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.

Zadnik Stirn L. and S. Drobne (Editors): The 9th International Symposium on Operational Research SOR'07, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2007, 460 pp. ISBN 978-961-6165-25-9.

Zadnik Stirn L., J. Žerovnik, S. Drobne and A. Lisec (Editors): The 10th International Symposium on Operational Research SOR'09, Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2009, 604 pp. ISBN 978-961-6165-30-3.

Proceedings SOR'11

Proceedings of the 11th International Symposium on OPERATIONAL RESEARCH

SOR '11

Dolenjske Toplice, Slovenia
September 28-30, 2011

Edited by:
L. Zadnik Stirn • J. Žerovnik • J. Povh • S. Drobne • A. Lisec

SOR '11 Proceedings

The 11th International Symposium on Operational Research in Slovenia

Dolenjske Toplice, SLOVENIA, September 28 - 30, 2011

Edited by:

L. Zadnik Stirn, J. Žerovnik, J. Povh, S. Drobne and A. Lisec

Slovenian Society Informatika (SDI)

Section for Operational Research (SOR)

Proceedings of the 11th International Symposium on Operational Research
SOR'11 in Slovenia, Dolenjske Toplice, September 28 - 30, 2011.

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 2011 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(082)
519.8:005.745(082)
519.81:519.233.3/.5(082)

INTERNATIONAL Symposium on Operational Research in Slovenia (11 ; 2011 ; Dolenjske Toplice)
SOR '11 proceedings / The 11th International Symposium on Operational Research in Slovenia, Dolenjske Toplice, Slovenia, September 28-30, 2011 ; [organiser] Slovenian Society Informatika (SDI), Section for Operational Research (SOR). - Ljubljana : Slovenian Society Informatika, Section for Operational Research, 2011

ISBN 978-961-6165-35-8
1. Slovensko društvo Informatika. Sekcija za operacijske raziskave
257839872

Proceedings of the 11th International Symposium on Operational Research in Slovenia (SOR'11) 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

Program Committee:

L. Zadnik Stirn, University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia, Chair
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia, Chair
Z. Babić, University of Split, Faculty of Economics, Department for Quantitative Methods, Split, Croatia
M. Bastič, University of Maribor, Faculty of Business and Economics, Maribor, Slovenia
M. Bogataj, University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož, Slovenia
K. Cechlarova, P.J. Šafarik University, Faculty of Science, Košice, Slovakia
T. Csendes, University of Szeged, Department of Applied Informatics, Szeged, Hungary
V. Čančer, University of Maribor, Faculty of Business and Economics, Maribor, Slovenia
S. Drobne, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
L. Ferbar, University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia
M. Gavalec, University of Hradec Králové, Faculty of Informatics and Management, Hradec Králové, Czech Republic
R. W. Grubbström, Linköping University, Linköping Institute of Technology, Linköping, Sweden
J. Jablonsky, University of Economics, Faculty of Informatics and Statistics, Praha, Czech Republic
P. Köchel, Chemnitz University of Technology, Faculty of Informatics, Chemnitz, Germany
J. Kušar, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia
L. Lenart, Institute Jožef Stefan, Ljubljana, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
L. Neralić, University of Zagreb, Faculty of Economics & Business, Zagreb, Croatia
I. Pesek, University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
J. Povh, Faculty of Information Studies, Novo mesto, Slovenia
M. S. Rauner, University of Vienna, Department of Innovation and Technology Management, Vienna, Austria
A. Schaerf, University of Udine, Department of Electrical, Management and Mechanical Engineering, Udine, Italy
M. Sniedovich, University of Melbourne, Department of Mathematics and Statistics, Melbourne, Australia
K. Šorić, University of Zagreb, Faculty of Economics & Business, Zagreb, Croatia
D. Škulj, University of Ljubljana, Faculty of Social Sciences, Ljubljana, Slovenia
P. Šparl, University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
T. Trzaskalik, Karol Adamiecki University of Economics, Department of Operational Research Katowice, Poland
B. Zmazek, University of Maribor, Faculty of Natural Sciences and Mathematics, Maribor, Slovenia
D. Yuan, Linköping University, Department of Science, Linköping, Sweden

Organizing Committee:

J. Povh, Faculty of Information Studies, Novo mesto, Slovenia, Chair
S. Drobne, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
J. Gabrič, Faculty of Information Studies, Novo mesto, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
K. Macedoni, Faculty of Information Studies, Novo mesto, Slovenia
L. Zadnik Stirn, University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia

Chairs:

Z. Babić, University of Split, Faculty of Economics, Split, Croatia
M. Bogataj, MEDIFAS, Nova Gorica, Slovenia
A. Brezavšček, University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
V. Čančer, University of Maribor, Faculty of Economics and Business Maribor, Maribor, Slovenia
Arie M. C. A. Koster, RWTH Aachen University, Lehrstuhl II für Mathematik, Aachen, Germany
J. Kušar, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia
A. Lisec, University of Ljubljana, Faculty of Civil Engineering and Geodesy, Ljubljana, Slovenia
Z. Lukać, University of Zagreb, Faculty Economics and Business, Zagreb, Croatia
U. Pferschy, Department of Statistics and Operational Research, University of Graz, Graz, Austria
S. Pivac, University of Split, Faculty of Economics, Split, Croatia
J. Povh, Faculty of Information Studies in Novo mesto, Novo mesto, Slovenia
P. Šparl, University of Maribor, Faculty of Organizational Sciences, Kranj, Slovenia
L. Zadnik Stirn, University of Ljubljana, Biotechical Faculty, Ljubljana, Slovenia
J. Žerovnik, University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia

Preface

The publication, Proceedings of The 11th International Symposium on Operations Research, called SOR'11, contains papers presented at SOR'11(http://sor11.fis.unm.si/) which was organized by Slovenian Society INFORMATIKA, Section for Operations Research (SSI-SOR) and Faculty of Information Studies (FIS), Novo mesto, Slovenia, and held in Dolenjske Toplice, Slovenia, from September 28 through September 30, 2011. In the Proceedings are published only the articles blindly reviewed and accepted by two independent reviewers. So, with the Proceedings SOR'11, the scientific activities of the symposium SOR'11 are available to all who participated in the symposium and to all who are interested in the contents and are considering to participate in the future SOR symposia.

SOR'11 stands under the auspices of the Slovenian Research Agency and PASCAL2, European Network of Excellence. The opening address was given by Prof. Dr. L. Zadnik Stirn, the President of the Slovenian Section of Operations Research, Mr. Niko Schlamberger, the President of Slovenian Society INFORMATIKA, Prof. Dr. Janez Povh, the dean of Faculty of Information Studies, Novo mesto, and presidents/representatives of Operations Research Societies from neighboring countries.

SOR'11 is the scientific event in the area of operations research, one of the traditional series of the biannual international OR conferences, organized in Slovenia by SSI-SOR. It is a continuity of ten previous symposia. The main objective of SOR'11 is to advance knowledge, interest and education in OR in Slovenia and worldwide in order to build the intellectual and social capital that are essential in maintaining the identity of OR, especially at a time when interdisciplinary collaboration is proclaimed as significantly important in resolving problems facing the current challenging times. Further, when joining IFORS and EURO, SSI-SOR agreed to work together with diverse disciplines, i.e. to balance the depth of theoretical knowledge in OR and the understanding of theory, methods and problems in other areas within and beyond OR. We are sure that SOR'11 creates the advantage of these objectives, contributes to the quality and reputation of OR with presenting and exchanging new developments, opinions, experiences in the OR theory and practice.

SOR'11 was highlighted by a distinguished set of six keynote speakers. Thus, the first part of the Proceedings SOR'11 comprises invited papers, presented by outstanding scientists: Professor Dr. Erling D. Andersen, MOSEK ApS, Denmark, Professor Dr. Walter Gutjahr, University of Vienna, Department of Statistics and Decision Support Systems, Austria, Professor Dr. Horst W. Hamacher, University of Kaiserslautern, Department of Mathematics, Germany, Professor Dr. Arie M.C.A. Koster, Lehrstuhl II für Mathematik, RWTH Aachen, Germany, Professor Dr. Zrinka Lukač, University of Zagreb, Faculty of Economics & Business, Croatia and Professor Dr. Ulrich Pferschy, University of Graz, Department of Statistics and Operations Research, Austria. The second part of the Proceedings includes 47 papers written by 100 authors and co-authors. Most of the authors of the contributed papers came from Slovenia (37), then from Croatia (29), Serbia (13), Bosnia and Herzegovina (4), Germany (5), Spain (3), Austria (2), Hungary (2), Iran (2), Uruguay (2) and Denmark (1). The papers published in the Proceedings are divided into sections: (the number of papers in each section is given in parentheses): Plenary Lectures (6), Graphs and their Applications (3), Production and Inventory (12), OR Applications in Telecommunication and Navigation Systems (3), Econometric Models and Statistics(6), Finance and Investments (6), Multiple Criteria Decision Making (6), Pascal2 session (3), Mathematical Programming and Optimization (3), Location and Transport (5).

The Proceedings of the previous ten International Symposia on Operations Research organized by Slovenian Section of Operations Research are indexed in the following secondary and tertiary publications: Current Mathematical Publications, Mathematical Review and MathScinet, Zentralblatt fuer Mathematik/Mathematics Abstracts, MATH on STN International and CompactMath, INSPEC. Also the present Proceedings SOR'11 will be submitted and is supposed to be indexed in the same basis.

On behalf of the organizers we would like to express our sincere thanks to all who have supported us in preparing this event - SOR'11. 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 prominent keynote speakers, to the members of the Program and Organizing Committees, to the referees who raised the quality of the SOR'11 by their useful suggestions, section's chairs, and to all the numerous people - far too many to be listed here individually - who helped in carrying out The 11th International Symposium on Operations Research SOR'11 and in putting together these Proceedings. At last, we appreciate the authors' efforts in preparing and presenting the papers, which made The 11th Symposium on Operational Research SOR'11 successful. The success of the scientific events at SOR'11 and the present proceedings should be seen as a result of joint effort.

Dolenjske Toplice, September 28, 2011

Lidija Zadnik Stirn
Janez Žerovnik
Janez Povh
Samo Drobne
Anka Lisec
(Editors)

Sponsors of the 11th International Symposium on Operational Research in Slovenia (SOR'11)

The following institution and network are gratefully acknowledged for their financial support of the symposium:

- *Slovenian Research Agency, The Republic of Slovenia*
- *PASCAL 2, European Network of Excellence*

Contents

Author index

The 11th International Symposium on
Operational Research in Slovenia

SOR ’11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Plenary Lectures

TEN YEARS OF EXPERIENCE WITH CONIC QUADRATIC OPTIMIZATION

Erling D. Andersen
MOSEK ApS, Fruebjergvej 3, Box 16 2100 Copenhagen O, Denmark
Email: e.d.andersen@mosek.com

Abstract: For about 10 years the software package MOSEK has been capable of solving large-scale sparse conic quadratic optimization (CQO) problems. Based on the experience gained with CQO during those 10 years we will present a few important applications of conic quadratic optimization, discuss properties of the CQO problem and give an overview of how MOSEK solve a CQO problem. We will also present numerical results demonstrating the performance of MOSEK on CQO problems and discuss future developments.

OPERATIONS RESEARCH METHODS IN THE PLANNING, CONTROL, AND ADAPTATION OF EVACUATION PLANS

Katharina Gerhardt, Horst W. Hamacher and **Stefan Ruzika**
Department of Mathematics, University of Kaiserslautern (Germany),
Email: {Gerhardt, Hamacher, Ruzika}@mathematik.uni-kl.de

Extended Abstract: One of the basic emergency measures in the case of (bomb) threats, attacks, large-scale accidents, and natural disasters is the evacuation of the affected buildings and regions. Here, the most important goal is to evacuate occupants, i.e. to take them away from the risk area and bring them to safety, as fast and reliable as possible. Unfortunately, several tragic scenarios are known from the past, for instance, the love parade disaster in Duisburg, Germany, of July 2010, where 19 people were killed and more than 300 injured while trying to leave the location in which the parade was to take place. It is not claimed that the application of operations research methods can avoid such tragic incidences, but it is reasonable to assume that they will yield insight to make them less likely to occur or help to mitigate the consequences.

The research project REPKA which is co-ordinated at the University of Kaiserslautern is primarily concerned with regional evacuation. In particular, the situation is considered where a large crowd has already left a building and must then be brought further away to safety. In order to reliably predict evacuation data like evacuation times or number of evacuees per given time period we propose an approach in which real-world data is sandwiched between a lower bound computed by optimization methods and upper bounds delivered by simulation. In our presentation we will focus on the following operations research issues.

- Representation of pedestrian movements in evacuation scenarios by dynamic network flows.
- Solution of maximal, earliest arrival, and quickest dynamic flows.
- Integration of location decisions with regard to the placement of emergency and commercial units.
- Control and adaptation of evacuation plans using the visualization tool of REPKA-Optimizer.
- Practical experience with the sandwich method applied to the evacuation of the Betzenberg, home of the Fritz-Walter football stadium in Kaiserslautern (Germany).

Reference:

REPKA webpage http://www.repka-evakuierung.de and further information therein

Acknowledgement
This paper is supported in part by the Federal Ministry for Education and Research (Bundesministerium für Bildung und Forschung, BMBF) as project REPKA under FKZ 13N9961 (TU Kaiserslautern).

A Note on the Utility-Theoretic Justification of the Pareto Approach in Stochastic Multi-Objective Combinatorial Optimization

Walter J. Gutjahr
Department of Statistics and Operations Research
University of Vienna

1 Introduction

For a precise mathematical formulation of stochastic multi-objective optimization (SMOO) problems, the *multi-objective* and the *stochastic* approach have been proposed [1, 2, 3]. The former reduces the problem in a first step to a deterministic multi-objective problem, whereas the latter starts by reducing it to a stochastic single-objective problem. In the case of a finite decision set, i.e., in *combinatorial* SMOO, the multi-objective approach is especially attractive, because it allows the computation of a list of Pareto-efficient solutions from which the decision maker (DM) can select her/his preferred solution without being forced to describe her/his preference structure or utility function in formal terms.

However, some authors have questioned the appropriateness of the multi-objective approach for the general case of a SMOO problem, see [3]. The aim of this note is to show that there are conditions under which the multi-objective approach (using the Pareto concept for solving the resulting deterministic multi-objective problem) can be proven to be well-founded in decision theory in the sense that it is consistent with expected-utility maximization. Two examples of such conditions will be given.

2 Problem Definition

We consider stochastic multi-objective combinatorial optimization (SMOCO) problems of the following form:

$$\max\ (f_1(x,\omega),\ldots,f_m(x,\omega)) \quad \text{s.t.} \quad x \in S. \tag{1}$$

Therein, S is a finite set of decisions, and ω denotes a random influence. More formally, for each x and j, the function $f_j(x,\cdot)$ is a random variable on a probability space $(\Omega, \mathcal{A}, P)$, and ω is an element of the sample space Ω. Each ω determines a specific realization of the random variables $f_j(x,\cdot)$, i.e., a random *scenario*. We shall identify ω with the scenario determined by it. Throughout the paper, we assume that the joint distribution of the random variables $f_j(x, \cdot)$ is known.[1]

Mathematically, the considered problem is still under-specified by (1). First, the formulation contains more than one objective, such that it is not yet clear what is meant by "maximize". Secondly, since each objective also depends on ω, it still needs to be defined what is done to obtain a *unique* evaluation of the objective(s).

In the literature, two ways have been described to re-formulate (1) in a precise sense (see [1, 2, 3]):

- The *multi-objective approach* applies operators as expectation, variance etc. to each component of the vector $(f_1(x, \omega), \ldots, f_m(x, \omega))$. In this way, it reduces the stochastic multi-objective problem to a deterministic multi-objective problem. After that, some method of treating a multi-objective problem is applied. We focus here on the case where the *expectation* operator is applied to each $f_j(x, \omega)$. This gives the multi-objective problem

$$\max\ (F_1(x), \ldots, F_m(x)) \quad \text{s.t.} \quad x \in S, \tag{2}$$

  where $F_j(x) = \mathrm{E}(f_j(x, \omega))$ $(j = 1, \ldots, m)$.

- The *stochastic approach* aggregates the random objectives $f_1(x, \omega), \ldots, f_m(x, \omega)$ by some aggregation function $\mathcal{A}$ and applies *then* an operator such as the expectation operator. E.g., it solves

$$\max\ \mathrm{E}(\mathcal{A}(f_1(x, \omega), \ldots, f_m(x, \omega))) \quad \text{s.t.} \quad x \in S. \tag{3}$$

  We see that now the order has been reversed: First, the original problem is reduced to a single-objective stochastic optimization problem, and then some method of stochastic optimization is applied to solve this resulting problem.

By the *Pareto approach*, we understand the multi-objective approach in the version given by (2) in the special case where as the solution concept for multi-objective optimization, the computation of the set of Pareto-efficient (or: non-dominated) solutions[2] is chosen. Thus, the Pareto approach reformulates problem (1) as the problem of determining all non-dominated vectors of expected objective function values.

Caballero et al. [3] argue that in general, the stochastic approach is more adequate than the multi-objective approach, because it takes dependencies between the objectives into account (cf. also Ben Abdelaziz [2]). However, the stochastic approach has the disadvantage that it forces the DM to unveil her preferences in advance (i.e., before computational solution) by specifying the aggregation function $\mathcal{A}$.

[1] A simple special case occurs if $\Omega = \{\omega_1, \ldots, \omega_n\}$ is finite. Then the joint distribution is defined by a vector of probabilities p_i assigned to the scenarios ω_i $(i = 1, \ldots, n)$ with $\sum_{i=1}^{n} p_i = 1$.

[2] A solution $x \in S$ is called Pareto-efficient with respect to objectives $F_1, \ldots, F_m$, if there is no other solution $y \in S$ with $F_j(y) \geq F_j(x)$ $\forall j$ and $\exists j : F_j(y) > F_j(x)$. We also extend this notion from the solution space to the objective space by calling the vector $(F_1(x), \ldots, F_m(x))$ Pareto-efficient, if solution x is Pareto-efficient.

In many practical situations, the DM is unable or unwilling to do that. Therefore, it makes sense to look for conditions under which the multi-objective approach in general and the Pareto approach in particular are justifiable by decision-theoretic considerations.

3 Utility-Theoretic Analysis

In deterministic multi-objective optimization, the attractiveness of the Pareto concept stems from the following observation, which is easy to prove: For each each vector $(F_1, \ldots, F_m)$ of objective functions on S and each utility function $u : \mathbb{R}^m \to \mathbb{R}$ that is increasing in each component, it holds that if solution $x^* \in S$ is optimal w.r.t. u in the sense that x^* maximizes $u(F_1(x^*), \ldots, F_m(x^*))$ on S, then $(F_1(x^*), \ldots, F_m(x^*))$ must be Pareto-efficient. In other words, it is safe to omit dominated solutions (for this purpose, the utility function needs not to be known!) since after this omission, the optimal solution w.r.t. the implicit utility function u can still be found in the Pareto set (the set of Pareto-optimal solutions), so the DM can inspect this set and choose the best solution according to her or his true preferences.

Unfortunately, the generalization of this assumption to problem (2) does not hold anymore in the general case: Simple examples show that the solution of the problem

$$\max \ \mathrm{E}(u(f_1(x), \ldots, f_m(x))) \quad \text{s.t.} \quad x \in S \tag{4}$$

needs not to be a Pareto-efficient solution of (2). The essential reason is that in general, the expectation operator cannot be interchanged with the utility function u. Thus, it can happen that the solution with maximum expected utility does not occur in the Pareto set. This does not necessarily imply that in these circumstances, the Pareto approach is inadequate[3]; it only means that it is not consistent with the principle of expected-utility maximization (sometimes called the Bernoulli principle).

In the following, we give two examples of situations where there is no conflict between the Pareto approach and expected-utility maximization. For this purpose, let us start with a definition.

Definition 1. Let $\mathcal{U}$ denote a set of increasing utility functions $u : \mathbb{R}^m \to \mathbb{R}$. We say that for an m-dimensional SMOCO problem $(f_1, \ldots, f_m)$ on S, the Pareto approach is *$\mathcal{U}$-consistent* with expected-utility maximization (short: $\mathcal{U}$-consistent), if for each $u \in \mathcal{U}$ and each $x^* \in S$, the validity of

$$\mathrm{E}(u(f_1(x^*, \omega), \ldots, f_m(x^*, \omega))) = \max_{x \in S} \ \mathrm{E}(u(f_1(x, \omega), \ldots, f_m(x, \omega)))$$

implies that $(\mathrm{E}(f_1(x^*, \omega)), \ldots, \mathrm{E}(f_m(x^*, \omega)))$ is Pareto-efficient.

[3] The Pareto approach could still be justified then by considering utilities of expectations instead of expected utilities.

Theorem 1. Let $f_j(x,\omega)$ be objective functions with $f_j(x,\omega) = f_j(x)$ deterministic for $j = 1,\dots,k$, and $f_j(x,\omega) \geq 0\ \forall x \in S$ for $j = k+1,\dots,m$ $(0 \leq k \leq m)$. Furthermore, let the class $\mathcal{U}$ consist of all utility functions u of the form

$$u(y_1,\dots,y_m) = \varphi(y_1,\dots,y_k) + \sum_{j=k+1}^{m} \psi_j(y_1,\dots,y_k)\, y_j,$$

where $\varphi(y_1,\dots,y_k)$ and $\psi_j(y_1,\dots,y_k)$ $(j = k+1,\dots,m)$ are increasing in $y_1,\dots,y_k$. Then for $(f_1,\dots,f_m)$, the Pareto approach is $\mathcal{U}$-consistent.

Proof. With $\bar{f}_j(x) = \mathrm{E}(f_j(x,\omega))$, we have

$$\begin{aligned} \mathrm{E}(u(f_1(x,\omega),\dots,f_m(x,\omega))) &= \varphi(f_1(x),\dots,f_k(x)) \\ &\quad + \sum_{j=k+1}^{m} \psi_j(f_1(x),\dots,f_k(x))\, \mathrm{E}(f_j(x,\omega)) \end{aligned}$$

$$= \varphi(\bar{f}_1(x),\dots,\bar{f}_k(x)) + \sum_{j=k+1}^{m} \psi_j(\bar{f}_1(x),\dots,\bar{f}_k(x))\, \bar{f}_j(x) = u(\bar{f}_1(x),\dots,\bar{f}_m(x)). \quad (5)$$

The function u is increasing. Therefore, by the observation that in the deterministic context, the optimality of $x^* \in S$ w.r.t. u implies the Pareto-efficiency of the vector of objective function values, for a maximizer x^* of (5), the vector $(\bar{f}_1(x^*),\dots,\bar{f}_m(x^*))$ must be Pareto-efficient. $\square$

It can be seen that in the sufficient condition given in Theorem 1 for $\mathcal{U}$-consistency, the set $\mathcal{U}$ needs not to be restricted to utility functions exhibiting *utility independence* in the sense of Keeney and Raiffa [5], i.e., utility functions u of the form $u(y_1,\dots,y_m) = \sum_{j=1}^{m} u_j(y_j)$.

Evidently, also the case of *linear* utility functions $u(y_1,\dots,y_m) = a_1 y_1 + \ldots + a_m y_m$ is covered by Theorem 1. As already outlined in [3], the multi-objective approach does not fall back behind the stochastic approach in this situation. However, it should be noted that in this case, only so-called *supported* solutions can be optimal, i.e., solutions that are optimal under a weighted sum of objectives with suitable weights. The conditions of Theorem 1, on the other hand, also extend to situations where under certain utility functions, the optimal solution is unsupported.

Example. Consider the case where the decision x determines a work plan for a project P. Each work plan x consumes a certain amount $r_1(x)$ of a resource (say, working time) and yields a (stochastic) profit $f_2(x,\omega)$. In total, B units of the resource are available. The remaining $f_1(x) = B - r_1(x)$ units of the resource can be used in a project Q, yielding there a (deterministic) profit $\varphi(f_1(x))$ that is increasing in $f_1(x)$. Note that φ needs not to be linear, and that the DM may shy away from providing an explicit mathematical representation of φ. Then, we can consider the bi-objective maximization problem with objective functions $f_1(x)$ and $f_2(x,\omega)$, where the first objective function is deterministic, the second is stochastic. The overall utility function will be of the form $u(f_1,f_2) = u_1(f_1) + u_2(f_2)$ with

$u_1(f_1) = \varphi(f_1)$ denoting the profit from project Q and $u_2(f_2) = f_2$ denoting the profit from project P. Evidently, we are within the conditions of Theorem 1. As a consequence, the Pareto approach will produce a set of solutions among which the solution with maximal expected utility occurs. An approach aggregating expectations by a weighted sum, on the other hand, may miss the solution with maximal expected utility even if all possible weights are considered. $\bowtie$

Utility functions can frequently be approximated by quadratic functions. The following theorem gives a second sufficient condition for consistency if all utilities are quadratic:

Theorem 2. Let f_j have values in some interval $[\alpha_j, \beta_j]$ $(j = 1, \ldots, m)$, and let the class $\mathcal{U}$ consist of all increasing utility functions of the form $u(y_1, \ldots, y_m) = \sum_{j=1}^{m} u_j(y_j)$ with

$$u_j(y_j) = a_j + b_j y_j + c_j y_j^2 \quad (j = 1, \ldots, m),$$

where u_j is increasing on $[\alpha_j, \beta_j]$. Furthermore, let the variance $\mathrm{var}(f_j(x, \omega))$ be independent of x, i.e., $\mathrm{var}(f_j(x, \omega)) = \sigma_j^2 \; \forall x \in S$. Then for $(f_1, \ldots, f_m)$, the Pareto approach is $\mathcal{U}$-consistent.

Proof.

$$\mathrm{E}(u(f_1(x, \omega), \ldots, f_m(x, \omega))) = \sum_{j=1}^{m} \left[a_j + b_j \,\mathrm{E}(f_j(x, \omega)) + c_j \,\mathrm{E}((f_j(x, \omega)^2)\right]$$

$$= \sum_{j=1}^{m} \left[a_j + b_j \,\mathrm{E}(f_j(x, \omega)) + c_j \, (\mathrm{E}(f_j(x, \omega)))^2\right] + \sum_{j=1}^{m} c_j \left[\mathrm{E}((f_j(x, \omega)^2) - (\mathrm{E}(f_j(x, \omega)))^2\right]$$

$$= u(\mathrm{E}(f_1(x, \omega)), \ldots, \mathrm{E}(f_m(x, \omega))) + \sum_{j=1}^{m} c_j \, \sigma_j^2,$$

where the last term is constant. Again by reference to the deterministic special case (as in the proof of Theorem 1), we conclude that for a maximizer x^* of u, the vector $(\mathrm{E}(f_1(x, \omega), \ldots, \mathrm{E}(f_m(x, \omega)))$ must be Pareto-efficient. $\square$

4 Conclusions

Two examples of sufficient conditions for the consistency of the Pareto approach with expected-utility maximization have been given. Although the conditions may look rather restrictive, a large class of practically relevant problems in multi-objective decision analysis is covered by them. Also other sufficient conditions may be provided, as well as relaxed versions guaranteeing approximate solutions to utility optimization problems on a certain approximation level. Results of this type are of interest, since during the last years, advances in the computational solution of SMOCO problems based on the Pareto approach by suitable meta-heuristic techniques have been made (cf. [4]).

References

[1] Ben Abdelaziz, F. L'efficacité en Programmation Multi-objectifs Stochastique. Thesis, Université de Laval, Québec.

[2] Ben Abdelaziz F. Solution approaches for the multiobjective stochastic programming. European Journal of Operations Research, in press.

[3] Caballero R, Cerda E, Muños MM, Rey L. Stochastic approach versus multiobjective approach for obtaining efficient solutions in stochastic multiobjective programming problems. European Journal of Operations Research 2004; 158; 633-648.

[4] Gutjahr WJ. Recent trends in metaheuristics for stochastic combinatorial optimization. Central European Journal of Computer Science 2011; 1; 58-66.

[5] Keeney R, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Trade Offs.* John Wiley and Sons, New York, 1976.

Robust Optimization of Telecommunication Networks

Arie M.C.A. Koster*

Abstract

Robust Optimization is an emerging field of Operations Research, focussing on dealing with data uncertainty in optimization problems. In this talk, we discuss a variety of robust optimization approaches and their application to both fixed and wireless telecommunication network design problems. In particular, we show how demand uncertainty can be incorporated in classical network design by LP duality. The robust optimization approach is evaluated with real-life Internet traffic data.

Keywords: network design, robust optimization, price of robustness, integer linear programming

Extended Abstract

Incorporating uncertainty within the mathematical analysis of operational research has been an effort since its very first beginnings. In the 1950s, Dantzig [13] introduced Stochastic Programming using probabilities for the possible realizations of the uncertain data. The main limitation of such probabilistic approaches is that the distribution of the uncertain data must be known a priori which is often not the case for "real-world" problems. Stochastic programming may also result in extremely hard to solve optimization problems.

A promising alternative to handle data uncertainties is the usage of so-called chance-constraints which were introduced in [24]. Chance-constrained programming is a one-stage concept for which the probability distribution of the uncertain data has to be known completely. The aim of this concept is to find the best solution remaining feasible for a given infeasibility probability tolerance. In [14] chance constraints for combinatorial optimization problems are studied.

In 1973, Soyster [26] suggested another approach based on implicitely describing the uncertain data introducing so-called uncertainty sets and establishing the concept of Robust Optimization. Using this framework we do not need any information about the probabilistic distribution of the uncertainty. Instead a solution is said to be robust if it is feasible for all realizations of the data in the given uncertainty set. In Robust Optimization we aim at finding the cost-optimal robust solution. This approach has been further developed by Ben-Tal and Nemirovski [3, 4, 5], Bertsimas and Sim [7], and others using different convex and bounded uncertainty sets. They introduce the concept of robust counterparts for uncertain linear programs. In [3] it is shown that these can be solved by deterministic linear programs or deterministic conic quadratic programs if the uncertainty set is polyhedral or ellipsoidal, respectively. Bertsimas and Sim [7] introduced a polyhedral uncertainty set that easily allows to control the price of robustness by varying the number Γ of coefficients in a row of the given linear program that are allowed to deviate from its nominal values simultaneously. By changing this parameter Γ the practitioner is enabled to regulate the trade-off between the degree of uncertainty taking into account and the cost of this additional feature.

Robust optimization is also a well known method in telecommunication network design. We distinguish between robust network design using static or dynamic routing which refers to the flexibility of flow to respond to the realization of the demand (while the capacity remains fixed). Static routing means that for every node pair the same paths are used with the same splitting independent of the realization of demand. Contrary, dynamic routing allows for full flexibility in rerouting the traffic if the demand changes. The concept of different routing schemes is strongly related to different levels of recourse in multi-stage

*RWTH Aachen University, Lehrstuhl II für Mathematik, Wüllnerstr. 5b, D-52062 Aachen, Germany, {koster,kutschka}@math2.rwth-aachen.de

stochastic and robust optimization [6]. We refer to [21] and [23] for a discussion on how to embed the two classical routing schemes in these more general frameworks. For general two-stage robust network design check [1]. We also note that recently there has been some progress in defining routing schemes in between static and dynamic, see for instance [2, 22, 23, 25].

In this presentation, we will discuss three recent applications of robust optimization to problems from telecommunications:

- The design of a backbone network under demand uncertainty is described by
  - an undirected connected graph $G = (V, E)$ representing a potential network topology,
  - on each of the links $e \in E$ capacity can be installed in integral units and costs κ_e per unit,
  - a set of commodities K represents potential traffic demands with for each commodity $k \in K$ a nodepair (s^k, t^k) and a demand value $d^k \geq 0$ for traffic from source $s^k \in V$ to target $t^k \in V$.

  The actual demand values are considered to be uncertain. The traffic for commodity k is realized by a splittable multi-path flow between s^k and t^k. Of course, the actual multi-commodity flow depends on the realization of the demand d. A *routing template* describes for every commodity the percental splitting of the traffic among the paths from s^k to t^k.

  Along the lines of the Γ-robustness approach of [7], we define for every commodity $k \in K$ a *nominal* demand value $\bar{d}^k$ and a *deviation* $\hat{d}^k$. It is assumed that $d^k \in [0, \bar{d}^k + \hat{d}^k]$ and, given a value $\Gamma \in \mathbb{Z}$, at most Γ commodities deviate from their nominal values simultaneously.

  The robust network design problem is to find a minimum-cost installation of integral capacities and a routing template for every commodity such that actual flow does not exceed the link capacities independent of the realization of demands. Theoretical and practical aspects of this problem are studied in a series of papers: [16, 17, 18, 19, 20]. In particular, the problem is formulated as an integer linear program, valid inenequalities are derived, and optimal network designs are evaluated on the basis of real-life traffic data.

- The green design of a wireless network under demand uncertainty. In wireless network planning we have to install a number of base stations (BSs) from a set S of BS candidates (including candidate sites at the same location but with different configurations) such that a set T of traffic nodes (TNs) can be assigned to the BSs. Each BS $s \in S$ has an available downlink (DL) bandwidth b_s, a basic power consumption p_s and a power consumption $\tilde{p}_s$ per TN served by s, whereas each TN $t \in T$ requests a data rate w_t. The spectral efficiency between any BS-TN pair (s, t) is denoted by e_{st}. This parameter gives the ratio between data rate and bandwidth. It incorporates, e. g., modulation and coding scheme that is supported by the associated signal-to-noise ratio (SNR). To establish a link from a BS to a TN the corresponding spectral efficiency must exceed the threshold $e_{\min}$. The amount of bandwidth that is allocated to TN t from BS s, if t is served by s, is $\frac{w_t}{e_{st}}$. The objective is to minimize the total amount of energy consumed by the deployed BSs and the number of TNs which are not served. A scaling parameter λ is needed to combine the two objectives reasonably. A TN can be assigned to at most one BS due to hard handover in future wireless networks.

  The required data rate of a TN varies over time due to mobility of the users and their behaviour (voice, web browsing, data transfer). Again, we assume that the data rate varies within an interval $[0, \bar{w}_t + \hat{w}_t]$ with $\bar{w}_t$ the nominal data rate and $\bar{w}_t + \hat{w}_t$ the peak data rate. In [12] the Γ-robustness approach is applied to this model, including valid inequalities to improve the performance of CPLEX. In a case study, we observed energy savings either by deploying less BSs or serving more TNs with the same number of BSs.

  A subproblem within this context is the robust knapsack problem [15]. In [8, 9] this classical problem is further generalized to include the possibility to exchange items in a second stage, the so-called recoverable robust knapsack problem. In [8], the case of discrete scenarios (i.e., a set of weight vectors) is studied, whereas in [9] the case of Γ-scenarios is studied.

- The design of reliable fixed broadband wireless networks is considered in [10, 11]. Here, the problem is similar to the design of backbone networks, except that radio links are used instead of optical fibres.

Hence, we cannot install multiple capacity units on a single link. The capacity of a link is determined by the chosen channel bandwidth and the bandwidth efficiency (modulation scheme). The quality of the signal varies over time resulting in a fluctuation of the provided capacity.

The design of such a network can be formulated as a chance-constrained optimization problem. Here, the probability that the provided capacity exceed the required bandwidth has to be close to 1. In case the probabilities of the links are independent, the problem can be reformulated as integer linear programming [10] and the performance of ILP solvers can again be improved by valid inequalities [11].

Acknowledgement

This extended abstract is based on research carried out in collaboration with many people. In particular, I would like to thank Grit Claßen, Manuel Kutschka, Napoleao Nepomuceno, and Christian Raack for the fruitful teamwork.

This work was partly supported by the Federal Ministry of Education and Research (BMBF grant 03MS616A, project *ROBUKOM - Robust Communication Networks*, `www.robukom.de`).

References

[1] A. Atamtürk and M. Zhang. Two-Stage Robust Network Flow and Design Under Demand Uncertainty. *Operations Research*, 55(4):662–673, 2007.

[2] W. Ben-Ameur. Between fully dynamic routing and robust stable routing. In *Design and Reliable Communication Networks, 2007. DRCN 2007. 6th International Workshop on*, pages 1–6. IEEE, 2007.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–14, 1999.

[4] A. Ben-Tal and A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 88:411–424, 2000.

[5] A. Ben-Tal and A. Nemirovski. Robust optimization - methodology and application. *Mathematical Programming*, 92:453–480, 2002.

[6] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[7] D. Bertsimas and M. Sim. The Price of Robustness. *Operations Research*, 52(1):35–53, 2004.

[8] C. Büsing, A. M. C. A. Koster, and M. Kutschka. Recoverable Robust Knapsacks: the discrete scenario case. *Optimization Letters*, 5:379–392, 2011. ISSN 1862-4472.

[9] C. Büsing, A. M. C. A. Koster, and M. Kutschka. Recoverable Robust Knapsacks: Γ-Scenarios. In J. Pahl, T. Reiners, and S. Voß, editors, *Network Optimization, 5th International Conference, INOC 2011, Hamburg, Germany, June 2011, Proceedings*, volume 6701. Springer, 2011.

[10] G. Claßen, D. Coudert, A. M. C. A. Koster, and N. Nepomuceno. Bandwidth allocation for reliable fixed broadband wireless networks. In *Proceedings of IEEE WoWMoM 2011*, 2011.

