculations

Table 1:

Correlation coefficients between the Eurozone and different EU member countries

Sources: Eurostat, own calculations

Euro Zone 12							
Country	*Corr*	*L*	d1	d2	d3	D4	d5
		d1	**0.953752**	0.004493	-0.003975	-0.003641	-0.003703
		d2	-0.006145	**0.934317**	0.009582	-0.016525	-0.013868
EU25	**0.9770**	d3	-0.007213	0.018677	**0.990372**	0.037329	0.042178
		d4	-0.004723	-0.027715	0.042398	**0.998219**	0.135373
		d5	-0.005428	-0.003962	0.046966	0.240061	**0.992227**
		d1	**0.959111**	0.004415	-0.003289	-0.003347	-0.003403
		d2	-0.003350	**0.945447**	0.014130	-0.017631	-0.015363
EU15	**0.9814**	d3	-0.006978	0.022870	**0.992213**	0.039622	0.042446
		d4	-0.004628	-0.027397	0.044016	**0.997837**	0.130394
		d5	-0.005373	-0.005861	0.047946	0.165739	**0.999982**
		d1	0.221615	-0.013512	0.004412	0.008193	0.005469
		d2	0.008791	**0.551204**	0.035231	-0.028576	-0.012460
Austria	**0.7267**	d3	-0.007923	-0.130513	**0.745668**	-0.003088	0.038970
		d4	6.98E-05	0.031206	-0.006458	**0.602028**	**0.592158**
		d5	-0.003051	-0.032062	0.023135	**0.771726**	-0.288128
		d1	0.244746	-0.026349	0.003470	0.018998	0.004229
		d2	0.002427	0.463686	-0.001840	0.007603	-0.017793
Belgium	**0.7511**	d3	-0.011423	0.022295	**0.855342**	0.068431	0.047890
		d4	0.004843	0.065799	-0.028445	0.237741	0.457648
		d5	-0.002934	-0.030143	0.024163	**0.766913**	-0.255445
		d1	**0.916335**	-0.008807	0.002924	0.007536	0.001821
		d2	-0.011282	**0.845426**	0.005632	-0.006739	-0.006604
Germany	**0.9143**	d3	-0.011483	-0.017649	**0.960187**	0.077957	0.065914
		d4	-0.003143	-0.004493	-0.000544	**0.822916**	**0.522826**
		d5	-0.001068	-0.005782	0.024373	**0.610699**	0.433593
		d1	0.232214	-0.023824	0.009215	0.015346	0.007185
		d2	-0.022159	0.195177	-0.000497	0.028376	0.019600
Spain	**0.7389**	d3	-0.011107	0.083164	0.485736	0.076802	0.056102
		d4	0.006014	0.068939	-0.040926	0.121846	0.486614
		d5	-0.002816	-0.041915	0.042982	**0.702226**	0.147526
		d1	0.404893	-0.007645	0.003548	0.007027	-0.000159
		d2	-0.004789	0.183430	0.020013	0.025999	0.010029
Finland	**0.6582**	d3	-0.009720	-0.071808	**0.867501**	0.007402	-0.036164
		d4	-0.003807	0.046464	-0.082226	0.162488	**0.733442**
		d5	0.003885	-0.000589	-0.011487	0.468575	-0.210425
		d1	**0.788831**	0.009152	-0.006027	-0.006696	-0.005213
		d2	0.003949	**0.931527**	0.012191	-0.028121	-0.020396
France	**0.8865**	d3	-0.012532	-0.011675	**0.806189**	0.090417	0.050699
		d4	0.003856	0.045745	-0.051823	**0.570423**	-0.004155
		d5	-0.004820	-0.034371	0.054166	**0.620719**	**0.672344**
		d1	0.034247	-0.003750	-0.005454	-0.002731	-0.000559
		d2	0.011503	0.466659	0.024766	-0.000583	0.000205
Ireland	**0.6219**	d3	0.006187	0.029541	**0.638025**	0.003253	0.016160
		d4	-0.044933	0.015634	0.188287	-0.006750	**0.516102**
		d5	0.043476	-0.002117	-0.091410	**0.839392**	0.347320
		d1	**0.797997**	-0.001365	-0.001164	0.001281	-0.002544
		d2	0.008453	**0.839144**	0.027627	-0.035797	-0.014922
Italy	**0.9045**	d3	-0.008898	0.065115	**0.907917**	0.081863	0.057536
		d4	0.001522	0.044549	-0.048528	0.296512	**0.720939**
		d5	-0.002909	-0.042378	0.045228	**0.689488**	0.189953
		d1	**0.610808**	-0.016453	-0.006174	0.009572	-6.17E-05
		d2	-0.003976	**0.717838**	-0.033273	-0.015649	-0.027906
Netherland	**0.7762**	d3	-0.008811	-0.023475	**0.736669**	0.032510	-0.002305
		d4	-0.000787	0.026985	0.017112	0.370969	**0.718304**
		d5	-0.001327	-0.013659	0.000779	**0.793836**	-0.128897
		d1	0.463952	0.040292	-0.023124	-0.031356	-0.017091
		d2	-0.036525	0.273216	-0.038937	0.017112	0.002676
Slovenia	0.4383	D3	-0.015467	-0.187438	0.339077	-0.008406	-0.002994
		d4	0.000994	0.044913	0.041432	0.264106	**0.666294**
		d5	0.000545	-0.007411	-0.006576	**0.640032**	-0.379520

Table 1: *continuoued*

United Kingdom	**0.5582**	d1	0.440933	-0.013967	0.004817	0.009926	0.005741
		d2	0.012807	0.246886	-0.008436	-0.007249	-0.004945
		d3	-0.010388	-0.171022	**0.764814**	-0.040809	-0.013740
		d4	-0.004015	0.037436	-0.065790	0.087530	**0.656455**
		d5	0.004993	0.006707	-0.025100	0.418329	-0.356038
Sweden	**0.7119**	d1	**0.750696**	-0.016912	0.007216	0.012551	0.005373
		d2	0.000834	0.426320	-0.014615	-0.016828	-0.033004
		d3	-0.007498	0.056066	**0.652777**	0.066009	0.063388
		d4	0.000829	0.044120	0.005872	0.485278	0.397688
		d5	0.002194	-0.003840	-0.010581	**0.603667**	-0.250228
Danemark	**0.5516**	d1	**0.744414**	-0.015639	0.003707	0.011546	0.003823
		d2	-0.016402	0.306012	-0.021566	-0.005414	-0.018197
		d3	-0.014390	-0.273655	0.328073	0.081528	-0.051045
		d4	-0.002282	0.006732	0.052323	**0.535333**	0.668383
		d5	0.003865	0.018804	-0.043349	0.410551	**-0.554542**
Czech Republic	-0.2224	d1	-0.183917	-0.003059	-0.000600	0.003314	0.000539
		d2	-0.032550	-0.450248	-0.129607	0.031627	-0.011535
		d3	0.008479	0.041839	-0.356920	-0.005123	0.098226
		d4	0.007126	0.016343	0.027367	-0.043548	0.174019
		d5	0.001969	0.012885	-0.031348	0.318478	**-0.686259**
Poland	0.4673	d1	0.018754	-0.006183	0.000475	0.004349	0.000865
		d2	0.012605	**0.838604**	0.058924	-0.040491	-0.009589
		d3	-0.001679	0.099140	**0.811693**	0.012552	0.086035
		d4	-0.001478	-0.038569	0.041335	**0.867092**	-0.083008
		d5	0.002920	0.017318	-0.031815	0.429657	-0.499836
Hungary	0.4016	d1	0.086229	0.006322	0.011257	0.000373	0.004616
		d2	-0.003471	**0.503662**	-0.010206	-0.012392	-0.021968
		d3	-0.007421	-0.028442	**0.597110**	0.062781	-0.068015
		d4	-0.006103	-0.022644	0.040922	**0.584100**	-0.202258
		d5	-0.003996	-0.017526	0.041996	-0.274862	**0.727780**
Slovakia	-0.2317	d1	0.282341	0.000638	0.011722	0.006080	0.006834
		d2	-0.023639	0.048875	-0.009270	0.019093	-0.010061
		d3	-0.009638	0.041192	0.413700	0.071956	0.099270
		d4	-0.006244	-0.011548	0.047270	-0.260726	**0.820830**
		d5	0.008413	0.031452	-0.059975	0.094142	**-0.659212**
Estonia	0.2371	d1	0.336771	0.006878	-0.001689	-0.004143	-0.001565
		d2	0.008066	**0.838602**	0.054474	-0.027623	-0.009407
		d3	-0.005963	-0.002362	**0.651685**	0.061304	-0.054919
		d4	0.001835	0.041663	-0.036439	-0.466674	**0.501049**
		d5	0.002269	0.015875	-0.033917	0.352139	**-0.629330**
Latvia	0.1174	d1	**0.515277**	0.004087	0.004177	-0.001845	0.000862
		d2	0.001597	**0.731896**	0.044161	-0.017963	0.000237
		d3	-0.007105	-0.032310	0.472825	0.060387	-0.074309
		d4	-0.005583	0.028042	-0.035941	-0.003753	**0.687280**
		d5	0.001031	0.013536	-0.024616	0.432807	**-0.511381**
Lithuania	-0.2773	d1	**0.591051**	0.004955	-0.003121	-0.005275	-0.002548
		d2	-0.000269	0.150302	-0.000887	0.001715	0.016680
		d3	-0.001932	-0.008769	0.300380	0.022602	-0.078080
		d4	-0.001439	0.015498	-0.057654	**-0.777292**	-0.131135
		d5	0.004703	0.013651	-0.045760	0.178073	**-0.844302**
Bulgaria	0.3525	d1	0.322482	-0.016906	0.005630	0.011514	0.005716
		d2	0.013024	**0.651294**	0.032749	-0.058798	-0.014800
		d3	-0.004951	0.116993	0.280922	0.075319	0.055107
		d4	0.004641	0.057039	-0.019272	0.481563	0.152503
		d5	-0.008909	-0.041416	0.077538	0.297245	**0.896321**
Romania	0.0186	d1	0.310376	-0.011776	0.003404	-0.009169	-0.010616
		d2	0.050558	**0.533389**	-0.154589	0.029973	-0.010387
		d3	0.011205	0.117284	**0.545343**	-0.054872	-0.009940
		d4	-0.010032	-0.030445	-0.214020	**-0.935397**	-0.313396
		d5	-0.017401	-0.026685	-0.138098	0.475455	**-0.713575**
Croatia	-0.1621	d1	0.241404	0.027172	0.004961	0.003391	-0.004352
		d2	0.025975	**0.735918**	0.113303	-0.007624	-0.006456
		d3	0.084702	0.140451	**0.568177**	-0.041696	-0.054095
		d4	0.025097	-0.019815	-0.207998	-0.257462	0.030048
		d5	-0.045304	0.000415	0.095360	**-0.837798**	-0.329175

4. Main results & conclusions

Numerically, different levels of synchronicity of the GDP growth time series can be presented by correlation coefficients. All correlation coefficients for different EU member countries are computed with respect to the Euro zone and the results are shown in Table 1. For each country the overall correlation coefficient was computed between that country GDP growth series and Euro zone GDP growth series (second column) together with correlation coefficients between the five MR components of the country GDP growth series and the five MR components of the Euro zone GDP growth series. The diagonal cells with the same frequency are shaded grey and for convenience the correlation coefficients with the absolute value above 0,5 are printed in bold typeface.

From the overall correlation coefficients four main different levels of synchronicity of business cycles can be seen. First there are big, old EU members with high synchronicity to the Euro zone and with the correlation coefficients values above 0,8 (e.g. Germany 0,91, Italy 0,90, France 0,89) The same high level of synchronicity can also be seen at almost all different same-frequency levels of GDP MR components. In the second group the smaller Euro zone economies can be found with correlation coefficients above 0,5 (e.g. Netherland 0,78, Belgium 0,75, Finland 0,65). Also in this group there are the old EU members not in the Euro zone with Sweden the most synchronous with the Euro zone (ρ=0,71). The third group is mainly composed of in 2004 new members of the EU with $0,1< \rho<0,5$. Among these Slovenia which joined the Euro zone in 2007 has only weak correlation with the last (ρ=0,43). The last group is composed of new EU members which have negative, although weak, overall correlation with the Euro zone (e.g. Czech Republic, Slovakia) and also the EU aspirant Croatia. The same results hold also at different MR component of GDP growth series.

It can be concluded that the Euro zone in many ways confirm OCA theory and that the most of the new members of the EU might experiences some difficullies if joining the Euro too early.

Some further research may include:

- Granger test of causality between different highly correlated componets;
- Coherence and phase shift computation between the same frequency components;
- VAR models for forecasting different frequency components...

5. Sources & literature

Burda M., Wyplosz C.: ***Macroeconomics: A European Text***. Oxford University Press, USA, 2005.

Chui C.K.: ***An Introduction to Wavelets***. Wavelet Analysis and Its Application (Volume 1), Academic Press, 1992.

Crowley P.M., Lee J.: ***Decomposing the co-movement of the business cycle:a time-frequency analysis of growth cycles in the euro area***. Bank of Finland Research, Discussion Papers, 2005.

EUROSTAT > Economy and finance > National accounts (including GDP) > Quarterly national accounts > GDP and main components (http://epp.eurostat.ec.europa.eu/)

Misiti M., Misiti Y., Oppenheim G., Poggi J.M.: ***Wavelet Toolbox 4 User's Guide.*** The MathWorks, Inc., 2007.

Percival D.B., Walden A.T.: ***Wavelet Methods for Time Series Analysis***. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2000.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Section 12
OR Communications

Classification and convergence of some stochastic algorithms

ARRAR K. Nawel & DJELLAB Natalia

Université Badji Mokhtar, Annaba
Faculté des Sciences, Département de Mathématique
kn_arrar@yahoo.fr & djellab@yahoo.fr

Abstract: We will present in this work a classification of most known metaheuristics by making a comparison between them, we will also be interested in convergence of two important algorithms which are the genetic algorithm and simulated annealing.

Key words: Markov Chains, simulated annealing, genetic algorithm, stochastic algorithms.

1 Introduction

Metaheuristics are generally iterative stochastic algorithms, which progress towards an optimum by sampling of an objective function. The successive iterations must make it possible to pass from a solution of bad quality to the optimal solution. The algorithm stops after having reached a criterion of stop, generally consisting of the attack of the assigned execution time or of a required precision.