[11] G. Claßen, D. Coudert, A. M. C. A. Koster, and N. Nepomuceno. A chance-constrained model & cutting planes for fixed broadband wireless networks. In J. Pahl, T. Reiners, and S. Voß, editors, *Network Optimization, 5th International Conference, INOC 2011, Hamburg, Germany, June 2011, Proceedings*, volume 6701, pages 37–42. Springer, 2011.

[12] G. Claßen, A. M. C. A. Koster, and A. Schmeink. Robust Planning of Green Wireless Networks. In *Proceedings of NetGCoop 2011*, 2011. to appear.

[13] G. B. Dantzig. Linear programming under uncertainty. *Management Science*, pages 197–206, 1955.

[14] O. Klopfenstein. Solving chance-constrained combinatorial problems to optimality. *Computational Optimization and Applications*, 45(3):607–638, 2010. ISSN 0926-6003.

[15] O. Klopfenstein and D. Nace. Valid inequalities for a robust knapsack polyhedron – Application to the robust bandwidth packing problem. In *Proc. of International Network Optimization Conference INOC*, 2009.

[16] A. M.C.A. Koster, M. Kutschka, and C. Raack. Cutset inequalities for robust network design. In *Proceedings of INOC 2011*, pages 118–123. International Network Optimization Conference, 2011.

[17] Arie M. C. A. Koster, Manuel Kutschka, and Christian Raack. Robust network design: Formulations, valid inequalities, and computations. ZIB Report 11-34, Zuse Institute Berlin, August 2011.

[18] Arie M.C.A. Koster and Manuel Kutschka. Network design under demand uncertainties: A case study on the abilene and geant network data. In *Tagungsband der 12. ITG-Fachtagung Photonische Netze*, volume 228, pages 154–161, Leipzig, Germany, 2011. VDE Verlag GmbH. http://www.vde-verlag.de/proceedings-de/453346030.html.

[19] Arie M.C.A. Koster, Manuel Kutschka, and C. Raack. Towards robust network design using integer linear programming techniques. In *Proceedings Next Generation Internet, NGI 2010*. IEEE Xplore, 2010. http://dx.doi.org/10.1109/NGI.2010.5534462.

[20] Arie M.C.A. Koster, Manuel Kutschka, and C. Raack. On the robustness of optimal network designs. In *Proceedings IEEE International Conference on Communications, ICC 2011*, pages 1–5. IEEE Xplore, 2011. http://dx.doi.org/10.1109/icc.2011.5962479.

[21] S. Mudchanatongsuk, F. Ordóñez, and J. Liu. Robust solutions for network design under transportation cost and demand uncertainty. *Journal of the Operational Research Society*, 59:652–662, 2008.

[22] A. Ouorou and J.-P. Vial. A model for robust capacity planning for telecommunications networks under demand uncertainty. In *6th International Workshop on Design and Reliable Communication Networks, 2007. DRCN 2007*, pages 1–4, 2007.

[23] M. Poss and C. Raack. Affine recourse for the robust network design problem: between static and dynamic routing. ZIB Report 11-03, Zuse Institute Berlin, February 2011.

[24] A. Prékopa. *Stochastic Programming*, volume 324 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[25] M. G. Scutellà. On improving optimal oblivious routing. *Operations Research Letters*, 37(3):197–200, 2009.

[26] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21:1154–1157, 1973.

METAHEURISTIC OPTIMIZATION

Zrinka Lukač
University of Zagreb, Faculty of Economics & Business
Trg J. F. Kennedy 6, 10000 Zagreb, Croatia
zlukac@efzg.hr

Abstract. Many real life problems are so complex in their nature that they cannot be efficiently and reliably solved within a reasonable amount of time by using traditional methods which guaranty the optimality of the solution. However, in recent years metaheuristics have proved to offer a promising approach to tackle this kind of problems. We give a short survey of the most important single-solution based metaheuristics for combinatorial optimization problems from the conceptual point of view and analyze their similarities and differences. The special attention is given to concepts of intensification and diversification, which are the two driving forces of any metaheuristics. For each metaheuristics presented, we study the mechanisms through which the intensification and diversification effects are achieved.

Keywords: metaheuristics, single-solution based methods, intensification, diversification

1 INTRODUCTION

Many real life optimization problems are so complex that traditional methods offer very little help in search of their solutions. Very often it is hard to tell how the solution looks like and where to look for the solution. At the same time the brute-force approach is out of the question because solution space is too wide. Metaheuristics offer a promising approach to tackle this kind of problems due to their ability to find acceptable solutions within a reasonable amount of computational time. The term metaheuristic was first introduced by F. Glover in [4] and it refers to master strategies that orchestrate the interaction between strategies which find locally optimal solutions and higher level strategies which make possible escaping from local optima and searching the whole solution space.

We give a survey of the most important single-based metaheuristics for combinatorial optimization problems as of today from the conceptual point of view and analyze their similarities and differences. Due to space limitation and extensiveness of the field, we consider single-solution based methods only, such as local search, tabu search, simulated annealing, greedy randomized adaptive search procedure, variable neighborhood search, iterated local search and guided local search.

Finding the right balance between intensification and diversification plays one of the key roles in design of any successful metaheuristic. Hereby diversification refers to the effective exploration of the whole search space, while the intensification refers to the thorough exploration of the promising regions which might contain good solutions. According to Glover and Laguna [6], the main difference between the intensification and diversification is that during an intensification stage the search focuses on examining neighbors of elite solutions, while the diversification stage encourages the search process to examine unvisited regions and to generate solutions that differ in various significant ways from those seen before. These two opposing principals are the driving forces of any metaheuristic. Therefore for each method presented we highlight the mechanisms through which the intensification and the diversification are achieved and controlled.

2 SINGLE-SOLUTION BASED METHODS

Single-solution based methods (also called trajectory methods) conduct the search by transforming a single solution by means of some iterative procedure. In general, after creating an initial solution, in each iteration they create a set of candidate solutions and replace the current solution by some candidate solution chosen according to a given evaluation criterion. Note that the new solution is not necessarily improving. The procedure stops once the predefined stopping criteria is met. Very often the effectiveness of the search procedure relies heavily on the definition of the neighborhood structure. In general, single-solution based methods are more intensification oriented.

2.1 Basic Local Search

Local search is very likely the oldest and the simplest metaheuristic method [9]. It starts from a given initial solution and in each iteration picks a neighbor which improves the objective function. The algorithm stops when all candidate neighbors are worse than the current solution. Different strategies may be applied when choosing a neighbor, such as best improvement (choose the best neighbor), first improvement (chose the first neighbor better than the current solution) and random selection (chose a new solution randomly from the set of neighbors who are improving the current solution). The main disadvantage of the algorithm is that it does not posses the ability to escape local optima. Therefore, the intensification component is the dominant one.

2.2 Simulated Annealing

Simulated Annealing [1,3,8,9] was one of the first metaheuristics to employ the idea of accepting solutions of a worse quality than the current one as a means of escaping from local optimum. It is inspired by the annealing process in metallurgy which involves heating and controlled cooling of a material so that atoms place themselves in a pattern that corresponds to the global energy minimum, thus reducing the defects in a material.

In the initialization phase, one has to set the initial temperature T as well as the temperature cooling schedule. They are of crucial importance for the performance of the algorithm. The search starts from some initial solution s. Next, at each iteration a new candidate solution s' is picked randomly from the set of neighbors of the current solution, $N(s)$. If s' improves the objective, then it becomes a new current solution. If not, it may still be accepted as a new current solution, but with acceptance probability p which is a function of the temperature parameter T and the difference of the objective function value in s' and s. The acceptance probability decreases as the temperature parameter decreases. The temperature parameter is updated at the end of each iteration according to the cooling schedule. The search stops once the stopping criterion is met.

The diversification and intensification mechanisms are controlled by the cooling schedule and the acceptance probability function. The diversification component is dominant at the beginning of the search since the acceptance probability is higher at higher temperatures. As the temperature decreases, the acceptance probability decreases and the bias moves towards the intensification. Therefore, the decrease of the temperature parameter drives the system from diversification to intensification, which in turn leads to the convergence of the system. At a fixed temperature, the acceptance probability decreases as the difference between the objective function value in s' and s increases.

2.3 Tabu Search

Tabu Search [1,3,4,5,6,9] was originally proposed by Glover [4] and is one of the most used metaheuristic methods. The main idea behind the search is the use of memory which enables both escaping from local optimum and diversification of the search. While simulated annealing may accept degrading solutions at any time, tabu search will do so only if it is trapped in local optima.

A short-term memory is implemented in the form of tabu list which contains a constant number of the most recently visited solutions or their attributes (depending on the definition of the search space). The length of a tabu list is controlled by a parameter called tabu tenure. At each iteration, the set of candidate solutions is restricted to the set of neighbors which are not tabu. The best of them is selected as a new solution. However, if the tabu list contains only some of the solution attributes, tabu list may be too restrictive in a sense that it may reject good solutions which have not yet been visited. In order to overcome this problem, a tabu move may be accepted if a set of predefined conditions (so called aspiration criteria) is satisfied. The tabu list is updated at the end of each iteration. The procedure stops once the stopping criteria are met.

Tabu Search behaves like a best improvement local search algorithm. The difference is that the use of tabu list enables avoiding cycles and escaping from local optima. The use of tabu list has both an intensification and diversification effect on the search. In every step it imposes restrictions on the set of possible solution, thus having a diversification effect. At the same time, such a restriction determines the set of neighbors and in that way influences the choice of the best neighbor, thus having an intensifying effect. The length of the tabu list determines the balance between these two effects. The shorter the tabu list, the lower the influence of the diversifying effect. The longer the tabu list, the higher the influence of the diversifying effect.

Intensification may also be achieved by means of medium term memory, which is very often represented in the form of the recency-based memory. For each solution (or its attribute), the recency based memory remembers the most recent iteration (or the number of successive iterations) it was involved in. The idea is to extract common features of elite solutions and then intensify the search around solutions which contain them. On the other hand, diversification may be achieved by means of long-term memory, which is very often represented in the form of frequency-based memory. For each solution (or its attribute), it memorizes how many times it has been visited. In this way it is possible to identify the regions which have already been explored and guide the search towards the unvisited regions.

2.4 Greedy Randomized Search Procedure (GRASP)

Greedy Randomized Search Procedure [1,2,5,9] is an iterative greedy heuristic for combinatorial optimization problems in which each iteration consists of a construction and a local search step. The construction step is applied first. It builds a feasible solution by adding one new component at a time. A new component is added to the partial solution by randomly choosing an element from the restricted candidate list, which consists of α best candidate elements whose addition to the existing partial solution does not destroy feasibility. Hereby, the elements are evaluated according to some greedy function, usually the cost of incorporating this element into the partial solution already being constructed. After a new component is added to the partial solution, the candidate list is updated and the incremental costs are reevaluated. Once a complete solution is built, a local search step is applied in order to improve the solution. Very often a simple local search is performed. The process iterates until a predefined number of iterations is performed. The output of the algorithm is the best solution found in the process.

Local search step has an intensifying effect on the search. On the other hand, diversification is achieved by means of restricted candidate list. Here, the length of the restricted candidate list, α, is a crucial parameter which determines the diversifying effect of the search. For α=1, the construction step is equivalent to a deterministic greedy heuristics, since the best element from the candidate list is always added to the solution. For α equal to the number of candidate elements, construction is completely random, thus achieving the maximum diversifying effect.

2.5 Variable Neighborhood Search (VNS)

Variable Neighborhood Search [1,3,5,7,9] is a metaheuristics method which conducts the local search by systematically changing the neighborhood structures. In that way, it explores increasingly distant neighborhoods of the current solution. It jumps from this solution to a new one if and only if a new solution improves the current solution. VNS is based on three simple observations. The first one is that a local optimum with respect to one neighborhood structure is not necessary so with another. The second one is that a global optimum is a local optimum with respect to all possible neighborhood structures. The third one is that for many problems local optima with respect to one or several neighborhood structures are relatively close to each other [7].

The basic VNS scheme can be described as follows. In the initialization phase, an initial solution s is generated and a finite number, k_{max}, of neighborhood structures to be used in search is defined. At the beginning of each iteration, the neighborhood counter k is set to 1. Then a shaking step begins. It refers to picking a random solution s' from the k-th neighborhood structure of the current solution s. Next, a local search step is applied with s' as the starting point. It should be noted that the local search step is independent of the neighborhood structures defined in the initialization phase. Let s'' denote the so obtained optimum. If s'' improves the objective, it becomes a new current solution s. Otherwise, the neighborhood counter is incremented by 1 and the shaking and the local search step are conducted again. This time shaking is conducted with respect to the new neighborhood structure. Once all the neighborhood structures are exhausted, i.e. when $k = k_{max}$, the stopping criteria are checked. The stopping criteria might be a maximum number of iterations, maximum number of iteration without improvement, a maximum CPU time allowed, etc. If the maximum number of neighborhood structures is exhausted and the stopping criteria still do not hold, the counter k is set to 1, and the whole process of running shaking and local search steps starts all over again. The output of the algorithm is the best solution found.

The key issue in design of VNS metaheuristic is the choice of neighborhood structures. Local search step contributes to the intensification of the search, while diversification is achieved through shaking stage and the process of changing neighborhoods in case of no improvements.

2.6 Iterated Local Search (ILS)

Iterated local search [1,3,5,9] is the most general of all the methods presented so far. It is a framework for some other metaheuristics (like VNS), but it can also incorporate other metaheuristics as its subcomponents. The main idea is very simple. It starts from some initial solution and applies local search on it. Then, at each iteration, it perturbs the obtained local optimum and applies local search on the perturbed solution. If the so obtained solution satisfies an acceptance criterion, it becomes a new solution. The process iterates until the stopping criterion is met.

The local search component of the ILS framework can be any of the single-solution based metaheuristics. Its contribution is towards the intensification of the search. The balance between the intensification and the diversification is achieved through choice of perturbation method and acceptance criterion. The choice of perturbation method is crucial. If a perturbation is too small, it might not be able to escape from local optimum and prevent cycling. Therefore small perturbations contribute to the intensification of the search. If a perturbation is too large, it might loose the memory of good properties of local optima already found. Large perturbations contribute to the diversification of the search. The acceptance criterion, on the other hand, acts as a counterbalance to perturbation. The effect of the acceptance criterion on the intensification and the diversification of the search is highly dependent on its definition. If only improving solutions in terms of objective function values are accepted, then it has a strong intensification effect. If any solution is accepted, regardless of its quality, then it has a strong diversifying effect. Acceptance criterion can take a range of values in-between these two extreme cases.

2.7 Guided local search (GLS)

Guided local search [1,3,5,9,10] is another general metaheuristics. It acts like an upper level strategy on top of other metaheuristics in order to improve their efficiency. Unlike all the other methods presented so far, its main approach to escaping local optima and guiding the search is by means of dynamically changing the objective function.

For a given optimization problem, a set of solution features is defined first, where solution features are solution characteristics used to discriminate between the solutions. A cost and a penalty is associated to each feature. In order to escape from local optimum, the method modifies the cost function by penalizing the unfavorable solution features. A penalty associated with each feature measures the importance of the feature. The higher the penalty, the higher the importance of the feature and the associated cost of having that feature in the solution. The method takes into account the current penalty value of the feature by considering the utility of penalizing that feature. If a feature is not present in the local optimum, then the utility of penalizing is 0. Furthermore, the higher the cost of the feature, the greater the utility of penalizing it. The more times the feature has been penalized, the lower the utility of penalizing it again. The penalty of the selected feature is always increased by 1, where the scaling of the penalty is normalized by some parameter λ.

The algorithm starts form some initial solution, with all the penalties initialized to 0. It applies some local search method until a local optimum with respect to the modified objective function (the original objective function plus the penalties) is achieved. Here the local search step can consist of some other metaheuristic. The feature utilities are computed next and the penalties of the features having the maximum utility are incremented. The process iterates until the stopping criteria are met. The output of the algorithm is the best solution found.

The efficiency of the search is affected by the choice of the solution features, their costs and parameter λ. The search will intensify in regions defined by lower costs of the features. On the other hand, diversification is achieved by penalizing the features of the local optima found so far. In that way, the search is diverted from searching the regions around local optima and directed towards the unexplored regions of the search space. For majority of problems, GLS is not very sensitive to the choice of λ. However, large values of λ will create greater diversification effect, while small values of λ will result in intensification of the search around local optima. The more effective the metehauristics used in the local search is, the lesser the value of λ and the lesser the values of penalties are needed.

3 CONCLUSION

Over the last couple of decades metaheuristics have proved to be a promising approach to tackle complex real life optimization problems. In this paper we have presented some of the most important single-solution based metaheuristics as of today. Special attention has been given to the principles of intensification and diversification, the two driving principals whose balance is crucial for design of any effective metaheuristics. We have presented each method from the conceptual point of view and highlighted the control mechanisms through which the intensification and diversification effects are achieved. It should be noted that the very same mechanism often has both an intensification and diversification effect on the search. In general, the higher the influence of the objective function criterion, the higher the intensification effect. On the other hand, diversification is achieved by using criteria other than objective function value. Therefore, in order to design an efficient problem specific metaheuristic, special attention should be given to finding the right balance between intensification and diversification effects.

References

[1] Blum, C., Roli, A., 2003. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. ACM Computing Surveys 35 (3), pp. 268-308

[2] Feo, T. A., Resende, M. G. C., 1995. Greedy randomized adaptive search procedures. Journal of Global Optimization 6, pp. 109–133

[3] Gendreau, M., Potvin, J. (Eds), 2010. *Handbook of Metaheuristics*, 2nd ed., Springer

[4] Glover, F. 1986. Future paths for integer programming and links to artificial intelligence. Computers & Operations Research 13, pp. 533–549

[5] Glover, F., Kochenberg, G. A. (Eds), 2003. *Handbook of Metaheuristics*, Kluwer Academic Publishers

[6] Glover, F., Laguna, M., 1997. *Tabu Search*. Kluwer Academic Publishers.

[7] Hansen, P., Mladenović, N., 2001. Variable neighborhood search: Principles and applications. European Journal of Operations Research 130, pp. 449–467

[8] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P., 1983. Optimization by simulated annealing. Science 220 (4598), pp. 671-680

[9] Talbi, E. G., 2009. *Metaheuristics: From design to implementation*. John Wiley & Sons

[10] Voudouris, C., Tsang, E. 1999. Guided local search. European Journal of Operations Research 113 (2), pp. 469–499

Mathematical Models and Solutions for Network Design Problems

Ulrich Pferschy

Department of Statistics and Operations Research, University of Graz, Universitaetsstrasse 15, A-8010 Graz, Austria.
pferschy@uni-graz.at

Abstract

We discuss two real-world scenarios of network design problems arising in the design of district heating systems and in the design of fiber optic networks. In both cases we can choose which customers to connect to the network, depending on the resulting profit, while the necessary connections incur construction costs. Maximizing the total profit minus the total cost yields the prize collecting Steiner tree problem (PCST) with additional connectivity constraints. We present different mathematical solution methods, namely cut-based models, multi-commodity flow formulations and path models. The computational behavior of these models will be illustrated.
Keywords: Network Design, Prize Collecting Steiner Tree, Branch-and-Cut

1 Introduction

Classical network design problems consider the most cost efficient way to connect a given set of customers to a network, often with additional conditions such as network reliability, capacity or degree constraints. A more complex problem arises if the owner of the network (e.g. a provider of public utilities) is free to choose which customers to connect. Indeed, it may be unprofitable to connect customers which pay a low prize for the network service but cause a huge connection cost because of their geographic position. Hence, we are faced with two decisions: On one hand, a subset of profitable customers has to be selected form a given ground set, on the other hand, the selected customers have to be connected by a network with lowest possible cost. In this situation we are faced with a trade-off between maximizing the sum of profits over all selected customers and minimizing the construction cost of the network. This leads to a *prize-collecting* objective function, which we studied extensively in Ljubic et al. [18]. The motivation of that paper was the connection of houses to a district heating system.

More formally, we formulate this problem as an optimization problem on an undirected graph $G = (V, E, c, p)$, where the vertices V are associated with profits, $p : V \to \mathbb{R}^+$, and the edges E with costs, $c : E \to \mathbb{R}^+$.

The graph corresponds to the topology of the customer vertices with edges representing street segments or potential cable connections. The cost c of each edge is the cost of establishing the connection, e.g., of laying a pipe or cable. The profit p of each vertex gives an estimate of the potential gain of revenue caused if the associated customer is connected to the network. Then we can formulate the *Linear Prize-Collecting Steiner Tree problem* (PCST) as finding a connected subgraph $T = (V_T, E_T)$ of G, $V_T \subseteq V$, $E_T \subseteq E$ that maximizes total profits minus construction costs, i.e.

$$profit(T) = \sum_{v \in V_T} p(v) - \sum_{e \in E_T} c(e) \quad . \tag{1}$$

A different objective function, where the ratio of profit over costs (resembling the return on investment) is maximized was considered by Klau et al. [14].

Obviously, every optimal solution T will be a tree. Note that vertices of the graph, which represent junctions of cables or streets but do not incur customer profits, can be easily represented as vertices with zero profit.

This problem was introduced in the literature (in a slightly modified form) by Segev [20] in 1987. Approximation algorithms were presented by Bienstock et al. [2], Goemans and Williamson [10], Johnson et al. [12] and Feofiloff et al. [7]

Preprocessing and reduction procedures were described by Duin and Volgenant [5] and Uchoa [21]. Metaheuristics are given by Canuto et al. [3] and Klau et al. [13]. Fischetti [8] and Goemans [9] studied the polyhedral structure of a closely related problem. Exact algorithms and lower bounds required for enumeration schemes were developed by Engevall et al. [6], Lucena and Resende [19] and more recently by Haouari et al. [11].

2 A Branch-and-Cut Algorithm

The currently best algorithmic framework for the exact solution of PCST was developed in Ljubic et al. [18], on which this article is partially based.

Their approach starts with a number of preprocessing steps to reduce the size of the given graph. These are a *Least-Cost Test* where each edge cost c_{ij} is replaced by the cost of the shortest path from i to j. The *Degree-l Test* to replace a connected set of junction vertices by a minimum spanning tree (see Uchoa [21] for extension of this idea), and a *Minimum Adjacency Test*, which merges two adjacent vertices i and j if

$$\min\{p_i, p_j\} - c_{ij} > 0 \text{ and } c_{ij} = \min_{it \in E} c_{it}.$$

The general approach of [18] is an integer linear programming formulation defined on a directed graph model and using connectivity inequalities corresponding to minimum weight cuts in the graph to guarantee connectivity of the solution. Related models were used by Wong [24] and Fischetti [8]. An undirected cut model was given by Aneja [1].

Assuming that E contains all arcs (i, j) and (j, i) we introduce binary variables $x \in \{0, 1\}^E$ and $y \in \{0, 1\}^V$ to represent the solution tree $T = (V_T, E_T)$ with the following interpretation:

$$x_{ij} = \begin{cases} 1 & (i,j) \in E_T \\ 0 & \text{otherwise} \end{cases} \quad \forall (i,j) \in E, \quad y_i = \begin{cases} 1 & i \in V_T \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in V, i \neq r \quad .$$

In a slightly simplified way the cut model for PCST can be written as follows:

$$
\begin{aligned}
(CUT) \qquad \max \quad & \sum_{i \in V} p_i y_i - \sum_{ij \in E} c_{ij} x_{ij} && && (2)\\
\text{subject to} \quad & \sum_{ji \in E} x_{ji} = y_i && \forall i \in V \setminus \{r\} && (3)\\
& x(\delta^-(S)) \geq y_k && k \in S, r \notin S, \forall S \subset V && (4)\\
& \sum_{ri \in E} x_{ri} \geq 1 && && (5)\\
& x_{ij}, y_i \in \{0,1\} && \forall (i,j) \in E, \forall i \in V \setminus \{r\} && (6)
\end{aligned}
$$

A given root vertex of the network is denoted by $r \in V$.

The objective function (2) corresponds directly to our goal as given in (1). Constraints (3) guarantee that an arc is incident to a vertex i if and only if i is included in the solution set V_T, while (5) enforces the inclusion of the root r. Finally, the crucial cut constraints (4) guarantee the connectedness of the solution tree. As usual, $\delta^-(S) = \{(i,j) \mid i \in \bar{S}, j \in S\}$ denotes the set of edges reaching into set S and starting from its complement. Note that disconnectivity would imply the existence of a cut S separating r and v which would clearly violate the corresponding cut constraint.

The classical branch-and-cut approach requires the separation of the cut constraints. This means that instead of enumerating all (exponentially many) cut sets S we compute only the *most violated* cut, i.e. the cut with smallest cut value. Such a min-cut computation is well-known to be equivalent to a *max-flow* problem. This can be done very efficiently by Cherkassky and Goldberg's maximum flow algorithm [4].

It turned out in the computational experiments (see [18] for details) that this straightforward approach fails to solve large test instances to optimality within reasonable running time. Thus, several algorithmic improvements were introduced.

First of all, it is beneficial to add more than one cut to the ILP model in each iteration. Such a *nested cut* approach leads to faster changes of the value of the LP-relaxation. Furthermore, instead of computing flows only from the root to each customer vertex, it is also convenient to compute *backcuts* arising from a hypothetical flow from the customer back to the root vertex.

Extensive computational experiments showed that the resulting algorithmic framework performs better than previous algorithms on benchmark instances from the literature and is able to solve problems with up to 60.000 edges to optimality within 10 minutes. Also real-world instances provided by a German telecom provider could be solved to optimality in reasonable running time.

3 Redundant Connectivity

A classical issue for network design problems is reliability. In particular, the failure of a single link of the network, i.e. the disruption of a cable by construction work or technical failure, should not disconnect any customer from the network root. Clearly, this requirement is fulfilled by a biconnected network structure where each vertex has two disjoint paths to the root. However, the construction of a biconnected network incurs extremely high costs. Thus, in many applications the network provider strives for a relaxed version of reliability. It may be acceptable that the disruption of a single edge of the network disconnects a limited set of customers from the network, as long as the vast majority of customer vertices remains connected. Such a scenario occurred in the design of fiber optic networks taking into account the last mile connections to the customers. This problem is currently a major issue for many telecommunication providers and was studied in the research project NETQUEST led by the Carinthia University of Applied Sciences and supported by the Austrian Research Promotion Agency (FFG), together with partners from industry.

A formal model of the above idea was described in Wagner et al. [22, 23] and Leitner et al. [16, 15] and works as follows:

For each customer vertex k we are given a distance value $b_{\max}(k)$ and require that vertex k is connected by a single path of length at most $b_{\max}(k)$ to a biconnected vertex, i.e. to a vertex with two disjoint paths to the root. This means that any failure of an edge will disconnect only a limited area of a certain diameter from the network. For many application scenarios this model offers a reasonable compromise between construction cost and network reliability. Of course, very important customers might still be biconnected by setting their distance value to 0.

A direct extension of the directed cut model sketched in Section 2 for this model was given by Wagner et al. [22]. However, modeling the single path of length $b_{\max}(k)$ turned out to be against the spirit of the cut idea. It requires to identify for each vertex whether it is single or biconnected and then establishes the length of the path connection to the nearest biconnected vertex. Computational experiments showed only moderate success of this approach.

The multi-commodity flow formulation introduced in Wagner et al. [23] models the connection of each customer vertex to the root by a different commodity within a multi-commodity flow problem. There are two different flows from the root to each vertex. Flows to biconnected vertices are required to be disjoint while flows on the last segment of length $b_{\max}(k)$ will coincide. This model was better suited to capture the essence of the relaxed connectivity.

Finally, a column generation approach was introduced by Leitner and Raidl [15] and further developed in Leitner et al. [17]. It is based on the implicit enumeration of the sets P_k for all customer vertices k consisting of all feasible connections of k to the root. Each element $p \in P_k$ is a fork-like structure consisting of two disjoint paths from r to a certain junction vertex z_k and a single path connecting z_k and k with length at most $b_{\max}(k)$. Binary assignment variables f_p^k assign a unique connection $p \in P_k$ to each customer vertex k.

Instead of enumerating the exponentially many fork connections they are generated iteratively in a branch-and-price framework (see Leitner et al. [17] for more details). It turns out that the associated pricing problem asks for the computation of the cheapest fork connection with positive edge costs. While this could be done very efficiently in polynomial time for the special cases of $b_{\max}(k) = \infty$ (single connection) and $b_{\max}(k) = 0$ (strictly biconnected), a mixed-integer linear programming model has to be solved for the general case. Alternatively, it can also be shown that the resulting problem is equivalent to an *elementary shortest path problem with resource constraints*.

Computational experiments show that the standard column generation approach suffers from primal degeneracy which worsens its performance considerably. Thus we applied stabilization techniques based on the selection of dual variables in the LP-relaxation (see Leitner et al. [16]). Computational results show that the branch-and-price approach performs reasonably well on small and medium sized instances while large test instances still pose an interesting challenge.

References

[1] Y.P. Aneja. An integer linear programming approach to the Steiner problem in graphs. *Networks*, 10:167–178, 1980.

[2] D. Bienstock, M.X. Goemans, D. Simchi-Levi, and D. Williamson. A note on the prize-collecting traveling salesman problem. *Mathematical Programming*, 59:413–420, 1993.

[3] S.A. Canuto, M.G.C. Resende, and C.C. Ribeiro. Local search with perturbations for the prize-collecting Steiner tree problem in graphs. *Networks*, 38:50–58, 2001.

[4] B.V. Cherkassky and A.V. Goldberg. On implementing push-relabel method for the maximum flow problem. *Algorithmica*, 19:390–410, 1997.

[5] C.W. Duin and A. Volgenant. Some generalizations of the Steiner problem in graphs. *Networks*, 17(2):353–364, 1987.

[6] S. Engevall, M. Göthe-Lundgren, and P. Värbrand. A strong lower bound for the node weighted Steiner tree problem. *Networks*, 31(1):11–17, 1998.

[7] P. Feofiloff, C.G. Fernandes, C.E. Ferreira, and J.C. Pina. Primal-dual approximation algorithms for the prize-collecting Steiner tree problem. *Information Processing Letters*, 103:195–202, 2007.

[8] M. Fischetti. Facets of two Steiner arborescence polyhedra. *Mathematical Programming*, 51:401–419, 1991.

[9] M.X. Goemans. The Steiner tree polytope and related polyhedra. *Mathematical Programming*, 63:157–182, 1994.

[10] M.X. Goemans and D.P. Williamson. The primal-dual method for approximation algorithms and its application to network design problems. In D. S. Hochbaum, editor, *Approximation algorithms for NP-hard problems*, pages 144–191. P. W. S. Publishing Co., 1996.

[11] M. Haouari, S.B. Layeb, and H.D. Sherali. The prize collecting steiner tree problem: Models and lagrangian dual optimization approaches. *Computational Optimization and Applications*, 40:13–39, 2008.

[12] D.S. Johnson, M. Minkoff, and S. Phillips. The prize-collecting Steiner tree problem: Theory and practice. In *Proceedings of 11th ACM-SIAM Symposium on Discrete Algorithms*, pages 760–769, San Francisco, CA, 2000.

[13] G.W. Klau, I. Ljubić, A. Moser, P. Mutzel, P. Neuner, U. Pferschy, and R. Weiskircher. Combining a memetic algorithm with integer programming to solve the prize-collecting Steiner tree problem. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2004)*, volume 3102 of *LNCS*, pages 1304–1315. Springer, 2004.

[14] G.W. Klau, I. Ljubić, P. Mutzel, U. Pferschy, and R. Weiskircher. The fractional prize-collecting Steiner tree problem on trees. In G. Di Battista and U. Zwick, editors, *ESA 2003*, volume 2832 of *LNCS*, pages 691–702. Springer, 2003.

[15] M. Leitner and G. Raidl. Strong lower bounds for a survivable network design problem. *Electronic Notes in Discrete Mathematics*, 36:295–302, 2010.

[16] M. Leitner, G. Raidl, and U. Pferschy. Accelerating column generation for a survivable network design problem. In M.G. Scutella, editor, *Proceedings of the International Network Optimization Conference (INOC)*, 2009.

[17] M. Leitner, G. Raidl, and U. Pferschy. Branch-and-price for a survivable network design problem. Technical Report TR-186-1-10-02, Department of Computer Graphics and Algorithms, Vienna University of Technology, 2010.

[18] I. Ljubic, R. Weiskircher, U. Pferschy, G.W. Klau, P. Mutzel, and M. Fischetti. An algorithmic framework for the exact solution of the prize-collecting steiner tree problem. *Mathematical Programming*, 105:427–449, 2006.

[19] A. Lucena and M.G.C. Resende. Strong lower bounds for the prize-collecting Steiner problem in graphs. *Discrete Applied Mathematics*, 141:277–294, 2004.

[20] A. Segev. The node-weighted Steiner tree problem. *Networks*, 17:1–17, 1987.

[21] E. Uchoa. Reduction tests for the prize-collecting Steiner problem. *Operations Research Letters*, 34:437–444, 2006.

[22] D. Wagner, U. Pferschy, P. Mutzel, G. Raidl, and P. Bachhiesl. A directed cut model for the design of the last mile in real-world fiber optic networks. In B. Fortz, editor, *Proceedings of the International Network Optimization Conference (INOC)*, pages 103/1–6, 2007.

[23] D. Wagner, U. Pferschy, P. Mutzel, G. Raidl, and P. Bachhiesl. A multi-commodity flow approach for the design of the last mile in real-world fiber optic networks. In *Operations Research Proceedings 2006*, pages 197–202. Springer, 2007.

[24] R.T. Wong. A dual ascent based approach for the Steiner tree problem in directed graphs. *Mathematical Programming*, 28:271–287, 1984.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section I:

Pascal 2 session

A SURVEY ON KNOWN UPPER BOUNDS FOR MULTICHROMATIC NUMBER FOR HEXAGONAL GRAPHS

Petra Šparl
University of Maribor, Faculty of Organizational Sciences
Kidričeva cesta 55a, SI-4000 Kranj, Slovenia
petra.sparl@fov.uni-mb.si
and
Institute of Mathematics, Physics and Mechanics
Jadranska 19, SI-1000 Ljubljana, Slovenia

Abstract

An optimization problem in the design of cellular networks is to assign sets of frequencies to transmitters to avoid unacceptable interference. Frequency assignment problem can be abstracted as a multicoloring problem on a weighted subgraph of the infinite triangular lattice, called hexagonal graph. In this paper a survey of known upper bounds for multichromatic number $\chi_m(G)$ in terms of weighted clique number $\omega_m(G)$ for hexagonal graphs and some open problems are presented.

Keywords: graph algorithm, approximation algorithm, graph coloring, frequency planning, cellular networks, local distributed algorithm

1 INTRODUCTION

A fundamental problem concerning cellular networks is to assign sets of frequencies (colors) to transmitters (vertices) in order to avoid unacceptable interferences [2]. The number of frequencies demanded at a transmitter may vary between transmitters. In a usual cellular model, transmitters are centers of hexagonal cells and the corresponding adjacency graph is a subgraph of the infinite triangular lattice. An integer $p(v)$ is assigned to each vertex of the triangular lattice and is called the *demand* of vertex v. The vertex-weighted graph induced on the subset of the triangular lattice of vertices of positive demand is called a *hexagonal graph*, and is denoted $G =$

(V, E, p). Hexagonal graphs arise naturally in studies of cellular networks, such as known Philadelphia examples [10]. A proper *n-[p]coloring* of G (also called *multicoloring*) is a mapping $f : V(G) \to 2^{\{1,\ldots,n\}}$ such that $|f(v)| \geq p(v)$ for any vertex $v \in V(G)$ and $f(v) \cap f(u) = \emptyset$ for any two adjacent vertices u and v. The least integer n for which a proper n-[p]coloring exists, denoted by $\chi_m(G)$, is called the *multichromatic number* of G. Another invariant of interest in this context is the *(weighted) clique number* $\omega_m(G)$, defined as follows: the weight of a clique of G is the sum of demands on its vertices and $\omega_m(G)$ is the maximal clique weight on G. Clearly, $\chi_m(G) \geq \omega_m(G)$.

Even for graphs with regular structure, such as hexagonal graphs, the problem of determining $\chi_m(G)$ is not trivial. It has been showed [6] that it is NP-complete to decide whether $\chi_m(G) = \omega_m(G)$ for an arbitrary hexagonal graph G. Hence, it is unlikely to expect that a polynomial time algorithm for computing $\chi_m(G)$ for an arbitrary hexagonal graph G can be devised. Therefore, it is of interest to study approximation algorithms for the discussed problem. In the last decade several results, reporting upper bounds for $\chi_m(G)$ in terms of $\omega_m(G)$ for hexagonal graphs, appeared in the literature.

In the continuation we will talk about so called distributed and k-local algorithms. An algorithm is *distributed* if it uses a distributed computation, such that for solving one big problem a set of interconnected processors is used. In the case of algorithms for graph multicoloring it means that we put processors in several vertices of a given graph such that each processor covers only a particular part of the graph. An algorithm for a graph multicoloring is *distributed* if computation of a single processor does not depend on the size of the graph.

An algorithm is a *k-local algorithm* if instead of a global information it needs only some local information of a given graph G. In the case of a graph multicoloring this means that every vertex $v \in G$, that has to be multicolored, uses only information about the demands of vertices whose graph distance from v is less than or equal to k.

The paper is organized as follows. In the next two sections the known results for multicoloring arbitrary hexagonal graphs and triangle-free hexagonal graphs are stated, respectively. In the last section some open problems are mentioned.

2 KNOWN RESULTS FOR ARBITRARY HEXAGONAL GRAPHS

In this section a survey of known upper bounds for weighted chromatic number in terms of weighted clique number for an arbitrary hexagonal graph G are given. The most of listed results are obtained by using rich structural properties of hexagonal graphs, especially the fact that there exists a natural 3-coloring of the vertices of the infinite triangular lattice, which gives rise to the partition of the vertex set of any hexagonal graph into three independent sets.

For arbitrary hexagonal graphs the upper bound $\chi_m(G) \leq (4/3)\omega_m(G)+C$, where C is an absolute constant, is the best known for both distributed and non-distributed model of computation. The bound was almost at the same time, but independently, proved by several authors [6, 7, 18]. All proofs are constructive, thus implying the existence of 4/3-approximation algorithms. But only the algorithm presented in [7, 8] is distributed and guarantees the $\left\lceil \frac{4\omega_m(G)}{3} \right\rceil$ bound. Later a distributed algorithm which gives a proper multicoloring of a hexagonal graph with at most $\left\lfloor \frac{4\omega_m(G)+1}{3} \right\rfloor$ colors, which is less than or equal to $\left\lceil \frac{4\omega_m(G)}{3} \right\rceil$, is given in [12]. Figure 1 shows two simple hexagonal graphs H and C_9, where numbers present demands of vertices. Note that $\omega_m(H) = 3$ and $\chi_m(H) = 4$, which means that the bound $\chi_m(G) \leq (4/3)\omega_m(G) + C$ is optimal for the general case.

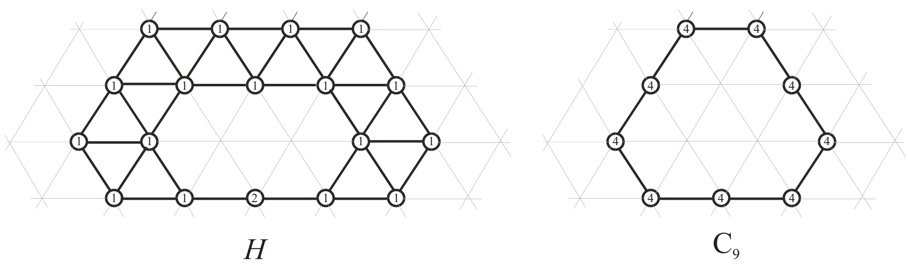


Figure 1: Two simple examples of hexagonal graphs.

An obvious drawback of the mentioned distributed algorithms is that each vertex needs one piece of global information, more precisely $\omega_m(G)$. Since so called k-local algorithms are more effective in the practice, several local algorithms for multicoloring hexagonal graphs appeared in the literature, where vertices can communicate to its neighbors to obtain some local information on graph G. A framework for studying distributed online assignment in cellular networks was developed in [5]. In particular 3-competitive 0-local, 3/2-competitive 1-local, 17/12-competitive 2-local and 4/3-competitive 4-local algorithm are presented. In [14] a 2-local algorithm with competitive ratio 4/3 is given. The best ratio for 1-local case for general hexagonal graphs was first improved to 13/9 in [1], later to 17/12 in [19], and finally to 7/5 in [16].

3 KNOWN RESULTS FOR TRIANGLE-FREE HEXAGONAL GRAPHS

It turned out that better bounds for weighted chromatic number in terms of weighted clique number can be obtained for triangle-free hexagonal graphs. The conjecture proposed by McDiarmid and Reed [6] is that $\chi_m(G) \leq (9/8)\omega_m(G) + C$, where C is an absolute constant, holds for triangle-free hexagonal graphs. It is not difficult to see that for the cycle C_9, depicted on the right picture of Figure 1, it holds $\omega_m(C_9) = 8$ and $\chi_m(C_9) = 9$, which means that the bound $9/8\omega_m(G)$ is the best possible for an arbitrary triangle-free hexagonal graph G.

Several algorithms reporting upper bounds for $\chi_m(G)$ for triangle-free hexagonal graphs can be found in the literature. The bound $\chi_m(G) \leq (7/6)\omega_m(G) + C$, where C is an absolute constant, is the best known for the triangle-free hexagonal graphs at the moment. The history of the results for triangle-free hexagonal graphs is the following. In [4] a distributed algorithm with competitive ratio 5/4 was presented. Later a 2-local distributed algorithm with the same ratio was given in [13], while an inductive proof for ratio 7/6 is reported in [3]. Better results are obtained for special sub-classes of triangle-free hexagonal graphs, where some particular configurations are forbidden. Namely, a 2-local 7/6-competitive algorithm for multicoloring triangle-free hexagonal graphs with no adjacent centers (i.e. vertices which has at least two neighbors in G, which are not on the same line) is given in [15] and a 1-local 4/3-competitive algorithm for multicoloring a similar sub-class of hexagonal graphs is presented in [20].

A special case of a proper multicoloring is when p is a constant function. For example, a 7-[3]coloring of a graph G is an assignment of three colors between 1 and 7 to each vertex $v \in G$. An elegant idea that implies the existence of a 14-[6]coloring is presented in [11]. Recently an algorithm to find a 7-[3]coloring of an arbitrary triangle-free hexagonal graph G was given in [9], which implies that $\chi_m(G) \leq (7/6)\omega_m(G) + C$. This provides a shorter alternative proof to the inductive proof of Havet [3] and improves the short proof of [11] that implied the existence of a 14-[6]coloring. Note that the algorithm in [9] is not linear because it uses a 4-coloring of a planar graph. This was improved very recently in [17], where a linear time algorithm for 7-[3]coloring of triangle-free hexagonal graphs is presented.

4 OPEN PROBLEMS

In the case of multicoloring problem for an arbitrary hexagonal graph the remaining question is whether one can find a 1-local algorithm with competitive ratio 4/3.

The 7-[3]coloring algorithm for multicoloring triangle-free hexagonal graphs presented in [17] is not distributed since the construction in one Lemma in the paper does not work in constant time if long paths exist. Therefore, it is an interesting question whether there exists a distributed algorithm for 7-[3]coloring of an arbitrary triangle-free hexagonal graph, using its rich structural properties.

At last, the conjecture of McDiarmid and Reed, that for triangle-free hexagonal graphs the inequality $\chi_m(G) \leq (9/8)\omega_m(G) + C$ holds, still remains opened.

References

[1] F.Y.L. Chin, Y. Zhang, and H. Zhu, *A 1-local 13/9-competitive Algorithm for Multicoloring Hexagonal Graphs*, Proc. of the 13th Annual International Computing and Combinatorics Conference (COCOON), volume 4598 of LNCS, pp. 526-536, 2007.

[2] W. K. Hale, *Frequency Assignment: Theory and Applications*, Proc. of the IEEE **68**(12):1497-1514, 1980.

[3] F. Havet, *Channel Assignment and Multicoloring of the Induced Subgraphs of the Triangular Lattice*, Discrete Mathematics **233**(1-3):219-231, 2001.

[4] F. Havet and J. Žerovnik, *Finding a Five Bicolouring of a Triangle-free Subgraph of the Triangular Lattice*, Discrete Mathematics **244**(1-3):103-108, 2002.

[5] J. Janssen, D. Krizanc, L. Narayanan, and S. Shende, *Distributed Online Frequency Assignment in Cellular Networks*, Journal of Algorithms **36**(2):119-151, 2000.

[6] C. McDiarmid and B. Reed, *Channel Assignment and Weighted Colouring*, Networks **36**(2):114-117, 2000.

[7] L. Narayanan and S. Shende, *Static Frequency Assignment in Cellular Networks*, Algorithmica **29**(3):396-410, 2001.

[8] L. Narayanan and S. Shende, *Corrigendum to Static Frequency Assignment in Cellular Networks*, Algorithmica **32**(4):697, 2002.

[9] I. Sau, P. Šparl and J. Žerovnik, *Simpler Multicoloring of Triangle-free Hexagonal Graphs,* Accepted in Discrete Mathematics.

[10] D. H. Smith, S. Hurley and S. M. Allen, *A new lower bound for the channel assignment problem*, IEEE Transactions on Vehicular Technology **49**(4):1265-1272, 2000.

[11] K. S. Sudeep and S. Vishwanathan, *A technique for multicoloring triangle-free hexagonal graphs*, Discrete Mathematics **300**(1-3):256-259, 2005.

[12] P. Šparl, S. Ubeda, and J. Žerovnik, *Upper bounds for the span of frequency plans in cellular networks*, International Journal of Applied Mathematics **9**(2):115-139, 2002.

[13] P. Šparl and J. Žerovnik, 2-local *5/4-competitive algorithm for multicoloring triangle-free hexagonal graphs*, Information Processing Letters **90**(5):239-246, 2004.

[14] P. Šparl and J. Žerovnik, *2-Local 4/3-Competitive Algorithm For Multicoloring Hexagonal Graphs*, Journal of Algorithms **55**(1):29-41, 2005.

[15] P. Šparl and J. Žerovnik, *2-local 7/6-competitive algorithm for multicoloring a sub-class of hexagonal graphs*, International Journal of Computer Mathematics **87**(9):2003-2013, 2010.

[16] P. Šparl, R. Witkowski and J. Žerovnik, *A 1-local 7/5-competitive Algorithm for Multicoloring Hexagonal Graphs*, Submitted for publication in Algoritmica.

[17] P. Šparl, R. Witkowski and J. Žerovnik, *A Linear Time Algorithm for 7-[3]coloring triangle-free Hexagonal Graphs*, Submitted for publication in Inf. Proc. Letters.

[18] S.Ubeda and J.Žerovnik, *Upper bounds for the span of triangular lattice graphs: application to frequency planing for cellular networks*, Research report No. 97-28, ENS Lyon, September 1997.

[19] R. Witkowski, *A 1-local 17/12-competitive Algorithm for Multicoloring Hexagonal Graphs*, Proc. of Fundamentals of Computation Theory, volume 5699 of LNCS, pp. 346-356 (2009).

[20] R.Witkowski, *1-local 4/3-competitive Algorithm for Multicoloring a subclass of Hexagonal Graphs,* Submitted for publication in Discrete Applied Mathematics.

A generalization of the ball packing problem

Janez Žerovnik
FME, University of Ljubljana,
Aškerčeva 6,
SI-1000 Ljubljana, Slovenia
and
Institute of Mathematics, Physics and Mechanics,
Ljubljana, Slovenia
janez.zerovnik@fs.uni-lj.si, janez.zerovnik@imfm.si

May 28, 2011

1 Introduction

The packing problems, and in particular ball (or sphere) packing problems, have applications in many areas of science and technology, for example in Nuclear technics, Chemistry, Physics, Computer science, Telecommunications, etc. (see [26], and many later papers, ommited due to space limitation) Different versions of the basic problem and different aspects are studied in available scientific literature. They are very interesting NP-hard combinatorial optimization problems; that is, no procedure is able to exactly solve them in deterministic polynomial time. Packing problems are encountered in a variety of real-world applications including production and packing for the textile, apparel, naval, automobile, aerospace, and food industries. They are bottleneck problems in computer aided design where design plans are to be generated for industrial plants, electronic modules, nuclear and thermal plants, etc. Packing problems consist of packing a set of geometric objects/items of fixed dimensions and shape into a region of predetermined shape while accounting for the design and technological considerations of the problem. The packing identifies the arrangement and positions of the geometric objects that determine the dimensions of the containing shape and reach the extremum of a specific objective function. However, the search for exact local extrema is time consuming without any guarantee of a sufficiently good convergence to optimum.

The literature on the ball packing and applications is enormous, the research is extensive, and there is a great variety of specific aspects that are elaborated. Later we will mention some work on the original ball packing problem related to the well-known Kepler's conjecture, whre the maximal density packing of uniform balls is considered. On the other hand, there is not so much known

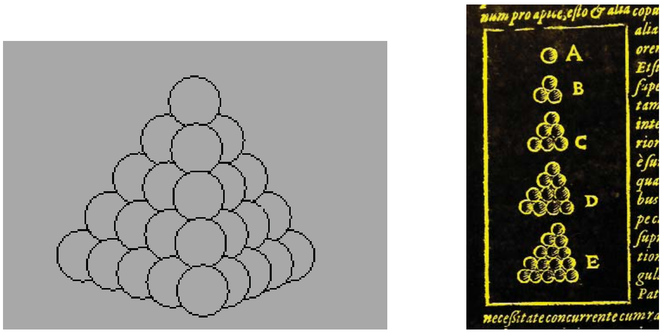


Figure 1: One of the two optimal arrangements and a diagram from Johannes Kepler's 1611 *Strena Seu de Nive Sexangula*.

about packing of balls with different sizes. Based on the searches through some available literature databases, it is clear that there is a number of simulation tools developed, and among them some recently developed simulation tools allow different shapes and different sizes of objects [8, 15, 16] .

In certain applications, it is important to achieve near maximal density with balls of different sizes, where in addition, size of the balls introduces a cost that may be a cost of production or depend on the availability. This seems to be a new optimization problem because we did not find any similar approach in literature search. We also provide some preliminary remarks on the maximal density bounds under some natural assumptions on the cost of balls.

2 History of the ball packing problem

The Kepler conjecture, named after Johannes Kepler, is a mathematical conjecture about sphere packing in three-dimensional Euclidean space. It says that no arrangement of equally sized spheres filling space has greater average density than that of the cubic close packing (face-centered cubic) and hexagonal close packing arrangements. The density of these arrangements is slightly greater than 74%. It may be interesting to note that the solution of Kepler's conjecture is included as a part of 18th problem in the famous list of Hilbert's problem list back in 1900 [29].

Recently Thomas Hales, following an approach suggested by Fejes Toth, published a proof of the Kepler conjecture. Hales' proof is a proof by exhaustion involving checking of many individual cases using complex computer calculations. For more details, see [11, 12, 7].

The basic problem of packing balls of equal size into infinite 3D space can be naturally generalized in many ways. Obviously, it is interesting to study maximal possible density when the region in 3D is bounded. In this case, given the shape (and size) of the region one asks how many balls of given size can be packed into the region. The second generalization is to allow balls of different sizes, where again there are many possibilities: there can be two, three or more ball sizes given, but one can also assume that the ball sizes are from an interval (approximating the realistic situation), or, furthermore, that the ball sizes are taken randomly according to some probability distribution.

This type of problems can be generaly regarded as optimization problems, and in most cases, they are studied as discrete optimization problems in the literature. As the problems are usually NP-hard, the optimal solutions are very difficult to find. In other words, there is no polynomial time algorithm that would guarantee optimality of the solution. Therefore many classical and modern metaheuristcs are applied for particular versions of the problem. The meta-heuristcs for integer and nonconvex programming include taboo search, simulated annealing, etc. [1]. The hardness of the problem is well illustrated by the fact that even two dimensional problems of packing circles are NP-hard, when restricted to circles of equal sizes and to simple rectangular or circular regions, see for example recent survey paper [13]. As usual, the best results are reported when general meta-heuristics are accompanied with clever hints that are based on properties of particular type of problem or instance (compare with [31]).

3 New optimization problem

From the point of view of Discrete optimization, or more generally, in view of Operational research, a possible approach to the problem is as follows.

A general problem in the very spirit of discrete optimization that may be of interest for reasons mentioned before is to allow different ball (or, circle) sizes, but at the same time control the distribution of sizes by introduction of cost, so that each circle (ball) has a cost $c(r)$ depending on the radius r. The motivation is that very small circles (balls) are expensive to produce and/or handle.

Of course, we also have the primary optimization objective, the maximization of density.

This naturally leads to the variety of optimization goals that may either be expressed as multicriterial optimization, or a single cost optimization where the cost function is a combination of the two criteria.

The general problem should apply to arbitrary region (volume), however already simple regions such as rectangle, circle or elipsis in 2D, and cube, polyhedra, ball, or ellipsoid in 3D may be both hard and interesting to consider.

3.1 FORMAL PROBLEM DEFINITION

General problem:

INPUT: An arbitrary region R in 3D, and a treasury of balls with cost $c(r)$ of a ball of radius r.

TASK: Pack the balls in R so that the volume of packed balls is maximised at minimal cost.

The general problem has to be defined more precisely, in particular one should define in more detail the following:

- regions,
- treasury of balls,
- objective function.

Objective function. There are obviously two conflicting goals in the task. This can be defined more precisely by introducing a goal function that would combine the cost of material (i.e. the balls) and the value (quality) of achieved density. Hence standard methods apply:

- multicriterial optimization (search for (un)dominated solutions)
- priorities: first objective, second objective
- design a single objective function AIM = VALUE - COST

Note that when single criteria optimization is considered, one also needs to evaluate the profit of densities in the same units as the ball cost.

Alternatively, one can formulate optimization tasks as follows:

- given target density X%, find minimal cost solution.
- given X units of money, find a mixture of balls that gives maximal density.
- provide maximal density with cost at most XX
- etc.

3.2 Theoretical bounds

When evaluating the quality of results (i.e. density) of the solutions provided by heuristics, it is very useful to have good bounds for the optimal solutions.

For example, recall that Hales' proof of Kepler's conjecture assures that the optimal packing density of equal sized balls is 74%, provided the volume is unbounded. The above densitiy applies to unbounded volumes, and are therefore may be good estimates (upper bounds) in case when the regions are "regular" in the sense that the surface (boundary) is not too complex and the ball raduis r is small in comparison to the diameter of the region. One indeed has to be very cautious

Clearly, Kepler's maximal density can be improved if balls of more than one size are allowed. Simple reasoning implies that densities arbitrary close to 100% can be obtained with a long enough sequence of allowed ball sizes. When restricted to only two balls sizes, a straightforward upper bound is $0.74 + 0.26 * 0.74 \approx 93\%$. Straightforward here means that it is intuitively "clear", however it is not at all clear that it can be easily proved formally. It seems natural that the cost of different balls will most likely enforce smallest possible difference between two ball sizes. This leads to a question, which seems to be tractable: find the densities with two ball sizes at fixed ratio between radii. Then, depending on the cost of balls, it may be easy to find the optimal solutions, in some cases. We elaborate this question in more detail elsewhere.

References

[1] Emile Aarts and Jan Karel Lenstra, Local Search in Combinatorial Optimization, Princeton University Press, Princeton 2003.

[2] Azeddine Benabbou, Houman Borouchaki, Patrick Laug, and Jian Lu, Sphere Packing and Applications to Granular Structure Modeling pages 1-18 Proceedings of the 17th International Meshing Roundtable, Garimella, Rao V. (Ed.), 2008.

[3] A. Benabbou, H. Borouchaki, P. Laug, and J. Lu, Geometrical modeling of granular structures in two and three dimensions. Application to nanostructures, International Journal for Numerical Methods in Engineering 80 (2009) 425454.

[4] J. Brest, J. Žerovnik, A heuristic for the asymmetric traveling salesman problem. In: The 6th Metaheuristics International Conference, August 22-26, 2005, Viena, Austria. MIC 2005. Wien: Universitaett Wien, [2005], 145-150. (See also: An approximation algorithm for the asymmetric traveling salesman problem, Ricerca Operativa 28 (1999) 59-67.)

[5] H.J. Brouwers, Particle-size distributions and packing fraction of geometric random packing, Physical Review E, Statistical, Nonlinear and Soft Matter Physics 74 (2006) 031309 (14 pages).

[6] Aleksandar Donev, Ibrahim Cisse, David Sachs, Evan A. Variano, Frank H. Stillinger, Robert Connelly, Salvatore Torquato and P. M. Chaikin, Improving the Density of Jammed Disordered Packings using Ellipsoids, Science 303 (2004) 990-993.

[7] Alireza Entezari, Towards Computing on Non-Cartesian Lattices, IEEE Transactions on visualization and computer graphics (submitted).

[8] M. Gan, N. Gopinathan, X. Jia and R.A. Williams Predicting Packing Characteristics of Particles of Arbitrary Shapes, KONA 22 (2004) 82-93.

[9] T. Gensane, Dense Packings of Equal Spheres in a Cube, Electronic J. Combinatorics 11, No. 1, R33, 2004.

[10] T. C. Hales, P. Sarnak, and M. C. Pugh, Advances in random matrix theory, zeta functions, and sphere packing PNAS u November 21, 2000 u vol. 97 u no. 24 u 1296312964 The Proceedings of the National Academy of Sciences USA (PNAS)

[11] T. Hales, Cannonballs and honeycombs, Notices of the American Mathematical Society 47 (2000) 440449.

[12] T. Hales, A proof of the Kepler conjecture, Annals of Mathematics. Second Series 162 (2005) 10651185.

[13] Mhand Hifi and Rym M'Hallah, A Literature Review on Circle and Sphere Packing Problems: Models and Methodologies, Advances in Operations Research, Volume 2009, Article ID 150624, 22 pages.

[14] H.M. Jaeger and S. R. Nagel, Physics of Granular States, Science 255 (1992) 1523-1531.

[15] X. Jia, M. Gan, R.A. Williams and D. Rhodes, Validation of a digital packing algorithm in predicting powder packing densities Powder Technology 174 (2007) 10-13.

[16] X. Jia and R. A. Williams, A packing algorithm for particles of arbitrary shapes, Powder Technology 120 (2001) 175-186.

[17] Jong Cheol Kim, David M. Martin and Chang Sung Lim, Effect of rearrangement on simulated particle packing, Powder Technology 126 (2002) 211-216 .

[18] B. Korte, J. Vygen, Combinatorial Optimization: Theory and Algorithms, Algorithms and Combinatorics 21, Springer, Berlin 2000.

[19] E.J.Maginn, Molecular simulation of ionic liquids: currecnt status and futute opportunities, J. Phys.: Condens. Matter 21 (2009) 373101 (17pp).

[20] Weining Man, Aleksandar Donev, Frank H. Stillinger, Matthew T. Sullivan, William B. Russel, David Heeger, Souhil Inati, Salvatore Torquato and P. M. Chaikin, Experiments on Random Packings of Ellipsoids, Physical Review Letters 94 (2005) 198001 (4 pages).

[21] V. V. Pavlov, Distinctive features of a crystal, rystal-like properties of a liquid and atomic quantum effects 13th International Conference on Liquid and Amorphous Metals IOP Publishing Journal of Physics: Conference Series 98 (2008) 052008

[22] Igor Pesek, Andrea Schaerf, Janez Žerovnik, Hybrid local search techniques for the resource-constrained project scheduling problem. Lect. notes comput. sci., vol. 4771, 57-68.

[23] Igor Pesek, Iztok Saje, Janez Žerovnik, Frequency assignment - case study. Part I, Problem definition. Frequency assignment - case study. Part II, Computational results, Proceedings of the 9th International Symposium on Operational Research in Slovenia, Nova Gorica, 2007, pp. 63-68 and 69-74.

[24] Nadge Reboul, Eric Vincens, and Bernard Cambou, A statistical analysis of void size distribution in a simulated narrowly graded packing of spheres, Granular Matter (2008) 10:457468.

[25] Roland Roth, Fundamental measure theory for hard-sphere mixtures: a review, J. Phys.: Condens. Matter 22 (2010) 063102 (18pp.).

[26] G.D.Scott and D.M. Kilgour, The density of random close packing of spheres, Brit. J. Appl. Phys. (J. Phsy. D) 2 (1969) 863866.

[27] Yu Shi and Yuwen Zhang, Simulation of random packing of spherical particles with different size distributions Appl Phys A (2008) 92: 621626.

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

[29] Benjamin H. Yandell, The Honors Class. Hilbertov Problems and Their Solvers, A K Peters, 2002.

[30] J. Q. Xu, R. P. Zou, and A. B. Yu, Analysis of the packing structure of wet spheres by Voronoi-Delaunay tessellation, Granular Matter (2007) 9:455463.

[31] J. Žerovnik, Simulated annealing type metaheuristics - to cool or not to cool, SOR '03 proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research, 2003, (6 pp.).

[32] Frederick S Hillier, Gerald J Lieberman, Introduction to Operations Research, McGraw-Hill New York 2005.

BALL PACKING WITH TWO DIFFERENT BALL SIZES: SOME BOUNDS BASED ON GEOMETRY

Janez Žerovnik
FME, University of Ljubljana,
Aškerčeva 6, SI-1000 Ljubljana, Slovenia
and
IMFM, Jadranska 19, SI-1000 Ljubljana, Slovenia
janez.zerovnik@fs.uni-lj.si, janez.zerovnik@imfm.si

Bojan Kuzma
FAMNIT, University of Primorska
Glagoljaška 8, SI-6000 Koper, Slovenia
and
IMFM, Jadranska 19, SI-1000 Ljubljana, Slovenia
bojan.kuzma@upr.si

Abstract: Densities of ball packing with two different sizes of balls is studied. For several ratios between the radii estimates for densities on unbounded regions are given.

Keywords: combinatorial optimization, ball packing, sphere packing, Kepler conjecture.

1 Introduction

The packing problems, and in particular ball (or sphere) packing problems, have applications in many areas of science and technology. In another short paper (this proceedings) we have introduced an optimization problem in which the goals are both maximizing the density and minimizing the cost that depends on the size of the balls. To the best of our knowledge, the optimization problem in this form has not been studied before. Besides experimental studies of particular versions of the problem (to be worked out elsewhere), it may also be important to study lower and upper bounds for the optimal solutions. The general problem is namely NP-hard and the same is expected for most of the specific problems.

The seemingly simple basic problem of packing balls of uniform size in unbounded 3D space has been a challenge for mathematicians for several centuries. The Kepler's conjecture, says that no arrangement of equally sized spheres filling space has greater average density than that of the cubic close packing (face-centered cubic) and hexagonal close packing arrangements. The density of these arrangements is slightly greater than 74%. It may be interesting to note that the solution of Kepler's conjecture is included as a part of the 18th problem in the famous list of Hilbert's problem list back in 1900 [4]. A proof, widely believed to be rigorous, appeared only recently [1]. A minor dispute over validity stems from the fact that the bulk of the proof consists of computer calculations and has spurred a new project Flyspeck to formally verify the calculations [2].

In this short note we report on some preliminary observations on densities that can be achieved with balls of two different sizes. The region is unbounded, and as we have no boundary, only the ratio between the radii of large and small balls is important. For several ratios we provide constructions and compute the densities. These densities may be seen as lower bounds for the maximal density that can be obtained with the ratio.

2 Basic geometry

First we give some constants (ratios) that are computed from the optimal packing of spheres of uniform size (see for example [3, 5]).

Definitions:

- R, size of the basic balls
- v, altitute of the equilateral triangle (three kissing balls centered at the vertices)
- h, distance between two layers in the optimal arrangement, height of a tetrahedron (four kissing balls centered at vertices)
- r_F, maximal radius of a ball that can be moved between three kissing balls
- d_T, distance from the centre of gravity of tetrahedron to a face
- d_O, distance from the centre of gravity of octahedron to a face
- r_T, maximum radius of a ball in tetrahedronal void
- r_O, maximum radius of a ball in octahedronal void

Formulas:

- $v = \sqrt{3}R \simeq 1.732R$
- $h = \frac{2\sqrt{6}}{3}R \simeq 1.633R$
- $r_F = (\frac{2\sqrt{3}}{3} - 1)R \simeq 0.155R$
- $d_T = \frac{\sqrt{6}}{6}R \simeq 0.408R$
- $d_O = \frac{\sqrt{6}}{3}R \simeq 0.816R$
- $r_T = (\frac{\sqrt{6}}{2} - 1)R \simeq 0.225R$
- $r_O = (\sqrt{2} - 1)R \simeq 0.414R$

In the optimal packing (both hexagonal close packing and cubic close packing) the number of octahedral voids is equal to the number of balls and the number of tetrahedral voids is twice this number.

3 Maximal density with two ball sizes

Based on the above ratios, we estimate the densities of the packings that are obtained by first packing the big balls in one of the optimal ways, and then filling the voids with the balls of smaller radius. First we consider the radii r_T, r_O, and $r_F = 0.154R$. (For simplicity, we will set $R = 1$, so for example $r_F = 0.154R = 0.154$.) The last radius, r_F, is small enough to fit into the narrowness between three kissing balls, and hence the balls of this size or smaller might flow through already rigid structure composed of the large balls. This may be an interesting feature when simulating random arrangements.

First we count how many smaller balls can be put into the voids between larger balls. Recall that in general the problem of positioning the optimal number of smaller balls into a bounded region is a difficult optimization problem. Let k_T denote the number of smaller balls that fit into the tetrahedral void, and let k_O denote the number of smaller balls that fit into octahedral void. The exact numbers of k_T and k_O are very difficult to find, near optimal constructions improving the bounds may be found by heuristic search. The numbers in the Table 1 for the second and the third radius are a rough estimate obtained by a simple algorithm. Design of a more sophisticated algorithm is in progress. We wish to emphasize that while the optimization problems that arise in the cases from Table 1 are hard, there is some hope to obtain good constructions by hand because the number of balls in the optimal solutions are rather small. On the other hand, when smaller radii (or better, larger ratios between radii) is considered, the only realistic approach is to design and run suitable heuristic algorithms to obtain near optimal solutions.

Table 1: Fitting the smaller balls into the tetrahedral and octahedral voids

smaller ball radius r	tetrahedral void k_T	octahedral void k_O
r_O	0	1
r_T	1	4
r_F =0.154	3	20

The figures from Table 1, more precisely the constructions behind, give rise to infinite arrangements of balls with higher densities than 74%. They may be understood as obvious lower bounds on densities with given ratio between sizes of large and small balls. Note that all the above constructions are regular in the sense that we find certain solutions how to fit a number of balls in both types of voids, sometimes assuming that each face of the void meets a ball but there is always at least half of the ball inside the void. Furthermore, we found certain arrangements and have no proofs (yet) that they are best possible, even under the assumptions mentioned.

We continue by a rough calculation of the densities obtained. These numbers are, as argued above, lower bounds on densities for each fixed ratio of ball sizes, but at the same time any of them gives a lower bound on the maximal density that may be achieved when two ball sizes are allowed.

Finally, we mention an obvious estimate of the maximal density that can be achieved with two ball sizes. Provided that the smaller balls are much smaller that the large balls, the voids are much larger than the smaller balls, and consequently one may believe that the density will be close to the optimal density on infinite region, hence the estimate

$$0.74 + 0.26 * 0.74 = 0.9324$$

However we have to be careful, because it is known that for certain region boundaries, it is possible to have higher density than 74%, so we can not claim that the estimate is indeed a proven upper bound on the maximal possible density on infinite region with two ball sizes.

In order to compute the lower bounds from our constructions we first compute the proportions of the number of larger and smaller balls. The number of small balls is computed as follows: roughly speaking, for each large ball we have two tetrahedral and one octahedral void (provided the region is infinite or at least large enough). Therefore for each large ball, we can use

$$n_2 = 2 * k_T + 1 * k_O$$

smaller balls, assuming $n_1 = 1$.

The densities are computed as follows: if the volume of a ball with $R = 1$ is taken as unit then a smaller ball has volume of r^3 units. The densities in the table are thus computed by the formula $0.74 + n_2 * r^3/0.74$.

Table 2: Proportions of number of balls in the constructions and respective densities

smaller ball radius r	large balls n_1	small balls n_2	density
r_O	1	1	0,8357
r_T	1	6	0,8322
0.154	1	26	0,8681
$r << 1$	1	$>> 1$	$\simeq 0,9324$

4 Concluding remarks

Constructions for ball packing of balls in unbounded regions with two allowed ball sizes were considered. It is clear that the densities higly depend on the ratio between ball sizes. However, even for each fixed ratio we have in general a difficult mathematical problem that in most cases does not allow analytical solution and it is therefore natural to consider the question as an optimization problem. These optmization problems are expected or known to be NP-hard, so it may be interesting to use heuristic methods to compute near optimal solutions.

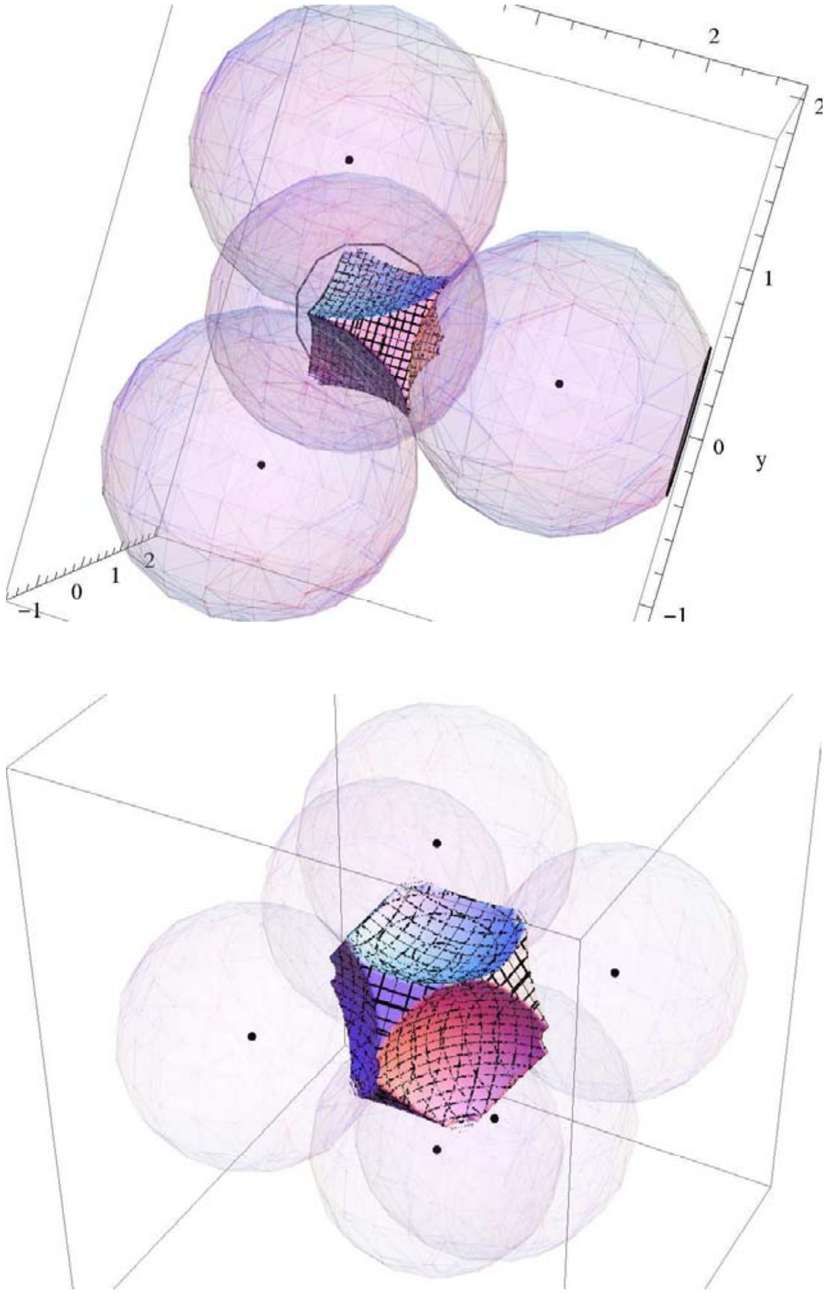


Figure 1: Tetrahedral and octahedral void with attached four, respectively, six, touching balls. Black dots are their centra.

On the other hand, although the preliminary results given here may lead to conjecture that for relative small proportions of radii the density need not uniformly increase, it is much less doubt that the density will uniformly increase when the radii proportion will increase to infinity, i.e. when the smaller radius will be very small.

Last but not least, the prelimiary analysis given here provides some ideas on the lower bounds on densities with certain assumptions mentioned in the text. Perhaps the most important is that we have assumed that in the optimal constructions the large balls are arranged according to one of the optimal arrangements and the smaller balls only cover the voids of this arrangements. If this assumption is dropped, there are many more possible arrangements. However, our lower bounds and the estimated upper bound remain valid.

References

[1] T. C. Hales, A proof of the Kepler conjecture. Ann. of Math. (2) 162 (2005), no. 3.

[2] T. C. Hales, J. Harrison, S. McLaughlin, T. Nipkow, S. Obua, R. Zumkeller, A revision of the proof of the Kepler conjecture. Discrete Comput. Geom. 44 (2010), no. 1, 1-34.

[3] P. Krishna and D. Padney, Close-Packed Structures, University college Cardiff, 1981.

[4] Benjamin H. Yandell, The Honors Class. Hilbert's Problems and Their Solvers, A K Peters, 2002.