Some metaheuristics are theoretically convergent under some conditions. It is then guaranteed that the total optimum will be found in a finished time, the probability of being done increase asymptotically with time. This guarantee amounts considering that the algorithm behaves at worst like a pure random research. In practice, the principal condition of convergence is to consider that the algorithm is ergodic, but we can be satisfied with quasi-ergodicity.

2 Classification of metaheuristics

- **Evolutionnary or not:** we can make the difference between the metaheuristics which take as a starting point natural phenomena and those which are not inspired any.
- **Trajectory and population:** Another way of classifying the metaheuristics is to distinguish those which work with a population of solutions of those which handle only one solution at the same time.
- **Statics and dynamic:** Metaheuristics can also be classified according to their manner of using the objective function.
- **Structures of vicinities:** Researchers prefer to classify the metaheuristics according to the number of structures of vicinities used.
- **Memory with short and long-term:** Certain metaheuristics make use of the history of research during optimization, whereas others do not have any report of the past.

In the description of principal metaheuristics, we will be based on the classification which distinguishes the methods of trajectory from the methods based on populations of solutions.

3 Local research or Methods of trajectory

The stochastic algorithms are primarily techniques of simulation of complex laws of probabilities on large-sized spaces. These measures can be arranged in two classes: measures of Boltzmann-Gibbs, and measures of Feynman-Kac. The first are defined on homogeneous space E, in term of an energy function $U : E \to [0, \infty)$, a parameter of temperature $\beta \in [0, \infty)$, and of a reference measure λ on E : $\mu_\beta(dx) = \frac{1}{Z_\beta} \exp\left[-\beta U(x)\right] \lambda(dx)$ with $Z_\beta = \lambda \exp\left[-\beta U\right]$

3.1 Simulated annealing method ([4])

Once fixed invariant measure, it remains has to judiciously choose a transition from probability of a Markov chain having such an asymptotic behavior while using is the transition from Métropolis-Hastings or the sampler of Gibbs.

3.1.1 Convergence of simulated annealing

The algorithm of simulated annealing that we will present is a method of random research of the total extremas of a numerical bounded function U. The random exploration of the space of state E is defined in term of a transition from probabilities $Q(x, dy)$ on E, reversible compared to measure λ. I.e., such as $\lambda(dx)Q(x, dy) = \lambda(dy)Q(y, dx)$, λ is necessarily an invariant measure of Q. The algorithm of simulated annealing is a nonhomogeneous Markovian algorithm. It is appeared as a Markov chain from which the transition kernel to each step $n \geqslant 1$ depends on a parameter of temperature $T(n) \in R+$.

- For $n = 0$, we simulate a random variable X_0, according to initial distribution η_0.
- To the step n, the transition $X_n \to X_{n+1}$ is broken up into a step of exploration, and a step of acceptance.

  1. The step of exploration consists in proposing a state Y_n with law $Q(X_n, .)$.
  2. The step of acceptance breaks up again into two under-steps:
  - If $U(Y_n) \leq U(X_n)$ we accept the state Y_n and we put $X_{n+1} = Y_n$.
  - If $U(Y_n) > U(X_n)$, then we carry out the following random choice

$$X_{n+1} = \begin{cases} Y_n & \text{with probability } e^{-\frac{1}{T(n)}(U(Y_n)-U(X_n))} \\ X_n & \text{with probability } 1 - e^{-\frac{1}{T(n)}(U(Y_n)-U(X_n))} \end{cases}$$

The stationary measure of homogeneous annealing (at constant temperature T) is thus given by the measure of Boltzmann-Gibbs $\eta^{[T]}(dx) = \frac{1}{\lambda(e^{-\frac{1}{T}U})} e^{-\frac{1}{T}U(x)}\lambda(dx)$

4 Methods based on the populations or Methods evolutionary

The majority of the models are founded on physical or biological principles. In other words, these algorithms copy the processes of evolution or training dictated by physical rules or resulting from the natural evolution. These models are formalized mathematically by Markov chain.

4.1 Genetic Algorithms

Genetic algorithms are algorithms of optimization (see [3]), based on techniques derived from the genetics and natural evolution. A genetic algorithm seeks the extrema of a function defined on a space of data. To use it, we must have the following elements

- A principle of coding the elements of population.
- A mechanism of generation of the initial population.
- A function to be optimized.
- Operators allowing to diversify the population during generations and to explore the state space. The purpose of the crossing operator recomposes genes of individuals existing in the population, the operator of change is to guarantee the exploration of the state space.
- Parameters of dimensioning: cut population, numbers total generations or criterion of stop, probabilities of application of the operators of crossing and change.

4.1.1 Modeling by Markov chain

This approach is most satisfactory as well on the mathematical level, as on that of modeling, the various operators being presented like "disturbing " a Markovian process representing the population to each step (see [1,2]). Here still it appears that only the operator of change is important, the crossing which can be completely absent.

We use a binary coding , P representing the number of bits used for coding. The function of evaluation, f is defined on space $E = \{0,1\}^P$ with values in $\mathbb{R}_+$. The problem is thus to locate the whole of maximum total of f , or, failing this, to find quickly and effectively areas of space where these maximums are located.

Let N the population size (fixed), let us note X_k population of the generation k : it is about a matrix $X_k = \left(X_k^1, X_k^2, \cdots, X_k^N\right)$ of E^N of which them N elements are chains of bits (chromosomes) of size P. The passage from generation k to generation $k+1$, i.e. from X_k to X_{k+1} breaks up into three steps:

$$X_k \overset{\text{mutation}}{\rightarrow} Y_k \overset{\text{crossing}}{\rightarrow} Z_k \overset{\text{selection}}{\rightarrow} X_{k+1}.$$

If $x = (x_1, \cdots, x_N)$ is an element of E^N and i a point of E, we will note

$$f(x) = f(x_1, \cdots, x_N) = \max\{f(x_i) : 1 \leq i \leq N\} \text{ and } x = \{x_k \in \arg\max f(x)\} \text{ et } [x] = \{x_k : 1 \leq k \leq N\}$$

4.1.2 Asymptotic convergence of the genetic algorithm.

The Markov chain (X_n^∞) without disturbance In the absence of disturbance, the studied process is a Markov chain $(X_n^\infty)_{n\geq 0}$ with the state space E^m. The writing ∞ reflect the fact that this process describes the behavior in extreme cases of our model, when all the disturbances vanishes. The probabilities of transition from these chains are

$$P\left(X_{n+1}^\infty = z / X_n^\infty = y\right) = \frac{1}{(card\widehat{y})^m} \prod_{k=1}^{m} 1_{\widehat{y}}(z_k)\, y(z_k) = \frac{1}{(card\widehat{y})^m} \prod_{i\in[z]} 1_{\widehat{y}}(i)\, y(i)^{z(i)}.$$

They are the individuals of the population X_{n+1}^∞ who are selected by chance (under the uniform distribution) and independantly among the elements of $\widehat{X}_n^\infty$ who are the best individuals of X_n^∞ granted to the fitness function f.

Let us suppose that the chain leaves the population initial $X_0 = x_0$. Then $\forall n \geq 1 \quad [X_n^\infty] \subset x_0$ almost surely, after a finished number of steps N, the chain is absorbed with the state (i) when i goes to $\widehat{x}_0$. In particular, if $\widehat{x}_0$ is reduced to the point i, the chain is absorbed instantaneously in (i).

Disturbed Markov chain $\left(X_n^l\right)$ The intensity of disturbance is controlled by an integer parameter l. At once that l grows towards infinity, the disturbances disappear gradually. The disturbed Markov chain $\left(X_n^l\right)$ is obtained through the overlapping of several chains $\left(U_n^l\right), \left(V_n^l\right)$ who represent the populations obtained successively by application of the operations of disturbances. More precisely, we break up the transition of X_n^l to X_{n+1}^l in three steps

$$X_n^l \overset{\text{mutation}}{\rightarrow} U_n^l \overset{\text{crossing}}{\rightarrow} V_n^l \overset{\text{selection}}{\rightarrow} X_{n+1}^l.$$

- $X_n^l \rightarrow U_n^l$: mutation

The operator of change is modelled by random disturbances independent of the individuals of the population X_n^l. Such a disturbance is described by a Markovien kernel p_l in space E, who is a definite function of $E \times E$ with values in $[0,1]$ checking $\forall\, i \in E \ \sum_{j\in E} p_l(i,j) = 1$. Transition probabilities from X_n^l to U_n^l are then as follows $P\left(U_n^l = u / X_n^l = x\right) = p_l(x_1, u_1) \cdots p_l(x_m, u_m)$.

This disturbance is small when the matrix $(p_l(i,j))_{(i,j)\in E\times E}$ approaches the matrix idendity. To ensure the disappearance of change when l tends towards the infinity, we impose

$$\forall\, i,j \in E \quad \lim_{l\to\infty} p_l(i,j) = \delta(i,j) \tag{1}$$

- $U_n^l \rightarrow V_n^l$: crossing

The operator of crossing is modelled by random disturbances independent of the formed couples of the consecutive individuals of the population $\left(X_n^l\right)$. As in the case of the change, such a disturbance is described by a Markovien kernel q_l in space $E\times E$, who is a defined function on $(E\times E)\times(E\times E)$ with values in $[0,1]$ checking $\forall\,(i_1, j_1) \in E\times E \quad \sum_{(i_2,j_2)\in E\times E} q_l((i_1,j_1),(i_2,j_2)) = 1$. Transition probabilities from U_n^l to V_n^l are $P\left(V_n^l = v / U_n^l = u\right) = \delta_m(u_m, v_m) \prod_{1\leq k\leq m/2} q_l((u_{2k-1}, u_{2k}), (v_{2k-1}, v_{2k}))$, where $\delta_m(i,j) = \delta(i,j)$ if m is odd (the last individual of the population remains unchanged after crossing) and $\delta_m(i,j) = 1$ if m is even.

To ensure the disappearance of the crossings when l goes to infinity we will impose $\forall\,(\,i_1, j_1) \in E \times E$, $\forall\,(\,i_2, j_2) \in E \times E$

$$\lim_{l\to\infty} q_l\left(\left(\,i_1, j_1\right), \left(\,i_2, j_2\right)\right) = \delta\left(i_1, j_1\right)\delta\left(i_2, j_2\right). \tag{2}$$

- $V_n^l \to X_{n+1}^l$: selection

With an aim of building the operator of selection, we use the function of selection F of order m defined on $\{1, ..., m\} \times \left(\mathbb{R}_+^*\right)^m$ with values in $[0, 1]$ satisfying for all $(f_1, ..., f_m)$ in $\left(\mathbb{R}_+^*\right)^m$

a) $\sum\limits_{k=1}^{m} F\left(k, f_1, ..., f_m\right) = 1$, and $f_1 \geq f_2 \geq ... \geq f_m \quad F\left(1, f_1, ..., f_m\right) \geq F\left(2, f_1, ..., f_m\right) \geq ... \geq F\left(m, f_1, ..., f_m\right)$.

b) $\sigma \in \sigma_m \quad \forall k \in \{1, ..., m\} : \quad F\left(\sigma\left(k\right), f_{\sigma(1)}, ..., f_{\sigma(m)}\right) = F\left(k, f_1, ..., f_m\right)$,

The value $F\left(k, f_1, ..., f_m\right)$ is the probability of the choice of f_k between $f_1, ..., f_m$. Now F_l is a selection function. The m individuals $X_{n+1}^{l,1}, ..., X_{n+1}^{l,m}$ who form the population X_{n+1}^l are selected by chance and independantly in the population V_n^l defined by the law F_l

$$\forall r \in \{1, .., m\} \quad \forall i \in E \quad P\left(X_{n+1}^{l,1} = i\right) = \sum_{h : V_n^{l,r} = i} F_l\left(h, f\left(V_n^l\right)\right).$$

whereas transition probabilities from V_n^l to X_{n+1}^l are

$$P\left(X_{n+1}^l = x / V_n^l = v\right) = \prod_{i \in [x]} \left(\sum_{k : v_k = i} F_l\left(k, f\left(v\right)\right)\right)^{x(i)} = \prod_{r=1}^{m} \sum_{k : v_k = x_r} F_l\left(k, f\left(v\right)\right).$$

The pressure of selection is maximum if individuals of X_{n+1}^l are selected by chance and uniformaly among the most suited individuals of V_n^l. The single selection function F_∞ who implements such a plan of selection is defined by $F_\infty\left(k, f\left(x\right)\right) = \frac{1}{card(\widehat{x})}$ i.e. that we obtained the uniform distribution on $\widehat{x}$. To ensure the disappearance of the selection of the individuals below the maximum fitness, we will force convergence from F_l to F_∞ in the unit $f\left(E\right)^m$

$$\forall x \in E \quad \forall k \in \{1, ..., m\} \quad \lim_{l\to\infty} F_l\left(k, f\left(x\right)\right) = F_\infty\left(k, f\left(x\right)\right). \tag{3}$$

Transition probability of the chain $\left(X_n^l\right)$

$$P\left(X_{n+1}^l = z / X_n^l = y\right) = \sum_{(u,v)\in(E^m)} P\left(X_{n+1}^l = z / V_n^l = v\right) P\left(V_n^l = y / U_n^l = u\right) P\left(U_n^l = u / X_n^l = y\right)$$

Conditions (1), (2) and (3) imply

$$\forall\left(y, z\right) \in E^m \times E^m \quad \lim_{l\to\infty} P\left(X_{n+1}^l = z / X_n^l = y\right) = P\left(X_{n+1}^\infty = z / X_n^\infty = y\right)$$

such as the transition probability of $\left(X_n^l\right)$ converges to $\left(X_n^\infty\right)$ when l goes to infinity. Thus the Markov chain $\left(X_n^l\right)$ seems a disturbance of the Markov chain $\left(X_n^\infty\right)$.

Generally we can say that simulated annealing gets generally a solution of good quality but requires a great number of parameters, the genetic algorithm is very powerful but difficult to manage and its effectiveness depends on the quality of coding.

References

[1] R. Cerf, "*Une théorie asyptotique des algorithmes génétiques.*" PhD thesis, Université Montpelier II, 1994.

[2] M. I. Freidlin et A. D. Wentzell. *"Random Perturbations of Dynamical Systems."* Springer-verlag, New-York, 1984.

[3] D. E. Goldberg. *"Genetic Algorithms."* Addison Wesley, 1989. ISBN: 0-201-15767-5.