[5] http://en.wikipedia.org/wiki/Tetrahedron

Section II:
Graphs and Their Applications

HOMOGENIZATION OF DIGITAL CADASTRE INDEX MAP IMPROVING GEOMETRICAL QUALITY

Marjan Čeh*, Frank Gielsdorf** and **Anka Lisec*****
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia, e-mail: mceh@fgg.uni-lj.si
**Technet GmbH, Berlin, Germany, e-mail: frank.gielsdorf@technet-gmbh.com
*** University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia, e-mail: anka.lisec@fgg.uni-lj.si

Abstract: In Slovenia, as well as in the comparable traditional cadastral societies, a digital land cadastre index map is a composition of digitized various analogue land cadastre maps and measurement data sets with different positional accuracy, depending on the scale of the map and on the methodology (quality) of data acquisition. The growing role of spatial data quality for GIS applications as well as the needs of advanced land administration systems, calls for improvement of the geometrical quality of land cadastre index maps, which is being associated also with the problem of heterogeneity of graphical data. For this purpose, the homogenisation process of digital land cadastre index maps is needed. In the article, we present the approach to the improvement of land cadastre index map geometrical quality, based on surveying measurements, where the basic principles of geodetic profession are to be respected (coordinate geometry, topology, adjustments, error propagation law etc.).

Keywords: cadastre map, homogenization, membrane method, proximity fitting.

1 INTRODUCTION

Efficiency and transparency of land administration systems are one of crucial importance in the society since land has always remained at the foundation of human life. Here, land administration system can be understood as conceptualization of rights, restrictions, and responsibilities related to people, policies and land (places), and it has been closely linked with the land evidences, land (cadastral) maps since ancient times (see [2] and [8]). Problems concerning effective land administration supporting suitable use of land resources are becoming more and more important all over the world. In recent time, there has been a revival of interest in the role and operation of land administration systems, which have been strongly influenced by development of information technology, in particular by development of geographical information technology. The core sub-system of the advanced land administration system has become a high quality digital data on land and real property organized in the multipurpose geographical (land) information systems GIS (LIS).

Focusing on the "parcel based cadastral societies", a common characteristic is the heterogeneous graphical subsystem of digital land cadastre data. Coordinates of land parcel border vertexes, are mainly derived from positional heterogeneous analogue maps that were digitized and geo-referenced. The analogue maps, which in Slovenia originate from the beginning of the 19th century, had been maintained throughout the centuries with manual technique of graphical adjustment. After digitalization of the heterogeneous analogue cadastral maps, integration of hundreds of analogue map sheets and mapping units (cadastral communities) has been executed that caused additional errors that were adjusted manually. Therefore improvement of geometrical accuracy of land cadastre index map has become a challenging issue in Slovenia. In addition, the integration of contemporary measurement data due to the permanent maintenance of the land cadastre index map is an important issue [1].

The resulting geometrical quality of digital land cadastral index map reflects the quality of underlying analogue cadastral maps and its integration, The problem of geometrical improvement of graphical subsystem of digital land cadastre is common for the traditional cadastral societies – not only Slovenia, but also Austria, Germany etc., and has been proven as serious problem in the field of spatial data analysis and spatial decisions since land cadastre index map has became widely used in the framework of different land administration systems (agriculture, spatial planning, utilities, environmental) through overlay in geographical information systems (GIS).

2 RESEARCH PROBLEM

The problem of measurement based spatial data is a wide problem in the field of GIS and spatial data infrastructure [9]. In general, GIS is designed to consider geometrical parameters (coordinates) as deterministic value. This is often misunderstanding of the nature of spatial data in GIS; coordinates are namely always calculated from the observations, which are redundant aleatory values. Consequently, coordinates are aleatory values as well as correlated. The accuracy of relative geometry is higher than the absolute accuracy – coordinates and relative measures are therefore not equivalent. As point positions remain unchanged in reality, new determined (measured) coordinates lead to a virtual displacement of the related point in GIS. Without any consideration of neighbourhood relationships to other points at all, the relative geometry between updated and unchanged points in GIS would be highly violated – such an integration of new coordinates are not to be accepted. Here, a virtual displacement of points the different accuracy of relative and absolute geometry has to be considered [4].

Spatial data integration has therefore become a widely used term, covering a variety of processes. In this paper it is used to describe methods that attempt to improve the geometrical accuracy of land cadastre vector dataset (index map) based on integration with more accurate measurement data at identical points. Land cadastre positional data are predominantly based on observations for direct mapping but the original measurements data are most often not stored. In some countries surveys for cadastral datasets, for example in the area of former Habsburg monarchy, started around 200 years ago. The quality of measurement equipment as well as methodology has improved dramatically since than. Therefore, the problem of integrating new measurements in the existing datasets appears. A transformation is necessary to fit the best results of the new measurements into the old dataset. Adding new measurements to the old datasets, by storing point coordinates only, does not improve the quality of the dataset.

Surveying measurements result in more accurate point coordinates providing the basis for improvement of spatial data – and its homogenisation. So called data conflation (spatial data integration) can involve shifting one dataset (land cadastre vector index map) to align with other target datasets (measured identical points). Another approach is Positional Accuracy Improvement (PAI), which is the main topic of our paper. Localized pockets of higher accuracy data (contemporary measurements of higher positional accuracy) are used to improve the positional accuracy of surrounding (neighbourhood) low level positional accuracy datasets, for example land cadastre vector index map. Rather than considering this higher accuracy data to be a target dataset, methods developed for PAI determine the best fit positioning solutions using positional information of both dataset; these methods are based on surveying adjustment theory and are also able to preserve geometric properties such as straight or parallel lines (see also [4][5]).

3 APPROACH (PROBLEM SOLVING)

The geometrical quality of the cadastral index map can be improved by integration of precise geodetic measurements, existing field book measurements as well as other sources of data (photogrammetric measurements etc.). The new measurements data might be the result of the sporadic cadastral data maintenance activities or result of systematic projects of mass measurements. The integration of new, more accurate spatial data into existing data sets should result in more accurate point coordinates (location) and should improve and homogenise geometrical accuracy of neighbouring cadastral (spatial) data. Here it should be stressed, that measured coordinates are random variables and they are stochastically dependent. If the coordinates of land plot boundary in GIS would be stochastically independent we could just exchange the less accurate coordinates by more accurate ones. But such an approach would neglect geometrical neighbourhood relationships. Here, wider areas (not only the measured land parcel) should gain from this accuracy improvement without loosing its internal geometrical quality. The current system for maintaining of land cadastre index map in Slovenia does not support such strategies, and update transformations lead to inconsistency because proximity fitting principle is avoided.

To keep neighbourhood relationships, proximity fitting methods should be applied. Spatial heterogeneity implies each location has intrinsic uniqueness, conditions vary from place to place. The lengths of the edges in network may be equal, but they have different relative interpretations within the clusters based on absolute distances. It is clear that relative proximity is more important than absolute proximity in geo-referenced settings and thus geospatial clustering [7].

Proximity fitting methods substitute a usual single system transformation. The result of a usual transformation can be seen as the first step of proximity fitting. At transformation phase an artificial coordinate differences between identical points, the connection points, and the new points are introduced into the adjustment. These ‘pseudo observations’ are weighted dependent on the distances. There are no direct neighbourhood relationships between the interpolated points. Additionally, the result depends on the number of identity points which create residuals. The method is not suitable to model direct neighbourhood relationships. The resulting displacements of the new points are dependent of the density and distribution of the identical points. In a second step the mapping approach is extended by the introduction of relative geometry information. Advanced methods use the Delaunay triangulation to model neighbourhood relationships directly. The resulting displacements are here independent of the density and distribution of identity points. Suggested approach is called ‘membrane method’. This method uses as functional model coordinate differences along the triangle sites, what leads to linear residual equations with a very stabile convergence behaviour. The stochastic model is derived from finite element methods, and it simulates the behaviour of a rubber membrane. The proximity fitting is run as an adjustment calculation [3][4].

3.1 Geometric data homogenization with adjustment techniques

The coordinates in GIS (cadastre index map) result from the evaluation of measured values. In a first step these measured values often were local coordinates of digitized analogue maps which were transformed into a global reference frame. These so determined global coordinates describe the geometry of the GIS objects unique whereby the coordinates have to be addressed as random variables. During the process of PAI new measured values with higher accuracy are introduced. The new measured values are redundant to the already existing coordinates. Therefore, the determination of new coordinates with improved positional accuracy is a typical adjustment problem. But measured values have two essential properties [4]:

- They are random variables. Because it is impossible to measure a value with arbitrary accuracy, which leads to the fact that any measured value contains some uncertainty.

- They are redundant. Commonly there exist more measured values then necessary to be able to calculate unique point coordinates.

A function of random variables results again in a random variable. Because of point coordinates are functions of measurement values they are like them random variables. The uncertainties contained in measured values lead necessarily to uncertainties in point coordinates. For the unique determination of a number of coordinates the exact same number of measured values is necessary.

3.1.1 Error Propagation for Linear Functions

The law of error propagation describes the propagation of accuracies for linear functions of random variables. Applying this law to an adjustment it is possible to calculate the standard deviations of the unknown parameters and those of the residual errors. This case can be explained with a simple example. Fig. 1 shows points in one dimensional coordinate system.

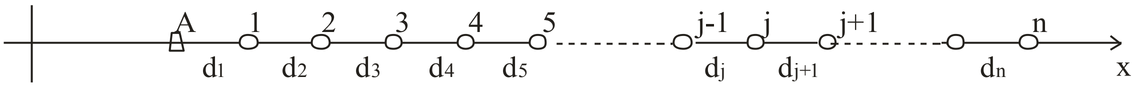


Figure 1: Points in a 1-dimensional coordinate system.

The coordinate of the control point A is known and fixed – A(x_A); based on measured distances d_i, we want to calculate the coordinate x_i of the new point *i:*

$$x_j = x_A + \sum_{i=1}^{j} d_i \tag{1}$$

According to the law of error propagation (see [3]), the standard deviations of x_i is:

$$\sigma_{x_i}^2 = \sum_{j=1}^{i} \sigma_{x_j}^2 \ . \tag{2}$$

3.1.2 Adjustment Considering the Correlations

We want to interpret the standard deviations of the distances d_i and those of the coordinate x_i. The standard deviation of the coordinate σ_{xi} represents the absolute accuracy of the coordinates in relation to the reference frame. On the other hand the standard deviations of the distances σ_{di} represent the relative accuracy of the coordinates related to each other. If two calculated random values are functions of partial the same random variable arguments they are stochastically dependent. The degree of their stochastic dependency is quantified by their covariance.

For one dimensional coordinate system (Fig. 1), the parameters x_{j-1} and x_j are stochastically dependent because they are functions of partial the same random variables. The distances $d_1 ... d_j$ are arguments of both functions (1). For this purpose generalisation of the law of error propagation is supposed. The general form of the law of error propagation can be represented in matrix notation. If there is a system of linear equations **F** describing the functional dependency of parameters x_j on the arguments d_i

$$X = F \cdot d \tag{3}$$

and the standard deviations of d_i are known then the variances and covariances of the parameters x_j can be calculated by:

$$C_{xx} = F \cdot C_{dd} \cdot F^T \quad (4)$$

In this formula the functional matrix **F** contains the coefficients of the linear functions (1). The matrix $\mathbf{C}_{dd}$ is called the covariance matrix of observations and contains the variances of observations on its principal diagonal and their covariances on its secondary diagonals. In the most common case of stochastically independent observations $\mathbf{C}_{dd}$ is a diagonal matrix. $\mathbf{C}_{xx}$ is the covariance matrix of the unknown parameters and contains their variances and covariances.

For the study case (Fig. 1) the $\mathbf{C}_{dd}$ matrix contains the variances of x_{j-1} and x_j as well as their covariance:

$$C_{x_{j-1}x_j} = \begin{bmatrix} \sigma^2_{x_{j-1}} & \mathrm{cov}(x_{j-1}, x_j) \\ \mathrm{cov}(x_{j-1}, x_j) & \sigma^2_{x_j} \end{bmatrix}. \quad (5)$$

If we solve the matrix equation for the general law of error propagation in a symbolic way then we get the expression:

$$\sigma^2_{x_{j-1}-x_j} = \sigma^2_{x_{j-1}} + \sigma^2_{x_j} - 2 \cdot \mathrm{cov}(x_{j-1}, x_j), \quad (6)$$

where

$$\mathrm{cov}(x_{j-1}, x_j) = \rho_{x_{j-1}x_j} \cdot \sigma_{\Delta x_{j-1}} \cdot \sigma_{\Delta x_j} = f(d_j). \quad (7)$$

This formula (6) which does not neglect the covariance between dependent random variables yields the right result (see [3]).

3.1.3 The positional accuracy improvement of cadastral index map

The problem of local geometrical distortions is illustrated with the following case. There is a GIS layer containing parcel boundaries of 4 boundary points. A surveyor may have determined coordinates (new high accurate measurements) of 4 boundary points. The positional standard deviations of the measurements are about 2 cm and substantially much more exact than the graphic coordinates in GIS layer. The introduction of the new coordinates would cause the following situation in the GIS (Fig. 2):

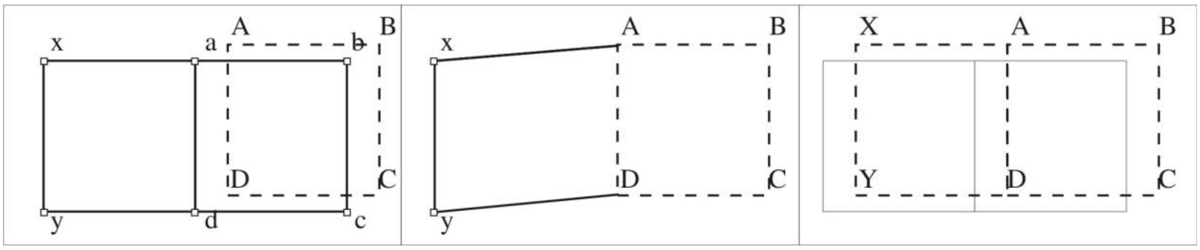


Figure 2: (Left) cadastral boundaries (solid lines) and boundaries determined by higher accuracy measurements (dashed lines). (Middle) polygon ***xady*** is distorted if points **abcd** are replaced by **ABCD**. (Right) consideration of neighbourhood.

In example the new determined points ***A, B, C*** and ***D*** are introduced in a map based cadastral data set. Neglecting the correlations leads to a massive distortion of the (left) parcel ***xady***. The simplest way to use the exact (high accurate) coordinates would be just to exchange them with the existing graphic coordinates. Unfortunately this simple action would cause the distortion of the adjacent parcel geometry! An obviously much better approach would keep the geometrical neighbourhood relations seen in the following picture (Fig. 2). The neighbourhood accuracy of two points which are descended from a digitized map is higher than their absolute positional accuracy. For this reason, the correlation is a function of the point distance. The smaller the point distance the greater the correlation.

The distance-dependent correlations of the coordinates have two essential reasons:

- The origin measurements were done with respect to the principle of neighborhood (e.g. tape distance measurements).
- The manual mapping was done according to the principle of neighbourhood as well.

Homogenization which models the distance-dependent correlations directly uses sophisticated adjustment algorithms for the calculation and analysis of coordinates. Firstly, topological neighbourhood information is determined. This is performed by a Delaunay triangulation over all GIS points (control points and new points) of the origin system. The triangle sides are used as carriers of neighbourhood information. All cadastral index map points are used for a Delaunay triangulation. Along the triangle sides, artificial coordinate difference observations are generated which are subsequently introduced in an adjustment calculation according to the least squares method. The observation values are derived from the coordinates in the origin system. The triangular net is acting like a homogeneous membrane; its elasticity is given by the weights following the map digitization accuracy. The remaining divergences of the control points are propagated thus on the new points (Fig. 4).

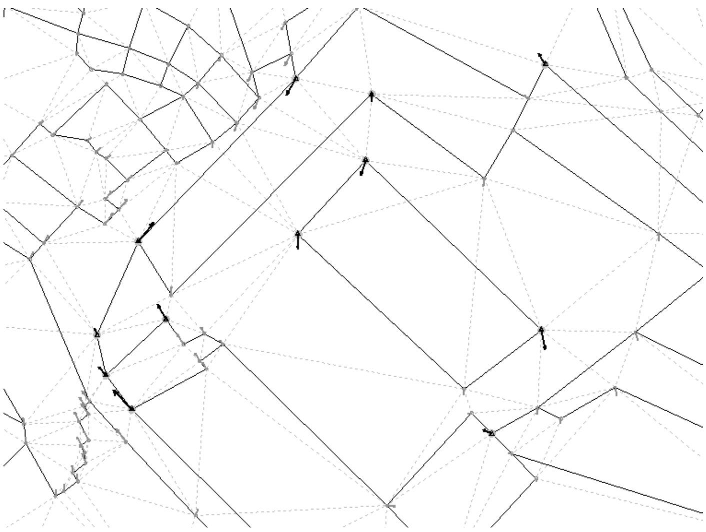


Figure 3: Homogenization: Residuals of the identical points (gray dots) are transmitted by triangle sides (dashed lines), the neighbourhood is considered.

4 CONCLUSIONS

The aim of this paper is to present method for homogenization of cadastre index maps. The method, suggested here, follows homogenization applying adjustment techniques, which uses simulated and real measurements. The method has been applied to analyse how heterogeneous cadastre map parts get fitted to the real positions by the use of point positions, relative measurements and geometrical constraints.

The significance of adjustment techniques for transformation problems is recognized long ago. To apply the least squares method following C.F. Gauß is the usual geodesic practice for two dimensional transformations with redundant identities from one Cartesian system into another. This classical adjustment method can be expanded to a simultaneous transformation of multiple systems, subsequently called ‘Interconnected Transformation’ which is not the subject of these paper.

If neighbourhood geometry is to be maintained through the inclusion of relative distances, as well as geometric and topological properties, there will be redundancy in the data integration process. It will not be possible to obtain a solution that perfectly meets every constraint. Instead, an adjustment problem is set up, for which an optimal solution can be determined using the method of weighted least squares. All of the available information is considered, and observations are weighted by their recorded accuracy values to determine the solution that best fits the datasets being integrated. Moreover, the method of least squares generates precision values for the calculated parameters, thereby enabling update of the positional accuracy of the upgraded dataset (see also [6]).

It is known that the best consideration of neighborhood relationships is warranted using proximity fitting adjustment methods where artificial observations between points are integrated. Advanced adjustment programs use for that task finite element methods based on triangles. Nevertheless, the real observations can completely be introduced in these proximity fitting adjustment processes.

References

[1] Čeh, M., Lisec, A., Ferlan, M., Šumrada, R., 2011. Geodetsko podprta prenova grafičnega dela zemljiškega katastra (The renovation of the land cadastre’s graphical part based on surveying principles). Geodetski vestnik 55(2): 257-268.

[2] Dale, P., Mclaughlin, J.D., 1999. Land Administration. Oxford etc., UK: Oxford University Press.

[3] Gielsdorf, F., 2010. Data Integration with Adjustment Techniques. Technet GmbH. (http://www.technet-gmbh.de/)

[4] Gielsdorf, F., Gruendig, L., Aschoff, B., 2004. Positional Accuracy Improvement – A Necessary Tool for Updating and Integrating of GIS Data. FIG Working week, May 22-27, 2004, Athens.

[5] Hope, S., Kealy, A., 2008. Using Topological Relationship to Inform a Data Integration Process. Transactions in GIS 12(2): 267-283.

[6] Hope, S., Gielsdorf, F., 2007. Quality Considerations for Optimal Positioning When Integrating Spatial Data. Spatial Sciences Institute International Biennial Conference 2007, Hobart.

[7] Ickjai, L., 2005. Geosaptial Clustering in data-rich environments: features and Issues. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.), Knowledge-Based Intelligent Information and engineering systems, 9th International Conference KES 2005 Proceedings, Part IV.

[8] Larsson, G., 1997. Land management – public policy, control and participation. Stockholm: Ljungflöts Offset.

[9] Navratil, G., Franz, M., Pontikakis, E., 2004, Measurement-Based GIS Revisited. 7th AGILE Conference on Geographic Information Science, April 29-May 1, 2004, Heraklion. (http://plone.itc.nl/agile_old/Conference/greece2004/papers/P-07_Navratil.pdf)

SOLVING JOB SHOP PROBLEMS IN THE CONTEXT OF HYPERGRAPHS

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

Abstract

A class of hypergraphs, called B-tails, is introduced. It is shown that a certain hypegraph can be assigned to any job shop problem in such a way that minimal B-tails correspond to optimal schedules.

Keywords: scheduling, job shop problem, hypergraph

1 Introduction

In coping with '*system nervousness*', i.e., the fact that a relatively small change of the environment can result in substantial change of the system [4], the use of *sensitivity analysis* (the determination of the bounds within which a given schedule remains optimal) has been proposed to resolve the problem, whether to reschedule or not [5].

The proposed approach is based on AND/OR graphs, and is applicable not only in 'pure' scheduling and rescheduling but also in *alternative process scheduling problems*, where the choice to be made is not only *when* to perform every activity, but also *which* of the available set of activities is to be used [1]. However, the drawback of this approach is that the AND/OR graphs produced in the process are quite large.

Here we show that scheduling can be discussed also in the environment of hypergraphs and that the corresponding hypergraph is substantially smaller.

2 Hypergraphs

An (oriented) *hypergraph* G is defined[1] 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 such a hypergraph $G_1 = (\mathcal{V}_1, \mathcal{A}_1)$ that $\mathcal{V}_1 \subset \mathcal{V}$ and $\mathcal{A}_1 \subset \mathcal{A}$. When convenient, we shall denote $\mathcal{V}_1 = \mathcal{V}(G_1)$ and $\mathcal{A}_1 = \mathcal{A}(G_1)$.

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 $u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \ldots, \mathcal{E}_{q-1}, u_q$, 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.*

If $u_1, \mathcal{E}_1, u_2, \mathcal{E}_2, \ldots, u_{q-1}, \mathcal{E}_{q-1}, u_q$ is a path, then $u_q, \mathcal{E}_{q-1}, u_{q-1}, \ldots u_2, \mathcal{E}_1, u_1$ will be called a *reversed path.*

3 B-tails

Let G be a hypergraph and $x \in \mathcal{V}(G)$.

Definition A hypergraph D is a *B-tail of* x if

1. for every $\mathcal{E} \in \mathcal{A}(D)$
   (a) in D there exists a path $\ldots, \mathcal{E}, \ldots, x$,
   (b) $|H(\mathcal{E}) \setminus F(D)| = 1$, unless $x \in H(\mathcal{E})$, when $|H(\mathcal{E}) \setminus F(D)| = 0$,
2. for every $t \in \mathcal{V}(D)$
   (a) if $t \notin B(G)$, then $|BS(t) \cap \mathcal{A}(D)| \geqslant 1$,
   (b) if $t \notin F(D)$, then $|BS(t) \cap \mathcal{A}(D)| \leqslant 1$,
   (c) $|BS(x) \cap \mathcal{A}(D)| \leqslant 1$,
3. there are no cycles in D.

[1]For details on hypergraphs, cf. [2].

Thus, to get a B-tail of a node x, one must select one of the hyperarcs from the back star of every node (starting at x), however, this must be done in such a way, that no cycles are completed and that for hyperarcs, which are not B-arcs, only one of the nodes from their head has successors in the B-tail.

As we shall see, B-tails are suitable for handling job shop problems, specifically, to any job shop a hypergraph can be assigned such that specific B-tails in it correspond to feasible schedules.

4 The job shop problem

The job shop problem consists of the following:

n jobs and m machines are given. A job is a set of operations, which are totally ordered by the precedence relation; each operation can be performed only on one of the machines. Thus, operations that take part in different jobs but must be performed on the same machine can not be performed simultaneously. The problem is to find a schedule of all operations so that the total duration time is minimal.

It is known that the job shop problem is $\mathcal{NP}$-complete [3]. Thus, it is at least as hard as any other problem that belongs to $\mathcal{NP}$ and therefore hard to solve.

With every job shop problem a *disjunctive graph* $G = (O, A, E)$ can be associated in the following way:

- the nodes (elements of O) represent operations;
- an arc $(o_i, o_j) \in A$ exists if and only if o_i precedes o_j (in one of the jobs);
- an edge $\{o_i, o_j\} \in E$ exists if and only if o_i and o_j must be processed on the same machine.

Thus, a disjunctive graph is partly directed and partly undirected.

Finally, two nodes — usually called source and sink — are added to G, together with an arc (source, n) for each $n \in O$ for which there are no incoming arcs in A and an arc (n,sink) for each $n \in O$ for which there are no outgoing arcs in A.

The graph is weighted by

$$c(o_i, o_j) = d(o_i),$$

where $(o_i, o_j) \in A$ and $d(o_i)$ is the duration of o_i. The cost of a path P is defined by[2]

$$c(P) = \sum_{(x,y)\in P} c(x, y)$$

[2] Alternatively one can weight the nodes:

$$w(o_i) = d(o_i).$$

An *orientation* Ω is an assignment $E \to O \times O$ such that

$$\Omega(\{o_i, o_j\}) \in \{(o_i, o_j), (o_j, o_i)\}$$

Thus, every orientation turns a disjunctive graph $G = (O, A, E)$ into a directed graph $(O, A \cup \Omega(E))$. If the latter is acyclic, it corresponds to a feasible schedule for the original job shop problem. Moreover, the one in which the maximal path (from source to sink) is minimal, corresponds to the optimal schedule, i.e., to the solution of the job shop problem.

Example 1 *Consider the following job shop problem: we have four machines* m_1*,* m_2*,* m_3 *and* m_4 *and two jobs* $J_1 = \{o_1, o_2, o_3\}$ *and* $J_2 = \{o_4, o_5, o_6\}$*. Operation* o_1 *must be performed on* m_1*, operations* o_2 *and* o_5 *on* m_2*, operations* o_3 *and* o_4 *on* m_3*, while* o_6 *must be performed on* m_4*; the processing times are* $d(o_1) = 5$*,* $d(o_2) = 4$*,* $d(o_3) = 2$*,* $d(o_4) = 6$*,* $d(o_5) = 3$ *and* $d(o_6) = 1$*.*

The corresponding disjunctive graph is in Fig. 2.

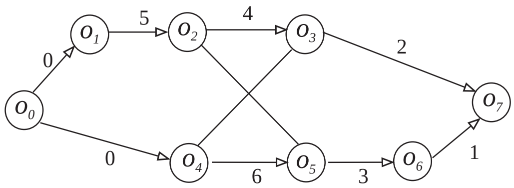


Figure 2

Orientation

$$\Omega_1 : \begin{cases} \{o_2, o_5\} \mapsto (o_2, o_5) \\ \{o_3, o_4\} \mapsto (o_3, o_4) \end{cases}$$

yields an acyclic directed graph, so that it corresponds to a feasible schedule o_1*,* o_2*,* o_3*,* o_4*,* o_5*,* o_6 *with makespan* 21*.*

We shall show that to any job shop problem a hypergraph can be assigned so that specific B-tails in it correspond to feasible schedules.

Starting with the disjunctive graph $G = (O, A, E)$ the corresponding hypergraph G_h is obtained in the following way:

- first form the hypergraph G' for which
  - $\mathcal{V}(G') = O$,

Of course, the cost of a path P is then

$$c(P) = \sum_{n \in P} w(n)$$

- for every $t \in \mathcal{V}(G')$ we have $BS(t) = \{(T_t, \{t\})\}$, where $T_t = \{u; (u,t) \in A\}$ (so that every node in G' has only one incoming hyperarc, which is a B-arc);

- then for every $t \in \mathcal{V}(G')$ such that t is incident to an edge in G, we
  - add to G' a new node $u_{t,t}$,
  - for every v such that $\{t, v\} \in E$ add to G' a new node $u_{t,v}$,
  - add to G' hyperarc $(\{t\}, \cup_{v \in E_t}\{u_{t,v}\})$, where $E_t = \{v; \{t,v\} \in E\}$;

- then for every $t \in \mathcal{V}(G')$ such that t is incident to an edge in G, we
  - for every $\mathcal{E} \in BS(t)$ do the following:
    * for every v such that $\{t, v\} \in E$ add to G' the hyperarc $(T(\mathcal{E}) \cup \{u_{v,t}\}, \{t\})$,
    * replace $\mathcal{E}$ with $(T(\mathcal{E}) \cup \{u_{t,t}\}, \{t\})$,

- for every edge-connected component C, add to G' the hyperarc $(\{\text{source}\}, \{u_{t,t}; t \in \mathcal{V}(C)\})$.

Clearly, the only nodes in G_h with more than one hyperarc in their back star are the nodes that are incident to an edge in G and the only hyperarcs which are not B-arcs are of the form $(\{t\}, \cup_{v \in E_t}\{u_{t,v}\})$ or $(\{\text{source}\}, \{u_{t,t}; t \in \mathcal{V}(C)\})$.

Example 2 *For instance, the hypergraph, corresponding to the disjunctive graph from Example 1, is depicted in Fig. 3.*

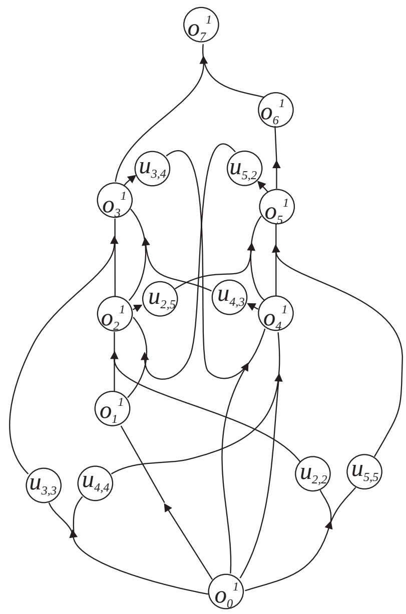


Figure 3

The following applies (due to space limitations we omit the proof):

Proposition 3 *B-tails of the node* sink *in hypergraph G_h correspond to feasible schedules for the job shop problem. In any particular schedule, operation o_i precedes operation o_j if and only if in the corresponding B-tail there is a path $o_i, \ldots, o_j$.*

The *cost of a B-tail Q* is defined as the highest of the costs of paths, completely contained in Q.

For the hypergraph G_h the cost of any path P is just the total duration of the operations lying on P. By Proposition 3 it is also the minimal time in which all these operations can be completed. Thus, if Q is a B-tail of sink, then its cost is equal to the makespan by the schedule, corresponding to Q.

Of course, the minimal B-tail of sink corresponds to the solution of the job shop problem.

An algorithm for searching for a minimal B-tail can be designed in the spirit of the A* algorithm so that an accurate heuristics can significantly improve effectiveness of the alghorithm (but this discussion must be omitted here).

5 Conclusion

We have demonstrated that hypergraphs are suitable for handling the job shop problem, specifically, that solving the job shop problem translates into searching for a minimal B-tail.

References

[1] Beck, J. C., Fox, M. S., 2000. Constraint-directed techniques for scheduling alternative activities. Artificial Intelligence 121, 211-250.

[2] C. Berge, C., 1989. Graphs and Hypergraphs. North-Holland,Amsterdam.

[3] Garey, M. R., 1979. Computers and Intractability. W. H. Freeman and Company, San Francisco.

[4] Grubbström, R. W., Tang, O., 2000. Modelling Rescheduling Activities in a Multi-Period Production-Inventory System. International Journal of Production Economics 68, No 2, 123-135.

[5] Hvalica, D., Bogataj, L.,2005. Sensitivity results considering rescheduling by AND/OR graphs. Int. j. prod. econ. [Print ed.] 93/94, 455-464.

An heuristic for the edge-survivable General Steiner Problem

Pablo Sartor Del Giudice and **Franco Robledo Amoza**
Instituto de Computación, Facultad de Ingeniería, Universidad de la República
Julio Herrera y Reissig 565 – Postal Code 11.300 – Montevideo, Uruguay
psartor@um.edu.uy; frobledo@fing.edu.uy

Abstract: The Generalized Steiner Problem with Edge-Connectivity constraints (GSP-EC) consists of computing the minimal cost subnetwork of a given feasible network where some pairs of nodes must satisfy edge-connectivity requirements and models the design of communications networks where connection lines can fail. In this paper we present an algorithm based on the GRASP metaheuristic to solve this version of the problem, known to be NP-Complete. Promising results are obtained when testing the algorithm over a set of heterogeneous network and optimal or near-optimal solutions are found.

Keywords: network design; optimization; graph theory; survivability; Steiner problem; GRASP.

1. INTRODUCTION

The design of communication networks often involves two antagonistic goals. One one hand the resulting design must bear the lowest possible cost; on the other hand, certain survability requirements must be met, i.e. the network must be capable to resist failures in some of its components. One way to do it is by specifying a connectivity level (a positive integer) and constraining the design process to only consider topologies that have at least that amount of disjoint paths (either edge or node disjoint) between each pair of nodes. In the most general case, the connectivity level can be fixed independently for each pair of nodes (heterogeneous connectivity requirements), some of them having even no requirement at all. This problem is known as Generalized Steiner Problem (GSP) [9] and is an NP-Complete problem [18]. Some references on the GSP and related problems are [1], [2], [3], [4], [6], [7], [10], [11] most of them using polyedral approaches and addressing particular cases (specific types of topology and or connectivity levels). Topologies verifying edge-disjoint path connectivity constraints ensure that the network can survive to failures in the connection lines; while node-disjoint path constraints ensure that the network can survive to failures both in switch sites as well as in connection lines. Finding a minimal cost subnetwork satisfying edge-connectivity requirements is modeled as a GSP edge-connected (GSP-EC) problem. Due to the intrinsic complexity of the problem, heuristic approaches have to be used to cope with general real-sized instances; this work presents one inspired in the ideas of recent work [13], [14], [16].

The remainder of this paper is organized as follows. Notation, auxiliary definitions and formal definition of the GSP-EC are introduced in Section 2. The GRASP metaheuristic and the particular implementation that we propose for the GSP-EC are presented in section 3. Experimental results obtained when applying the algorithms on a test set of GSP-EC instances with up to one hundred nodes and four hundred edges are presented in Section 4. Finally conclusions are presented in Section 5.

2. PROBLEM FORMALIZATION AND DEFINITIONS

We use the following notation to formalize the GSP-EC: $G = (V, E, C)$: simple undirected graph with weighted edges; V: Nodes of G; E: Edges of G; $C: E \rightarrow \mathfrak{R}^{+}$: edge weights; $T \subseteq V$: Terminal nodes (the ones for which connectivity requirements exist); $R: R \in Z^{|T| \times |T|}$: Symmetrical integer matrix of connectivity requirements with $r_{ii} = 0\ \forall i \in T$.

The set V models existing sites among which a certain set E of feasible links could be deployed, being the cost of including a certain link in the solution given by the matrix C. The set T models those sites for which at least one connectivity requirement involving other site has to be met; these requirements are specified using the matrix R. Nodes in the set $V \setminus T$ (named Steiner nodes) model sites that can potentially be used (because doing so reduces the total topology cost or because it is impossible to avoid using them when connecting a given pair of terminals) but for which no requirements exist. Using this notation the GSP-EC can be defined as follows:

Definition 2.1: **GSP-EC**. Given the graph G with edge weights C, the teminals set T and the connectivity requirements matrix R, the objective is to find a minimum cost subgraph $G_T = (V, E_T, C)$ where every pair of terminals i, j is connected by r_{ij} edge-disjoint paths.

3. THE "GRASP" METAHEURISTIC

GRASP (Greedy Randomized Adaptive Search Procedure) is a metaheuristic that proved to perform very well for a variety of combinatorial optimization problems. A GRASP is an iterative "multistart local optimization" procedure which performs two consecutive phases during each iteration: ***Construction Phase*** (ConstPhase): it builds a feasible solution that chooses (following some randomized criteria) which elements to add from a list of candidates defined with some greedy approach; ***Local Search Phase*** (LocalSearchPhase): it explores the neigborhood of the feasible solution delivered by the Construction Phase, moving consecutively to lower cost solutions until a local optimum is reached. Typical parameters are the size of the list of candidates, the amount of iterations to run $MaxIter$ and a seed for random number generation. After having run $MaxIter$ iterations the procedure returns the best solution found. Details of this metaheuristic can be found in [12].

3.1 Construction Phase Algorithm

The algorithm, shown in Figure 1, is an adaptation of the one found in [13] to the edge-connected case. It proceeds by building a graph which satisfies the requirements of the matrix R, starting with an edgeless graph and adding one new path in each iteration to the solution G_{sol} under construction. The matrix $M = (m_{ij})$ records the amount of connection requirements not yet satisfied in G_{sol} between the terminal nodes i, j; the sets P_{ij} will record the r_{ij} disjoint paths found for connecting the nodes i, j. One improvement over the previous algorithm is to alter the costs of the matrix C to introduce random and enable the chance to build an optimal solution no matter what the problem instance is. We have proven that a sufficient condition to ensure this is that all edges have their costs altered independently from the others and the altered costs take values in $(0, +\infty)$ with any probability distribution that assigns non-zero probabilities to any open subinterval of $(0, +\infty)$. Loop 3-15 is repeated until all terminal nodes have their connectivity requirements satisfied, or until for a certain pair of terminals i; j, the algorithm fails to find a path a certain number of times MAX ATTEMPT. In each iteration, one pending connection requirement is chosen and the shortest-path is computed considering a modified cost matrix C' where edges already introduced in the solution under construction during previous iterations have cost zero, enabling edge-reusing among different pairs of terminals. Finally, the algorithm ends by returning the feasible solution G_{sol} together with the path set P which "certifies" that the requirements specified by R were met.

3.2 Local Search Phase Algorithms

Any local search algorithm needs a precise definition of the neighbourhood concept; we propose two different ones, which we will chain inside our suggested LocalSearchPhase algorithm. They are defined in terms of the structural decomposition of graphs in "key-node" and "key-paths" [13] plus a new structural component that we define below.

Definition 3.2.1 Key-star: Given a GSP-EC instance, a feasible solution G_{sol} and any of its nodes v, the key-star associated to v is the subgraph of G_{sol} obtained through the union of all key-paths having v as an endpoint.

Definition 3.2.2 Path-Based Local Search Neighbourhood1: Our first neighbourhood is based on the replacement of any key-path k by another key-path with the same endpoints, built with any edge from the feasible connections graph G (even some of G_{sol}), provided no connectivity levels are lost when reusing edges. Let k be a key-path of a certain solution G_{sol} and P a set of paths which "certificates" its feasibility (as the one returned by ConstPhase). We will denote by $J_k(G_{sol})$ the set of paths $\{p \in G_{sol}: k \subseteq p\}$. These are the paths which contain the key-path k. We will also denote by $\chi_k(G_{sol})$ the edge set $\bigcup_{q=i..j \in J_k(G_{sol})} E(P_{ij} \setminus q)$. These are the edges that, if used to replace the key-path k in P (obtaining a path set P') would turn to be shared by some paths from G_{sol} with the same endpoints, thus invalidating the resulting set P' as a feasibility certificate. The algorithm LocalSearchPhase1 (shown in Figure 1) then considers the replacement of key-paths k by other paths p such that $cost(p) < cost(k)$ and the edges of p are chosen from the set $(E \setminus \chi_k(G_{sol})) \cup k$.

Definition 3.2.3 Key-Star-Based Local Search Neighbourhood2: This is a second neighbourhood based on the replacement of key-stars, which frequently allows to improve feasible solutions that are locally optimal when only considering Neighbourhood1. In the case of the GSP-NC, as no node sharing is allowed among disjoint paths, all key-stars are trees (named key-trees); a key-tree replacement neighbourhood for the GSP-NC can be found in [13], [14]. Due to the possibilty of sharing nodes among edge-disjoint paths, when working with GSP-EC problems, we must work with key-stars. Unlike [13], [14] we will allow the root node to be a terminal node in order to get a broader neighbourhood. In the GSP-NC any key-tree can be replaced by any tree with the same leaves with no loss of connectivity levels. In the GSP-EC, if the replacing structure is also a keystar the same holds true; but it does not for other general structures (non-star trees included). We propose an algorithm that given a key-star k, deterministically seeks for the lowest cost replacing key-star k' able to "repair" the paths from P broken when removing the edges of k. For allowing as much reusing of edges as possible, we can extend our previous definition of $J_k(G_{sol})$ and $\chi_k(G_{sol})$ to consider key-stars k instead of key-paths. Figure 1 presents the LocalSearchPhase2 algorithm, making use of a BestKeyStar algorithm shown in Figure 2. Given a keystar k we denote by θ_k its root node; by ψ_k the set of its leaf nodes; and by $\hat{\delta}_{k,m}$ (being m the root node of k or one of its leaves) the highest amount of key-paths that join m in k with any other node that is root or leaf in k. Figure 2 also depicts the process of determining which the best key-star to replace a given one is. It illustrates (a) the feasible graph G with a key-star that keep4s connected the leaf nodes t, u, v; (b) the graph G' obtained after adding the virtual nodes w linked to t, u, v by the appropriate amount of edges, and a "candidate" root node z; (c) the shortest paths found to connect z and w (using a polynomial-time algorithm minimum-cost k-edge-disjoint paths algorithm like the one in [5]; and (d) the new key-star obtained after removing the virtual node w.

```
Procedure LocalSearchPhase1(G, C, T, S)
improve ← TRUE
κ ← k-decompose(S)
while improve do
   improve ← FALSE
   for all kpath k ∈ κ with endpoints u, v do
      G′ ← the subgraph induced from G by E(k) ∪ (E \ χ_k(S))
      C′ ← (c′_ij)/c′_ij = 0 if (i, j) ∈ S \ k; c′_ij = c_ij otherwise
      k′ ← shortest-path(G′, C′, u, v)
      if cost(k′, C′) < cost(k, C′) then
         improve ← TRUE
         update S : ∀p ∈ J_k(S)(p ← (p \ k) ∪ k′)
         if ∃z ∈ V(k′), z ∉ {u, v}, degree(z) ≥ 3 in S then
            remove-cycles(J_k(S))
            κ ← k-decompose(S)
         else
            κ ← κ \ {k} ∪ {k′}
         end if
      end if
   end for
end while
return S
```

```
Procedure LocalSearchPhase2(G, C, T, S)
improve ← TRUE
κ ← k-decompose(S)
while improve do
   improve ← FALSE
   for all kstar k ∈ κ do
      [k′, newCost] ← BestKeyStar(G, C, T, S, k)
      if newCost < cost(k, C) then
         improve ← TRUE
         replace k by k′ in all paths from S
         κ ← k-decompose(S)
         abort for all
      end if
   end for
end while
return S
```

Figure 1: Algorithms for Local Search Phase (1 and 2)

```
Procedure BestKeyStar(G, C, T, S, k)
G′ ← the subgraph induced from G by E(k) ∪ (E \ χ_k(S))
C′ ← (c′_ij)/c′_ij = 0 if (i, j) ∈ S \ k; c′_ij = c_ij otherwise
add a "virtual node" w to G′
Ω ← ψ_k
if θ_k ∈ T then
   Ω ← Ω ∪ {θ_k}
end if
for all m ∈ Ω do
   add δ̂_k,m parallel edges (w, m) to G′ with cost 0
end for
c_min ← 0; k_min ← k
for all z ∈ V(G) do
   k′ ← simult-shortest-paths(G′, δ_G,w, z, w)
   if k′ has δ_G,w paths ∧ cost(k′, C′) < c_min then
      c_min ← cost(k′, C′); k_min ← k′
   end if
end for
return [k_min, c_min]
```

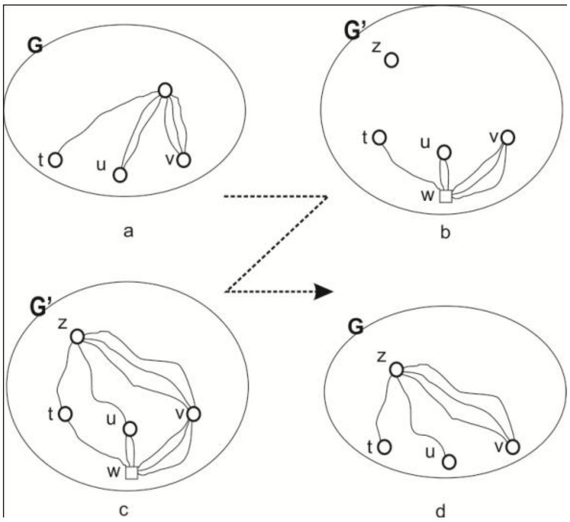


Figure 2: Algorithm for determination of the best key star to replace a given one

3.3 GRASP algorithm description

Finally we can put together the pieces and build a GRASP algorithm to solve the GSP-EC using a compound local search phase that will operate by applying key-path movements until no further improvements are possible followed by a (single) key-star replacement, and so on. The cycle is repeated until no further improvements are found with none of both movements.

4. PERFORMANCE TESTS

This section presents the results obtained after testing our algorithms with twenty-one test cases. The algorithms were implemented in C/C++ and tested on a 2 GB RAM, Intel Core 2 Duo, 2.0 GHz machine running Microsoft Windows Vista. For every instance we ran 100

GRASP iterations. To our best knowledge, no library containing benchmark instances related to the GSP-NC nor GSP-EC exists; we have built a set of twenty-one test cases that are based in cases found in the following public libraries: ***steinlib*** [8]: instances of the Steiner problem; in many cases the optimal solution is known, in others the best solution known is available; ***tsplib*** [15]: instances of diverse graph theory related problems, including a “Traveling Salesman Problem” section. The main characteristics of the twenty-one test cases are shown in Table 1 where we show the amount of nodes (V), feasible edges (E), terminal nodes (T), Steiner (non terminal) nodes (St), the level of edge-connectivity requirements (one, two, three or mixed) (Redund.) and the optimal costs when available (Opt). Source data of the twenty-one instances as well as the best solutions found are available in [17]. Computational results of the tests can be also seen in Table 1. Here follows the meaning of each column: (Reqs.) total amount of requirements satisfied by the best solution found; (t(ms)) the average running time in milliseconds per iteration; (Cost) the cost of the best solution found; (%LSI) “local search improvement” – the percentage of cost improvement achieved by the local search phase when compared to the cost of the solution delivered by the construction phase, for the best solution found.

Case	V	E	T	St	Redund.	Opt	Reqs.	t(ms)	Cost	%LSI
b01-r1	50	63	9	41	1-EC	82	36	77	82	3.0
b01-r2	50	63	9	41	2-EC	NA	42	80	98	3.4
b03-r1	50	63	25	25	1-EC	138	300	2611	138	10.6
b03-r2	50	63	25	25	2-EC	NA	378	3108	188	4.1
b05-r1	50	100	13	37	1-EC	61	78	298	61	9.2
b05-r2	50	100	13	37	2-EC	NA	144	1389	120	5.2
b11-r1	75	150	19	56	1-EC	88	171	1477	88	13.8
b11-r2	75	150	19	56	2-EC	NA	324	4901	180	3.4
b17-r1	100	200	25	75	1-EC	131	300	6214	131	10.2
b17-r2	100	200	25	75	2-EC	NA	531	15143	244	3.0
cc3-4p-r1	64	288	8	56	1-EC	2338	28	388	2338	10.0
cc3-4p-r3	64	288	8	56	3-EC	NA	84	2221	5991	4.6
cc6-2p-r1	64	192	12	52	1-EC	3271	66	2971	3271	2.4
cc6-2p-r2	64	192	12	52	2-EC	NA	132	4801	5962	10.2
cc6-2p-r123	64	192	12	52	1,2,3-EC	NA	140	6317	8422	9.8
hc-6p-r1	64	192	32	32	1-EC	4003	496	25314	4033	6.8
hc-6p-r2	64	192	32	32	2-EC	NA	992	28442	6652	3.5
hc-6p-r123	64	192	32	32	1,2,3-EC	NA	957	26551	7930	5.2
bayg29-r2	29	406	11	18	2-EC	NA	110	975	6856.88	4.6
bayg29-r3	29	406	11	18	3-EC	NA	165	2413	11722	4.2
att48-r2	48	300	10	38	2-EC	NA	90	1313	23214	13.0
Averages							**265**	**6524**	**-**	**6.7**

Table 1: Test cases and results

5. CONCLUSIONS

In this work we overcame the problems introduced by edge-disjointess (when compared to node-disjointness on previously proposed algorithms). Our algorithm GRASP-GSP was shown to find good quality solutions to the GSP-EC when applied to a series of heterogeneous test cases with up to 100 nodes and up to 406 edges. For all cases with known optimal cost the algorithm was able to find solutions with costs no more than 0.74% higher than the optimal cost. Significant cost reductions averaging 6.7% are achieved after applying the local search phase over the greedy solutions built by the construction phase. Execution times were comparable to the ones of previous similar works like [13], [14] for the node-connected version of the GSP. In all cases the maximum possible amount of connection requirements (allowed by the topology of the original graph G and the requirements in R) was reached and always returning edge-minimal solutions.

References

[1] Agrawal, Klein, Ravi. When trees collide: An approximation algorithm for the generalized steiner problem on networks. SIAM Journal on Computing, 24(3):440–456, 1995.

[2] Baïou. Le problème du sous-graphe Steiner 2-arête connexe: Approche polyédrale. PhD thesis, Université de Rennes I, Rennes, France, 1996.

[3] Baïou, Mahjoub. Steiner 2-edge connected subgraph polytope on series-parallel graphs. SIAM Journal on Discrete Mathematics, 10(1):505 – 514, 1997.

[4] Baïou. On the dominant of the Steiner 2-edge connected subgraph polytope. Discrete Applied Mathematics, 112(1-3):3 – 10, 2001.

[5] Bhandari. Optimal physical diversity algorithms and survivable networks. In Computers and Communications, 1997. Proceedings, Second IEEE Symposium on, pages 433 –441, July 1997.

[6] Coullard, Rais, Wagner, Rardin. Linear-time algorithms for the 2-Connected Steiner Subgraph Problem on Special Classes of Graphs. Networks, 23(1):195 – 206, 1993.

[7] Kerivin, Mahjoub. Design of Survivable Networks: A survey. Networks, 46(1):1 – 21, 2005.

[8] Koch. Konrad-Zuse-Zentrum für Informationstechnik Berlin. Steinlib test data library. Available at http://steinlib.zib.de/steinlib.php.

[9] Krarup. The generalized steiner problem. Technical report, DIKU, University of Copenhagen, 1979.

[10] Mahjoub, Pesneau. On the Steiner 2-edge connected subgraph polytope. RAIRO Operations Research, 42(1):259–283, 2008.

[11] Monma, Munson, Pulleyblank. Minimum-weight two connected spanning networks. Mathematical Programming, 46(1):153 – 171, 1990.

[12] Resende, Ribeiro. Greedy randomized adaptive search procedures. In F. Glover and G. Kochenberger, editors, Handbook of Metaheuristics, pages 219–249. Kluwer Academic Publishers, 2003.

[13] Robledo, Canale. Designing backbone networks using the generalized steiner problem. In Design of Reliable Communication Networks, 2009. DRCN 2009. 7th International Workshop on, volume 1, pages 327 – 334, October 2009.

[14] Robledo. GRASP heuristics for Wide Area Network design. PhD thesis, IRISA, Université de Rennes I, Rennes, France, february 2005.

[15] Ruprecht-Karls-Universitat Heidelberg. Tsplib network optimization problems library. Available at http://comopt.ifi.uniheidelberg.de/software/TSPLIB95.

[16] Sartor. Problema general de Steiner en grafos: resultados y algoritmos GRASP para la versión arista-disjunta. M.Sc. thesis, Universidad de la República, Montevideo, Uruguay, 2011.

[17] Universidad de Montevideo. GSP-EC test set with best results found. Available at http://www2.um.edu.uy/psartor/grasp-gsp-ec.zip.

[18] Winter. Steiner problem in networks: A survey. Networks., 17(2):129–167, 1987.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section III:

OR Applications in Telecommunication and Navigation Systems

LP MODEL FOR DAY-AHEAD PLANNING IN ENERGY TRADING

Marinović Minja, Milan Stanojević and **Dragana Makajić Nikolić**
University of Belgrade, Faculty of Organizational Sciences
Jove Ilića 154, 11000 Belgrade, Serbia
marinovic75709@fon.bg.ac.rs, milans@fon.rs, gis@fon.rs

Abstract: In this paper, the problem of day-ahead planning in the trading section of energy trading companies has been considered. It has been assumed that the demand and supply are arranged and that the additional MW and transmission capacity can be purchased. The problem is observed as directed multiple-source and multiple-sink network and then represented by LP mathematical model of total daily profit maximization subject to flow constraints. A numerical example is presented to illustrate the application of the model.

Keywords: energy trading, day-ahead planning, network flow, LP model.

1 INTRODUCTION

During XX century, power production and transmission were carried out between monopolistic public power companies. In the last twenty years, electricity markets have been deregulated allowing customers to choose their provider and new producers. The concept of a single European electricity market foresees seamless competition throughout the electricity supply chain, within and between EU Member States and adjoining countries. Customers enjoy a choice between competing electricity retailers, who source their requirements in liquid and competitive wholesale markets, irrespective of national or control area boundaries. Market participants actively compete to meet the demand in their own countries and to supply electricity across borders into neighborhood markets. Cross-border trading and supply is an integral part of this competition, as market participants enjoyed non-discriminatory access to interconnected transmission lines. Energy Trading Companies (ETCs) are buying transmission capacity from Transmission System Operators (TSOs) [1]. TSOs consistently release to the market a truly maximum amount of cross-border transmission capacity. ETCs are trying to manage the risks associated with fluctuating prices through buying and selling gas and electricity contracts. Both traders and end-users apply financial instruments such as futures, options and derivatives to protect their exposures to prices and to speculate on price fluctuations.

There are few ways of trading electricity but two main ways are via the telephone in bilateral transactions (so called “Over The Counter“ or OTC, usually through the intermediation of a broker), or it is traded through futures markets such as Nordpool or the EEX. Some key factors influencing energy prices include geopolitical factors, global economic growth, short term weather impacting demand, supply disruptions from maintenance or unexpected outages, fuel price movements and product swapping in response to relative prices [2].

The literature concerning different issues in energy trading and transmission is extensive. Kristiansen in [5] analyses the auction prices at the cross-border annual, monthly and daily capacity auction in area of energy trading in case of Denmark and Germany. Triki et al. [10] consider the multiple interrelated markets for electricity and propose a multi-stage mixed-integer stochastic model for capacity allocation strategy in a multi-auction competitive market. A generalized network flow model of the national integrated energy system that incorporates the production, transportation of coal, natural gas, and electricity storage with respect to the entire electric energy sector of the U.S. economy is proposed in [9]. The authors have formulated the multi period generalized flow problem as an optimization model in which the total costs are minimized subject to energy balance constraints. The problem of energy allocation between spot markets and bilateral contracts is formulated as a general portfolio optimization quadratic programming problem in [6]. The proposed methodology with risky assets can be applied to a market where pricing, either zonal or nodal, is adopted to mitigate transmission congestion. Purchala et al. in [8] propose a zonal network model, aggregating individual nodes within each zone into virtual equivalent nodes, and all cross-border lines into equivalent border links. Using flow-based modeling, the feasibility of the least granularity zonal model where the price zones are defined by the political borders, is analyzed. The authors in [3] consider multiple-source multiple-sink flow network systems such as electric and power systems. They observe the problem in which resources are transmitted from resource-supplying (source) node(s) to resource-demanding (sink) node(s) through unreliable flow networks. Nowak et al. [7] analyzed the simultaneous optimization of power production and day-ahead power trading and formulated it as a stochastic integer programming model.

The rest of the paper is organized as follows. In section 2 are description of the main assumptions of the observed problem. The LP mathematical model for day-ahead planning is presented in Section 3. Section 4 is dedicated to the experimental results to illustrate the model, along with comments on the main results of its simulation. Conclusions along with perspectives regarding further work are finally presented in section 5,.

2 PROBLEM DESCRIPTION

The focus of this paper is on electricity trading from the perspective of the ETC. The main task of the trading section is to optimize the portfolio of energy products, ensuring clients' demands are met, whatever the circumstances. The trading section also enables companies to respond to the ever-changing state of the region's transmission grid and production capacities. The organization of this section is traditional. The trading department deals with spot and longer term arrangements. The scheduling and portfolio Management department makes schedules, takes care about cross border capacities allocations and creates paths in order to optimize the whole portfolio, managing different energy sources, customers in different countries and cross border energy flows and costs. Finally, the settlements department is dealing with invoicing, preparing deal confirmations and statistics necessary for all local ETCs.

Trading section of ETC is dealing with at least two optimization problems: first one is long term planning and consist determination of transmission capacity that will be used for next period, while the second one is day-ahead planning and represents finding the optimal routes that will satisfy the short term demand considering the available transmission capacity. Subject of this paper is day-ahead problem.

The assumptions of the observed problem are:

- Daily demand is known, and all arranged demands must be satisfied.
- Energy selling prices are known for each country.
- Daily supply is known and some of them are arranged.
- Energy buying prices are known for each country.
- If there is a surplus or shortage of arranged supply, it will be traded through futures markets.
- Available daily capacity is purchased and presented in MW.
- It is possible to buy extra daily transmission capacity if it is necessary.
- Profit presents difference between selling price and cost of buying energy and transmission capacity.

Decision that should be made is: where and how much ETC should buy and on which ways that energy should be transferred so that they maximize total daily profit.

3 MODEL FORMULATION

Described problem originally can be modeled as directed multiple-source and multiple-sink network. All ETCs from one country are represented by one node (Figure 1).

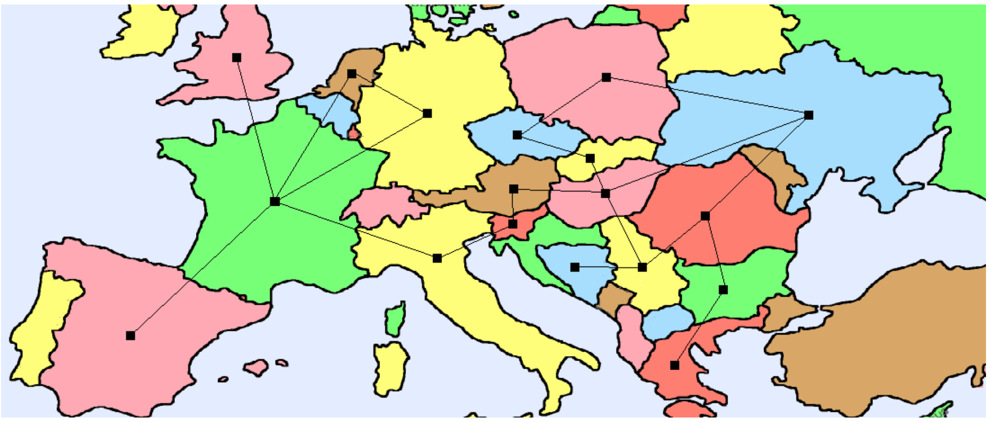


Figure 1: Cross-border energy flow presented as network

Obtained network then can be converted into single-source and single-sink problem by introducing supersource *s* and supersink *t*. [4]

The notation used to define sets, parameters and decision variables is as follows.

- Sets
  - V – set of all nodes
  - A - set of arcs presenting existance of purchased capacity between nodes,
  - B - set of arcs presenting existance of additional capacity between nodes,
- Parameters
  - c_{ij} – daily capacity in MW of arc $(i,j)\in$A, $i\neq s$, $j\neq t$;
  - c_{sj} – daily offer in MW of supplier j, $(s,j)\in$A;
  - c_{it} – daily demand in MW of buyer i, $(i,t)\in$A;
  - d_{ij} – flow price in €/MW on arc $(i,j)\in$A, $i\neq s$, $j\neq t$:
  - d_{sj} – purchase price in €/MW from supplier j, $(s,j)\in$A;
  - d_{it} – selling price in €/MW for buyer i, $(i,t)\in$A;
  - p_{ij} – daily capacity in MW which is possible to buy on arc (i,j) $\in$B, $i\neq s$, $j\neq t$;
  - q_{ij} – flow price for additional capacity in €/MW on arc $(i,j)\in$B,
  - q_{sj} – purchase price for additional supply in €/MW from supplier j, $(s,j)\in$B;
  - r – daily total arranged demand $r = \sum_{(i,t)\in A} c_{it}$
- Variables:
  - x_{ij} – optimal flow in purchased arc $(i,j)\in$A
  - y_{ij} – optimal flow in additional arc $(i,j)\in$B

Using given notation, the LP mathematical model for day-ahead planning can be stated as:

$$(\max) f(x,y) = \sum_{(i,t)\in A} d_{it}x_{it} - \sum_{(i,j)\in A,\, j\neq t} d_{ij}x_{ij} - \sum_{(i,j)\in B} q_{ij}y_{ij} \tag{1}$$

Subject to.

$$\sum_{(s,j)\in A} x_{sj} + \sum_{(s,j)\in B} y_{sj} = r \tag{2}$$

$$\sum_{(i,t)\in A} x_{it} = r \tag{3}$$

$$\sum_{(i,j)\in A} x_{ij} + \sum_{(i,j)\in B} y_{ij} = \sum_{(j,i)\in A} x_{ji} + \sum_{(j,i)\in B} y_{ji} \tag{4}$$

$$0 \le x_{ij} \le c_{ij},\ (i,j)\in A \tag{5}$$

$$0 \le y_{ij} \le p_{ij},\ (i,j)\in B \tag{6}$$

Objective function (1) maximizes the total profit. Equations (2-4) represent standard constraints for value of the flow. Constraints (5-6) refer to the purchased and additional arcs capacities.

The optimal values of variables x_{sj}, $(s,j)\in$A and y_{sj}, $(s,j)\in$B represent the amount of MW that should be transmitted from regular and additional supplier *s*. Variables, x_{it}, $(i,t)\in$A represent the amount of MW which should be delivered to the buyer *i*. The rest of the variables, x_{ij}, $(i,j)\in$A, $i\neq s$, $j\neq t$ and y_{ij}, $(i,j)\in$B, $i\neq s$, $j\neq t$ are related to the optimal flow between purchased and additional capacities, respectively.

4 A NUMERICAL EXAMPLE

A hypothetical example is as follows: Suppose that an ETC is in the middle of the day-ahead planning and that there are arranged trading with three buyers and three suppliers. Energy can be transmitted using 19 purchased capacities. Input data are given in Table 1. The main values in the table refer to energy flow while values in parentheses represent prices. The daily supplies and purchase prices from suppliers are given in the first row and the daily demands and selling prices for buyer are given in the last column of the table. The rest of the data represent available daily flow capacities and prices.

Table 1: Demand, supply and purchased transmission capacities and prices

	s	1	2	3	4	5	6	7	8	9	*t*
s					30(16)	50(14)			60(15)		
1			50(2)		45(2.5)						
2				30(3)			35(2)				
3											
4						30(4)		40(1)			
5		20(3)					30(1.5)	20(5)	30(3.5)		
6				50(4)		25(2)					20(20)
7						25(3.5)			25(2)		30(25)
8						15(4.5)	30(3)			35(1.5)	
9							50(4)		15(2.5)		60(22)
t											

Since there is a surplus of arranged supply, ETC must sell that amount of energy through futures markets. Also, if it is necessary, ETC can purchase additional daily transmission capacity.

The corresponding data are given in Table 2.

Table 2: Additional transmission capacities and prices

	s	1	2	3	4	5	6	7	8	9	t
s											
1			30(12)		25(12)						
2				40(10)			20(7)				
3							30(8)				
4						25(14)		15(6)			
5		10(13)						25(10)	40(9.5)	45(11)	
6				30(8)		25(7)					
7						25(11)			15(7)		
8						35(15)				25(12)	
9							40(13)		10(9)		
t											

GNU Linear Programming Kit (GLPK) has been used for modeling and solving. GLPK is an open source software for solving linear and mixed integer mathematical programming problems. The optimal solution obtained for the described problem by using GLPK is given in Table 3 and represented graphically in Figure 2.

Table 3: Optimal solution

Arcs flow	Optimal value	Maximum value
x[s,4]	30	30
x[s,5]	45	50
x[s,8]	35	60
x[4,7]	30	40
x[5,6]	20	30
x[8,9]	35	35
x[6,t]	20	20
x[7,t]	30	30
x[9,t]	60	60
y[5,9]	25	45

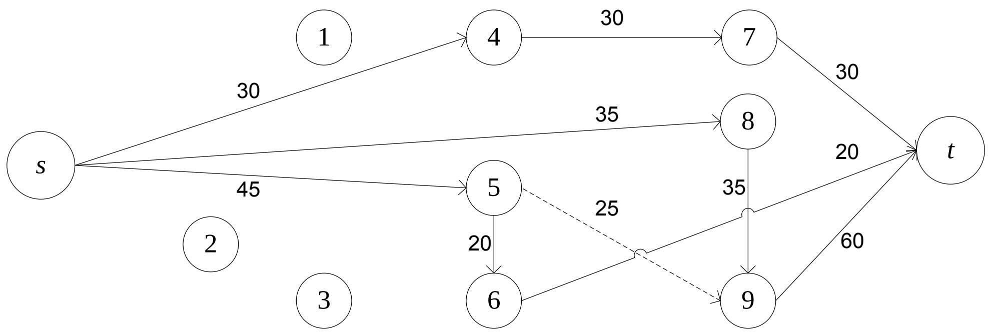


Figure 2: Graph representation of the optimal solution

Total daily profit is 447.5€, 5MW from the supplier 5 and 25MW from the supplier 8 will be sold through futures markets. In order to transmit arranged demand it is necessary to buy additional 25MW of transmission capacity from node 5 to node 9.

5 CONCLUSIONS

The optimization problems that appear in the trading section of Energy Trading Companies have been considered in this paper and one of them, day-ahead energy planning, has been formulated as linear problem. Day-ahead energy planning implies finding the optimal routes that will satisfy the daily demand using the purchased and additional energy transmission capacity. A hypothetical example has been presented in this paper. Since developed model is linear, it can be used to solve real life problems of large dimensions. As a topic of further research, the trading through futures markets can be taken into consideration. Another topic of future research is modeling long term planning strategy of Energy Trading Companies.

ACKNOWLEDGMENTS

This research was partially supported by the Ministry of Education and Science, Republic of Serbia, Projects numbers TR32013 and TR35045.

References

[1] EFET, Electricity markets harmonisation - New edition, http://www.efet.org/Position_Papers/Electricity_Market,_Position_Papers,_European_Level_5492.aspx?urlID2r=72

[2] HARCOURT, Harcourt strategy sheet – Energy & Power Trading, http://www.harcourt.ch/cms/hauptseite/zeigeBereich/11/gibDatei/1088/26-energypowertrading.pdf

[3] Hsieh, C. C., Chen, Y. T., 2005. Resource allocation decisions under various demands and cost requirements in an unreliable flow network *Computers & Operations Research* Vol. 32 pp. 2771–2784

[4] Hu, T.C., Shing, M.T., 1982. "*Combinatorial Algorithms*", Dover Publication, Inc., NY

[5] Kristiansen, T., 2007. An assessment of the Danish–German cross-border auctions. *Energy Policy* Vol. 35, pp. 3369–3382

[6] Liu, M., Wu, F. F., 2007. Portfolio optimization in electricity markets, *Electric Power Systems Research* Vol. 77 pp. 1000–1009

[7] Nowak, M. P. Schultz, R. D., Westphalenm M., A., 2005. Stochastic Integer Programming Model for Incorporating Day-Ahead Trading of Electricity into Hydro-Thermal Unit Commitment, *Optimization and Engineering*,Vol. 6, pp. 163–176

[8] Purchala, K.; Haesen, E.; Meeus, L.; Belmans, R., 2005. Zonal network model of European interconnected electricity network, CIGRE/IEEE PES, International Symposium New Orleans, pp. 362 - 369

[9] Quelhas, A., Gil, E., McCalley, J. D., Ryan, S. M., 2007. A Multiperiod Generalized Network Flow Model of the U.S. Integrated Energy System: Part I—Model Description. *IEEE Transactions on Power Systems* Vol. 22 No. 2 pp. 829 – 836

[10] Triki, C., Beraldi, P., Gross, G., 2005. Optimal capacity allocation in multi-auction electricity markets under uncertainty. *Computers & Operations Research* Vol. 32 pp. 201–217

GNSS ORBIT RE-CONSTRUCTION USING WAVELET NEURAL NETWORKS

Polona Pavlovčič Prešeren and **Bojan Stopar**
University of Ljubljana, Slovenija
polona.pavlovcic@fgg.uni-lj.si, bojan.stopar@fgg.uni-lj.si

Abstract: This paper presents a slightly different approach of setting continuous GNSS (Global Navigation Satellite System) orbit from discrete precise ephemerides, as contained in GNSS processing software packages, where the orbit problem is solved using known mathematical tools such as numerical integration and polynomial interpolation. This time we present solution based on the use of wavelet neural networks (WNN), which is proved to be universal approximator. Numerical studies showed that WNN employement in GNSS orbit re-construction can be useful alternative to known solutions in GNSS orbit processing strategies.

Keywords: GNSS orbit re-construction, ephemerides, numerical integration, polynomial inerpolation, wavelet neural networks

1 INTRODUCTION

Processing of GNSS-observations is based on two essential input data: observations and ephemerides. The latter serve to determine the positions of GNSS-satellites at any time since in the GNSS-processing satellites are treated as points of known positions. Satellites are moving along the orbit and circle the Earth in a half of sidereal day. This means that satellite positions are time-dependent quantities with large amount of positional data.

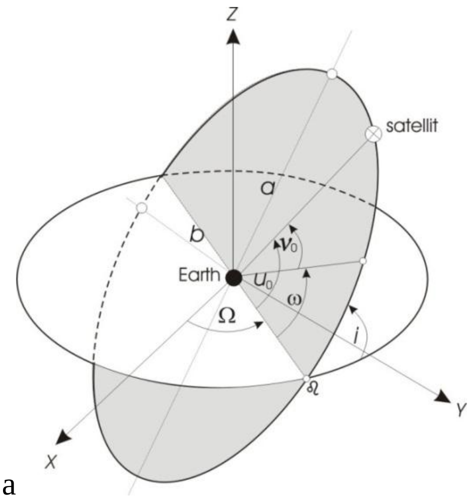
To reduce this amount of data satellite positions are presented in a compressed form and termed as ephemerides. GNSS satellites broadcast information in the form of broadcast ephemerides, which are pre-prepared in master control station and uploaded to the satellites. In case of NAVSTAR GPS (Global Positioning System) technology broadcast ephemerides are presented by Keplerian orbital elements, but in the case of GLONASS (rus. Globalnaja Navigacionnaja Sputnikovaja Sistema) technology by discrete satellite positions. In case of more precise ephemeris data, known as precise ephemerides, provided by IGS (International GPS Service) and IGLOS (International GLONASS Service) over the web, data include positions and clock corrections of all active satellites in a 15-minute tabular form.

Using both broadcast and precise ephemerides requires numerical methods for orbit re-construction. This means that in GNSS-processing orbit function has to be re-constructed from ephemeris data and only from the resulting continuous function user can calculate the positions of a satellite at specific time.

1.1 GNSS ephemeris data form

Broadcast and precise ephemerides use different data and therefore each uses different re-construction method.

GNSS control segment derives GNSS broadcast ephemeris data by extrapolating the estimated orbit for 1 to 14 days into the future. Parameters in broadcast ephemerides are obtained from a curve fitting to the predicted satellite orbit over an interval of 4 to 6 hours. GPS broadcast ephemerides contain orbital parameters (a, e, i, Ω, ω, u_0) and specific coefficients of perturbed orbit motion, which take into account orbit perturbation derived from Earth's gravity field, solar pressure, attraction from Sun and Moon. Orbital parameters (Figure 1) are used to compute orbit function in an inertial frame, which means that is rotated to terrestrial frame (i.e. Earth fixed). Since Keplerian elements in broadcast ephemerides only describe the satellite's orbit over the interval of applicability, they should't be treated as true Keplerian elements of satellite motion. A sample of GPS-broadcast orbit for one satellite is given in Figure 2.



a

Figure 1: Keplerian orbital parameters [1]

a	is semimajor axis of the orbit
e	is numerical eccentricity, computed from b: e $= \sqrt{a^2 - b^2}$
i	inclination of the orbital plane
Ω	is right ascension of the ascending node
ω	is argument of perigee
u_0	argument of latitude

```
     2              NAVIGATION DATA                        RINEX VERSION / TYPE
DAT2RIN 2.35b       pp                 03MAY00 14:12:11 GMTPGM / RUN BY / DATE
                                                           COMMENT
     .3446D-07   .0000D+00  -.1192D-06   .1192D-06         ION ALPHA ——→ ionospheric coefficient α
     .1413D+06  -.3277D+05  -.6554D+05  -.3932D+06         ION BETA  ——→ ionospheric coefficient β
     .390798504668D-13  .931322574615D-08   503808        9 DELTA-UTC: A0,A1,T,W
        13                                                 LEAP SECONDS
satellit date       time                                   END OF HEADER
 1 99 10 27 10  0  0.0  .103783793747D-03  .159161572810D-11  .000000000000D+00 → clock parameters (a0,a1,a2)
     .690000000000D+02  .245937500000D+02  .478448500729D-08 -.174965979949D+01 → IODE, Crs[m], Δn[rad/s], M0[rad]
     .118091702461D-05  .467325572390D-02  .686571002007D-05  .515434567070D+04 → Cuc[rad],e,Cus[rad],sqrt(A)[sqrt(m)]
     .295200000000D+06  .670552253723D-07  .154613932826D+01  .484287738800D-07 → Toe[s],Cic[rad],Ω[rad],Cis[rad]
     .958350471281D+00  .246468750000D+03 -.175988173183D+01 -.813141013444D-08 → i0[rad],Crc[m],ω[rad],Ω'[rad/s]
     .321441960776D-09  .100000000000D+01  .103300000000D+04  .000000000000D+00 → IDOT[rad/s],L2 code,GPS week,P2 label
     .700000000000D+01  .000000000000D+00 -.325962901115D-08  .325000000000D+03 → precision [m],state,TGD[S],IODC
     .290700000000D+06                                                          → broadcast time (GPS week)
```

Figure 2: GPS navigation message with one set of broadcast ephemerides

Since in geodesy many applications have required orbits of better accuracy than those from broadcast ephemerides, in 1992 IGS started to produce precise ephemerides. The precise ephemerides's problem of arrival time delay was partly solved in 1996 with rapid ephemerides (IGR) and eventually in 2001 with ultra-rapid ephemerides (IGU). IGU, IGR and IGS final ephemerides are distinguished according to criteria of time delay of access and accuracy of data. However, comparing to broadcast all sets of precise ephemerides have better accuracy that is constant for the whole interval, but these data are available in smaller discrete time intervals [2]. Precise ephemerides are packaged in daily (or one day and a half in case of IGU) SP3 files and contain discrete positions (x,y,z) and clock rate's for all satellites at 15-minute interval.