[4] B. Ycart, "*Modèles et Algorithmes Markoviens.*" Mathématiques & Applications 39, Springer, 2002.

FUZZY MULTIPLE OBJECTIVE MODELS FOR FACILITY LOCATION PROBLEMS

Mehmet Can
Faculty of Arts and Social Sciences,
International University of Sarajevo, Paromlinska 66,
71000 Sarajevo, Bosnia and Herzegovina
E-mail: mcan@ius.edu.ba

Abstract: There are a variety of efficient approaches to solve crisp multiple objective decision making problems. However in the real life the input data may not be precisely determined because of the incomplete information. This paper deals with a multple objective facility location problem using the algorithm developed by Drezner and Wesolowski.

Keywords: fuzzy decision making, multi objective decision, fuzzy goal programming, facility location problem.

1 INTRODUCTION

In a standard multiple goal programming, goals and constraints are defined precisely. Fuzzy goal programming has the advantage of allowing for the vague aspirations of decision makers, which are quantified by some natural language rules [1-18].

To our knowledge, first R. Narasimhan [15] introduced fuzzy set theory into objective programming. Since then many achievements have been added to the literature. In the following, an approach for solving fuzzy multiple goal problems will be presented, and its application to a facility location problem will be discussed.

2 MULTIPLE FUZZY GOAL PROGRAMMING

In a multiple goal programming problem, the optimal realization of multiple objectives is desired under a set of constraints imposed by a real life environment. If the goals and constraints are all expressed with equalities, we have a completely symmetric formulation:

$$\text{find } \mathbf{x} \text{ such that } \mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}. \tag{1}$$

where **x** is the vector of variables, **b** is the vector of the goals and available resources, and **A** is the matrix of the coefficients. In the cases when the decision maker is not precise in goals and restrictions, the linguistic statements such as "around **b**" will be used. In this case the above crisp goal programming problem becomes:

$$\text{find } \mathbf{x} \text{ such that } \mathbf{Ax} = \tilde{\mathbf{b}},\ \mathbf{x} \geq \mathbf{0}. \tag{2}$$

where the fuzzy components b_i of the fuzzy vector **b** can be represented by, for example, triangular fuzzy membership function (Figure 1):

$$\mu_i(z) = \begin{cases} \left(z - (b_i - d_{i1})\right)/d_{i1}, & b_i - d_{i1} \leq z \leq b_i, \\ \left((b_i + d_{i2}) - z\right)/d_{i2}, & b_i \leq z \leq b_i + d_{i2}, \\ 0, & \text{elsewhere}. \end{cases} \tag{3}$$

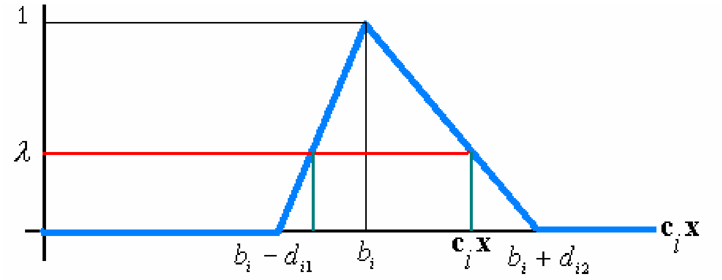


Figure 1. Fuzzy components b_i of the fuzzy vector **b.**

To have a membership number at least λ, $\mathbf{c}_i\mathbf{x}$ must remain in the interval

$$b_i - d_{i1} + \lambda d_{i1} \le \mathbf{c}_i\mathbf{x} \le b_i + d_{i2} + \lambda d_{i2} \tag{4}$$

that is

$$(\mathbf{c}_i\mathbf{x} - (b_i - d_{i1}))/d_{i1} \ge \lambda, \ ((b_i + d_{i2}) - \mathbf{c}_i\mathbf{x})/d_{i2} \ge \lambda. \tag{5}$$

Hence the above fuzzy goal programming problem is the maximum satisfaction problem of the fuzzy equations, and this goal can be achieved by the solution of the below crisp linear programming problem described by Lai, and Wang [14]:

Max λ such that for all *i*,

$$(\mathbf{c}_i\mathbf{x} - (b_i - d_{i1}))/d_{i1} \ge \lambda, \ ((b_i + d_{i2}) - \mathbf{c}_i\mathbf{x})/d_{i2} \ge \lambda. \tag{6}$$

3 A FACILITY LOCATION PROBLEM

Bhattacharya, J.R. Rao, and R.N. Twari [2] have used fuzzy goal programming to locate a single facility on a plane bounded by a convex polygon under three objectives:

i. Maximize the minimum distances,
ii. Minimize the maximum distances from the facilities to the demand points,
iii. Minimize the sum of all transport costs.

Let $P_i = (a_i, b_i)$, $i = 1,2,...,m$ be the locations of demand points, $S = (x, y)$ is the location of the new facility, and X is the set of feasible points for new facility. Then,

$$\text{Max } g_1(x,y) = \min_i (|x - a_i| + |y - b_i|)$$
$$\text{Min } g_2(x,y) = \max_i (|x - a_i| + |y - b_i|)$$
$$\text{Min } g_3(x,y) = \Sigma_i w_i (|x - a_i| + |y - b_i|)$$
(7)
$$\text{Such that} \quad c_{j1}x + c_{j2}y \le c_{j3}, \ j = 1,2,...,n, \ \ (x,y) \in X.$$

where w_i's denote the cost per unit distance between the new facility S and demand points $P_i = (a_i, b_i)$. To describe the distances the taxicab geometry or city block distance is used since the problem is considered in an urban area. Euclidean distance could also be used.

The same problem can also be formulated as follows:

Find $S = (x, y)$ such that

$$
\begin{aligned}
& g_1 \geq g_1^0 \\
& g_2 \leq g_2^0 \\
& g_3 \leq g_3^0 \\
& |x - a_i| + |y - b_i| \geq g_1,\ \forall i \\
& |x - a_i| + |y - b_i| \leq g_2,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n, \quad (x, y) \in X.
\end{aligned} \tag{8}
$$

where g_1^0, g_2^0, g_3^0 are the three goals prescribed by the decision maker. One may use positive ideal solution (g_1^*, g_2^*, g_3^*) to represent the goals and tolerances of fuzzy goals may be the differences of the positive and negative ideal solutions (g_1^-, g_2^-, g_3^-).

Positive and negative ideal solutions are the solutions of the following problems:

g_1^*: Max g_1 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \geq g_1,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned}
$$

g_1^-: Min g_1 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \geq g_1,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned}
$$

g_2^* : Min g_2 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \leq g_2,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned}
$$

g_2^-: Max g_2 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \leq g_2,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned}
$$

g_3^*: Min g_3 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \leq g_3,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned}
$$

g_3^-: Max g_3 such that

$$
\begin{aligned}
& |x - a_i| + |y - b_i| \leq g_3,\ \forall i \\
& c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.
\end{aligned} \tag{9}
$$

Then using the values of (g_1^*, g_2^*, g_3^*) and (g_1^-, g_2^-, g_3^-), the fuzzy limitations for the goals $(\tilde{g}_1, \tilde{g}_2, \tilde{g}_3)$ are obtained as follows.

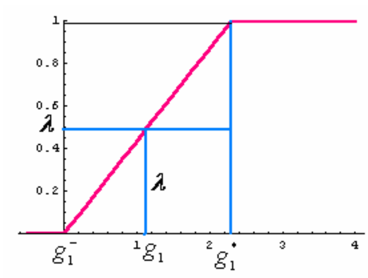


Figure 2. Fuzzy goal $\tilde{g}_1$.

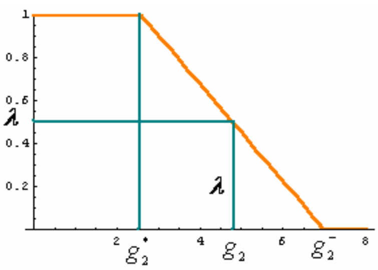


Figure 3. Fuzzy goal $\tilde{g}_2$.

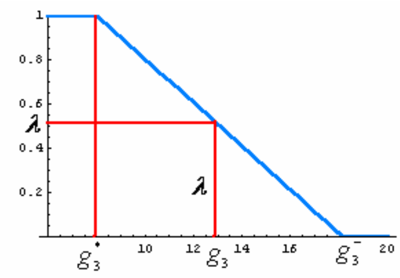


Figure 4. Fuzzy goal $\tilde{g}_3$.

With these fuzzy goals, problem (8) can be expressed as:

find $S = (x, y)$ Such that

$$g_1 \geq \tilde{g}_1,\ g_2 \leq \tilde{g}_2,\ g_3 \leq \tilde{g}_3 \tag{10}$$
$$|x - a_i| - |y - b_i| \geq g_1,\ \forall i,\ |x - a_i| - |y - b_i| \leq g_2,\ \forall i$$
$$c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.$$

To transform the problem (10) into a crisp problem with only one objective, we get the $\lambda -$cuts:

Maximize λ such that

$$g_1 - g_1^- \geq \lambda(g_1^* - g_1^-),\ -g_2 + g_2^- \geq \lambda(-g_2^* + g_2^-),\ -g_3 + g_3^- \geq \lambda(-g_3^* + g_3^-) \tag{11}$$
$$|x - a_i| + |y - b_i| \geq g_1,\ \forall i,\ |x - a_i| + |y - b_i| \leq g_2,\ \forall i$$
$$c_{j1}x + c_{j2}y \leq c_{j3},\ j = 1,2,...,n,\ (x, y) \in X.$$

3 AN APLICATION

Let $\{P_i\} = \{(0, 2), (2, 1), (3, 4)(1, 3.5)(2.5, 2)(2, 4)(1, 0)(.5, .5)\}$ be the locations of the eight demand points, $S = (x, y)$ is the location of the new facility. Let unit costs per unit distance between the new facility S and the demand point $P_i(a_i, b_i)$ be $\{w_i\} = \{0.7, 2.1, 1.5, 1., 1., 1., 1., 1.\}$. In this case the values of (g_1^*, g_2^*, g_3^*) are found to be $(1.5, 3.0, 19.0)$ and the values of (g_1^-, g_2^-, g_3^-) are found to be $(0.0, 7.0, 35.2)$.

Hence in the feasible region $X = \{(x, y) | 4x + 57 \leq 20, 8x + 3y \leq 24, x, y \geq 0\}$, the problem (16) becomes:

Maximize λ such that

$$g_1 \geq 1.5\lambda,\ g_2 \leq 7 - 4.\lambda,\ g_3 \leq 35.2 - 16.2\lambda$$
$$x + y - g_1 \geq c3,\ x - y + g_1 \leq c2,\ x + y + g_1 \leq c1$$
$$x + y + g_2 \geq c1,\ x - y + g_2 \geq c2,\ x + y - g_2 \geq c3,\ x - y - g_2 \geq c4$$
(12)
$$4x + 5y \leq 20,\ 8x + 3y \leq 24.\ x, y \geq 0.$$

Where $c1 = \max_i(a_i + b_i),\ c2 = \max_i(a_i - b_i),\ c3 = \min_i(a_i + b_i),\ c4 = \min_i(a_i - b_i)$.

The solution of the above problem is found to be $x = 1.696,\ y = 2.642,\ \lambda = 0.831$. For this optimum supply point $S = (1.31, 2.23)$, one has

$$g_1(x, y) = 1.419,\ g_2(x, y) = 3.458,\ g_3(x, y) = 20.838. \tag{13}$$

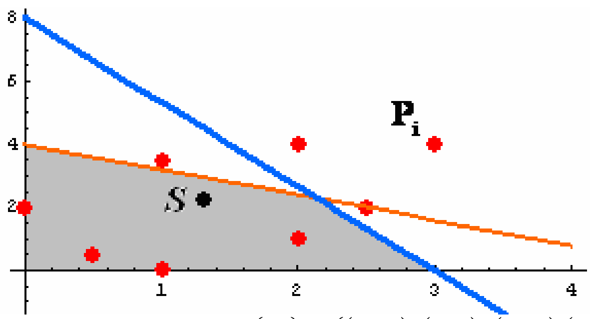


Figure 5. The three demand points $\{P_i\} = \{(0,\ 2),(2,1),(3,\ 4)(1,3.5)(2.5,\ 2)(2,\ 4)(1,\ 0)(.5,\ .5)\}$, and the new optimum supply point $S = (1.31, 2.23)$.

4 DISCUSSION

In this article, to deal with and optimization problem with three objectives, a method due to Narasimhan, R. is used. The maximum and minimum solutions of each sub problems are found, then using this information, goals are transformed into fuzzy equalities. Then a crisp symmetric optimization problem is obtained by $\lambda -$cuts. The number of the demand points can be increased to represent a real world problem easily.

References

[1] Abo-Sinna, M.A., Multiple objective (fuzzy) dynamic programming problems: a survey and some applications, *Applied Mathematics and Computation* 157/3 (2004) 861-888

[2] Bhattacharya, J.R. Rao, and R.N. Twari, Fuzzy multi-criteria facility location, *Fuzzy Sets and Systems* 51 (1992) 277-287.

[3] Buckley, J.J., Multiobjective possibilistic linear programming, *Fuzzy Sets and Systems*, 35 (1990) 23-28.

[4] Chanas, S., D. Kutcha, Multiobjective programming in optimization of the interval objective function-a generalized approach, *European Journal of Operational Research* 94 (1996) 594-598.

[5] Deb, K.,Multi-Objective Optimization using Evolutionary Algorithms, John Wiley & Sons, England (2001).

[6] Dey, J.K., S. Kar, and M. Maiti, An interactive method for inventory control with fuzzy lead-time and dynamic demand, *European Journal of Operational Research* 167 (2004) 381-397.

[7] Eatman, J.L., and Sealey, Jr., A multiobjective linear programming model for commercial bank balance sheet management, *Journal of Bank research* 9 (1979) 227-236.

[8] French, S., Interactive multiobjective programming: Its aims, applications, and demands, *Journal of Operational Research Society* 30 (1984) 824-837.

[9] Hannan, E.L., On the efficiency of the product operator in fuzzy programming with multiple objectives, *Fuzzy Sets and Systems* 2 (1979) 259-262.

[10] Hannan, E.L., Linear programming with multiple fuzzy goal, *Fuzzy Sets and Systems* 6 (1981) 235-248.

[11] Hannan, E.L., Fuzzy decision making with multiple objective and discrete membership functions, *International Journal of man-machine Studies* 18 (1983) 49-54.

[12] Hwang, C.L., S.R. Paidy, and K. Yoon, Mathematical Programming with multiple objectives: a tutorial, Computers and Operations research, 7 (1980) 5-31.

[13] Ishibuchi, H., and H. Tanaka, Multiobjective programming in optimization of the interval objective function, *European Journal of Operational Research* 48 (1990) 219-225.

[14] Lai, Y.J., and Hwang, C.L. *Fuzzy Multi Objective Decision Making*, Springer, 2nd ed. (1996).

[15] Narasimhan, R., Goal programmimg in a fuzzy environment, *Decision Sciences* 11 (1980) 325-338.

[16] Li, X., B. Zhang and H. Li, Computing efficient solutions to fuzzy multiple objective linear programming problems, *Fuzzy Sets and Systems* 157/10 (2006) 1328-1332.

[17] Rommenfanger, H.,and R.Slowinski, Fuzzy linear programming with single or multiple objective functions, In Slowinski R.(Ed), Fuzzy Sets in Decision Analysis, Kuwler, Boston, (1998)

[18] Sakawa, M., and H. Yano, Multiobjective fuzzy linear regression analysis for fuzzy input-output data, *Fuzzy Sets and Systems* 47 (1992) 173-181.

INVENTORY MANAGEMENT IN SUPPLY CHAIN CONSIDERING QUANTITY DISCOUNTS

Anton Čižman
University of Maribor, Faculty of organizational sciences
Kranj, Kidričeva cesta 55a
E-mail: anton.cizman@fov.uni-mb.si

Abstract: Inventory, transportation, facilities and information are four major drivers that can improve the performance of any supply chain in terms of responsiveness and efficiency. An important supply chain driver is inventory which is a major source of cost and thus has significant impact on the supply chain profitability. The paper shows the decision support model for planning optimal cycle inventory in the case of all-units quantity discount which is illustrated by practical example using POM-QM analytical software tool.

Keywords: management, supply chain, simulation, decision support system, inventories, optimization, POM-QM software

1. Introduction

A Supply Chain (SC) consists of all stages involved, directly or indirectly, in fulfilling a customer request. The SC not only includes the manufacturer and suppliers, but also transporters, warehouses, retailers, and customers themselves [1, 4, 6]. A SC is dynamic and involves the constant flow of information, product, and funds between different stages. Each stage of the SC performs different processes and interacts with other stages of the supply chain.

The objective of every SC is to maximize the overall value generated. For most commercial SCs, value will be strongly correlated with SC profitability, the difference between the revenue generated from the customer and the overall cost across the supply chain. SC success should be measured in terms of supply chain profitability and not in terms of the profits at an individual stage. All flows of information, product and funds generate costs within the SC. Therefore, Supply Chain management (SCM) involves the management of all flows between and among stages in supply chain to maximize total profitability.

The purpose of the paper is to show how decision support system (DSS) that utilizes the basic economic order quantity (EOQ) model is able to improve SC profitability by means of reducing total inventory costs in the case of quantity discounts [2, 3, 5]. In the first part of the paper we present shortly basic features of SCM, in the second part of this paper we have focused on the quantitative decision support model for Inventory Management and then an example is given using POM-QM software [4, 7] to illustrate the applicability of such models in practice.

2. Basic features of SCM

SCM is the process of planning, implementing, and controlling the operations of the SC with the purpose to satisfy customer requirements as efficiently as possible. SCM can be defined as the management of upstream and downstream relationships with suppliers and customers to deliver greater customer value at *less cost* to the supply chain as a whole. Thus the focus of SCM is upon the management of relationships in order to achieve a more profitable outcome for all parties in the chain. SCM encompasses the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. A typical SC may involve a variety of stages, such as:

• Customers
• Retailers
• Wholesalers/distributors
• Manufacturers
• Component/raw material suppliers

Successful SCM requires several decisions relating to the flow of information, product, and funds. These decisions fall into three categories or phases: design, planning, and operation, depending on the frequency of each decision and the time frame over which a decision phase has an impact [1, 4].

The strategic fit requires that a company achieves the balance between responsiveness and efficiency in its SC that best meets the needs of the company's competitive strategy. Strategic fit means that both the competitive and SC strategy have the same goal. A company can improve SC performance in terms of responsiveness and efficiency by means of examination the four drivers of supply chain performance: inventory, transportation, facilities, and information. These drivers not only determine the supply chain's performance in terms of responsiveness and efficiency, they also determine whether strategic fit is achieved across the supply chain.

3. Cycle inventory management considering quantity discounts

Inventory is a major source of costs in a SC, and it has a huge impact on responsiveness. Inventory exists in the supply chain because of a mismatch between supply and demand. An important role that inventory plays in the supply chain is to increase the amount of demand that can be satisfied by having the product ready and available when the customer wants it. Another significant role inventory plays is to reduce cost by exploiting any economies of scale that may exist during both production and distribution. Inventory is spread throughout the SC from raw materials to work in process to finished goods that suppliers, manufacturers, distributors, and retailers hold.

One assumption of the most basic version of the EOQ model is that the cost of the item is not affected by the order size. *Quantity discounts* sometimes are offered for externally purchased items. In addition, economies of scale may result in different unit costs for different production lot sizes when items are produced internally. In this paper we present the decision support model for planning EOQ when all unit quantity discounts are available.

The procedure for finding the best order quantity in this type of situation [2, 5] is as follows:

1. Consider the lowest price, and solve the basic EOQ formula for the EOQ at this price. If the EOQ is feasible, this is the best quantity, so stop; otherwise go to step *2*.
2. Solve for the EOQ for the next higher price. If this EOQ is feasible, proceed to step *4*.
3. If the EOQ is not feasible, repeat step 2 until a feasible EOQ is found.
4. Compute the Total Costs (TC) for the feasible EOQ and for all the greater quantities where the price breaks occur. Select the quantity with the lowest TC.

3.1 The decision support model for planning EOQ in the case of quantity discounts

The decision support model includes four components: the user-interface, the modeling base, the database, and solution techniques [2, 6]. The integration of this four components by means of POM-QM software which is a user-friendly package for quantitative methods and production/operations management [7], is presented in Fig. 1.

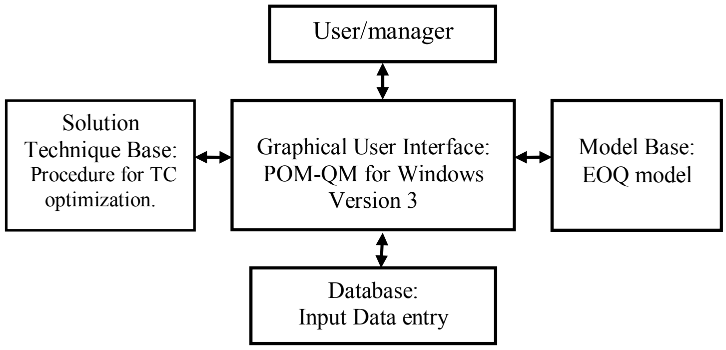


Fig. 1: The structure of a decision support model

The Problem: Drugs Online (DO) is an online retailer of prescription drugs and health supplements. Vitamins represent a significant percentage of their sales. Demand for vitamins is 10.000 bottles per month. DO incurs a fixed order placement, transportation, and receiving cost of 100 € each time an order for vitamins is placed with the manufacturer. DO incurs a holding cost of 20 percent per year. The price charged by the manufacturer varies according to the all unit discount pricing schedule shown. Evaluate the number of bottles that the DO manager should order in each lot [1].

Input data and results of the problem solution using POM-QM analytical software tool are given in Tab.1 and Fig. 4.

Table 1: Input data and results of total inventory cost optimization

Input Data				***Results***	
Parameter	**Value**			**Parameter**	**Value**
Demand rate(D)	**120000**	**xxxxxxx**	**xxxxxxx**	***Optimal order quantity (Q*)***	***10000***
Setup/Ordering cost(S)	**100**	**xxxxxxx**	**xxxxxxx**	**Maximum Inventory Level (Imax)**	**10000**
Holding cost(H)	**20%**	**xxxxxxx**	**xxxxxxx**	**Average inventory**	**5000**
				Orders per period(year)	**12**
	From	**To**	**Price**	**Annual Setup cost**	**1200**
1	**0**	**4999**	**3**	**Annual Holding cost**	**2920**
2	**5000**	**9999**	**2,96**		
3	**10000**	**999999**	**2,92**	**Unit costs (PD)**	**350400**
				Total Cost (€)	***354520***

The results of optimal solution show that DO manager should order 10.000 bottles each time (12 times per year) to fulfill the annual customer demand. This EOQ assures the minimum total annual relevant costs ***354.520,00 €.*** The results also show how optimization techniques can easily be applied by means of user-friendly software tool POM-QM for Windows Version 3 for solving the problem of the balance between efficiency (profit) and the responsiveness (customer demands) in SC.

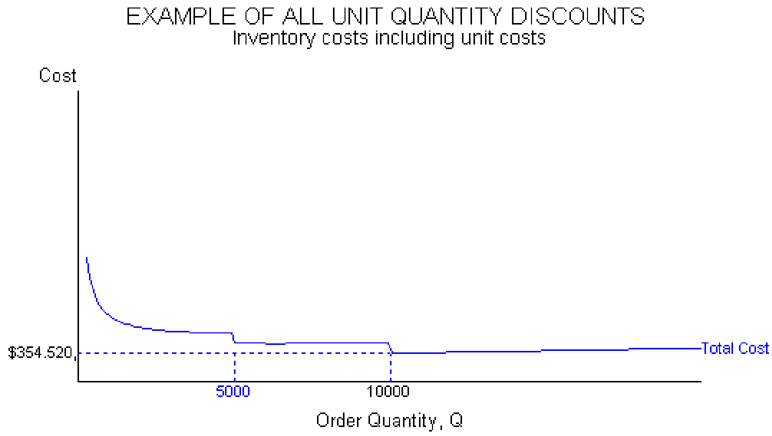


Figure 2: Total Costs vs. order quantity with price breaks

4. Conclusion

If the manufacturer in preceding example sold all bottles for 3€, it would be optimal for DO to order in lots of 6.324 bottles. The quantity discount is an incentive for DO to order in larger lots of 10.000 bottles, raising both the cycle inventory and the flow time. The impact of the discount is further magnified if DO works hard to reduce its fixed ordering cost from *S =100* € to *S* = 4€. The optimal lot size in the absence of a discount would be 1.265 bottles. In the presence of all unit quantity discounts, the optimal lot size will still be 10.000 bottles. In this case, the presence of quantity discounts leads to an eight-fold increase in average inventory as well as flow time at DO. This means that in many SCs, quantity discounts contribute more to cycle inventory then fixed ordering costs.

Pricing schedules with all unit quantity discounts encourage retailers to increase the size of their lots to take advantage of price discounts, which adds to the average inventory and flow time in a supply chain. This increase in inventory raises a question about the value that all unit quantity discounts offer in the supply chain.

We can conclude that quantity discounts can be valuable in SC for two following reasons: *improved coordination in SC and extraction of surplus through price discrimination.*

References

1. Chopra, S., P. Meindl (2004, 2001): Supply Chain Management: Strategy, Planning and Operation, Prentice Hall, New Jersey.
2. Čižman, A. (2002): Logistični management v organizaciji, Moderna organizacija, Kranj.
3. Čižman, A. (2003): Učinkovit management zalog – pomemben strateški cilj podjetja, Organizacija, 36, str. 242-249.
4. Čižman, A.: uporaba programa POM-QM za planiranje cikličnih zalog v oskrbovalni verigi, Zbornik posvetovanja, Dnevi slovenske informatike, Portorož, Slovenija, 11.-13. april 2007.
5. Dilworth, James, B. (1996), Operations Management, McGraw-Hill.
6. Lambert, M. Douglas, Stock, R. James, Ellram, M. Lisa (1998): Fundamentals of Logistics Management, McGraw-Hill.
7. Weiss, J. Howard (2005), POM-QM for Windows, Version 3, Software for Decision Sciences, Pearson Prentice Hall, New Jersey, http://www.prenhall.com/weiss.

ECONOMETRIC MODEL OF INVESTMENT AS PART OF CROATIAN GDP

Fran Galetić, Faculty of Economics and Business Zagreb, fgaletic@efzg.hr
Nada Pleli, Faculty of Economics and Business Zagreb, npleli@efzg.hr

Abstract: Gross domestic product (GDP) consists of: consumption, investment, government spending and net exports. Investment is a part that depends on two elements: interest rate and production. The models we have developed show the impact of these variables on Croatian investment. At the end there is a model that includes both these variables in the calculation of investment.

Keywords: GDP, investment, Croatia, interest rate, production

1. INTRODUCTION

Investment is one of four parts of the usual calculation of gross domestic product. Croatian GDP is calculated and analyzed very often, but there are few analysis based on econometric models. Due to this, we wanted to make a model that best describes elements that influence investment: production and interest rate. The model is based on historical data from 1997 to 2004.

2. ELEMENTS OF GDP

Gross domestic product (GDP) of any country is usually divided on its components: consumption, investment, government spending and net exports. So, the GDP equation is normally written as:

$$Y = C + I + G + NX \quad (1)$$

Consumption (C) is referred to goods and services purchased by consumers. It is always the largest component of GDP.

The second component is investment (I). It is the sum of two components. The first one is called nonresidential investment – this is purchase of new plants and machines by firms. The second one, residential investment, is the purchase of real estates by people. These two types are calculated together into „investment“ because they both refer to the future. The term investment is used in economy to refer to the purchase of new capital goods, such as machines, buildings or houses.[1]

The third component of GDP is government spending (G). It refers to all goods and services purchased by the state or local governments.

Net export (NX) is the difference between import (Q) and export (X). Import is defined as purchases of foreign goods and services by domestic consumers, while export refers to purchases of domestic goods and services by foreigners. If exports exceed imports, a country is running a trade surplus. That means that the trade balance is positive. The other case is by negative trade balance – in that case a country has a trade deficit.

3. INVESTMENT

All models have two types of variables. Variables that depend on other variables in the model and therefore are explained within the model are called endogenous. Other variables are not explained in the model – they are called exogenous.

---

[1] Blanchard: Macroeconomics

Investment depends mostly on two variables: production and interest rate.