1.2 GNSS orbit re-construction

GPS broadcast orbits are re-constructed from ephemerides using a method, proposed in ICD-GPS-200C, Table 20-IV 1997 [3].

GNSS data processing softwares use the interpolation functions for orbit re-construction using discrete tabular data (numerical integration is not used very often, since SP3 precise orbits include only positional data). However, those functions have restrictions since they can be used only for central interpolation, while the accuracy of the calculation is not good enough at the beginning and end of the interval. The solution of the problem are presented by providing more functions for definition area, known as several successive interpolations [4]. However, the problem is still not resolved at the beginning and end of the computational area (Figure3). Interpolation by polynomials is done for each poisition component separately. Most used Lagrange's interpolation has a better performance at the boundaries (as trigonometric) which makes it more convinient for real time applications [5].

Here we present a different approach of satellite orbit construction from 15-minute tabular data, allowing identification of only one function throughout the interval. This approach is based on WNN learning.

2 WAVELET NEURAL NETWORKS

Wavelet neural networks [6] are a special case of feedforward networks, that combine two traditionally separate theories into optimal functional tool for solving non-linear problems. The first wavelet theory is used for optimization of theorems of approximation and scaling, while feed-forward networks keep properties of universal approximation and effective learning.

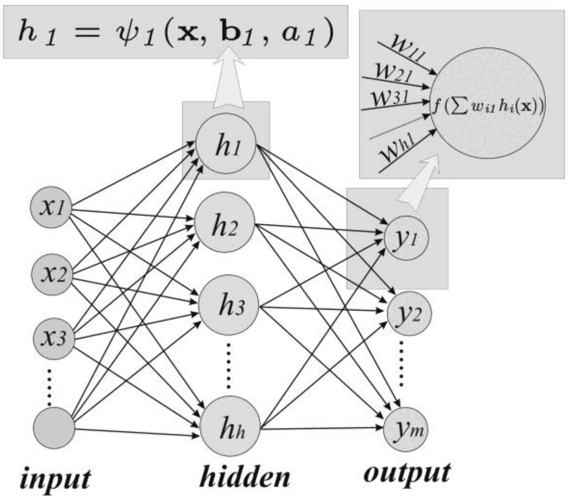


Figure 3: Wavelet neural network architecture

Wavelet networks are three-layer networks, i.e. they consist from a single hidden layer in which activation functions are dilated ($b_j \in \mathbb{R}^n$) and translated ($a_j \in \mathbb{R}$) versions of a single function, called *mother wavelet function*. *Daughter wavelets* are formed by [7]:

$$\psi_j(\mathbf{x}) = a_j^{-d/2}\psi\left(\frac{\mathbf{x}-b_j}{a_j}\right) \tag{1}$$

where **x** consists of known data in the input layer and vector **y** consists of known data in the output layer.

And the WNN output ***f*** is given by:

$$\boldsymbol{f} = \sum_{j=1}^{m} \boldsymbol{w}_j \psi_j(\mathbf{x}) \tag{2}$$

ψ_j can be any suitable basis function, such as Gaussian, Gaussian wavelet, Mexican hat or Morlet function.

Gaussian:	$\psi(\mathbf{x}) = \exp(-\mathbf{x}^2)$	(3)
Gaussian wavelet:	$\psi(\mathbf{x}) = (-\mathbf{x}) \cdot \exp\left(-\frac{\mathbf{x}^2}{2}\right)$	(4)
Mexican hat:	$\psi(\mathbf{x}) = (n - 2\mathbf{x}^2) \cdot \exp(-\mathbf{x}^2),\ \mathrm{n} = \dim(\mathbf{x})$	(5)
Morlet:	$\psi(\mathbf{x}) = \cos(5\mathbf{x}) \cdot \exp\left(-\frac{\mathbf{x}^2}{2}\right)$	(6)

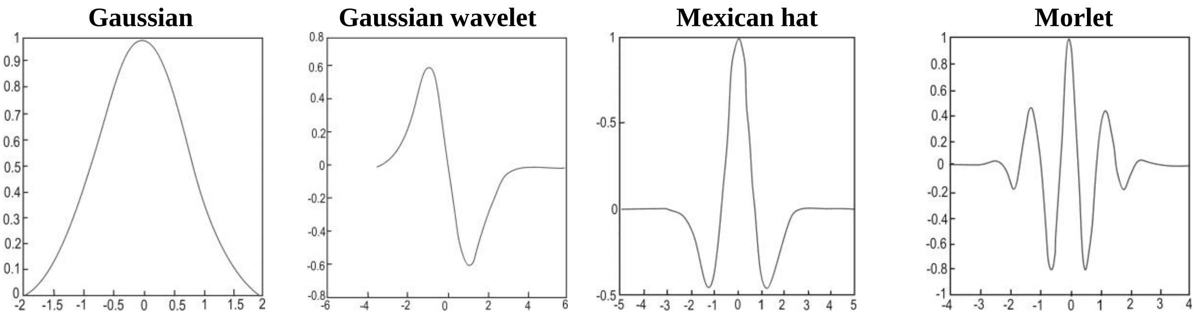


Figure 4: Wavelet functions as types of activation functions

WNN learning is done by two steps: first by fixing dilation and translation parameters, and next by finding weights solving the system of linear equations. The translation parameters can be chosen randomly from the input data, while the dilation parameters are usually fixed. The network output **O** for all the input data can be expresseed as:

$$\mathbf{O} = \mathbf{\Psi W} \tag{7}$$

where:

$$\mathbf{\Psi} = \begin{bmatrix} \psi(x_1, a_1, b_1) & \cdots & \psi(x_1, a_n, b_n) \\ \vdots & \ddots & \vdots \\ \psi(x_n, a_1, b_1) & \cdots & \psi(x_{n1}, a_n, b_n) \end{bmatrix} \tag{8}$$

The weights matrix is solved by pseudoinversion:

$$\mathbf{W} = \mathbf{\Psi}^{+}\mathbf{O},\ \ \mathbf{\Psi}^{+} = (\mathbf{\Psi}^{\boldsymbol{T}}\mathbf{\Psi})^{-\mathbf{1}}\mathbf{\Psi}^{\mathbf{T}} \tag{9}$$

For the evaluation of WNN learning mean square function is used:

$$E(f) = \tfrac{1}{2}\textstyle\sum_{i=1}^{n}\big(y_i - f(x_i)\big)^2 \tag{10}$$

3 NUMERICAL SIMULATIONS

The WNN, used for orbit approximation, consisted of one neuron in the input layer (time) and three neurons in the output layer, representing the three components of the precise broadcast orbit corrections (□x, □y, □z).

Two experiments have been done to verify the performance of WNN: the first experiment is to determine the applicability of WNN for function approximation, and second the performance of WNN in prediction.

WNN learning was evaluated using Eq. 10 and further compared to positions computed from other (polynomial inetrpolation) orbit computation methods. Approximation results were evaluated using root mean square function as the error criteria:

$$RMSE = \sqrt{\tfrac{1}{n} \cdot \textstyle\sum_{i=1}^{n}(f_i - y_i)^2} \tag{11}$$

where f_i is the WNN output and y_i known value. Smaller *RMSE*, better the accuracy.

In Tables 1-3 experimental results of WNN in approximating with different types of activation functions in hidden layer (in our case wavelet functions) are presented. Minimal and maximal deviations are shown, as well as RMSE as final error criteria.

Table 1: WNNs results [m] (x-component)

Basis functions	min	max	$RMSE$
Gaussian	-0.067	0.255	0.080
Gaussian wavelet	**-0.037**	**0.006**	**0.009**
Mexican hat	-0.082	0.005	0.019
Morlet	-0.112	0.008	0.026

Table 2: WNNs results [m] (y-component)

Basis functions	min	max	$RMSE$
Gaussian	-0.002	0.032	0.007
Gaussian wavelet	-0.002	0.033	0.007
Mexican hat	-0.017	0.002	0.004
Morlet	**-0.013**	**0.002**	**0.003**

Table 3: WNNs results [m] (z-component)

Basis functions	min	max	$RMSE$
Gaussian	-0.162	0.018	0.039
Gaussian wavelet	-0.175	0.020	0.042
Mexican hat	-0.037	0.003	0.009
Morlet	**-0.026**	**0.002**	**0.006**

As seen WNNs approximate function efficiently, which probably results due to the fast oscillating characteristics of the wavelet functions. Among all the wavelet basis functions used, Gaussian function yields the lowest accuracy. For the x-component best results gives the Gaussian wavelet, while for y and z component Morlet function is best choice in above presented situation.

WNN can be used also for prediction. Simulations showed method performed well for 20 minutes after last known precise broadcast orbit correction used for training. Figure 4 shows different WNNs performances, i.e. deviations in radial componet. As seen from Figure 4 after deviations after 20 minutes of last known data dont' exceed 2 cm. This is an important ascpect especially in situations when the user cannot get new ephemeris data for example during the reception gaps.

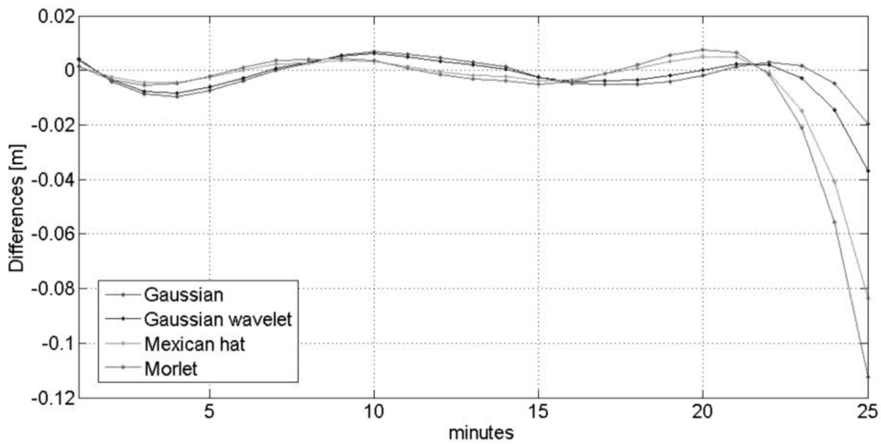


Figure 4: WNNs' performance in prediction (radial component)

4 CONCLUSIONS

In this paper WNN is addressed to estimate the GNSS-orbit function. This product could support GNSS software packages especially near the end of the precise ephemerides. The advantage of such orbit function construction is the ability to determine unique orbit function for the whole are of given points (i.e. satellite's positions), but also allows to use orbit function outside the area (when using interpolation polynomials extrapolation can be critical).

References

[1] Rothacher, M., Mervart, L., 1998. Bernese GPS Software Version 4.2. Astronomical Institute, University of Bern, Switzerland.

[2] Spofford, P.R., Remondi, B.W., 1999. The National Geodetic Survey Standard GPS Format SP3.

[3] Neta. B., Sagovac, C. P, Danielson, D. A., Clynch, J. R., 1996. Fast interpolation for Global Positioning System (GPS) Satellite Orbits. In: Proc. AIAA/AAS Astrodynamics Specialist Conference, San Diego, CA, Paper Number AIAA 96-3658.

[4] Schenewerk, M., 2003. A brief of basic GPS orbit interpolation strategies. GPS Solutions 6(4), pp. 265–267.

[5] Zhang, Q., Benveniste, A., 1992. Wavelet Networks. IEEE Transcastions on neural networks, Vol. 3, No.6. 1992

[6] Zhang, Q., 1997. Using wavelet network in nonparametric estimation. IEEE Trans. Neural Networks, 8(2):227–236.

[7] ICD-GPS-200C. http://www.navcen.uscg.gov/gps/geninfo

APPLICATION OF DATA MINING IN TELECOMMUNICATIONS COMPANIES

Željko V. Račić
University of Banja Luka, Faculty of Economics
Majke Jugovića 4, 78000 Banja Luka, Bosnia and Herzegovina&Republic of Srpska
zeljko.racic@efbl.org

Abstract: Application of Business Intelligence and Data Mining is necessary in process of obtaining competitive advantage and survival in the market. Application of tools and methods of data search results lead to a decrease of loss of income and indicates how a company should manage its business, ie. how to make the organization more "intelligent", and how to make its human resources to be knowledge society. In our work we will use the demographic group and REFII model, with use of a software solution Time Explorer, developed in Visual FoxPro tool.

Keywords: Business Intelligence, Data Mining, Knowledge Discovery from Databases, REFII model, Customer Relationship Management

1 INTRODUCTION

Telecommunications companies face a tough competition today, the demands of the Internet, the possibilities of broadband services (teleconferencing, videoconferencing, video on demand) and the ever increasing demands. By monitoring the evolution of the telecommunications industry, more and more is evident that many telecommunications companies are moving away from a business model based on the strategy of development of infrastructure / products / services and accept a business model based on the strategy focused on customer service. Business Intelligence (BI) means a category of funds data analysis, reporting, querying, that in a business process it can help to use out large amounts of data that are synthesized in valuable information which will be based on prudent business decisions. Business intelligence involves the following four factors: the decision support system (DSS-Decision Support System), Data Warehouse and / or Data Mart, Best Analytical Processing and data mining applications. For the successful operation of modern companies need to focus right information at the appropriate parts of the company at the right time. It is required to digitize all the processes in the organization and make the organization of "intelligent one", and its human resources to their staff knowledge.

To get the large amount of information provided from data that have business significance and serve as decision support, it is necessary to perform data mining analysis, or conduct the process of discovering knowledge from data (KDD - Knowledge Discovery from Databases). Data Mining is the analysis (often very large) of data sets with the goal of finding unexpected relations and patterns in data sets or summary view of data, so that the owner or user of data provides new, understandable and useful information. Data mining plays a major role in every aspect of CRM. Only by applying data mining techniques, can be a great hope for the transformation of numerous company records in the user database into a kind of a comprehensive image of its users. An important prerequisite for successful analysis of DM's is"purity" of the data from the database, and great attention should be paid to the DM model building and selection of appropriate techniques. The choice of data mining techniques is just one of the phases in the generic data mining model.This paper presents the process of DM analysis on the example of Telekom Srpske segmentation. There are several DM techniques that can be used for the segmentation of the user. Segmentation of users is also known as clustering technique. Clustering is a research DM technique.

2 CUSTOMER SEGMENTATION ANALYSIS USING DATA MINING

Identifying characteristics of users is the target's activity analysis and gives views of the users, which directly involve the business strategy for CRM (customer relationship management). CRM is a business model based on a strategy focused to customer service. Strategy focused on customer service lies at the basis of the available modern information technologies that enable rapid and efficient understanding of the needs and behavior of service users, so that understanding is to be as better as possible used to increase the competitiveness of the company. CRM in telecommunications companies as a result of the convergence of services, provides a view of the user - one call, one view, one bill.

Necessary conditions of successful business strategy for telecommunications companies are understanding users' behaviour and proper management of relationships with these customers. The reason for this is also illustrated by the following data:

1. 5% savings in operational costs, the company can raise profits in the range of 25-60%,
2. 35% of users using only one telecom service or have only one job with telecom, potentially are ready to go in the competition,
3. about 20% of customers make 80% profit to telecom,
4. five times the higher cost of acquiring a new customer than retain an existing.

Using the DM model we can get a model for the segmentation of customers to identify groups with similar characteristics that will help better understanding of the user. To know who our customers are, is a good starting point for understanding the rapid changes in the market for telecommunications industry. Customer segmentation is the basis for taking targeted action for each user group. Thus, for example, we offer free minutes to a particular group of users such as users of telecom services are mostly used on weekends or those who have the most incoming calls.

Customer segmentation can be done based on different criteria. It can be simple criteria such as age, gender, demographic characteristics or a combination of these. The DM deals with the segmentation even though it actually works on clustering. Clustering means grouping of users with similar characteristics while simultaneously trying to maximize the differences between these groups. The goal of segmentation in the DM is to discover groups of users that come from the data that the company has about them, instead of ordering a group on the basis of deciding which features are the most important. DM tells what business rules should be, instead of what is thought it should be. There are two DM techniques for segmentation of users: demographics and neuronal clustering. Neuronal clustering is mainly used when the majority of variables are numerical, while demographic clustering is mainly used when the most variables are categorical, because it can operate with categorical values not converting them into numeric. These techniques are complementary and can be used together. Neuronal clustering involves determining the maximum number of clusters. In our case, we can identify four clusters. The number of clusters is base for CRM marketing department of the company, and it is very difficult to use a large number of clusters and data arising from the same. The decision about a number of clusters depends on where and how to use the results of segmentation and should be discussed with the marketing perspective. There are software's for this scaling of input data (IBM's Intelligent Miner for Data Mining). The main difference between the demographic and neuronal clustering is that the demographic clustering automatically determines the number of clusters, whereas for neural clustering is needed to decide how many clusters are to be performed. We will describe how a generic DM model can be used to define groups of users in the telecom industry and how this can be used in the CRM system.

3 CUSTOMER SEGMENTATION TELEKOM SRPSKE (CASE STUDY)

From a large set of questions about users' behaviours of telephone services from Telekom Srpske, and for which we shall give an answer using the DM analysis, we recommend the following: How are the users of telephone services grouped due to: the frequency of calls by days of a week, total call duration and total number of spent pulse? The data in CDR format are taken from the headquarters in Banja Luka for a period of six months of the 2011. In order to obtain answers to the above question, we used the DM technique - cluster analysis. In this case, the entire available data set (total traffic), the purpose of exploration, cluster analysis was performed using a neural network (Kohonen Feature Map - Kohonen self-organizing map).

Cluster analysis was based on the variables: the frequency and duration of calls by call type (local, international ...), the days of the week and hours in the day when the calls were completed, a specified number of clusters was four. This is just one of many possible combinations. The clusters are shown in Figure 1:

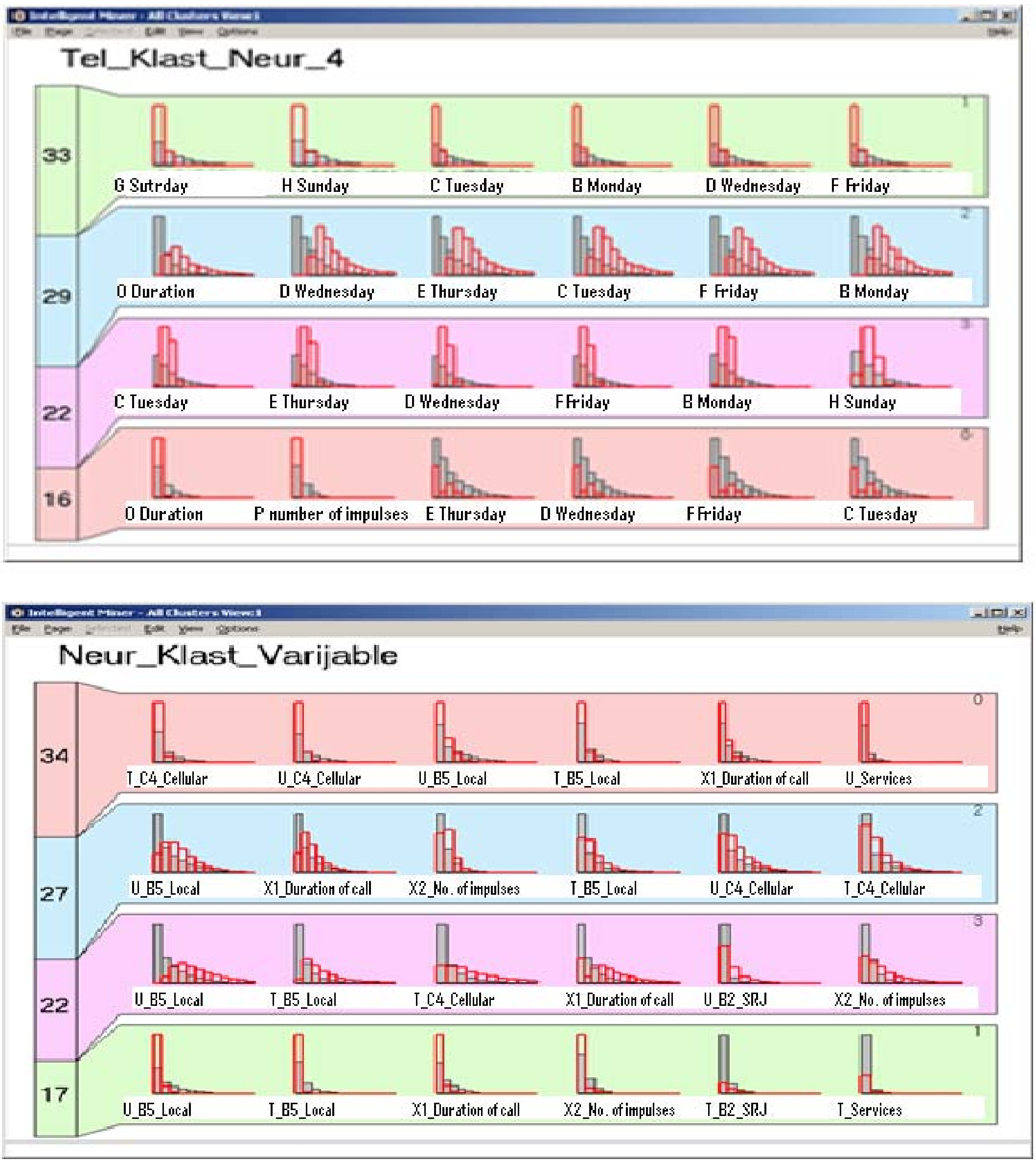


Figure 1: Cluster analysis of frequency and duration of calls by call type

The variable "total number of impulses" contains rough information about the profitability of customers, and it is easy to see that each cluster 1 and cluster 2 are very different in profitability. This confirms the distribution of the total number of impulses in these clusters, with Figure 2.

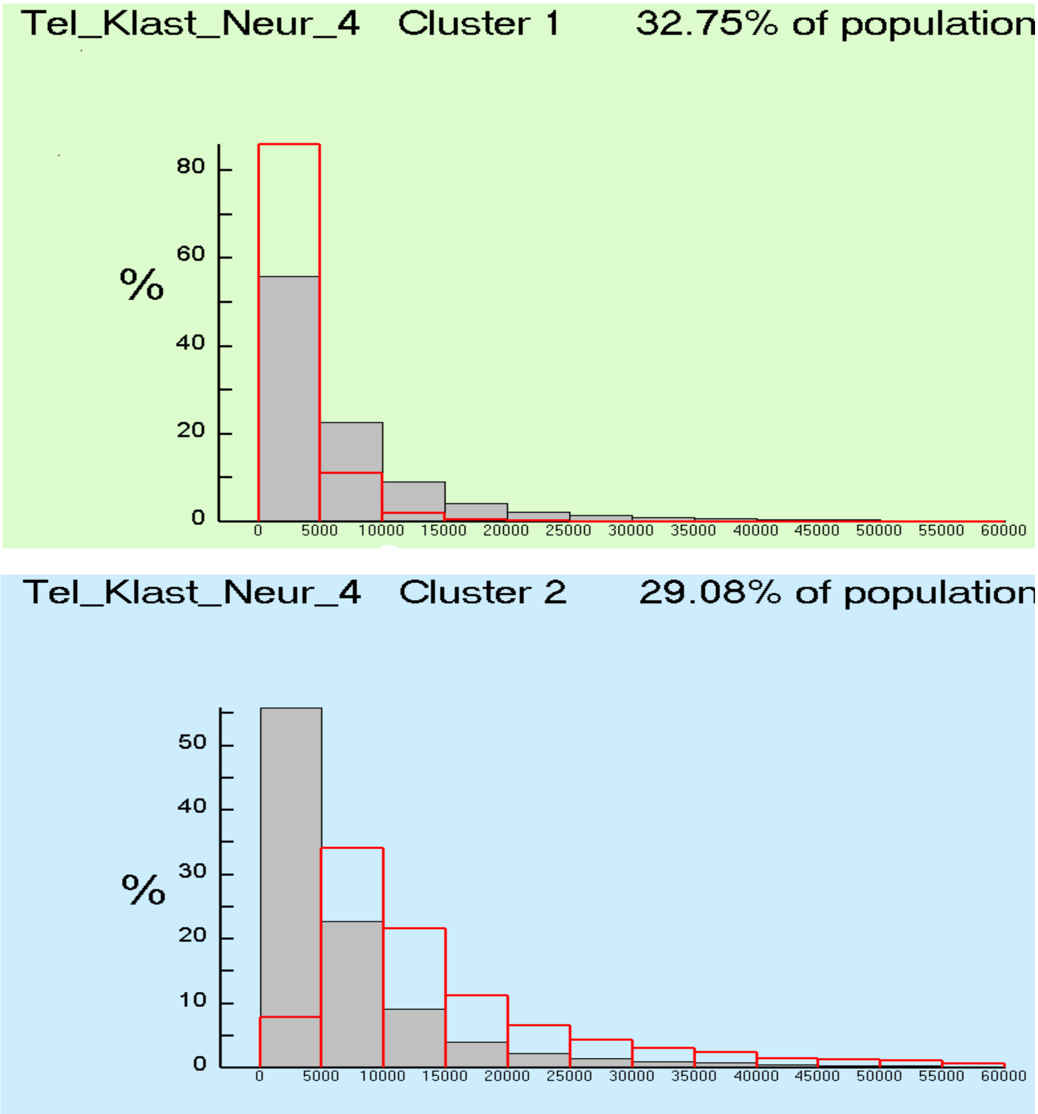


Figure 2: The distribution of pulses from Cluster 1 and Cluster 2

Characteristics of Cluster 1 are (33% of users belong to this cluster):

1. Local calls are above average frequency and duration in certain periods of time (early in a day during the whole week),
2. low level of use of services,
3. low level of other types of calls and to other networks.

From the description of Cluster 1, we can conclude that it is the largest group of users of low profitability. Behaviour based on the method of a call indicates that this is the home users (natural persons), mostly unemployed (students, pensioners), with low levels of use of services that are charged. Marketing efforts directed towards this group of users whose consumption is highly correlated with the prices of services, should concern the campaign to retain customers and prevent their leaving to the competition (it is five times more expensive to attract new users than to retain existing). These users need to be offered discounts during the day when level of calling is below average (the introduction of service-type happy hour - cheaper or free calls during certain periods of the day). Also they need to offer new services with discounts.

Characteristics of Cluster 2 are (29% of users belong to this cluster):

1. Local calls are evenly spread throughout the time frame (except Saturday and Sunday) uniformed frequency and duration,
2. Calls to MOBI'S network of evenly spread throughout the time framework of uniformed frequency and duration (except Saturday and Sunday)
3. All calls are above average frequency and duration in the middle of all working days.

The characteristics of this cluster proved the claim that about 20% of users made 80% profit of telecommunications company. This means that this is a very profitable segment of customers (business users) with high purchasing power. Marketing campaign for this segmentation group should concern keeping profitable customers and increase profits brought by these users (a campaign to increase the loyalty of profitable customers, and cross-selling and up-selling through service packages and business solutions).

4 CONCLUSIONS

Data mining and extraction of hidden knowledge from large databases is a powerful new technology with great potential to help companies to focus on the most important information in their databases. The most common areas of use of data mining, KDD ie. (Knowledge Data Discovery) are: telecommunications, customer retention and to prevent their transition to competition, profiling of users, the techniques of direct marketing, TV, business with credit cards, credit card fraud, investment analysis, risk management, banking, bioinformatics, the collection of loans, education, chemical industry, pharmaceutical industry, insurance, industrial processes - detecting defects, manufacturing, marketing, e-commerce.

Application areas where the greatest benefits are achieved are: improving relationships with customers (CRM), customer retention and preventing their transition to competition, the detection of fraud (Fraud Detection) to detect and predict failures (Fault Detection and Prediction).

CRM system can help to build customer loyalty through: better service with a view to the user (one call, one view, one bill); better marketing with analysis of the users (targeted campaigns using analytical CRM tools for in-depth analysis of customer data in order to find models and usage of indicators, rather than "spray and pray" campaigns) to identify potential churn-era and proactive action, linking behaviour, "user-in focus" with rewards for employees. Keeping existing customers is more valuable than attracting new ones, because efforts to customers' retention avoid direct acquisition costs.

Until recently, the market of the Republika Srpska, Telekom Srpske, was protected by the state and, generally, there was no need for technology to support decision making and improving relationships with customers because customers had no choice. However, by the emergence of more operators, it leads to rapid maturation of the market of Republika Srpska, and BiH. However, due to specific conditions in the environment, the market is still not mature in the true sense of the word.

In order that Telekom Srpske remains the leading operator in Republika Srpska (or become in BiH), the company's strategy must be constantly improving the CRM function, quality improvement and expansion of services and providers. Since data mining plays a major role in every aspect of CRM, just by applying data mining techniques Telekom Srpske can hope for a better understanding of the user. The primary goal of Telekom Srpske management must be the creation of information infrastructure (decision support systems) that will provide the necessary support to data meaning, such as data warehouse, database supplied from many operational systems and the inspection of the entire company with one point of view.

Current practices show that the implementation of data mining in the business of the company brings a quick return on investment, actively applying the knowledge gained and the most important is advantage in a competitive environment.

References

[1] Amor, D. 2002. The E-Business Revolution, Prentice Hall, New York.

[2] Atkinson, R.D. i dr., 2002. The 2002 New Economy Index, PPI, Washington D.C.

[3] Blenkhorn, L. D. 2005. Fleisher, S. C. Outsource Competitive Intelligence? A Viable Option, Competitive Intelligence Magazine, 8/br. 6.

[4] Inmon, B., William, H. 1992. Building the data warehouse, John Wiley & Sons, Inc.

[5] Liautaud, B. 2001. E-Business Intelligence: Turning Information into Knowledge into Profit, McGraw-Hill, New York.

[6] Linoff, G., Berry, J.A. 2000. Mastering Data Mining-The Art and Science of Customer Relationship Management, John Wiley & Sons, Inc.

[7] Pyle, D. 1999. Data preparation for Data Mining, Morgan Kaufmann Publishers.

[8] Pyle, D. 2003. Business Modelling and Data Mining, Morgan Kaufmann Publishers.

[9] Rudd, P.O. 2001. Data Mining Cookbook - Modelling Data for Marketing, Risk, and Customer Relationship Management, Wiley.

[10] Solomon, M.R., Rabolt N.J. 2004. Consumer Behaviour in Fashion, Prentice Hall.

[11] Sachs, R.J., 1993. Macroeconomics, Prentice Hall, New York.

[12] Turban, E., McLean, E., Wetherbe, J. 2004. Information Technology for Management, John Wiley & Sons, Inc.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section IV:

Mathematical Programming and Optimization

TRACE OPTIMIZATION USING SEMIDEFINITE PROGRAMMING

Kristijan Cafuta
University of Ljubljana, Faculty of Electrical Engineering
Tržaška 25, 1000 Ljubljana, Slovenia, E-mail: kristijan.cafuta@fe.uni-lj.si
Igor Klep
University of Maribor, Faculty of Natural Sciences and Mathematics
Koroška 160, 2000 Maribor, Slovenia
University of Ljubljana, Faculty of Mathematics and Physics
Jadranska 19, 1111 Ljubljana, Slovenia, E-mail: igor.klep@fmf.uni-lj.si
Janez Povh
Faculty of Information Studies in Novo Mesto
Novi trg 5, 8000 Novo mesto, Slovenia, E-mail: janez.povh@fis.unm.si

Abstract: In this paper we present the algorithm and its implementation in the software package `NCSOStools` for finding sums of Hermitian squares and commutators decompositions for polynomials in noncommuting variables. It is based on noncommutative analogs of the classical Gram matrix method and the Newton polytope method, which allow us to use semidefinite programming.

Keywords: noncommutative polynomial, sum of squares, semidefinite programming, Matlab toolbox, free positivity.

1 INTRODUCTION

The main question studied in this paper is whether a given real polynomial in noncommuting variables (nc polynomial) can be decomposed as a sum of Hermitian squares and commutators. Using semidefinite programming (SDP) one can obtain numeric evidence and may ask if an exact proof (or certificate) for the answer based on an algorithm can also be given.

1.1 Motivation

The interest in finding decompositions of an nc polynomial as a sum of Hermitian squares and commutators is based on the following fact. If such a decomposition exists, the nc polynomial is necessarily trace-positive, i.e. all of its evaluations at tuples of symmetric matrices have nonnegative trace. Following Helton's seminal paper [5], this can be considered as a specific case of *free real algebraic geometry* in the study of positivity of nc polynomials. Much of today's interest in real algebraic geometry is due to its powerful applications. For instance, the use of sums of squares and the truncated moment problem for polynomial optimization on $\mathbb{R}^n$ established by Lasserre and Parrilo [9, 8, 12, 11] is nowadays a common fact with applications to control theory, mathematical finance or operations research. A nice survey on connections to control theory, systems engineering and optimization is given by Helton, McCullough, de Oliveira, Putinar [3]. Applications of the free case to quantum physics are explained e.g. by Pironio, Navascués, Acín [13] who also consider computational aspects related to sums of Hermitian squares (without commutators).

As a consequence of this surge of interest in free real algebraic geometry and sums of (Hermitian) squares of nc polynomials we have developed `NCSOStools` [1] – an open source Matlab toolbox for solving such problems using *semidefinite programming* (SDP). As a side product our toolbox implements symbolic computation with noncommuting variables in Matlab.

2 PRELIMINARIES

2.1 Words, nc polynomials and involution

Fix $n \in \mathbb{N}$ and let $\langle \underline{X} \rangle$ be the set of *words* in the n noncommuting letters $X_1, \ldots, X_n$ (including the empty word denoted by 1). We consider linear combinations $\sum_w a_w w$ with $a_w \in \mathbb{R}$, $w \in \langle \underline{X} \rangle$ of words in the n letters $\underline{X}$ which we call *nc polynomials*. The set of nc polynomials is actually a free algebra, which we denote by $\mathbb{R}\langle \underline{X} \rangle$. An element of the form aw where $a \in \mathbb{R} \setminus \{0\}$ and $w \in \langle \underline{X} \rangle$ is called a *monomial* and a its *coefficient*. The length of the longest word in an nc polynomial $f \in \mathbb{R}\langle \underline{X} \rangle$ is the *degree* of f and is denoted by $\deg f$. The set of all nc polynomials of degree $\leq d$ will be denoted by $\mathbb{R}\langle \underline{X} \rangle_{\leq d}$. If an nc polynomial f involves only two variables, we write $f \in \mathbb{R}\langle X, Y \rangle$ instead of $\mathbb{R}\langle X_1, X_2 \rangle$.

We equip $\mathbb{R}\langle \underline{X} \rangle$ with the *involution* $*$ that fixes $\mathbb{R} \cup \{\underline{X}\}$ pointwise and thus reverses words, e.g. $(X_1 X_2^2 X_3 - 2X_3^3)^* = X_3 X_2^2 X_1 - 2X_3^3$. Hence $\mathbb{R}\langle \underline{X} \rangle$ is the $*$-algebra freely generated by n symmetric letters. Let $\operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$ denote the set of all *symmetric elements*, that is,

$$\operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle := \{ f \in \mathbb{R}\langle \underline{X} \rangle \mid f = f^* \}.$$

2.2 Sum of Hermitian squares and commutators

An nc polynomial of the form $g^* g$ is called a *Hermitian square* and the set of all sums of Hermitian squares will be denoted by Σ^2. Clearly, $\Sigma^2 \subsetneq \operatorname{Sym} \mathbb{R}\langle \underline{X} \rangle$. The involution $*$ extends naturally to matrices (in particular, to vectors) over $\mathbb{R}\langle \underline{X} \rangle$. For instance, if $V = (v_i)$ is a (column) vector of nc polynomials $v_i \in \mathbb{R}\langle \underline{X} \rangle$, then V^* is the row vector with components v_i^*. We use V^t to denote the row vector with components v_i.

Example 2.1 *The polynomial $f = X^2 - X^2 Y - Y X^2 + Y X^2 Y + X Y^2 X$ is a sum of Hermitian squares, in fact, $f = (X - XY)^*(X - XY) + (YX)^*(YX)$.*

The next notation we introduce is cyclic equivalence which resembles the fact that we are interested in the trace of a given polynomial under matrix evaluations.

Definition 2.2 *An element of the form $[p, q] := pq - qp$, where p, q are polynomials from $\mathbb{R}\langle \underline{X} \rangle$, is called a* commutator. *Nc polynomials $f, g \in \mathbb{R}\langle \underline{X} \rangle$ are called* cyclically equivalent *($f \overset{\text{cyc}}{\sim} g$) if $f - g$ is a sum of commutators:*

$$f - g = \sum_{i=1}^{k} [p_i, q_i] = \sum_{i=1}^{k} (p_i q_i - q_i p_i) \text{ for some } k \in \mathbb{N} \text{ and } p_i, q_i \in \mathbb{R}\langle \underline{X} \rangle.$$

It is clear that $\overset{\text{cyc}}{\sim}$ is an equivalence relation. The following remark shows how to test it.