Firms faced with high sales need to increase their production, so they have to buy new plants and equipment. Contrary to this, firms with low sales do not have such need, so they don't spend much on investment. This is the reason why production is positive correlated to investment.

$$I = I(Y^{+}) \quad (2)$$

When a firm is taking the decision to invest, it must consider the interest rates. To buy new plant or equipment, the firm must borrow money by either taking a loan from the bank or issuing bonds. If the interest rate is high, it is less likely that the firm will borrow the needed money. At a high interest rate, the profits from new equipment will not be high enough to cover interest payments. So interest rate is negatively correlated to investment.

$$I = I(i^{-}) \quad (3)$$

Now we can write the whole investment relation:

$$I = I(Y^{+}, i^{-}) \quad (4)$$

The signs indicate the correlation between the variable and investment. The plus by Y indicates that production is positively correlated to investment – an increase in production leads to an increase in investment. The minus sign by interest rate means a negative correlation – an increase in interest rate leads to a decrease in investment.

4. INVESTMENT IN CROATIA

The data for Croatia are collected in the publications and web-sites of Croatian Central Bureau of Statistics and Croatian National Bank.

Table 1: *Investment, nominal interest rate and GDP in Croatia from 1997 to 2004*

Year	Investment	Interest rate	GDP
1997.	29936	6,00	123811
1998.	32066	5,90	137604
1999.	32956	7,57	141579
2000.	33281	6,40	152519
2001.	36984	5,90	165639
2002.	44105	5,55	181231
2003.	56662	4,50	198422
2004.	60513	4,50	212826

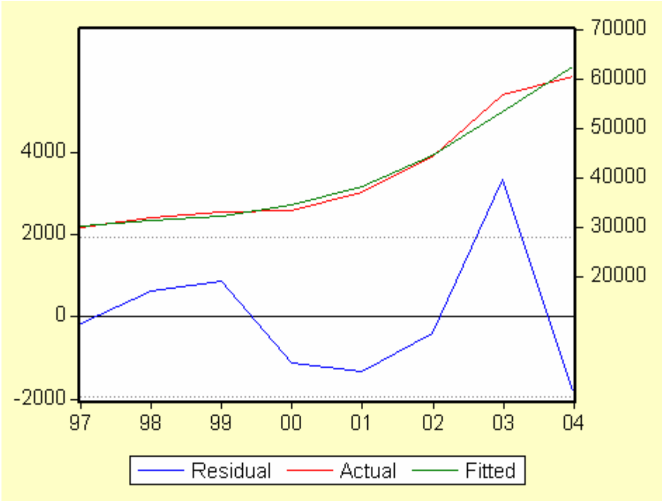
Source: DZS and HNB (8. and 9.)

The analysis made by EViews shows the following results:

Dependent Variable: I
Method: Least Squares
Sample: 1997 2004
Included observations: 8
I = C(1) + C(2)*GDP + C(3)*GDP*GDP

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	77605.36	26610.02	2.916397	0.0332
C(2)	-0.817761	0.322656	-2.534465	0.0522
C(3)	3.50E-06	9.53E-07	3.675969	0.0144

R-squared	0.980864	Mean dependent var	40812.88
Adjusted R-squared	0.973209	S.D. dependent var	11820.45
S.E. of regression	1934.753	Akaike info criterion	18.25334
Sum squared resid	18716355	Schwarz criterion	18.28313
Log likelihood	-70.01338	Durbin-Watson stat	2.451301



The model :

$$I = a + b_1 \cdot GDP + b_2 \cdot GDP^2 \qquad (5)$$

is applied to Croatian values and with the parameters we get

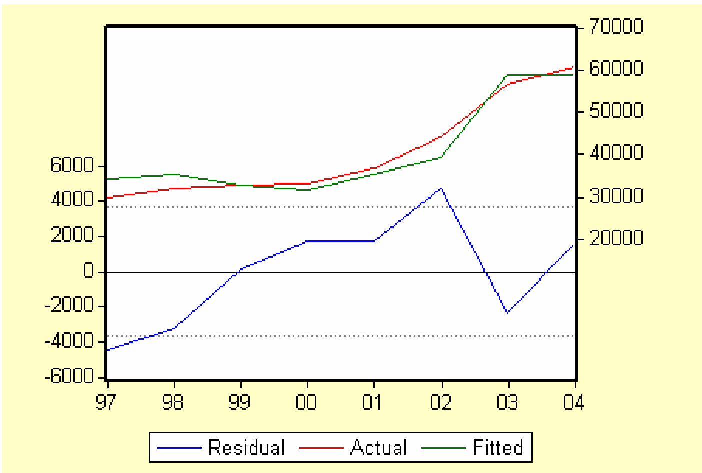
$$I = 77605{,}36 - 0{,}817761 \cdot GDP + 0{,}0000035 \cdot GDP^2 \qquad (6)$$

R-square is 98%, so the model is very good.
In the same way we calculate the interest rate:

Dependent Variable: I
Method: Least Squares
Sample: 1997 2004
Included observations: 8
I = C(1) + C(2)*INT + C(3)*INT*INT

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	268456.8	41275.46	6.504028	0.0013
C(2)	-69172.30	14143.03	-4.890912	0.0045
C(3)	5025.338	1192.441	4.214327	0.0084

R-squared	0.932742	Mean dependent var	40812.88
Adjusted R-squared	0.905839	S.D. dependent var	11820.45
S.E. of regression	3627.183	Akaike info criterion	19.51030
Sum squared resid	65782286	Schwarz criterion	19.54009
Log likelihood	-75.04119	Durbin-Watson stat	1.350391



$$I = 268456,8 - 69172,3 \cdot INT + 5025,34 \cdot INT^2 \quad (7)$$

R-square is 93%, so the model is good. All variables are significant at alpha = 5%.
Now let's find the regression equation that consists of both interest rate and GDP as variables that influence investment. The equation that best fits is

$$I = 138242,9 - 39834,18 \cdot INT + 2940,92 \cdot INT^2 + 0,195165 \cdot GDP \quad (8)$$

This is shown on the following analysis:

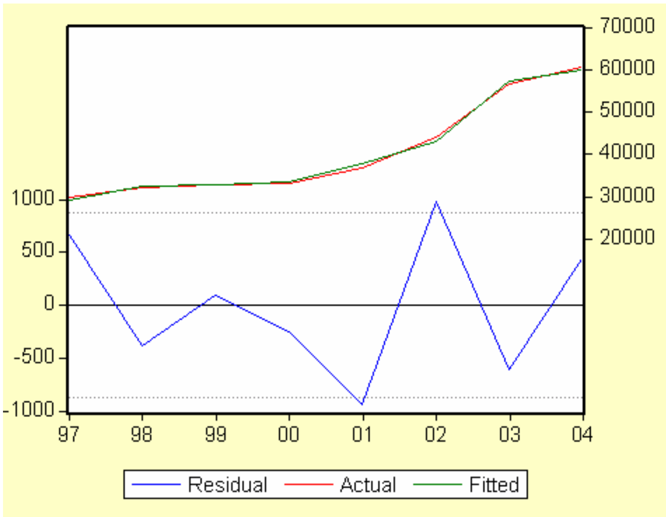
Dependent Variable: I
Method: Least Squares
Sample: 1997 2004
Included observations: 8
I = C(1) + C(2)*INT + C(3)*INT*INT + C(4)*GDP

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	138242.9	17488.86	7.904625	0.0014
C(2)	-39834.18	4704.088	-8.467994	0.0011
C(3)	2940.920	368.3337	7.984392	0.0013
C(4)	0.195165	0.021554	9.054877	0.0008

R-squared	0.996871	Mean dependent var	40812.88
Adjusted R-squared	0.994525	S.D. dependent var	11820.45
S.E. of regression	874.6383	Akaike info criterion	16.69235
Sum squared resid	3059968.	Schwarz criterion	16.73207
Log likelihood	-62.76940	Durbin-Watson stat	3.013540

All four parameters are significant regarding the level of confidence of 1%. R-square is 99,69%, which suggests that the model is excellent.

This can also be seen on the following graph showing actual and expected values.



5. CONCLUSION

In this paper we have shown the model of investment in Croatia. First we developed a model based on production, and then another model based on interest rates. At the end we showed the model based on both production and interest rates. The model (7)

$$I = 138242,9 - 39834,18 \cdot INT + 2940,92 \cdot INT^2 + 0,195165 \cdot GDP$$

has very high R-square, which means that it is very good picture of reality.

References and bibliography

1. Babić, M. (2000) Makroekonomija, Mate, Zagreb
2. Baltagi, B.H; Ed. (2003) A Companion to Theoretical Econometrics, Blackwell Publishing
3. Blanchard, O. (1997) Macroeconomics, Prentice-Hall, Inc.
4. Družić, I. et al. (2003) Hrvatski gospodarski razvoj, Politička kultura i Ekonomski fakultet Zagreb
5. Favero, C.A. (2001) Applied Macroeconometrics, Oxford University Press
6. Gartner, M. (2002) Macroeconomics, Financial Times/Prentice Hall
7. Mankiw, N.G. (2002) Macroeconomics, 5th edition, Worth Publishers
8. Central Bureau of Statistics publications and web www.dzs.hr
9. Croatian National Bank publications and web www.hnb.hr

PREEMPTIVE FUZZY GOAL PROGRAMMING IN FUZZY ENVIRONMENTS

J. Jusufovic*, A. Omerovic*, Mehmet Can**
**Faculty of Economics and Business Administration*
***Faculty of Arts and Social Sciences,*
International University of Sarajevo, Paromlinska 66,
71000 Sarajevo, Bosnia and Herzegovina
E-mails: jjusufovic@ius.edu.ba amir1608@gmail.com mcan@ius.edu.ba

Abstract: There are a variety of efficient approaches to solve crisp multiple objective decision making problems. However in the real life the input data may not be precisely determined because of the incomplete information. This paper deals with a method which can be applied to solve fuzzy multi objective production marketing problems.

Keywords: fuzzy goal programming, fuzzy environments

1 INTRODUCTION

In a standard multiple goal programming, goals and constraints are defined precisely. Fuzzy goal programming has the advantage of allowing for the vague aspirations of decision makers, which are quantified by some natural language rules.

To our knowledge, first R. Narasimhan (1980) introduced fuzzy set theory into objective programming. Since then many achievements have been added to the literature. In the literature there are several approaches for solving fuzzy goal programming, and their application to the production marketing problem. Among these methods there are the ones using preemptive fuzzy goal programming, interpolated membership function, weighted additive model, preference structure on aspiration levels, and nested priority.

2 MULTIPLE FUZZY GOAL PROGRAMMING

In a multiple goal programming problem, the optimal realization of multiple objectives is desired under a set of constraints imposed by a real life environment. If the goals and constraints are all expressed with equalities, we have a completely symmetric formulation

Find $\mathbf{x}$
Such that $\mathbf{Ax} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}.$ (1)

Where $\mathbf{x}$ is the vector of variables, $\mathbf{b}$ is the vector of the goals and available resources, and $\mathbf{A}$ is the matrix of the coefficients. In the cases when the decision maker is not precise in goals and restrictions, the linguistic statements such as “around **b**” will be used. In this case the above crisp goal programming problem becomes

Find $\mathbf{x}$ such that $\mathbf{Ax} = \tilde{\mathbf{b}},\ \mathbf{x} \geq \mathbf{0}.$ (2)

Where the fuzzy components b_i of the fuzzy vector **b** can be represented by, for example, triangular fuzzy numbers:

$$\mu_i(z) = \begin{cases} \left(z - (b_i - d_{i1})\right)/d_{i1}, & b_i - d_{i1} \leq z \leq b_i, \\ \left((b_i + d_{i2}) - z\right)/d_{i2}, & b_i \leq z \leq b_i + d_{i2}, \\ 0, & \text{elsewhere}. \end{cases} \quad (3)$$

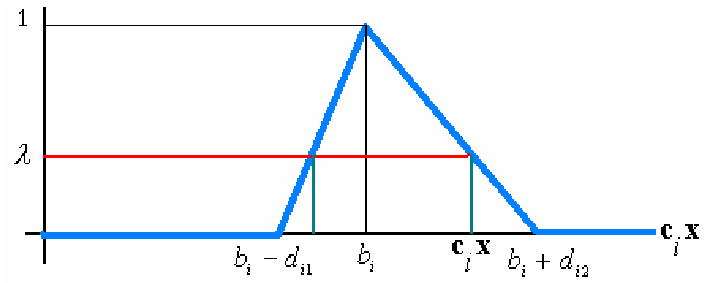


Figure 1. Fuzzy components b_i of the fuzzy vector **b.**

To have a membership number at least λ, $\mathbf{c}_i\mathbf{x}$ must remain in the interval

$$b_i - d_{i1} + \lambda d_{i1} \le \mathbf{c}_i\mathbf{x} \le b_i + d_{i2} + \lambda d_{i2} \tag{4}$$

that is

$$(\mathbf{c}_i\mathbf{x} - (b_i - d_{i1}))/d_{i1} \ge \lambda,\ ((b_i + d_{i2}) - \mathbf{c}_i\mathbf{x})/d_{i2} \ge \lambda. \tag{5}$$

Hence the above fuzzy goal programming problem is the maximum satisfaction problem of the fuzzy equations, and this goal can be achieved by the solution of the below crisp linear programming problem Lai, and Wang (1996).

Max λ such that for all *i*,

$$(\mathbf{c}_i\mathbf{x} - (b_i - d_{i1}))/d_{i1} \ge \lambda,\ ((b_i + d_{i2}) - \mathbf{c}_i\mathbf{x})/d_{i2} \ge \lambda. \tag{6}$$

$\lambda \in [0,1]$, and $\mathbf{x} \ge \mathbf{0}$.

In some decision problems, some goals are so important that unless these goals are reached, the decision maker would not consider the achievement of other goals. The method of differentiating goals according their importance is called preemptive fuzzy goal programming

3 PREEMPTIVE FUZZY GOAL PROGRAMMING

Let us assume the existence of **K** priority levels in the fuzzy goal programming problem (2). The problem is then partitioned in **K** sub problems which can be transformed into a standard goal programming problem (6).

The goal set $G_r(x)$ has higher priority then the goal set $G_s(x)$ if $r < s$. After ordering goals, we will solve the first sub problem by considering the first priority goals $(G_1(x))$ only:

Find **x**

such that $g_{1i}(\mathbf{x}) = \tilde{b}_{1i},\quad i = 1,2,..,m_1,\quad \mathbf{x} \ge \mathbf{0},\ \ g_{1i}(\mathbf{x}) \in G_1(\mathbf{x})$ (7)

where m_1 is the number of the goals in the set of first priority goals $(G_1(x))$. Next the goals in the second priority level $(G_2(x))$ will be considered; under the condition that achievement of the first sub problem is satisfied:

Find **x**

such that $g_{1i}(\mathbf{x}) = b_{1i} - d_{1i}(g_{1i}^{-} - g_{1i}^{*}),\quad i = 1,2,..,m_1,\quad \mathbf{x} \ge \mathbf{0},\ \ g_{1i}(\mathbf{x}) \in G_1(\mathbf{x}),$ (8)

$g_{2i}(\mathbf{x}) = \tilde{b}_{2i},\quad i = 1,2,..,m_2,\quad \mathbf{x} \ge \mathbf{0},\ \ g_{2i}(\mathbf{x}) \in G_2(\mathbf{x}),$

where g_{1i}^{-}, g_{1i}^{*} are the maximum and minimum solutions of the first sub problem. Similarly we can solve the third problem under the condition that full achievements of the first and second sub problems are preserved:

Find $\mathbf{x}$

such that $g_{1i}(\mathbf{x}) = b_{1i} - d_{1i}(g_{1i}^{-} - g_{1i}^{*}), \quad i = 1,2,..,m_1, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{1i}(\mathbf{x}) \in G_1(\mathbf{x}),$

$$g_{2i}(\mathbf{x}) = b_{2i} - d_{2i}(g_{2i}^{-} - g_{2i}^{*}), \quad i = 1,2,..,m_2, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{2i}(\mathbf{x}) \in G_2(\mathbf{x}), \tag{9}$$

$$g_{3i}(\mathbf{x}) = \tilde{b}_{3i}, \quad i = 1,2,..,m_3, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{3i}(\mathbf{x}) \in G_3(\mathbf{x}),$$

where g_{2i}^{-}, g_{2i}^{*} are the maximum and minimum solutions of the second sub problem. This procedure is then repeated until all priority levels are finished.

One can also allow some tolerances in the solution of the sub problems. For example the problem (9) with some tolerances becomes:

Find $\mathbf{x}$

such that

$$g_{1i}(\mathbf{x}) = b_{1i} - d_{1i}(g_{1i}^{-} - g_{1i}^{*}) - p_{1i}, \quad i = 1,2,..,m_1, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{1i}(\mathbf{x}) \in G_1(\mathbf{x}),$$

$$g_{2i}(\mathbf{x}) = b_{2i} - d_{2i}(g_{2i}^{-} - g_{2i}^{*}) - p_{2i}, \quad i = 1,2,..,m_2, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{2i}(\mathbf{x}) \in G_2(\mathbf{x}),$$

(10)

$$g_{3i}(\mathbf{x}) = \tilde{b}_{3i}, \quad i = 1,2,..,m_3, \quad \mathbf{x} \geq \mathbf{0}, \quad g_{3i}(\mathbf{x}) \in G_3(\mathbf{x}),$$

where p_{1i}, p_{2i} are allowable tolerances.

4 AN APLICATION: THE PRODUCTION-MARKETING PROBLEM

Assume decision maker consider sale goals only after the profit goal is absolutely achieved. Then the optimization problem is divided into two sub problems. The first problem deals with the first priority goal, profit. Assume profit is computed through the formula

$$p(x, y) = 80x + 40y$$

in dollars, when monthly sales are (x, y) items from each kind of products, and the fuzzy goal is:

The first problem is:

Find (x, y)

such that $p(x, y) = \tilde{b}_{11}, \quad x, y \geq 0,$

(11)

where $\tilde{b}_1$ is the fuzzy profit goal with $b_{11} = 7000, d_{11} = 1000, d_{12} = 2000.$

This problem is transformed into the crisp linear programming problem:

Maximize λ such that

$$p(x, y) \geq 6000 + 1000\lambda$$

$$p(x, y) \leq 9000 - 2000\lambda$$

(12)

$$0 \leq \lambda \leq 1, \ x, y \geq 0 .$$

The solution of the above problem is found to be $x = 87.5,\ y = 0,\ \lambda = 1$. For this production plan one has $p(x, y) = 7000$ as expected.

Next consider the second priority level of sales

Find (x, y)

such that $p(x, y) = 7000,$

$x = \tilde{b}_{21},$

$y = \tilde{b}_{22},$

$x, y \geq 0.$

(13)

where $\tilde{b}_{21}$ is the fuzzy sell of the first product with $b_{21} = 60,\ d_1 = 10, d_2 = 20$, and $\tilde{b}_{22}$ is the fuzzy sell of the second product with $b_{22} = 40,\ d_1 = 10, d_2 = 20.$

This problem is transformed into the crisp linear programming problem:

Maximize λ such that

$p(x, y) = 7000$

$x \leq 80 - 20\lambda$

$x \geq 50 + 10\lambda$

$y \leq 60 - 20\lambda$

$y \geq 30 + 10\lambda$

(14)

$0 \leq \lambda \leq 1,\ x, y \geq 0$.

By the use of the Linear Programming package under MATHEMATICA, the solution of the above problem is found to be $x = 65,\ y = 45,\ \lambda = 0.75$.

4 DISCUSSION

In this article, to deal with and optimization problem with objectives in two importance levels, the method of preemptive fuzzy goal programming is used. The solution of the first sub problem is found, using this information, the goal in the second priority level is considered; under the condition that achievement of the first sub problem is satisfied. Using $\lambda -$cuts, fuzzy goals are transformed into fuzzy equalities. The obtained crisp linear programming problems are solved by the Linear Programming package under MATHEMATICA. The number of the priority levels can be increased to represent a real world problem without any difficulty.

References

Narasimhan, R., Goal programmimg in a fuzzy environment, *Decision Sciences*, 11 (1980) 325-338.

Bhattacharya, J.R. Rao, and R.N. Twari, Fuzzy multi-criteria facility location, *Fuzzy Sets and Systems*, 51 (1992) 277-287.

Lai, Y.J., and Hwang, C.L. *Fuzzy Multi Objective Decision Making*, Springer, (1996).

GENETIC DISTANCE AND PHYLOGENETIC ANALYSIS (BOSNIA, SERBIA, CROATIA, ALBANIA, SLOVENIA)

Naris Pojskic, Faruk Berat Akcesme[1]
International university of Sarajevo, Faculty of Engineering and Natural Science, Paromlinska 66, 7100 Sarajevo, Bosnia and Herzegovina
[1]e-mail addresses: farberak@yahoo.com

Abstract: In this paper we present several models of genetic distances and phylogenetic analysis. We choose five populations which are closed to each other. We determine determined tree microsatalite loci; we get this information from ALFRED. (The Allele frequency database). According to this chosen genes we measured the genetic distance between this chosen countries and we make an analysis of phylogenetic trees.

Keywords: genetic distance, phylogenetic analysis

1. INTRODUCTION

To measure the genetic distance and make a phylogenetic analysis we should determine;

- Average heterozygosity* and its standard error for each population.
- Standard errors of standard genetic distances.
- Gene diversity and its associate parameters.
- Distances between populations.
- Standard genetic distances between populations.

There are couple of programs which can be use for our task. We used to DISPAN (Genetic Distance and Phylogenetic Analysis) designed by Tatsuya Ota and the Pennsylvania State University.

We determined tree microsatalite loci which are D7S820, CSF1R and D3S1358. We inserted those loci` s alleles values in a matrix and after this application with using DISPAN` s command we enter the input. According to DISPAN` s output we are trying to make a comment.

2. METHOD

The program is written in a C language so we created a directory. To do this, we typed the following:

C:\MD DISPAN

We typed gene frequencies for each locus in the same order for all population, and we followed by the number of genes sampled. (i.e., two times the number of diploid individuals sampled).

We were careful that sum of the gene frequencies is not below 0.9989 or above 1.0011. The gene frequencies for different population were presented in the same order.

When we were finished with designed to our input, we started with DISPAN;

* heterozygous: An organism is a *heterozygote* or is *heterozygous* at a locus or gene when it has different alleles occupying the gene` s position in each of the homologues chromosomes,

```
Microsoft Windows [Version 6.0.6000]
Copyright (c) 2006 Microsoft Corporation.  All rights reserved.

C:\Users\user>cd..

C:\Users>cd..

C:\>cd dispan

C:\DISPAN>gnkdst -ds -finput_file.dat
 gnkdst: essential paramters registered in the file input_file.dat
         the number of populations = 5
         the number of loci = 3
         the number of registered loci = 3
         the maximum number of alleles = 9

         O.K. ?  { Y>es, N>o } :y
 prepgn: reading @Locus 1:D3S1358
 prepgn: reading @Locus 2: D7S820
 prepgn: reading @Locus 3: CSF1R

C:\DISPAN>_
```

We have input file which names `input_file'. In our commend window we can check our input file, which are given by DISPAN.

We had an opportunity to check our input file in DISPAN (the number of population, the number of loci and the maximum number of alleles was shown.

For getting phylogenetic tree DISPAN has a special commend which donated by `tn'.

```
Microsoft Windows [Version 6.0.6000]
Copyright (c) 2006 Microsoft Corporation.  All rights reserved.

C:\Users\user>cd..

C:\Users>cd..

C:\>cd dispan

C:\DISPAN>gnkdst -ds -finput_file.dat -tn
 gnkdst: essential paramters registered in the file input_file.dat
         the number of populations = 5
         the number of loci = 3
         the number of registered loci = 3
         the maximum number of alleles = 9

         O.K. ?  { Y>es, N>o } :y
 prepgn: reading @Locus 1:D3S1358
 prepgn: reading @Locus 2: D7S820
 prepgn: reading @Locus 3: CSF1R
 gnkdst: done for a NJ tree with standard genetic distances

C:\DISPAN>treeview_
```

With this commend DISPAN became ready for drawing the phylogenetic tree.

3. RESULT

Our task was completed successfully.

We managed:

- Average heterozygousity and its standard error
- Standard genetic distances
- Standard error of standard genetic distances

(As you see below we have all information which we needed)

Average heterozygousity and its standard error
(population 1) BOSNIA: 0.784241 ñ 0.017878
(population 2) SERBIA: 0.763635 ñ 0.023092
(population 3) CROATIA: 0.763501 ñ 0.037001
(population 4) ALBANIA: 0.763794 ñ 0.025516
(population 5) SLOVENIA: 0.781067 ñ 0.026097

All loci Gst 0.033423 Ht 0.792647 Hs 0.766155

matrix: Standard genetic distances

	1	2	3	4
2	0.2288			
3	0.1286	0.1179		
4	0.2064	0.1042	0.1951	
5	0.0722	0.0584	0.0729	0.1307

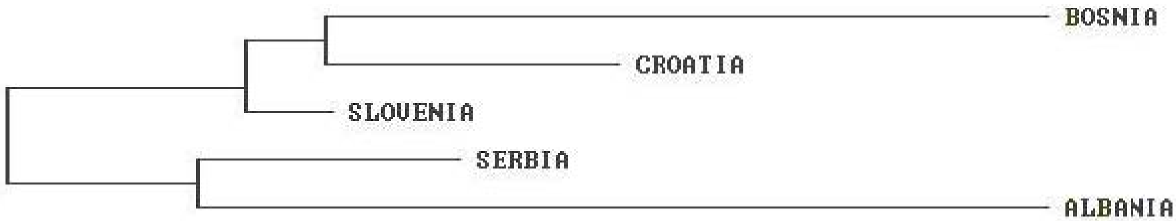
matrix: Standard error of standard genetic distances

	1	2	3	4
2	0.1512			
3	0.0273	0.0945		
4	0.0882	0.1010	0.0722	
5	0.0390	0.0463	0.0455	0.0840

matrix: DA distances

	1	2	3	4
2	0.1724			
3	0.0928	0.0743		
4	0.1576	0.1011	0.1685	
5	0.0797	0.0550	0.0428	0.1401

We get the phylogenetic tree also!



4. DISCUSSION

As a result there is no big differentiation according to observed microsatalite loci (D7S820, CSF1R and D3S1358) between those populations. (Bosnia, Serbia, Croatia, Albania, Slovenia).

Our phylogentic tree structure was shown the relationship with a group of this population. Phylogenetic trees are widely used to study the relationship among living species and genes.

We should not assume that how it has became apparent that trees are commonly misunderstood, leading to confusion about the concept of a common ancestry.

According to this result we can make a comment with this microsatalite loci differentiation in those population but we can not make a comment about ancestor of this population.

The biggest similarities is between Bosnia and Croatia, Albania is a little bit separated from the other countries but not so much from Serbia. The differences between Slovenia and Bosnia-Croatia are less than Slovenia and Serbia-Albania.

All loci heterozygosity is 0.03; the differences of locus is 3% percent among the this population, and we can conclude that this all five tested locus have the same root.

REFERENCES

Efron (1982) The jackkife, the bootstrap, and other resampling plans. CBMS-NSF. Regional conference series in applied mathematics. No 38. Society for industrial and applied mathematics. Philadelphia, PA.

Felsenstein (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.

Nei, M. (1972) Genetic distances between populations. Am. Nat. 106:283-292.

Nei, M. (1973) Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci., USA 70:3321-3323.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Nei, M., Tajima, F., and Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. J. Mol. Evol. 19:153-170.

Saitou, N. and Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic tree. Mol. Biol. Evol. 4:406-425.

Sneath, P.H.A. and Sokal. R.R. (1973) Numerical Taxonomy. Freeman, San Francisco.

TESTING A COMPUTER VISION ALGORITHM AS AN ALTERNATIVE TO SIGNPOST TECHNOLOGY FOR MONITOR TRANSIT SERVICE RELIABILITY

Roman Starin and Dejan Paliska
University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož, Slovenia
Roman.Starin@fpp.uni-lj.si, Dejan.Paliska@fpp.uni-lj.si

Abstract: This article presents a laboratory test results of computer vision/dead reckoning based monitoring for transit service reliability in urban areas. The traditional dead reckoning and signpost technologies for position determination, which many agencies use, suffers from a number of limitations, including the drift of the dead reckoning system and the incapability of the signpost system to locate a vehicle on a continuous basis. In an urban area a GPS based system also may have problems in positioning; i.e., signal reflect from buildings and other reflective surface or building canyons and overpasses can block some or all of a satellite signal. To overcome these limitations, we suggest an integrated positioning system, consisting of a dead reckoning unit coupled with a computer vision system. In the article the results of testing the suggested computer vision algorithm is also presented.