Remark 2.3

(a) For $v, w \in \langle \underline{X} \rangle$, we have $v \overset{\text{cyc}}{\sim} w$ if and only if there are $v_1, v_2 \in \langle \underline{X} \rangle$ such that $v = v_1 v_2$ and $w = v_2 v_1$. That is, $v \overset{\text{cyc}}{\sim} w$ if and only if w is a cyclic permutation of v.

(b) Nc polynomials $f = \sum_{w \in \langle \underline{X} \rangle} a_w w$ and $g = \sum_{w \in \langle \underline{X} \rangle} b_w w$ ($a_w, b_w \in \mathbb{R}$) are cyclically equivalent if and only if for each $v \in \langle \underline{X} \rangle$,

$$\sum_{\substack{w \in \langle \underline{X} \rangle \\ w \overset{\text{cyc}}{\sim} v}} a_w = \sum_{\substack{w \in \langle \underline{X} \rangle \\ w \overset{\text{cyc}}{\sim} v}} b_w. \tag{1}$$

Example 2.4 *We have* $2X^2Y^2X^3 + XY^2X^2 + XY^2X^4 \overset{\text{cyc}}{\sim} 3YX^5Y + YX^3Y$ *as*

$$\begin{aligned}2X^2Y^2X^3 + XY^2X^2 + XY^2X^4 - (3YX^5Y + YX^3Y) = \\ = [2X^2Y, YX^3] + [XY, YX^4] + [XY, YX^2].\end{aligned}$$

Let

$$\Theta^2 := \{f \in \mathbb{R}\langle \underline{X}\rangle \mid \exists g \in \Sigma^2 : \; f \overset{\text{cyc}}{\sim} g\}$$

denote the convex cone of all nc polynomials cyclically equivalent to a sum of Hermitian squares. By definition, the elements in Θ^2 are exactly nc polynomials which can be written as sums of Hermitian squares and commutators.

Example 2.5 *Consider the nc polynomial* $f = X^2Y^2 + XY^2X + XYXY + YX^2Y + YXYX + Y^2X^2$. *This* f *is of the form*

$$\begin{aligned} f &= (XYXY + YXYX + XY^2X + YX^2Y) + 2XY^2X + \textit{(sum of commutators)} \\ &= (XY + YX)^*(XY + YX) + 2(YX)^*(YX) + \textit{(sum of commutators)}, \end{aligned}$$

hence we have $f \in \Theta^2$ *taking the polynomials* $g_1 = (XY + YX)$, $g_2 = \sqrt{2}YX$ *in the certificate.*

2.3 Semidefinite programming

Semidefinite programming (SDP) is a subfield of convex optimization concerned with the optimization of a linear objective function over the intersection of the cone of positive semidefinite matrices with an affine space. More precisely, given symmetric matrices C, $A_1, \ldots, A_m \in \mathbb{SR}^{s\times s}$ and a vector $b \in \mathbb{R}^m$, we formulate a *semidefinite program in standard primal form* (in the sequel we refer to problems of this type by PSDP) as follows:

$$\begin{array}{lrcl} \inf & \langle C, G\rangle & & \\ \text{s.t.} & \langle A_i, G\rangle & = & b_i, \quad i = 1, \ldots, m \\ & G & \succeq & 0. \end{array} \tag{PSDP}$$

Here $\langle \_ , \_ \rangle$ stands for the standard scalar product of matrices: $\langle A, B\rangle = \operatorname{tr}(B^tA)$. The dual problem to (PSDP) is the *semidefinite program in the standard dual form*

$$\begin{array}{ll} \sup & \langle b, y\rangle \\ \text{s.t.} & \sum_i y_iA_i \preceq C. \end{array} \tag{DSDP}$$

Here $y \in \mathbb{R}^m$, and the difference $C - \sum_i y_iA_i$ is usually denoted by Z.

The mentioned matrix C is arbitrary. One can use $C = I$, a commonly used heuristic for matrix rank minimization. Often however, a solution of *high-rank* is desired. Then $C = 0$ is used, since under a strict feasibility assumption the interior point methods yield solutions in the relative interior of the optimal face, which is in our case the whole feasibility set. If strict complementarity is additionally provided, the interior point methods lead to the analytic center of the feasibility set [4]. Even though these assumptions do not always hold for the instances of SDP we construct, in our computational experiments the choice $C = 0$ in the objective function almost always gave a solution of higher rank than the choice $C = I$.

The relevance of SDP increased with the ability to solve these problems efficiently in theory *and* in practice. Given an $\varepsilon > 0$ we can obtain by interior point methods [10] an ε-optimal solution with polynomially many iterations, where each iteration takes polynomially many real number operations (provided that both (PSDP) and (DSDP) have non-empty interiors of feasible sets and we have good initial points).

There exist several general purpose open source packages (cf. SeDuMi , SDPA , SDPT3) which can efficiently find ε-optimal solutions. If the problem is of medium size (i.e., $s \leq 1000$ and $m \leq 10.000$), these packages are based on interior point methods, while packages for larger semidefinite programs use some variant of the first order methods. Nevertheless, once $s \geq 3000$ or $m \geq 250000$, the problem must share some special property, otherwise state-of-the-art solvers will fail to solve it for complexity reasons.

3 COMPUTATIONAL ALGORITHM

In this section we discuss an algorithm based on the Gram matrix method for testing the membership in Θ^2. The algorithm with the aid of semidefinite programming is presented in Section 3.2.

3.1 The tracial Gram matrix method

Testing whether a given $f \in \mathbb{R}\langle \underline{X} \rangle$ is an element of Σ^2 or Θ^2 can be done efficiently by using semidefinite programming as first observed in [7, Section 3]. The method behind it is a variant of the *Gram matrix method* and is based on the following proposition, which is a natural extension of the results for sums of Hermitian squares (cf. [5, Section 2.2] or [6, Theorem 3.1 and Algorithm 1]), which are in turn variants of the classical result for polynomials in commuting variables due to Choi, Lam and Reznick ([2, Section 2]; see also [11]).

Proposition 3.1 *Let W be the vector of all words $w \in \langle \underline{X} \rangle$ satisfying $2\deg(w) \leq \deg(f)$, where $f \in \mathbb{R}\langle \underline{X} \rangle$. Then*

(a) $f \in \Sigma^2$ if and only if there exists a positive semidefinite matrix G such that

$$f = W^* G W; \tag{2}$$

(b) $f \in \Theta^2$ if and only if there exists a positive semidefinite matrix G such that

$$f \overset{\text{cyc}}{\sim} W^* G W; \tag{3}$$

Moreover, given a positive semidefinite matrix G of rank r satisfying (2) or (3), respectively, one can construct nc polynomials $g_1, \ldots, g_r \in \mathbb{R}\langle \underline{X} \rangle$ such that $f = \sum_{i=1}^r g_i^* g_i$ or $f \overset{\text{cyc}}{\sim} \sum_{i=1}^r g_i^* g_i$, respectively.

The matrix G satisfying (2) is called *a Gram matrix* for f, while matrix G satisfying (3) is called *a* (*tracial*) *Gram matrix* for f.

For an nc polynomial $f \in \Sigma^2$ or $f \in \mathbb{R}\langle \underline{X} \rangle$ the (tracial) Gram matrix is *not* unique, hence determining whether $f \in \Sigma^2$ or $f \in \Theta^2$ amounts to finding *a* positive semidefinite Gram matrix from the affine set of all (tracial) Gram matrices for f. Problems like this can (in theory) be solved *exactly* using quantifier elimination. However, this only works for problems of small size, so a *numerical* approach is needed in practice. Thus we turn to semidefinite programming.

3.2 Sums of Hermitian squares and commutators and SDP

In this subsection we present a *conceptual algorithm* based on SDP for checking whether an nc polynomial $f = \sum_{w \in \mathbb{R}\langle \underline{X} \rangle} a_w w$ of degree $\leq 2d$ is cyclically equivalent to a sum of Hermitian squares. Following Proposition 3.1 we must determine whether there exists a positive semidefinite matrix G such that $f \overset{\text{cyc}}{\sim} W^* G W$. This is a semidefinite feasibility problem in the matrix variable G, where the constraints $\langle A_i, G\rangle = b_i$ are essentially equations (1). Note that since $w^* \overset{\text{cyc}}{\not\sim} w$ in general, these constraints need not be symmetric. As we restrict our attention to polynomials which are cyclically equivalent to symmetric polynomials (the others are clearly not in Θ^2), we may always merge the equations corresponding to a particular word and its involution. We formalize this lesson as follows:

Proposition 3.2 *If $f = \sum_w a_w w \in \Theta^2$ then for every $v \in \langle \underline{X}\rangle$,*

$$\sum_{w \overset{\text{cyc}}{\sim} v} a_w = \sum_{w \overset{\text{cyc}}{\sim} v^*} a_w. \tag{4}$$

Corollary 3.3 *Given $f \in \mathbb{R}\langle \underline{X}\rangle$ we have:*

(1) *if f does not satisfy* (4), *then $f \notin \Theta^2$;*

(2) *if f satisfies* (4), *then we can determine whether $f \in \Theta^2$ by solving the following SDP with only symmetric constraints:*

$$\begin{array}{lrcl} \inf & \langle C, G\rangle & & \\ \text{s. t.} & \sum\limits_{\substack{p,q,\ p^*q \overset{\text{cyc}}{\sim} v \\ \vee\ p^*q \overset{\text{cyc}}{\sim} v^*}} G_{p,q} & = & \sum\limits_{w \overset{\text{cyc}}{\sim} v} (a_w + a_{w^*}), \quad \forall v \in W \\ & G & \succeq & 0. \end{array} \tag{$\text{CSOHS}_{\text{SDP}}$}$$

The constraints in ($\text{CSOHS}_{\text{SDP}}$) are $\langle A_v, G\rangle = b_v$, where $b_v = \sum_{w \overset{\text{cyc}}{\sim} v}(a_w + a_{w^*})$ and $A_v = A_{v^*}$ is the symmetric matrix defined by

$$(A_v)_{p,q} = \begin{cases} 2; & \text{if } p^*q \overset{\text{cyc}}{\sim} v \ \&\ p^*q \overset{\text{cyc}}{\sim} v^*, \\ 1; & \text{if } p^*q \overset{\text{cyc}}{\sim} v \ \&\ p^*q \overset{\text{cyc}}{\not\sim} v^*, \\ 0; & \text{otherwise.} \end{cases}$$

The conceptual algorithm to determine whether a given polynomial is cyclically equivalent to a sum of Hermitian squares (the *Gram matrix method*) is now as follows:

INPUT: *$f \in \mathbb{R}\langle \underline{X}\rangle$ with $f = \sum_{w \in \langle \underline{X}\rangle} a_w w$, where $a_w \in \mathbb{R}$.*

STEP 1: *If f does not satisfy* (4), *then $f \notin \Theta^2$.* **Stop.**

STEP 2: *Construct W.*

STEP 3: *Construct data A_v, b_v, C corresponding to* ($\text{CSOHS}_{\text{SDP}}$).

STEP 4: *Solve* ($\text{CSOHS}_{\text{SDP}}$) *to obtain G. If it is not feasible, then $f \notin \Theta^2$.* **Stop.**

STEP 5: *Compute a decomposition $G = R^t R$.*

OUTPUT: *Sum of Hermitian squares cyclically equivalent to f: $f \overset{\text{cyc}}{\sim} \sum_i g_i^* g_i$, where g_i denotes the i-th component of RW.*

Algorithm 1: The Gram matrix method for finding Θ^2-certificates.

Example 3.4 *We conclude this presentation considering the polynomial*

$$f = S_{8,2}(X,Y) = X^6Y^2 + X^5YXY + X^5Y^2X + X^4YX^2Y + X^4YXYX + X^4Y^2X^2 + \\ + X^3YX^3Y + X^3YX^2YX + X^3YXYX^2 + X^3Y^2X^3 + X^2YX^4Y + X^2YX^3YX + \\ + X^2YX^2YX^2 + X^2YXYX^3 + X^2Y^2X^4 + XYX^5Y + XYX^4YX + XYX^3YX^2 + \\ + XYX^2YX^3 + XYXYX^4 + XY^2X^5 + YX^6Y + YX^5YX + YX^4YX^2 + \\ + YX^3YX^3 + YX^2YX^4 + YXYX^5 + Y^2X^6.$$

To prove that $f \in \Theta^2$ with the aid of `NCSOStools`, *proceed as follows:*

(1) *Define two noncommuting variables:*

```
>> NCvars x y
```

(2) *For a numerical test whether $f \in \Theta^2$, run*

```
>> params.obj = 0;
>> [IsCycEq,G,W,sohs,g,SDP_data] = NCcycSos(BMV(8,2), params);
```

where our polynomial f is constructed using `BMV(8,2)`. *This yields a* floating point *Gram matrix G*

$$G = \begin{bmatrix} 3.9135 & 2.0912 & -0.1590 & 0.9430 \\ 2.0912 & 4.4341 & 1.0570 & -0.1298 \\ -0.1590 & 1.0570 & 4.1435 & 1.9088 \\ 0.9430 & -0.1298 & 1.9088 & 4.0865 \end{bmatrix}$$

for the word vector

$$W = \begin{bmatrix} X^3Y & X^2YX & XYX^2 & YX^3 \end{bmatrix}^t.$$

The rest of the output: `IsCycEq` $= 1$ *since f is (numerically) an element of Θ^2;* `sohs` *is a vector of nc polynomials g_i with $f \overset{\text{cyc}}{\sim} \sum_i g_i^* g_i =$* `g`; `SDP_data` *is the SDP data for* (CSOHS$_{\text{SDP}}$) *constructed from f.*

(3) *To round and project the obtained floating point solution* `G`, *feed* `G` *and* `SDP_data` *into* `RprojRldlt`*:*

```
>> [Grat,L,D,P,err]=RprojRldlt(G,SDP_data,true)
```

This produces a rational Gram matrix `Grat` *for f with respect to W and its LDU decomposition $PLDL^tP^t$, where P is a permutation matrix, L lower unitriangular, and D a diagonal matrix with positive entries. Finally, the obtained rational sum of Hermitian squares certificate for $f = S_{8,2}(X,Y)$ is*

$$f \overset{\text{cyc}}{\sim} \sum_{i=1}^{4} \lambda_i g_i^* g_i$$

for

$$\begin{aligned} g_1 &= X^3Y + \frac{1}{2}X^2YX + \frac{1}{4}YX^3 \\ g_2 &= X^2YX + \frac{1}{3}XYX^2 - \frac{1}{6}YX^3 \\ g_3 &= XYX^2 + \frac{13}{22}YX^3 \\ g_4 &= YX^3 \end{aligned}$$

and

$$\lambda_1 = 4, \quad \lambda_2 = 3, \quad \lambda_3 = \frac{11}{3}, \quad \lambda_4 = \frac{105}{44}.$$

4 Conclusions

In this paper we consider polynomials in noncommuting variables which can be decomposed into a sum of Hermitian squares and commutators. Such nc polynomials are cyclically equivalent to a sum of Hermitian squares and are trace-positive. In the first part of the paper we present a systematic way to find such a decomposition of a given nc polynomial using computer algebra system. The main part of the method, which is a variant of the classical Gram matrix method, is a construction of semidefinite programming feasibility problem. We also present some examples illustrating our results. We conclude with a demonstration of how to use the proposed algorithm for practical problem solving implemented in the computer algebra system `NCSOStools`.

References

[1] K. Cafuta, I. Klep, and J. Povh. NCSOStools: a computer algebra system for symbolic and numerical computation with noncommutative polynomials. *Optim. Methods Softw.*, 26(3):363–380, 2011. `http://ncsostools.fis.unm.si/`.

[2] M. Choi, T. Lam, and B. Reznick. Sums of squares of real polynomials. In *K-theory and algebraic geometry: connections with quadratic forms and division algebras (Santa Barbara, CA, 1992)*, volume 58 of *Proc. Sympos. Pure Math.*, pages 103–126. Amer. Math. Soc., Providence, RI, 1995.

[3] M. de Oliveira, J. Helton, S. McCullough, and M. Putinar. Engineering systems and free semi-algebraic geometry. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 17–61. Springer, New York, 2008.

[4] M. Halická, E. de Klerk, and C. Roos. On the convergence of the central path in semidefinite optimization. *SIAM J. Optim.*, 12(4):1090–1099, 2002.

[5] J. Helton. "Positive" noncommutative polynomials are sums of squares. *Ann. of Math. (2)*, 156(2):675–694, 2002.

[6] I. Klep and J. Povh. Semidefinite programming and sums of hermitian squares of noncommutative polynomials. *J. Pure Appl. Algebra*, 214:740–749, 2010.

[7] I. Klep and M. Schweighofer. Sums of Hermitian squares and the BMV conjecture. *J. Stat. Phys*, 133(4):739–760, 2008.

[8] J. B. Lassere. *Moments, positive polynomials and their applications*, volume 1 of *Imperial College Press Optimization Series*. Imperial College Press, London, 2009.

[9] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01.

[10] 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.

[11] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003.

[12] P. Parrilo and B. Sturmfels. Minimizing polynomial functions. In *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, volume 60 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 83–99. Amer. Math. Soc., Providence, RI, 2003.

[13] S. Pironio, M. Navascués, and A. Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM J. Optim.*, 20(5):2157–2180, 2010.

On semidefinite programming based heuristics for the graph coloring problem

Igor Đukanović*, Jelena Govorčin†, Nebojša Gvozdenović‡1, Janez Povh†2

* Univerza v Mariboru, Ekonomsko-poslovna fakulteta, Razlagova 14, 2000 Maribor
† Fakulteta za informacijske študije v Novem mestu, Novi trg 5, 8000 Novo mesto, Slovenia
‡ Univerzitet u Novom Sadu, Ekonomski fakultet Subotica, Segedinski put 9-11, 24000 Subotica, Serbia

email: igor.mat@uni-mb.si,jelena.govorcin@fis.unm.si, nebojsa.gvozdenovic@gmail.com,janez.povh@fis.unm.si

Abstract: The Lovász theta number is a well-known lower bound on the chromatic number of a graph G, and $\Psi_K(G)$ is its impressive strengthening. We apply semidefinite programming formulation of both functions to obtain suboptimal (matrix) solutions in a polynomial time. These matrices carry valuable information on how to color the graph. The resulting graph coloring heuristics utilizing these two functions are compared on medium sized graphs.

Keywords: semidefinite programming, graph coloring problem, boundary point method

Math. Subj. Class. (2010): Primary 90C22, 13J30; Secondary 14P10, 47A57, 08B20

1 Introduction

Let $G(V,E)$ be a *simple* (i.e. loopless and undirected) graph, where V denotes the set of vertices of G, and E the set of edges of G. Since we represent graphs by matrices, we assume that *vertex set* is $V = \{1, 2, ..., n\}$. Symmetric matrix X is positive semidefinite (denoted by $X \succeq 0$), if all its eigenvalues are non-negative.

Map $c : V \to \{1, ..., C\}$ such that

$$ij \in E \Rightarrow c(i) \neq c(j) \tag{1}$$

is a graph C-coloring. Subset $S \subseteq V$ is *stable*, if no pair of vertices in S is connected by an edge, i.e. $i, j \in S \Rightarrow ij \notin E$. Vertices colored by the same color i therefore form a stable set $c^{-1}(i)$. So graph C-coloring forms a partition of the vertex set into C stable sets, and the *graph coloring problem* is to find such a partition with the smallest cardinality $\chi(G)$.

2 Some formulations of $\chi(G)$ and their SDP relaxations

Given a simple graph $G(V,E)$, $V = \{1, ..., n\}$, and a C-coloring c we can assign to each stable set $c^{-1}(i)$, $1 \leq i \leq C$, its *characteristic* vector $x^i \in \{0,1\}^n$, defined coordinate-wise by

$$x^i_j = \begin{cases} 1, & \text{if } c(j) = i, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The *coloring matrix* $X := \sum_i x^i (x^i)^\top$ has the following obvious properties

[1]Supported by the Serbian Ministry of Education and Science (projects III 44006 and OI 174018)
[2]Supported by the Slovenian Research Agency (bilateral project no. BI-SR/10-11-040).

$$
\begin{aligned}
x_{ij} &\in \{0,1\} \text{ for all } i,j \in V, && (3)\\
x_{ii} &= 1 \text{ for all } i \in V, && (4)\\
x_{ij} &= 0 \text{ for all } ij \in E, && (5)\\
\tilde{X} &\succeq 0, && (6)
\end{aligned}
$$

and $\operatorname{rank}(X) = C$. In fact any $X \succeq 0$ satisfying (3)-(5) represents a legal coloring[3].

Following standard homogenization procedure ([7, 5, 9] we rather add to the matrix X the *zeroeth* row and column (which do not represent any vertex), and study the matrices

$$
\tilde{X} = \begin{bmatrix} C & e^\top \\ e & X \end{bmatrix} := \sum_{i=1}^{C} \begin{bmatrix} 1 \\ x^i \end{bmatrix} \cdot \begin{bmatrix} 1 \\ x^i \end{bmatrix}^\top \tag{7}
$$

where e denotes the vector of all ones.

Obviously $\tilde{X} \succeq 0$. In fact, it is easy to see [3] that

$$
C = \operatorname{rank}(\tilde{X}) = \min\{t | \begin{bmatrix} t & e^\top \\ e & X \end{bmatrix} \succeq 0,\ X \text{ feasible for } (3) - (5)\}. \tag{8}
$$

Therefore

$$
\begin{aligned}
\chi(G) \quad &= \quad \min\ t \\
\text{s.t.} \quad & \quad \tilde{X} = \begin{bmatrix} t & e^\top \\ e & X \end{bmatrix} \succeq 0 \\
& \quad \tilde{X} \text{ feasible for } (3) - (5).
\end{aligned} \tag{GCP}
$$

Removing binary constraint (3) yields the Lovász Θ function([7, 3, 4])

$$
\begin{aligned}
\Theta(G) \quad &= \quad \min\ t \\
\text{s.t.} \quad & \quad \tilde{X} = \begin{bmatrix} t & e^\top \\ e & X \end{bmatrix} \succeq 0, \\
& \quad x_{ij} = 0 \text{ for all } ij \in E, \\
& \quad x_{ii} = 1 \text{ for all } i \in V,
\end{aligned} \tag{Θ}
$$

a well-known relaxation of the chromatic number.

Let $\tilde{X}$ be an optimal solution of (Θ). Karger, Motwani and Sudan [6] defined a heuristic with the best theoretical upper bound on the number of colors needed to legally color a graph. A pair $i, j \in V$ is likely to be colored by the same color by their heuristic only when x_{ij} is large.

A recursive variant of their heuristic performs well on medium sized graphs [3], and benefits from strengthening SDP bound (Θ) toward $\chi(G)$. Therefore in the next section we introduce $\Psi_K(G)$, an impressive strengthening of $\Theta(G)$ based on moment matrix idea[4].

3 Function $\Psi_K(G)$

Let $K \subseteq V$ be a clique, i.e. a set of pairwise connected vertices. W.l.o.g. we may assume $K = \{1, \ldots, k\}$, i.e. each pair of the first k vertices of V forms an edge, for some $k < n$. We define for the given coloring c the *moment coloring vectors* $y^i \in \{0,1\}^{1+n+k(1+n)}$, $1 \le i \le C$ based on the clique K as follows

$$
y^i := [1, x_1^i, \ldots, x_n^i, x_1^i, x_1^i x_1^i, x_1^i x_2^i, \ldots, x_1^i x_n^i, \ldots x_k^i, x_k^i x_1^i, \ldots, x_k^i x_n^i]^\top \tag{9}
$$

where x_j^i are from (2).

For easier referencing to elements of y^i we label components of y^i by monomials of order 2 in (commutative) variables $z_0 := 1, z_1, \dots, z_n$:

$$\underbrace{z_0^2, z_0 z_1, \dots, z_0 z_n}_{\text{0-th block}}, \underbrace{z_1 z_0, z_1^2, z_1 z_2, \dots, z_1 z_n}_{\text{1-st block}}, \dots, \underbrace{z_k z_0, z_k z_1, \dots, z_k z_n}_{\text{k-th block}}.$$

For any n-tuple of values $(x_1^i, \dots, x_n^i)$ we have $y^i_{z_p z_q} = y^i_{z_q z_p}$, $0 \le p, q \le n$. Vectors $[x_1^i, \dots, x_n^i]$ have actually only $0-1$ values and since we fix $z_0 = 1$ we also have the following equations for every y^i:

$$y^i_{z_p^2} = y^i_{z_0 z_p}$$

These equations imply that several components from y^i can be deleted. In numerical implementations they need to be deleted, but here we keep them as the larger matrix (10) has nicer structure making the arguments simpler.

Resembling (7) we can construct the matrix

$$Y := \sum_{i=1}^{C} y^i (y^i)^\top \succeq 0. \tag{10}$$

It has a $(k+1) \times (k+1)$ block structure

$$Y = \begin{bmatrix} Y^{00} & Y^{01} & \dots & Y^{0k} \\ Y^{10} & Y^{11} & \dots & Y^{1k} \\ \vdots & \vdots & \ddots & \vdots \\ Y^{k0} & \dots & \dots & Y^{kk} \end{bmatrix}$$

where blocks are actually indexed by $z_0, \dots, z_k$ and each block is a $(n+1) \times (n+1)$ matrix (indexed again by $z_0, \dots, z_n$). Hence y^{pq}_{st} denotes the $z_s z_t$-th element in $z_p z_q$-th block. Notice that $Y^{00} = \tilde{X}$ from (7).

The elements of the matrix Y satisfy the following obvious properties

$$\begin{aligned} y^{ij}_{pq} &\in \{0,\ 1\}\ \forall\, i, j, p, q,\ i + j + p + q \neq 0 && (11) \\ y^{ij}_{pq} &= 1, \ \text{ if all the vertices in } \{i, j, p, q\} \text{ bear the same color} && (12) \\ y^{ij}_{pq} &= 0, \ \text{ if } \{i, j, p, q\} \text{ contains an edge,} && (13) \\ y^{ij}_{pq} &= y^{i'j'}_{p'q'} \text{ if } z_i z_j z_p z_q \ = \ z_{i'} z_{j'} z_{p'} z_{q'}. && (14) \end{aligned}$$

Since the index 0 does not make a difference, we also have

$$Y^{0i} = Y^{ii},\ 1 \le i \le k. \tag{15}$$

Meanwhile the "off-diagonal" blocks $Y^{ij} = Y^{ji}$, $1 \le i < j \le k$, equal 0 as $ij \in E$ since i, j belong to the clique K.

So Y has the following block structure:

$$Y = \begin{bmatrix} Y^0 & Y^1 & Y^2 & \dots & Y^k \\ Y^1 & Y^1 & 0 & \dots & 0 \\ Y^2 & 0 & Y^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ Y^k & \dots & \dots & \dots & Y^k \end{bmatrix}$$

where $Y^i := Y^{ii} = Y^{0i}$.

Though Y is a large matrix the condition $Y \succeq 0$ is easy to check as (due to a simple manipulation by a certain unitary matrix [4])

$$Y \succeq 0 \iff Y^0 - \sum_{i=1}^{k} Y^i \succeq 0 \text{ and } Y^i \succeq 0 \text{ for each } 1 \leq i \leq k.$$

After relaxing the binary constraint (like from (GCP) to (Θ)) we define

$$\begin{array}{rcl} \Psi_K(G) & = & \min\ t \\ \text{s.t.} & & y^0_{00} = t, \quad y^0_{ii} = 1, \quad \forall i \in \{1,...,n\} \\ & & y^i_{pq} = 0, \quad \text{whenever } \{i,p,q\} \text{ contains an edge} \\ & & y^i_{pq} = y^{i'}_{p'q'}, \quad \text{whenever } z_i z_p z_q = z_{i'} z_{p'} z_{q'} \\ & & Y^0 - \sum_{i=1}^{k} Y^i \succeq 0, \quad Y^i \succeq 0 \forall i \in \{1,...,k\}. \end{array} \tag{Ψ_K}$$

Now $\Theta(G) \leq \Psi_K(G) \leq \chi(G)$.

Let $\bar{Y}$ be an optimal solution of (Ψ_K). It turns out that from the principal block $\bar{Y}^0$ one can extract better information about graph coloring than from an optimal solution $\tilde{X}$ of (Θ).

Yet another interpretation of a large element $\bar{y}^i_{pq}$, $1 \leq i \leq k$, inspired by (12) is that all three vertices i, p and q should be colored by the same color.

4 Heuristics

For a given graph $G(V, E)$ we numericaly solve semidefinite program (Θ). Let $\tilde{X}$ be the computed matrix solution, and x_{ij} its largest off-diagonal entry. Then with large probability there exists an optimal partition of V into $\chi(G)$ stable sets such that i and j belong to the same stable set (the same element of this partition). In fact typpically there exists many such optimal partitions. Therefore we want to color vertices i and j by the same color. But finding a coloring c of the graph $G(V, E)$ such that $c(i) = c(j)$ is equivalent task to finding a coloring c' of a graph $G' = G/_{\{i,j\}}$ with one vertex less. In the graph $G/_{\{i,j\}}$ vertices i and j are merged into new vertex i' where the neighborhood of i' s the union of the neighborhoods of vertices i and j. Let c' be by this algorithm recursively obtained coloring of G'. We define $c(j) := c(i) := c'(i')$ while for all other vertices $c := c'$.

This recursive heuristic (with details and fine tuning explained in [3]) is very competitive especially in terms of quality on medium sized graphs. It has been observed already in [3] that it benefits from replacing (Θ) by its strengthening toward (GCP).

So we replace $\Theta(G)$ by $\Psi_K(G)$ and likewise merge the two vertices i and j corresponding to the largest off-diagonal entry $\bar{y}^0_{ij}$ in the numerical optimal matrix solution $\bar{Y}^0$ of (Ψ_K). The later program though quite large can efficiently be solved by a very robust boundary point method [10, 8].

Additional advantage of solving (Ψ_K) is that we can also merge three vertices l, p, q corresponding to the largest off-diagonal element y^l_{pq} in the blocks $Y^l, l = 1, ..., k$, (see (12)) an approach that we plan to investigate further.

5 Preliminary numerical results

There is an extensive numerical evidence [4] that $\Psi_K(G)$ strengthens $\Theta(G)$ a lot when K is a largest or an almost largest clique. It is relatively cheap to efficiently find a suboptimal clique [1], however computational complexity increases roughly $|K|$ times when we switch from $\Theta(G)$ to $\Psi_K(G)$. As in our recursive heuristic we need to compute such SDP bound on $\chi(G)$ almost $n = |V|$ times we have investigated how the bound improves with and increase in the size of a clique. A rough demonstration is presented in the Table 1. Notice that all our results are obtained on eight random graphs $G_1, ..., G_8$ with 20, 40, 60 or 80 vertices and with edge density 0.4 or 0.6. These graphs are exactly the toughest for a competitive interior point method approach (see [2]).

graph	n	$\|E\|$	$\Theta(G)$	$\Psi_2(G)$	$\Psi_3(G)$	$\Psi_4(G)$
G_1	20	81	4.9999	4.9999	4.9999	4.9999
G_2	20	98	5.9999	5.9998	5.9999	5.9999
G_3	40	304	6.0029	6.0057	6.0372	6.1217
G_4	40	489	9.2509	9.4696	9.5133	9.5882
G_5	60	694	6.9999	6.9880	7.0772	7.1620
G_6	60	1062	10.6542	10.9544	11.0745	11.1591
G_7	80	1221	7.4595	7.7838	7.9111	8.0381

TABLE 1. Comparison of $\Theta(G)$ with $\Psi_K(G)$ for K a clique of size 2,3 and 4

These examples in Table 1 demonstrate that Ψ_K is indeed a significant strengthening of Θ. However, the improvements are consistent with general observations from discrete optimization: we have to put a lot of effort for a small strengthening of such bound.

There are also time complexity issues. Computing approximate colorings by applying Ψ_K is very time consuming, e.g. applying Ψ_4 on G_7 takes approximately 10 hours on laptop with two 2GBh processors and 2 Gb Ram.

The Table 2 compares colorings obtained by applying Θ and Ψ_K. Coloring small graphs by applying the recursive variant of Karger-Motwani-Sudan's heuristics typically finds an optimal $\chi(G)$-coloring and on medium graphs an almost optimal one [3] - a hard to bit result. By comparing the lower bound in Table 1 with the number of colors in Table 2 notice that the colorings of G_1, G_2 and G_3 are indeed optimal, while already on the graph G_7, a graph with just 80 vertices, we were able to improve the coloring impressively by using two colors less - a rare instance and hard to obtain result.

On the other hand notice an unusual result in Table 1 for the graph G_5 where for a stronger bound $\Psi_K(G_5)$ with K an edge, i.e. $|K| = 2$, we have computed its approximation which is smaller then $\Theta(G_5)$. The problem is due to the fact that for boundary point method we do not have a simple duality-gap-like stopping criterium. Likewise the numerical results for graphs G_5 and G_6 in Table 2 where stronger bounds occasionaly produced worse colorings are just a bit unusual. This kind of numerical unstability is not very unusual for this heuristic approach, and it should be possible to avoid by more carefull choice of stopping criterium.

6 Conclusions

Replacing Lovász theta function $\Theta(G)$ by its impressive strengthening $\Psi_K(G)$ can improve colorings obtained by Karger-Motwani-Sudan-motivated heuristics (see [3]) impressively

graph/bound	Θ	Ψ_2	Ψ_3	Ψ_4
G_1	5	5	5	5
G_2	6	6	6	6
G_3	7	7	7	7
G_4	11	11	11	11
G_5	9	10	10	9
G_6	13	14	13	13
G_7	12	11	11	10

TABLE 2. Number of colors used by heuristics depending on the bound Θ or Ψ_K, $|K| = 2, 3, 4$

already on quite small graphs. Tradeoff between computational complexity (size of the clique K, stopping criteria for computing Ψ_K by the boundary point method, number of merged vertices (2 or 3)) versus quality of the produced coloring need to be investigated further.

As numerical evidence shows (see G_5, G_6, G_7 in Table 2) Ψ_K can jump over an integer thus improving the Θ bound on the chromatic number by 1. Such improvement can to a much more serious extent be exploited in a branch and bound algorithm for finding an optimal coloring. Thus we plan to substantially improve the speed reported in [11] where Θ was applied in a branch and bound exact coloring algorithm.

References

[1] S. Burer, R.D.C. Monteiro, and Y. Zhang. Maximum stable set formulations and heuristics based on continuous optimization. *Math. Program. Ser. A*, 94:137—-166, 2002.

[2] I. Dukanovic and F. Rendl. Semidefinite programming relaxations for graph coloring and maximal clique problems. *Mathematical Programming*, 109:345–365, 2007.

[3] I. Dukanovic and F. Rendl. A semidefinite programming-based heuristic for graph coloring. *Discrete Applied Mathematics*, 156(2):180 – 189, 2008.

[4] N. Gvozdenović and M. Laurent. Computing semidefinite programming lower bounds for the (fractional) chromatic number via block-diagonalization. *SIAM J. Optim.*, 19(2):592–615, 2008.

[5] N. Gvozdenović, M. Laurent, and F. Vallentin. Block-diagonal semidefinite programming hierarchies for 0/1 programming. *Oper. Res. Lett.*, 37(1):27–31, 2009.

[6] D. Karger, R. Motwani, and M. Sudan. Approximate graph coloring by semidefinite programming. *Journal of the Association for Computing Machinery*, 45:246–265, 1998.

[7] L. Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:1–7, 1979.

[8] J. Malick, J. Povh, F. Rendl, and A. Wiegele. Regularization methods for semidefinite programming. *SIAM J. Optim.*, 20(1):336–356, 2009.

[9] J. Povh and F. Rendl. Copositive and semidefinite relaxations of the quadratic assignment problem. *Discrete Optim.*, 6(3):231–241, 2009.

[10] J. Povh, F. Rendl, and A. Wiegele. A boundary point method to solve semidefinite programs. *Computing*, 78(3):277–286, 2006.

[11] A. T. Wilson. *Applying the Boundary Point Method to an SDP Relaxation of the Maximum Independent Set Problem For a Branch And Bound Algorithm.* New Mexico Institute of Mining and Technology, Socorro, New Mexico, 2009.

EXPERIMENTAL COMPARISON OF BASIC AND CARDINALITY CONSTRAINED BIN PACKING PROBLEM ALGORITHMS

Maja Remic
University of Ljubljana, Faculty of Computer and Information Science
Tržaška 25, SI-1000 Ljubljana, Slovenia
maja.remic@siol.net

Gašper Žerovnik
Jožef Stefan Institute, Reactor Physics Division
Jamova 39, SI-1000 Ljubljana, Slovenia
gasper.zerovnik@ijs.si
and
Janez Žerovnik
Institute of Mathematics, Physics and Mechanics,
Jadranska 19, SI-1111 Ljubljana, and
University of Ljubljana, Faculty for Mechanical Engineering,
Aškerčeva 6, SI-1000 Ljubljana, Slovenia
janez.zerovnik@imfm.uni-lj.si

Abstract: Bin packing is an NP-hard optimization problem of packing items of given sizes into minimum number of capacity-limited bins. Besides the basic problem, numerous other variants of bin packing exist. The cardinality constrained bin packing adds an additional constraint that the number of items in a bin must not exceed a given limit N_{max}. The FFD, RFF and Zhang's bin packing problem algorithms are compared to the three cardinality constrained bin packing problem specific algorithms on random lists of items with 0%, 10%, 30% and 50% of large items. The behaviour of all algorithms when N_{max} increases is also studied. Results show that all three specific algorithms outperform the general algorithms on lists with low percentage of big items. One of the specific algorithms performs better or equally even on lists with high percentage of big items and is therefore of significant interest. The behaviour when N_{max} increases shows that all three specific algorithms can be used for solving the general bin packing problem as well.