Keywords: computer vision algorithm, transit service reliability, public transport, monitoring system reliability

1 INTRODUCTION

Maintaining reliable service is important for both transit passengers and transit providers. The literature generally supports the ability of a transit system with high-quality service to attract more users, as well as for poor service to encourage more automobile use [1], [2] [3], [4], [5] [6], [7]. Due to the importance of transit service reliability, the agencies use different technologies to measure reliability, technologies commonly known as AVL (Automatic Vehicle Location) systems. In an urban area the AVL system may have location tracking problems, especially if the system is based on the Global Positioning System (GPS) technology. Building canyons and overpasses can block some or all satellite signals. Interference from wireless and radio communications as well as reflections of the GPS signal from buildings and other reflective surfaces make the utilization of GPS problematic at best. Many AVL systems use additional positioning systems such as dead reckoning and map matching techniques and signpost to maintain the location of a vehicle where GPS fails, increasing the costs of tracking vehicles.

2 DATA COLLECTION TECHNOLOGY

Prior to the availability of GPS, the most common form of AVL chosen by transit agencies was the signpost/dead reckoning system in which a series of radio beacons are placed along the bus routes. The identification signal transmitted by the signpost is received by a short range communication device on the bus. Since the location of each signpost is known, the location of the bus at the time of passing the signpost is determined. The distance traveled since passing the last signpost is measured by odometer sensor, while the vehicle direction information is obtained from the gyroscope. However, this method is limited because signposts are placed at fixed locations. Thus, changes in bus routes could require the installation of additional signposts. Additionally, the system is incapable of tracking vehicles that stray off-route.

Another problem is the density of the signposts. Unless the density of signposts is sufficiently high, the dead-reckoned position error could be unacceptably large. However, the positioning error can be reduced significantly by reducing the signpost separations, which increases costs, and of course the system becomes defunct if the route changes.

Computer vision can provide an alternative to signpost with almost the same positioning accuracy. The advantage of using computer vision is that it does not require any equipment placed along the bus route, and if the bus route changes only the image database must be updated. Global localization systems based on computer vision could be a good solution for localization vehicles in public transit.

3 COMPUTER VISION AS AN ALTERNATIVE TO SIGNPOST TECHNOLOGY

Images acquired by the cameras can provide enough information for determining the position of a vehicle in the urban environment. To achieve this, a two stage process can be used. The first stage is acquiring a database of objects and locations of a particular area. The second stage is recognition by matching to the closest model in the database. This problem is interesting as a navigation task and also as an example of an object recognition problem. The class of buildings has many similarities and demands the techniques which are capable of fine discrimination between instances of the class.

In the past, several authors have written about the problem of building recognition. In [9] authors for matching suggested use of descriptors associated with interest regions. Authors in [11] achieved recognition by matching line segments and their associated descriptors. False matching was prevented by imposing the epipolar geometry constraint. In [12] an alternative approach on context-based place recognition was proposed.

For this type of global localization system a good database of objects and their features is needed. Objects in the database are obtained by extracting images of significant objects from the recorded environment along the transportation line. To avoid problems related with view angle, the camera mounted in the vehicle should be in a rectangular position with respect to driving direction. Objects in the database must be related with a coordinate system using readily available map data.

In real world scenes object recognition requires local image features which are unaffected by partial occlusion and must be at least partially invariant with respect to illumination and 3D projective transforms. To identify a specific object between alternatives the features must be distinctive enough. The success of an object recognition system is highly dependent on finding such image features. For finding image features we suggest a method for image feature generation called the Scale Invariant Feature Transform (SIFT) which is more explicitly described in [9]. The basic idea of this method is to transform an image into a large collection of local feature vectors. Each of these vectors is invariant with respect to image translation, rotation and scaling, and partially invariant with respect to illumination changes and affine or 3D projection. Key locations in scale space are identified by looking for locations that are maxima or minima of a difference of Gaussian function. To generate a feature vector each point is used. The features achieve partial invariance with respect to local variations. The resulting feature vectors are called SIFT keys. The SIFT keys are then used in a nearest neighbor approach to indexing to identify candidate object models. The keys that correspond to a potential model are first identified through a Hough transform table. After that identification through a least squares fit is used. At least 3 keys must correspond to the model to establish that there is a strong likelihood for the presence of the object.

For faster recognition objects in the database should be ordered by location. The object we expect to be first for recognition should be first for matching. Already recognized objects should be marked and not used for further recognition until the vehicle starts the same route from the beginning.

When the object is recognized, the system on the vehicle sends the position to the central part of the system where position and time are entered into the database and shown on the map.

3.1 Laboratory algorithm test

The most important aspect of the SIFT approach is that it enables a large number of features that cover the image with a full range of different scales and a large collection of local feature vectors.

The first step in the algorithm testing process requires building an object dataset. For this purpose 30 reference images of different objects along the bus route were acquired. In the next step the SIFT features were extracted from the previously acquired images and stored in the database for each reference image. Since the size of reference images was 800x600 it was possible to attain about 2500 feature vectors for each reference image.

To simulate at different light conditions and different angle perspectives that occur during driving, another set of images for each object was acquired in the second process stage. These new images differ from the first set in illumination levels and perspective angles. For each new image the SIFT features were extracted and later compared with the features of reference images saved in the database.

The next step in the process was recognition by matching. When features of the second set of images were compared with features of reference images saved in the database, we came across many incorrect matches. To avoid this problem, clusters of at least 3 features were used. The correct matches were filtered from the full set of matches by identifying subsets of key points that agreed on the object's location, scale and orientation. These clusters were then further verified one by one. After verification the probability of matches was computed for each particular cluster of features.

The lines in Figure 1 represent the matching pairs of features. The object was marked as correctly matched when the new image features corresponded with at least 50 reference image features, which was a constraint that we determined ahead of time. All matches with a distance ratio between the closest and second closest neighbour of each key point greater than 0.8 were rejected. The use of this distance ratio enabled us to eliminate 90% of the incorrect matches and less than 5% of the correct matches.

Fig. 1. Matching features on compared images

4 CONCLUSIONS

In our simulation the SIFT features approach was tested. During laboratory tests we established that using SIFT features is a good solution for monitoring transit service reliability, even though there are a few limitations. This approach proved to be relatively independent of illumination and angle changes. The comparing of newer images with reference images was robust enough due to a large number of feature vectors. In addition, the application runs almost in real-time, a result of the efficient computation of the feature vectors.

In the future we propose optimizing the search of feature vectors in a large database. We also propose that images can be taken every second since the speed of public transit vehicles is relatively low.

References

[1] Transit Cooperative Research Program (TCRP), "A Handbook for Measuring Customer Satisfaction and Service Quality", TCRP Report 47. Washington, DC: Transportation Research Board, National Research Council, 1999.

[2] J. Bates, P. Polak, J. Jones and A. Cook, "The Valuation of Reliability for Personal Travel", Transportation Research, Part E, No. 37, 2001, pp. 191-229.

[3] P. Prioni and D. Hensher, "Measuring Service Quality in Scheduled Bus Services", Journal of Public Transportation, vol. 3, 2000, pp. 51-74.

[4] P. Welding, "The Instability of Close Interval Service", Operational Research Quarterly, No. 8, 1957, pp. 133-148.

[5] M Turnquist, "A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Times" Transportation Research Record, 663, 1978, pp. 70-73.

[6] L. Bowman and M. Turnquist, "Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops", Transportation Research, Part A, vol. 15, 1981, pp. 465-471.

[7] N. Wilson, D. Nelson, A. Palmere, T. Grayson and C. Cederquist, "Service Quality Monitoring for High Frequency Transit Lines", Paper presented at the 71st Annual Meeting of the Transportation Research Board, Washington, DC, 1992.

[8] H. Mohring, J. Schroeter and P. Wiboonchutikula, "The Values of Waiting Time, Travel Time, and a Seat on the Bus", Rand Journal of Economics, No.18 (1), 1987, pp. 40-56.

[9] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features", In International Conference on Computer Vision (ICCV’99), 1999, pp. 1150–1157.

[10] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adapative Window: Theory and Experiment", IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(9), 1994, pp. 920–932.

[11] Y. Dufournaud, C. Schmid, and R. Horaud, "Matching Images with Different Resolutions", In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’00), 2000, pp. 612–618.

[12] Z. Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 2000, pp. 1330–1334.

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Appendix:

Authors' addresses

Addresses of SOR'07 Authors

(The 9th International Symposium on OR in Slovenia, Nova Gorica, SLOVENIA, September 26 – 28, 2007)

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
1.	Faruk Berat	Akcesme	International University of Sarajevo, Faculty of Engineering and Natural Science	Paromlinska 66	71000	Sarajevo	Bosnia and Harzegovina	farberak@yahoo.com
2.	Josip	Arnerić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	jarneric@efst.hr
3.	Nawel K.	Arrar	Université Badji Mokhtar Annaba, Faculté des Sciences,Département de Mathématiques	BP. 12	23000	Annaba	Algerie	kn_arrar@yahoo.fr
4.	Hossein	Arsham	University of Baltimore		MD21201-5779	Baltimore	Maryland, USA	harsham@ubalt.edu
5.	Jan	Babič	Jožef Štefan Institute	Jamova 39	1000	Ljubljana	Slovenia	jan.babic@ijs.si
6.	Zoran	Babić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	babic@efst.hr
7.	Ana Gabriela	Babucea	University of Târgu-Jiu, Faculty of Economics				Romania	babucea@utgjiu.ro
8.	Vlasta	Bahovec	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	bahovec@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
9.	Peter	Bajt		Unec 82	1381	Rakek	Slovenia	bajtpeter@volja.net
10.	Majda	Bastič	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	majda.bastic@uni-mb.si
11.	Alfonzo	Baumgartner	Faculty of Electrical Engineering, University of Osijek	Kneza Trpimira 2b	31000	Osijek	Croatia	alfonzo.baumgartmer@ et.fos.hr
12.	Ivo	Bićanić	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	ibicanic@efzg.hr
13.	Hans Joachim	Böckenhauer	ETH – Zentrum, Informationstechnologie und Ausbildung, CAB F16	Universitätsstras se 6	8092	Zürich	Switzerland	hjb@inf.ethz.ch
14.	Ivan	Bodrožić	University of Split, Faculty of Theology	Zrinsko Frankopanska 19	21000	Split	Croatia	
15.	Ludvik	Bogataj	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	ludvik.bogataj@ ef.uni-lj.si
16.	Marija	Bogataj	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	marija.bogataj@ ef.uni-lj.si
17.	Bernhard	Böhm	Vienna University of Technology, Institute for Mathematical Models and Economics		1040	Vienna	Austria	bernhard.boehm@ tuwien.ac.at
18.	Valter	Boljunčič	Juraj Dobrila University of Pula, Department of Economics and Tourism «Dr.Mijo Mirković»	Preradovičeva 1	52100	Pula	Croatia	vbolj@efpu.hr
19.	Immanuel	Bomze	TU Vienna	Universitätsstras se 5	A - 1010	Vienna	Austria	immanuel.bomze@ univie.ac.at

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
20.	Darja	Boršič	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	darja.borsic@uni-mb.si
21.	Zina	Boussaha	Department of Mathematics, Faculty of Sciences University of Annaba	BP 12	23000	Annaba	Algeria	boussaha_z@yahoo.fr
22.	Lidija	Bradeško	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	lidija.bradesko@ fs.uni-lj.si
23.	Andrej	Bregar	University of Maribor, Faculty of Electrical Engineering and Computer Science	Smetanova 17	2000	Maribor	Slovenia	andrej.bregar@uni-mb.si
24.	Mehmet	Can	International University of Sarajevo, Faculty of Arts and Social Sciences	Paromlinska 66	71000	Sarajevo	Bosnia and Harzegovina	mcan@ius.edu.ba
25.	Boris	Cota	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	bcota@efzg.hr
26.	Jesus	Crespo-Cuaresma	University of Innsbruck				Austria	jesus.crespo-cuaresma@ubk.ac.at
27.	Vesna	Čančer	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	vesna.cancer@uni-mb.si
28.	Anton	Čižman	University of Maribor, Faculty of Organizational Sciences Kranj	Kidričeva 55a	4000	Kranj	Slovenia	anton.cizman@ fov.uni-mb.si
29.	Mirjana	Čižmešija	University of Zagreb, Faculty Economics and Business,	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	mcizmesija@efzg.hr
30.	Draženka	Čižmić	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	dcizmic@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
31.	Daniela-Emanuela	Danacica	University of Târgu-Jiu, Faculty of Economics				Romania	danutza@utgjiu.ro
32.	Vesna	Dizdarević	Promo + d.o.o.	Perčeva 4	SI-1000	Ljubljana	Slovenia	promoplus@siol.net
33.	Natalija	Djellab	Department of Mathematics, Faculty of Sciences University of Annaba	BP 12	23000	Annaba	Algeria	djellab@yahoo.fr
34.	Matevž	Dolenc	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	mdolenc@itc.fgg.uni-lj.si
35.	Samo	Drobne	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	samo.drobne@ fgg.uni-lj.si
36.	Ksenija	Dumičić	University of Zagreb, Graduate School of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	kdumicic@efzg.hr
37.	Nataša	Erjavec	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	nerjavec@efzg.hr
38.	Liljana	Ferbar	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	liljana.ferbar@ef.uni-lj.si
39.	Fran	Galetić	Faculty of Economics and Business of Zagreb	Trg J. F. Kennedya 6	10000	Zagreb	Croatia	fgaletic@efzg.hr
40.	Martin	Gavalec	Department of Information Technologies, Faculty of Informatics and Management, University Hradec Králové	Rokitanskehó 62	50003	Hradec Králové	Czech Republic	martin.gavalec@uhk.cz
41.	Janez	Grad	University of Ljubljana, Faculty of Administration	Gosarjeva ulica 5	1000	Ljubljana	Slovenia	janez.grad@fu.uni-lj.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
42.	József	Györkös	University of Maribor, Faculty of Electrical Engineering and Computer Science	Smetanova 17	2000	Maribor	Slovenia	jozsef.gyorkos@uni-mb.si
43.	Željko	Hocenski	Faculty of Electrical Engineering, University of Osijek	Kneza Trpimira 2b	31000	Osijek	Croatia	zeljko.hocenski@ et.fos.hr
44.	Juraj	Hromković	ETH – Zentrum, Informationstechnologie und Ausbildung, CAB F16	Universitätsstrasse 6	8092	Zürich	Switzerland	juraj.hromkovic@ inf.ethz.ch
45.	Roman	Hušek	University of Economics	W. Churchilla 4	13067	Praha	Czech Republic	husek@vse.cz
46.	Dušan	Hvalica	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	dusan.hvalica@ ef.uni-lj.si
47.	Tibor	Illes	Eötvös Loránd University of Science, Department of Operations Research	Pázmány Péter sétány 1/c		Budapest	Hungary	illes@math.elte.hu
48.	Josef	Jablonsky	Department of Econometrics, University of Economics		130 67	Praha	Czech Republic	jablon@vse.cz
49.	Gašper	Jaklič	University of Ljubljana, Institute of Mathematics, Physics and Mechanics	Jadranska 19	1000	Ljubljana	Slovenia	gasper.jaklic@ fmf.uni-lj.si
50.	Matjaž B.	Jurič	University of Maribor, Faculty of Electrical Engineering and Computer Science	Smetanova 17	2000	Maribor	Slovenia	matjaz.juric@uni-mb.si
51.	Elza	Jurun	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	elza@efst.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
52.	Jasmin	Jusufović	International University of Sarajevo, Faculty of Economics and Business Administration	Paromlinska 66	71000	Sarajevo	Bosnia and Herzegovina	jjusufovic@ius.edu.ba
53.	Alenka	Kavkler	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	alenka.kavkler@ uni-mb.si
54.	Robert	Klinc	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	rklinc@itc.fgg.uni-lj.si
55.	Peter	Köchel	Chemnitz University of Technology, Faculty of Informatics, Chair of Modelling & Simulation	Straße der Nationen 62	09107	Chemnitz	Germany	peter.koechel@ informatik.tu-chemnitz.de
56.	Robert	Kunst	University of Vienna				Austria	robert.kunst@ univie.ac.at
57.	Nataša	Kurnoga Živadinović	University of Zagreb, Faculty Economics and Business,	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	nkurnoga@efzg.hr
58.	Janez	Kušar	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	janez.kusar@fs.uni-lj.si
59.	Lado	Lenart	Jožef Štefan Institute	Jamova 39	1000	Ljubljana	Slovenia	lado.lenart@ijs.si
60.	Andrej	Lisec	University of Maribor, Faculty of Logistics	Hočevarjev trg 1	8270	Krško	Slovenia	andrej.lisec@posta.si
61.	Anka	Lisec	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	anka.lisec@fgg.uni-lj.si
62.	Zrinka	Lukač	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	zlukac@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
63.	Robert	Manger	University of Zagreb, Department of Mathematics	Bijenička cesta 30	10000	Zagreb	Croatia	manger@math.hr
64.	Marija	Marinović	University of Rijeka, Faculty of Philosophy	Omladinska 14	51000	Rijeka	Croatia	marinm@ffri.hr
65.	Ivan	Martinić	University of Zagreb, Faculty of Forestry	Svetošimunska 25	10000	Zagreb	Croatia	martinic@sumfak.hr
66.	Miklavž	Mastinšek	University of Maribor, Faculty of Economics and Business	Razlagova 14	2000	Maribor	Slovenia	mastinsek@uni-mb.si
67.	Gregor	Miklavčič	Bank of Slovenia	Slovenska 35	1000	Ljubljana	Slovenia	gregor.miklavcic@bsi.si
68.	Dubravko	Mojsinović	Privredna banka Zagreb d.d.	Račkoga 6	10000	Zagreb	Croatia	dubravko.mojsinovic@pbz.hr
69.	Marianna	Nagy	Eötvös Loránd University of Science, Department of Operations Research	Pázmány Péter sétány 1/c		Budapest	Hungary	nmariann@cs.elte.hu
70.	Boris	Nemec	HIT d.d.	Delpinova 7a	5000	Nova Gorica	Slovenia	boris.nemec@hit.si
71.	Luka	Neralić	Faculty of Economics, University of Zagreb	Trg J. F. Kennedya 6	10000	Zagreb	Croatia	lneralic@efzg.hr
72.	Maciej	Nowak	The Karol Adamiecki University of Economics in Katowice, Department of Operations Research	Ul. 1. Maja 50	40-287	Katowice	Poland	nomaci@ae.katowice.pl
73.	A.	Omerovič	International University of Sarajevo, Faculty of Economics and Business Administration	Paromlinska 66	71000	Sarajevo	Bosnia and Herzegovina	amir1608@gmail.com

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
74.	Elif	Oyuk	International University of Sarajevo	Paromlinska 66	71000	Sarajevo	Bosnia and Herzegovina	eoyuk@ius.edu.ba
75.	Dejan	Paliska	University of Ljubljana, Faculty of Maritime Studies and Transport	Pot pomorščakov 4	6320	Portorož	Slovenia	dejan.paliska@ fpp uni-lj.si
76.	Václava	Pánková	University of Economics	W. Churchilla 4	13067	Praha	Czech Republic	pankova@vse.cz
77.	Mirjana	Pejić Bach	University of Zagreb, Graduate School of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	mpejic@efzg.hr
78.	Tunjo	Perić		Komedini 1	1000	Zagreb	Croatia	tunjo.peric1@zg.t-com.hr
79.	Igor	Pesek	IMFM	Jadranska 19	1000	Ljubljana	Slovenia	igor.pesek@ imfm.uni-lj.si
80.	Snježana	Pivac	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	spivac@efst.hr
81.	Ján	Plavka	Department of Mathematics, Faculty of Electrical Engineering and Informatics, University of Košice	B. Nemcovej 32	04200	Košice	Slovak Republic	jan.plavka@tuke.sk
82.	Nada	Pleli	Faculty of Economics and Business of Zagreb	Trg J. F. Kennedya 6	10000	Zagreb	Croatia	npleli@efzg.hr
83.	Naris	Pojskić	International University of Sarajevo, Faculty of Engineering and Natural Science	Paromlinska 66	71000	Sarajevo	Bosnia and Harzegovina	naris.pojskic@gmx.net

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
84.	Marko	Potokar	Bankart d.o.o.	Celovška 150	1000	Ljubljana	Slovenia	marko.potokar@ bankart.si
85.	Janez	Povh	School of Business and Management	Na Loko 2	8000	Novo mesto	Slovenia	janez.povh@ guest.arnes.si
86.	Mirjana	Rakamarić Šegić	Politechnic of Rijeka	Vukovarska 58	51000	Rijeka	Croatia	mrakams@veleri.hr
87.	Viljem	Rupnik	INTERACTA, LTD, Business Information Processing	Parmova 53	1000	Ljubljana	Slovenia	viljem.rupnik@siol.net
88.	Iztok	Saje	Mobitel d.d.	Vilharjeva 23	1000	Ljubljana	Slovenia	iztok.saje@mobitel.si
89.	Sebastian	Sitarz	Institute of Mathematics, University of Silesia in Katowice	Ul. Bankowa 14	40-007	Katowice	Poland	ssitarz@ux2.math.us. edu.pl
90.	Marko	Starbek	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	marko.starbek@ fs.uni-lj.si
91.	Roman	Starin	University of Ljubljana, Faculty of Maritime Studies and Transport	Pot pomorščakov 4	6320	Portorož	Slovenia	roman.starin@ fpp uni-lj.si
92.	Leen	Stougie	Eindhoven University of Technology, Department of Mathematics	PO Box 513	5600 MB	Eindhoven	The Netherlands	leen@win.tue.nl
93.	Nataša	Šarlija	University of Osijek, Faculty of Economics	Gajev trg 7	31000	Osijek	Croatia	natasa@efos.hr
94.	Ksenija	Šegotić	University of Zagreb, Faculty of Forestry	Svetošimunska 25	10000	Zagreb	Croatia	segotic@sumfak.hr
95.	Petra	Šparl	University of Maribor	Smetanova 17	2000	Maribor	Slovenia	petra.sparl@uni-mb.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
96.	Mario	Šporčić	University of Zagreb, Faculty of Forestry	Svetošimunska 25	10000	Zagreb	Croatia	sporcic@sumfak.hr
97.	E.	Tacgin	International University of Sarajevo	Paromlinska 66	71000	Sarajevo	Bosnia and Herzegovina	tacgin@ius.edu.ba
98.	Tamás	Terlaky	McMaster University, Department of Computing and Software			Hamilton, Ortario	Canada	terlaky@mcmaster.ca
99.	Dragan	Tevdovski	Facilty of Economics - Skopje, University Ss. Cyril and Methodius – Skopje	Blvd. Krste Misirkov bb		Skopje	Macedonia	dragan@ eccf.ukim.edu.mk
100.	Katerina	Tovsevska	Facilty of Economics - Skopje, University Ss. Cyril and Methodius – Skopje	Blvd. Krste Misirkov bb		Skopje	Macedonia	katerina@ eccf.ukim.edu.mk
101.	Tadeusz	Trzaskalik	The Karol Adamiecki University of Economics in Katowice, Department of Operations Research	Ul. Bogucicka 14	40-587	Katowice	Poland	ttrzaska @ae.katowice.pl
102.	Žiga	Turk	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	zturk@itc.fgg.uni-lj.si
103.	Robert	Volčjak	Economic Institute of the Law School	Prešernova 21	SI-1000	Ljubljana	Slovenia	robert.volcjak@eipf.si
104.	Ilko	Vrankič	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	ivrankic@efzg.hr
105.	Danijel	Vukovič	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	danijel.vukovic@ uni-mb.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
106.	Kangzhou	Wang	Lanzhou Polytechnical College, Department of Basic Science			Lanzhou	China	kanzhou.wang@ hotmail.com
107.	Lidija	Zadnik Stirn	University of Ljubljana, Biotechnical Faculty	Večna pot 83	1111	Ljubljana	Slovenia	lidija.zadnik@bf.uni-lj.si
108.	Lyudmyla	Zahvoyska	Department of Ecological Economics, National University of Forestry and Wood Technology	Gen Chuprynky Str., 103,	79057	Lviv	Ukraine	zld@forest.lviv.ua
109.	Karel	Zimmermann	Faculty of Mathematics and Physics	Malostranske nam 25	11800	Prague	Czech Republic	zimm@ ms.kam.mff.cuni.cz
110.	Janez	Žerovnik	IMFM	Jadranska 19	1000	Ljubljana	Slovenia	janez.zerovnik@ imfm.uni-lj.si

The 9th International Symposium on
Operational Research in Slovenia

SOR '07

Nova Gorica, SLOVENIA
September 26 - 28, 2007

Appendix:

Sponsors' Notices

ASO LJUBLJANA
AUSTRIAN SCIENCE AND RESEARCH
LIAISON OFFICE LJUBLJANA

Austrian Science and Research Liaison Office (ASO) Ljubljana

The Austrian Science and Research Liaison Office Ljubljana has been established in October 1990 as branch office of the Vienna based Austrian Institute for East and Southeast European Studies to foster scientific co-operation between Austria and Slovenia. ASO Ljubljana has been reorganised in March 2004 and is since that time part of the Centre for Social Innovation (ZSI) in Vienna. ASO Ljubljana receives its funding mainly from Austrian Federal Ministry of Science and Research (bm:wf) and partially also from Ministry of Higher Education, Science and Technology of Republic of Slovenia.

ZSI is in charge of coordination of activities of ASO Ljubljana and ASO Sofia with bm:wf as well as of coordination with regard to national, bilateral and international initiatives and programmes. The Austrian Science and Research Liaison Offices in Ljubljana and Sofia support the science policy of Austria in South Eastern Europe which is co-ordinated on European level with projects and initiatives like SEE-ERA.net www.see-era.net , Information Office of the Steering Platform on research for the Western Balkans www.see-science.eu , etc.

Some highlights of ASO Ljubljana work:

ASO Ljubljana has initiated and co-organised together with Ministry of Higher Education, Science and Technology of Republic of Slovenia, Austrian Federal Ministry of Education, Science and Culture, Hellenic Ministry of Development in February 2005 in European Parliament the international conference "**Participation of Western Balkan Countries in EU RTD Framework Programmes**" www.aso.zsi.at/de/slo/veranstaltung/190.html

In November 2005 ASO Ljubljana and UNESCO Office in Venice organized in cooperation with European Association of Research Managers and Administrators EARMA a **Training Seminar on International Project Management for Research Managers from South-east European countries** in Ljubljana, from 9 to 11 November, 2005. 21 participants were chosen from 260 applications from both governmental agencies and academia and came from all of the Balkan countries as well as Bulgaria, Romania and Turkey and the host country Slovenia.

In November 2002 ASO Ljubljana organised a Round table on "**Challenges for RTD co-operation with non-candidate countries in South-eastern Europe**" at the official FP6 Launching conference in Brussels

In September 2006 ASO Ljubljana organised together with UNESCO Office in Venice and Slovenian Ministry of Higher Education, Science and Technology the International conference and Ministerial Roundtable "**Why invest in science in SEE countries?**" http://investsciencesee.info/

Contact:
Austrian Science and Research Liaison Office Ljubljana (ASO) / Avstrijski znanstveni institut v Ljubljani/ Österreichisches Wissenschaftsbüro Ljubljana, Dunajska 104; SI-1000 Ljubljana; Slovenija
e-mail: aso-ljubljana@zsi.at; homepage: www.aso.zsi.at
tel (office): 00 386 (0) 1 5684 168 fax: 00 386 (0) 1 5684 169

The Hit group has successfully created a unique universe of services aimed at the entertainment of its guests. This universe is a source of pride and a continuous challenge for the Hit staff. With its gaming and entertainment centres and its tourist resort accommodation facilities, Hit ranks among Europe's largest entertainment providers.

The Perla, Park, Aurora, Korona and other entertainment centres from the **Hit Stars** chain are intended for guests who wish to add colour to their lives with new and interesting experiences. Hit's approach is well-rounded, classifying it among Europe's top gaming providers: in Hit centres you can enjoy your evening with games of chance and exquisite cuisine and then await the morning hours in top hotels.

The Larix, Grand Hotel Prisank, Kompas and other hotels in Kranjska Gora, as well as the Maestral in Montenegro, are however intended for guests who wish to spoil themselves and enjoy their free time amidst tranquil nature. The **Hit Holidays** chain thus builds upon outstanding tourist locations, combined with accompanying wellness services and sport facilities, superb cuisine and above all, comfortable hotels.

hit
universe of fun

www.hit.si