Keywords: cardinality constrained bin packing problem, approximation algorithms, comparison, FFD, RFF, Zhang's algorithm.

1 INTRODUCTION

The cardinality constrained bin packing problem can be described as follows: A list I of n items with specified size (or weight/volume/etc.) x_i has to be arranged into bins of limited capacity C_{max} and maximum number of item per bin (cardinality constraint) N_{max} to minimize the number of bins m used. The minimization problem has been shown to be NP-hard [8].

The cardinality constrained bin packing problem can, for example, be applied to optimization of the spent nuclear fuel deposition in deep repository [7]. Swedish concept of deep repository in hard rock [3] is currently seriously regarded in Slovenia as an option for nuclear power plant Krško decommissioning program [6]. Motivated by this application, several heuristics for the cardinality constrained bin packing problem with N_{max} =4 were designed in [8]. The new heuristics were compared against an obvious adaptation of the first fit decreasing (FFD) algorithm and were proven to clearly outperform the FFD algorithm on the datasets of interest. It is well known that FFD algorithm is a good approximation algorithm for the general bin packing problem (more precisely, it is well known [2] that FFD always gives a solution with at most 11/9 $OPT(I)$ + C bins, where $OPT(I)$ stands for the value of optimal solution). Therefore, it is natural to ask how the generalized versions of the new algorithms behave on the bin packing problem with arbitrary cardinality constraints. A preliminary experimental study is presented in this report which shows that the obvious adaptations of the new algorithms are competitive also on the general cardinality constrained bin packing problem.

2 ALGORITHM ADJUSTMENTS

In this paper, three algorithms (referred to as Algorithm 1, 2, and 3) from Ref. [8] (see Appendix) are experimentally compared to First Fit Decreasing (FFD) [1], Refined First Fit (RFF) [4] and Zhang [5] algorithms. Since the latter three are designed for the original bin packing problem, the following modifications had to be performed in order to adapt to the additional cardinality constraint:

- For any item, the FFD algorithm chooses the first bin with both enough space (capacity constraint) and at least 1 empty slot (cardinality constraint).
- Similarly, the RFF algorithm places each item in the first possible bin of the suitable group.
- Zhang algorithm closes the (active or additional) bin also when the cardinality constraint is reached.

All input data for all algorithms have been sorted by size, even though for RFF it is not necessary.

3 EXPERIMENTAL COMPARISON OF ALGORITHMS

First, the quality of solutions, obtained by different algorithms, was experimentally compared for different input data distributions. The main purpose of this investigation was to asses the performance of algorithms from [8], specifically designed for the cardinality constrained bin packing, relative to the (suitably adapted) algorithms for the general bin packing problem. Similar experiment was already presented in [8], with two significant differences. In this paper, RFF and Zhang algorithms were added for the comparison. Furthermore, in the present experiments additional instances with different fractions of large items were regarded.

3.1 Experiments with fixed $N_{max} = 4$

As in [8], default values of $N_{max} = 4$ items per bin and capacity $C_{max} = 1$ were adopted. Four different types of input data of length $n = 100$ were used:

- lists without large items (i.e. items, larger than $C_{max}/2$),
- lists with 10% large items,
- lists with 30% large items, and
- lists with 50% large items.

All input lists were generated by default Octave random number generator, using uniform distribution with compositions given above. The input data distributions are shown on Fig. 1. The experiment has been repeated 10000 times with input data generated from different random seeds.

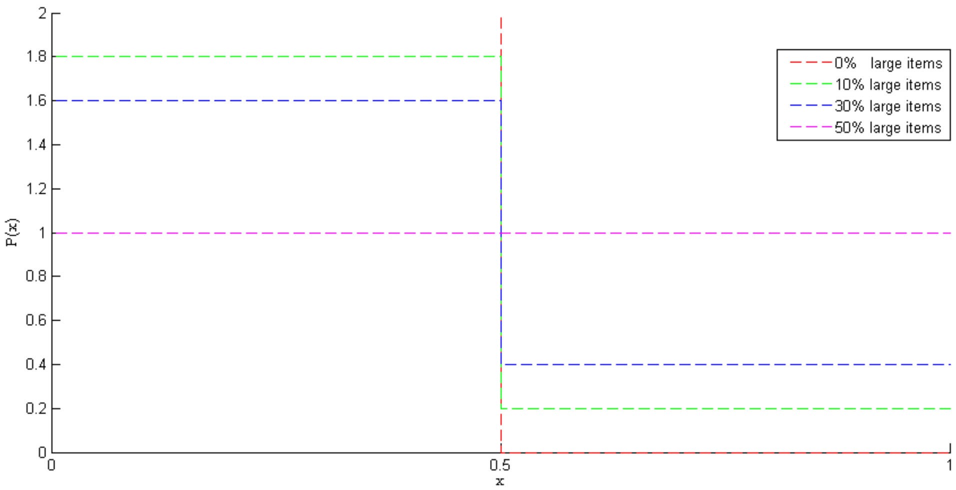


Figure 1: Probability density *P(x)* over item size *x* for different types of input data.

The theoretical lower bound *MIN(I)* for the solutions of cardinality constrained bin packing problem can be expressed as:

$$MIN(I) = \max\left(\left\lceil \frac{\sum_{i=1}^{n} x_i}{C_{\max}} \right\rceil, \left\lceil \frac{n}{N_{\max}} \right\rceil\right), \tag{1}$$

which is a useful and conservative approximation for the optimal solution $OPT(I) \geq MIN(I)$. The quality of the obtained solution *A(I)* is measured by $A(I)/OPT(I)$ and can be estimated by $A(I)/MIN(I)$. The values for all algorithms, averaged over 10000 instances, are presented in Table 1.

Table 1: Average values of the *A(I) / MIN(I)* ratio for selected algorithms at different lists.

A(I) / MIN(I)	***0% large items***	***10% large items***	***30% large items***	***50% large items***
FFD	1.1273 ± 0.0033	1.0933 ± 0.0036	**1.0348 ± 0.0018**	**1.0418 ± 0.0018**
RFF	1.3318 ± 0.0041	1.3199 ± 0.0045	1.3251 ± 0.0038	1.3323 ± 0.0034
Zhang	1.3606 ± 0.0034	1.3530 ± 0.0040	1.2680 ± 0.0026	1.2581 ± 0.0024
Algorithm 1	**1.0262 ± 0.0026**	1.0698 ± 0.0024	1.1421 ± 0.0034	1.2059 ± 0.0038
Algorithm 2	1.0493 ± 0.0039	**1.0552 ± 0.0030**	**1.0504 ± 0.0019**	**1.0440 ± 0.0017**
Algorithm 3	**1.0256 ± 0.0024**	**1.0571 ± 0.0022**	1.0992 ± 0.0022	1.1434 ± 0.0024

The results clearly show the influence of the fraction of large items on the quality of the solution for individual algorithms. The Zhang and FFD algorithms work much better with significant fractions of large items, while the opposite can be stated for Algorithm 1 and 3. For RFF and Algorithm 2, the sensitivity of the quality of solutions on large item fractions is insignificant.

Direct comparison between algorithms shows that Zhang and RFF in general give much worse solutions than FFD. For RFF that kind of behavior is expected since it is (in contrast to other five algorithms) basically an online algorithm. Worse performance of the Zhang algorithm is probably due to linear time complexity. All three algorithms from [8] perform significantly better than others for small (up to 10%) fractions of large items, whereas for 50% of large items, only Algorithm 2 is comparable to FFD. This is expected since Algorithm 2 was designed for input data with significant fraction of large items [8] while Algorithms 1 and 3 were designed for input data with (almost) no large items. Let us mention at this point that Algorithm 3 is random number based [8], therefore its solution may be improved significantly when taking advantage of multi-start mode.

3.2 Increasing the cardinality constraint

The purpose of the second experiment was to examine the behaviour of the algorithms when increasing the cardinality constraint N_{max}. When N_{max} is converged to the total number of items n, the cardinality constraint becomes irrelevant, consequently the cardinality constrained bin packing problem converges to the original bin packing problem in this limit.

In this experiment, list length $n = 100$ and capacity $C_{max} = 1$ was used throughout. Two different types of input data lists were used: lists without large items and with 50% large items (Fig. 1), which were generated in exactly the same way as in the first experiment. The cardinality constraint was changed between $N_{max} = 2$ and $N_{max} = n/4$ since the latter was experimentally observed to be enough to achieve the convergence. The experiment was repeated 1000 times with different input data, each generated with different random number sequence. The results are shown in Figs. 2 and 3.

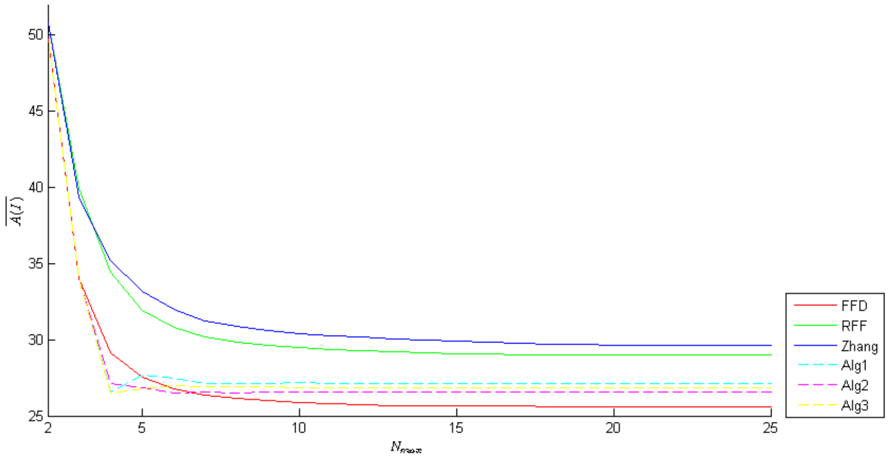


Figure 2: Average number of bins as a function of the cardinality constraint for lists without large items.

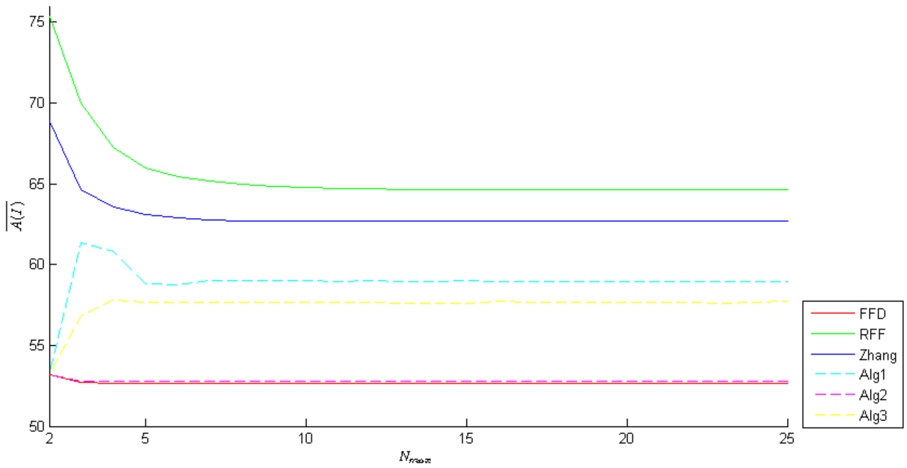


Figure 3: Average number of bins as a function of the cardinality constraint for lists with 50% of large items.

Expectedly, the number of bins used in most cases decreases with the increase of N_{max}, up to the point where N_{max} no longer has impact and the capacity C_{max} of the bin becomes the limiting factor in Eq. (1). The speed of convergence depends on the percentage of large items in the list. When the percentage is high, the impact of N_{max} is smaller, meaning faster convergence. This is evident in lists with 50% large items (Fig. 3).

For lists without large items, all three algorithms for solving the cardinality constrained bin packing problem perform best at $N_{max} = 4$, which is the setting the algorithms were developed for, and deteriorate slightly as N_{max} increases. Similarly, for lists with 50% large items, both Algorithm 1 and 3 experience unexpected swings at $N_{max} < 5$. This behaviour can be explained by the functioning of those two algorithms, since at small N_{max}, larger items are considered first.

It is notable that the increasing of N_{max} has little influence over the final number of bins used by each algorithm. Even at the non-limiting N_{max}, the algorithms for solving the general bin packing problem find no better solutions than the specific algorithms, the FFD algorithm being the only exception and performing slightly better on lists without large items.

4 CONCLUSION

As the experiments have shown, all three (cardinality constrained bin packing problem) specific algorithms perform better compared to the general (bin packing) algorithms on lists without large items or with a low percentage of large items, while having a similar time complexity. In such cases, the use of specific algorithms is recommended. In case of input data with significant fractions of large items, the general FFD algorithm perform best, closely followed by the specific Algorithm 2.

All specific algorithms can be generalized to the basic bin packing problem by sufficiently increasing the cardinality constraint. Surprisingly or not, the quality of solutions is hardly compromised. The best performer is again Algorithm 2, therefore it can be regarded as a good alternative to the FFD algorithm even for the basic bin packing problem.

References

[1] Coffman, E. G., Garey, M. R., Johnson, D. S., 1997. Approximation algorithms for bin-packing: A survey. In: Approximation Algorithms for NP-Hard Problems, PWS, Boston, pp. 46-93.

[2] Korte B, Vygen J., 2000. Combinatorial Optimization, Theory and Algorithms. Springer-Verlag: Berlin.

[3] Milnes AG., 2002. Guide to the documentation of 25 year of geoscientific research (1976-2000). TR-02-18, VBB Anlaggning AB, Swedish Nuclear Fuel and Waste Management Co.

[4] Yao, A. C., 1980. New algorithms for bin packing. Journal of ACM, vol. 2, issue 27, 207-227.

[5] Zhang, G. Cai, X., Wong, C. K., 2000. Linear-time-approximation algorithms for bin packing. Operation Research Letters, vol. 5, issue 26, pp. 217-222.

[6] Železnik N, Mele I, T. Jenko T, Lokner V, Levanat I, Rapić A., 2004. Program razgradnje NEK in odlaganja NSRAO in IJG (Program of NPP Krško Decommissioning and SF \& LILW Disposal). ARAO-T-1123-03, ARAO (Agency for Radwaste Management), Ljubljana, Slovenia, and APO (Agency for Hazardous Waste), Zagreb, Croatia.

[7] Žerovnik, G., Snoj, L., Ravnik, M., 2009. Optimization of Spent Nuclear Fuel Filling in Canisters for Deep Repository. Nuclear Science and Engineering, vol. 163, issue 2, pp. 183-190.

[8] Žerovnik, G., Žerovnik, J., 2010. Constructive heuristics for the canister filling problem. Central European Journal of Operations Research, DOI:10.1007/s10100-010-0164-5.

5 APPENDIX

While FFD, RFD and Zhang's algorithm are well known, the Algorithms 1, 2 and 3 first appear in a very recent publication [5], and therefore we give a brief outline here.

Generic Algorithm
quicksort X;
$m := 0$;
while $X \neq \emptyset$ **do begin**
    set aux := $\emptyset$; $C := C_{max}$;
    for k: = K **down to** 1 **do begin**
        x := CHOOSE(X); (*)
        if x is defined **then** $X := X \setminus \{x\}$; $C := C - x$; aux := aux $\cup$ $\{x\}$ **endif**;
    endfor;
    $m := m + 1$;
    R_n = aux;
endwhile;
return $R = \{R_1, R_2, \ldots, R_n\}$.

The algorithms differ only in implementation of the function CHOOSE.

CHOOSE in **Algorithm1**:

CHOOSE(X) := x, the largest element x of X that satisfies $x \leq C/k$;

CHOOSE in **Algorithm2**:

if $k = K$ **then** CHOOSE(X) := x, the largest element of X that satisfies $x \leq$ C;
if $k < K$ **then** CHOOSE(X) := x, the largest element of X that satisfies $x \leq C/k$;

CHOOSE in **Algorithm3**:

if $k >$ Floor(K/2) **then** CHOOSE(X) := x, a random element of X that satisfies $x \leq C$;
if $k \leq$ Floor(K/2) **then** CHOOSE(X) := x, the largest element of X that satisfies $x \leq C/k$;

Remark: Clearly if CHOOSE is simply

CHOOSE(X) := x, where x is largest element of X that satisfies $x \leq C$;

we have the algorithm which puts each element in the first bin in which there is enough room. If K is large enough, this variant is equivalent to the FFD Algorithm.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section V:

Multiple Criteria Decision Making

INVESTIGATION OF THE AGGREGATION-DISAGGREGATION APPROACH TO MULTI-CRITERIA NEGOTIATIONS: CONSOLIDATION OF SIMULATION AND CASE STUDIES

Andrej Bregar
Informatika d.d., Vetrinjska ulica 2, 2000 Maribor
andrej.bregar@informatika.si

Abstract: The characteristics of a state-of-the-art outranking relation and dichotomic sorting based aggregation-disaggregation procedure for multi-criteria negotiations are uniformly analysed with two research methods – a simulation study and a case study. Results are compared according to a standard holistic model for the evaluation of group decision-making methods and systems, which consists of several key quality factors, such as ability to direct the decision-making process, credibility of the decision, robustness, and cognitive complexity. Both studies consistently prove that the observed approach to automated negotiations is efficient.

Keywords: Multi-criteria decision analysis, Decision support systems, Negotiations, Group decision-making, Consensus seeking, Preference aggregation and disaggregation, Outranking relation, Sorting, Simulation experiments, Case studies, Comparative studies

1 INTRODUCTION

To aid conflicting and opposing decision-makers in reaching consensual or compromise solutions, many methods and decision support systems for negotiations and group decision-making have been introduced [10]. They are based on various preference models. Among the most widely used are the approaches that capture preferential information in the form of pseudo-criteria to construct outranking relations on pairs of alternatives. These methods and systems include group PROMETHEE [8], ELECTRE TRI for groups [6], ELECTRE-GD [11] and the collective preorder inference method [9]. On the other end, state-of-the-art approaches to negotiations and group decision-making aim at improving human-computer interaction and efficiency of outcomes by applying the aggregation-disaggregation analysis [13]. They are, however, mainly based on the utility function [12] and oriented graphs [7]. Recently, two methods that combine the principles of group decision-making, outranking and aggregation-disaggregation have been introduced. The interactive dichotomic sorting procedure for multi-agent consensus seeking is the result of our research work [4], while the group version of the IRIS system has been developed at the LAMSADE laboratory [5].

It is necessary to determine the characteristics and efficiency of the outranking relation based aggregation-disaggregation methods that are applied to multi-criteria negotiations and group decision-making processes, especially in the context of convergence, autonomous guidance, conflict resolution, robustness and cognitive complexity. For this purpose, our consensus seeking procedure has been evaluated with a simulation study [1] and tested with a case study [3]. However, the results of the latter have neither been assessed according to a formal evaluation model nor systematically compared with the results of the former. The goal of the paper is thus to:

1. uniformly analyse the results of the case study and the simulation study with regard to a holistic model for the evaluation of group decision-making methods/systems [2];
2. compare the results of both studies, and consequently determine, if they expose any divergence in the characteristics of the consensus seeking procedure.

The rest of the paper is organized as follows. Section 2 provides a brief description of the evaluated method. Section 3 presents the simulation model, while Section 4 summarizes the case study. In Section 5, experimental and case based results are presented, analysed and compared. Finally, Section 6 gives a resume and some directions for further work.

2 DICHOTOMIC SORTING PROCEDURE FOR GROUP DECISION ANALYSIS

The procedure is based on the ELECTRE TRI outranking method [14], which is slightly modified so that preferences are modelled in a symmetrically-asymmetrical manner in the neighbourhood of the reference profile *b*. The purpose of the profile is to divide the set of alternatives into two exclusive categories – all acceptable choices are sorted into the positive class C^+, while unsatisfactory ones are the members of the negative class C^-. The decision-maker has to provide six preferential parameters for each criterion x_j, including the value of the profile $g_j(b)$, the importance weight w_j, and the thresholds of preference (p_j), indifference (q_j), discordance (u_j) and veto (v_j). Additionally, he can also specify the upper and lower allowed limits of these parameters, which constrain their automatic adjustment in the process of unification with the common opinion of the group. In order to reduce the cognitive load and enable a rational convergence of individual judgements towards the consensual solution, several mechanisms are applied:

- Preferences may be specified with fuzzy variables or by holistic assessments.
- The most discordant decision-maker who has to conform to the collective opinion is identified by computing the consensus and agreement degrees.
- Several robustness metrics reveal if preferences of an individual are firmly stated.
- The centralized agent negotiation architecture and protocol eliminate the need for a human moderator and minimize the activity of each decision-maker.
- An optimization algorithm is implemented for the purpose of automatic preference unification, which is based on the inference of parameter values.

3 SIMULATION STUDY

In order to evaluate the efficiency of the aggregation-disaggregation approach to group consensus seeking with a simulation study, several independent variables have been defined:

- Number of criteria may be $n \in \{4, 7, 10\}$. Number of observed alternatives is fixed to $m = 8$ because only the $m : n$ ratio is significant. Thus, three fundamental situations are considered: $n > m$, $n \approx m$ and $n < m$.
- Number of decision-makers may also be $o \in \{4, 7, 10\}$. Situations with less than four decision-makers are irrelevant because the experimental study focuses on multilateral rather than bilateral negotiations.
- Preferential parameters of the decision model are obtained in the following way:
  - The referential profile is sampled from the normal distribution N(50, 15).
  - Thresholds are calculated relatively to the profile. The deviations are sampled from the set of real values {0.2, 0.4, 0.6, 0.8, 1} corresponding to linguistic modifiers {very weak, weak, moderate, strong, very strong}.
  - Criteria importance weights are sampled from the uniform distribution on the [0, 1] interval. Afterwards, they are normalized to sum to 1.
  - Criteria-wise values of alternatives are sampled from the uniform distribution on the [0, 100] interval.
  - The cut-level is $\lambda = 0.5$, which is the most common value.
  - The sensitivity threshold is set to $\psi = 0.3$.

An experimental combination is hence determined by the $<n, o>$ pair. For each of nine possible combinations, 20 samples have been generated. This number of simulation cases is sufficient to obtain statistically significant results, and is small enough to cope with a high complexity of applied nonlinear optimization algorithms. In the original simulation study, four dependent variables have been observed: (1.) ability to reach a compromise, (2.) ability of autonomous guidance and conflict resolution, (3.) convergence of opinions, and (4.) robustness of the consensual decision. For the purpose of consolidation with the case study, additional dependent variables have been introduced according to the standard evaluation model, as is thoroughly explained in Section 5. Consequently, post-processing and additional synthesis of experimental data have been performed.

4 CASE STUDY

A case from the information and communications technology domain is studied. It deals with the selection of a software development subcontractor. Five quantitative criteria are considered: payment, time required for the realization of the project, number of successfully finished projects in the past, experience with similar applications, and availability of suitable technologies. There are eight alternatives (subcontractors) available. To hide their identity, symbols a_i are used. Similarly, six non-autocratic decision-makers in the group are denoted with DM_k. They have agreed on the values of two technical parameters. The cut-level, which determines the outranking relation between two alternatives, is set to $\lambda = 0.5$. The robustness threshold, which disallows an alternative to be sorted into a different class, is set to $\psi = 0.3$. Four group members have directly specified the initial values and lower/upper limits of the profile, thresholds and weights. Two have reduced their cognitive load in the following way:

- They have set the initial values of the reference profile only. By disregarding the allowed limits, they have forced the optimization algorithm to search the entire space of criteria values.
- They have uniformly specified the initial values and limits of all thresholds.
- DM_2 has defined the limits of threshold intervals, while the initial values have been automatically calculated as centre points.
- DM_4 has specified all threshold values and limitations relatively to the profile with linguistic modifiers from the set {very weak, weak, moderate, strong, very strong}.

Each iteration of the dichotomic sorting based aggregation-disaggregation procedure for consensus seeking consists of several steps:

1. At first, fuzzy and strict outranking relations are calculated. The former are expressed with values on the [0, 1] interval. The latter are derived with the λ-cut.
2. Robustness degrees $r^k(a_i)$ are computed. Each shows to what extent the k-th decision-maker's preferential parameters must deviate for the category of the i-th alternative to change. The bigger adjustment that is required, the more robustly a_i is evaluated from the perspective of DM_k. If the robustness exceeds the ψ threshold, then the decision-maker or his agent should not be asked to conform to the majority opinion.
3. Alternatives are rank ordered according to received votes. For a single alternative this is the number of outranking relations stating that it is an acceptable choice.
4. Since alternatives can differ considerably with regard to compromise votes, the group gets directed on the basis of consensus and agreement degrees. The agreement degree ζ^k shows to what extent DM_k is in concordance with the collective opinion of the whole group. DM_k with the lowest ζ^k must generally reconsider his preferences in order to conform to the group. However, if some of his assessments are robust, which means that the $r^k(a_i) \geq \psi$ inequality holds true for at least one alternative, someone else is selected for conformation. The decision

support system thus verifies for each decision-maker, whether the required reassignment of alternatives is feasible.

In the first iteration, each decision-maker is able to sort at least one alternative with low agreement and robustness degrees to a different category. Hence, a matrix of required strict outranking relations is induced from the initial outranking matrix by changing the relation either from 0 to 1 or vice versa for each alternative with $\zeta_i < 0.5$ and $r(a_i) < \psi$. The most discordant decision-maker DM_2 is able to reassign alternatives a_1, a_3, a_4 and a_7, but he is not allowed to change the category of a_2 because of its high robustness $r^2(a_2) = 0.399 > 0.3 = \psi$.

DM_2	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
Initial fuzzy relations	0.688	0.725	0.600	0.750	0.700	0.267	0.588	0.900
Initial strict relations	1	1	1	1	1	0	1	1
Robustness degrees	0.263	0.399	0.053	0.194	0.345	0.541	0.288	0.724
Agreement degrees	0.200	0.400	0.200	0.000	0.600	1.000	0.000	0.600
Required strict relations	0	1	0	0	1	0	0	1
Reassignments	1	0	1	1	0	0	1	0

The optimization algorithm is now applied to reassign alternatives and to simultaneously infer new values of preferential parameters. As these values do not violate specified interval constraints, a new negotiation iteration is started.

The overall consensus degree rises from 0.458 to 0.625 in the second iteration, and in the final iteration, it reaches 0.917. Seven iterations are needed to find a consensual solution. All subsequent iterations, except the second, proceed identically as the first, which means that the negotiation procedure operates totally automatically. Only in the second iteration, the decision-maker's constraints are violated, so he must manually check newly inferred values of preferential parameters and confirm their acceptability.

5 COMPARATIVE ANALYSIS OF RESULTS

The case study does not refer to the same cases as the simulation. As has been explained in Section 3, 180 different samples (cases) have been randomly generated for nine possible experimental combinations of the latter. On the contrary, exactly one specific case has been defined within the scope of the former. However, the results of both studies are directly comparable because they are analysed according to the same set of criteria that constitute a formal model for the evaluation of group decision-making methods and systems [2]. Since the model is holistic and complex, only a subset of its criteria and metrics is considered. It is structured on Figure 1.

All metrics are quantitative. They can be directly applied to data of both studies. They hence enable a consistent comparison of simulation and case study results. For most metrics, results are computed as scalars ($M_{1.1}$, $M_{1.2}$, $M_{1.3}$, $M_{2.3}$, $M_{3.2}$, $M_{3.3}$, $M_{4.1}$, $M_{4.2}$, $M_{5.1}$, $M_{5.2}$, $M_{6.1}$, $M_{6.2}$, $M_{7.1}$, $M_{7.2}$, $M_{7.3}$, $M_{7.4}$). Other metrics produce vectors or matrices. Where appropriate, the comparability of results for various experimental combinations $<n, o>$ of the simulation study is determined with a one-way ANOVA analysis of variance. The $M_{5.1}$ and $M_{5.2}$ metrics are operationalized with the Kemeny-Snell distance between ordinal rank orders. Exact definitions of metrics and general criteria are omitted. The interested reader should refer to the related literature [1, 2].

The results of scalar metrics are listed in the table. They provide a direct comparison of simulation and case studies. It can be stated that the dichotomic sorting based aggregation-disaggregation procedure for multi-criteria group consensus seeking is efficient with regard to both research methods. The following conclusions can be drawn:

1. The best alternative is always unique. It has a large distance to the second best one, and especially to the set of all other suboptimal choices. The distance considerably increases when a compromise solution iteratively progresses to the final consensual solution.

Group maintenance (ability to direct the process of group decision-making)
- Ability to reach a compromise
  - $M_{1.1}$: Percentage of cases in which the best alternative is not unique
  - $M_{1.2}$: Distance from the best to the second best alternative
  - $M_{1.3}$: Distance from the best alternative to the subset of all suboptimal alternatives
  - $M_{1.4}$: Average degree of compromise for the *i*-th best alternative
  - $M_{1.5}$: Average number of alternatives ranked in the *k*-th place
  - $M_{1.6}$: Average robustness of alternatives ranked in the *k*-th place
- Ability of autonomous guidance and conflict resolution
  - $M_{2.1}$: Progression of the compromise solution over negotiation iterations
  - $M_{2.2}$: Agreement degrees of individuals that are reached in each iteration
  - $M_{2.3}$: Absolute increase of the calculated agreement degrees
- Convergence of opinions
  - $M_{3.1}$: Monotonous iterative progression of the consensus degree
  - $M_{3.2}$: Improvement of the consensus degree compared to the initial compromise
  - $M_{3.3}$: Number of iterations

Analysis
- Credibility of analysis with regard to robustness
  - Improved richness of discriminating information (consensus versus compromise)
    - $M_{4.1}$: Difference in the distance from the best to the second best alternative
    - $M_{4.2}$: Difference in the distance from the best alternative to all other alternatives
    - $M_{4.3}$: Change of the average number of votes for the *i*-th best alternative
    - $M_{4.4}$: Change of the average number of alternatives ranked in the *k*-th place
    - $M_{4.5}$: Change of the average robustness of alternatives ranked in the *k*-th place
  - Sensitivity to changes in the problem structure
    - $M_{5.1}$: Number of rank reversals when an existing alternative is discarded
    - $M_{5.2}$: Percentage of rank reversal cases when an existing alternative is discarded
  - Sensitivity to changes in the decision-making group
    - $M_{6.1}$: Number of rank reversals when a decision-maker leaves the group
    - $M_{6.2}$: Percentage of rank reversal cases when a decision-maker leaves the group
- Complexity of analysis
  - $M_{7.1}$: Total number of preferential parameters
  - $M_{7.2}$: Quantity of inputs that are required for the first negotiation iteration
  - $M_{7.3}$: Number of manual adjustments of parameters in each subsequent iteration
  - $M_{7.4}$: Amount of data analysed in each iteration

Figure 1: Applied criteria and metrics of the evaluation model

2. The overall agreement and consensus degrees increase drastically from the first to the last iteration of the problem solving process. This means that opinions of decision-makers are considerably more unified at the end than at the beginning of negotiation. Convergence is fast, since maximally seven iterations are required to reach unanimity with regard to the assessments of alternatives. Hence, it is unlikely that any decision-maker would drop out of the group because of personal time constraints, motivation or difficulties in communication.
3. Robustness of assessments of alternatives increases over subsequent iterations of the decision-making process. Although the best alternative is already sufficiently distant from suboptimal alternatives in the initial compromise solution, the distance is even larger in the case of the final consensual solution.

4. Decision analysis is unsensitive to changes in the problem structure. If an alternative is excluded from the set of available choices, this has no effect on the evaluation of remaining alternatives. Their rank order is thus preserved.
5. Decision analysis is not significantly sensitive to any changes in the decision-making group. If an individual stops participating in the consensus seeking process, this has a minor or even negligible influence on the evaluation of alternatives. Their rank order changes only slightly, in a way that rank reversals never occur. This means that it is impossible for the preference between two alternatives to revert ($\neg[a\,P\,b \Rightarrow b\,P\,a]$). It can only happen that preference gets exchanged with indifference ($a\,P\,b \Rightarrow b\,I\,a$), or vice versa ($a\,I\,b \Rightarrow b\,P\,a$). This conclusion is not only verifiable by a formal proof, but also by observing the discrepancy between the intensity and frequency of rank order changes ($M_{6.1}$ versus $M_{6.2}$).
6. Cognitive load of participants of the negotiation process is relatively high only in the first iteration, when initial preferences have to be set. In all subsequent iterations, the negotiation support system operates almost automatically. In the simulation study, no human interaction was required, while in the case study, the decision-makers had to continuously observe just a small set of crucial parameters, and were asked to make a minimal amount of manual adjustments.

Metric	Case study	Simulation study
$M_{1.1}$	0 %	0 %
$M_{1.2}$	0.333	0.276
$M_{1.3}$	0.905	0.646
$M_{2.3}$	0.295	0.386
$M_{3.2}$	0.459	0.399
$M_{3.3}$	7	(average, min, max) = (4.150, 2, 7)
$M_{4.1}$	0.333	0.165
$M_{4.2}$	0.548	0.131
$M_{5.1}$	0	0
$M_{5.2}$	0 %	0 %
$M_{6.1}$	0.010	(0.015, 0.008, 0.026)
$M_{6.2}$	19.048 %	(26.081 %, 14.381 %, 31.833 %)
$M_{7.1}$	132	maximally $n \cdot (m + 18)$
$M_{7.2}$	110	maximally $n \cdot (m + 18)$
$M_{7.3}$	0.167	0
$M_{7.4}$	6.194	0

The results of the simulation study and the case study are consistent. A relatively high deviation can be observed only with regard to the $M_{4.2}$ metric. However, even in this case it is evident that both studies consistently detect an improvement of the consensual solution over the initial compromise solution.

Similarly, the comparability of results for various experimental combinations $<n, o>$ of the simulation study is also determined with regard to several metrics – $M_{1.2}$, $M_{1.3}$, $M_{6.1}$ and $M_{6.2}$. A one-way ANOVA analysis of variance returns p-values 0.611, 0.248, 0.108 and 0.292, respectively. No statistically significant difference among experimental combinations can hence be found for the usual threshold $\alpha = 0.05$, nor even for the less common $\alpha = 0.1$.

Figure 2 shows a relatively fast monotonous increase of total consensus degrees towards the highest possible value of 1. Although the upper limit is not utterly reached, this is not a drawback, because rubust assignments of alternatives must prevent decision-makers to fully conform to the collective opinion of the group. Analogous to scalar metrics, $M_{3.1}$ reveals that the observed method exhibits similar characteristics according to both research studies.

Figure 3 demonstrates an iterative progression of the compromise solution towards the consensual solution in the case study. Decision-makers eventually agree on sorting six non-optimal alternatives into the negative category C^-, and assigning the optimal alternative a_8 to the positive category C^+. Seven out of eight alternatives thus converge towards the upper or lower limit of votes, respectively. Only a_5 is not subjected to convergence because of its high robustness. The negotiation mechanism is consequently not allowed to propose its irrational reassignment.

For the case study, metric $M_{2.2}$ is depicted on Figure 4. It shows that agreement degrees of individuals generally increase, yet may instantaneously fall in certain cases. This happens when a group member adjusts his preferential parameters in order to unify with the collective opinion of the whole group. The discordance of one or more other decision-makers can then consequently decrease.

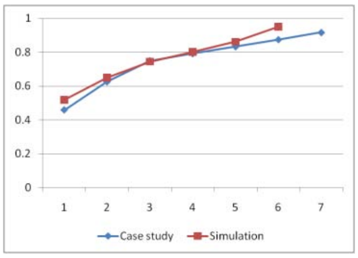


Figure 2: Iterative increase of the consensus degree

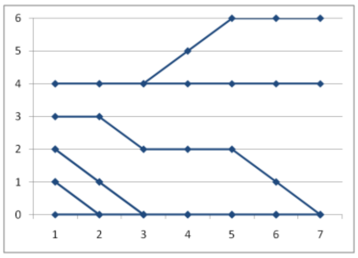


Figure 3: Progression of the compromise solution

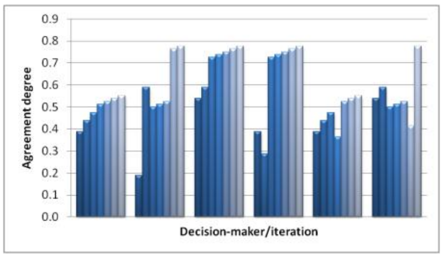


Figure 4: Agreement degrees of decision-makers for each iteration

Figures 5, 6 and 7 correspond to metrics $M_{1.4}$, $M_{1.5}$, $M_{1.6}$, $M_{4.3}$, $M_{4.4}$ and $M_{4.5}$. It can be seen that the weak ordinal rank order of alternatives is more efficient in the last than in the first iteration of the consensus seeking procedure. The implication is that the robustness of assessments, as well as the discrimination among acceptable and unsatisfactory alternatives improve. Similar results may be observed for the simulation study (above) and the case study (below).

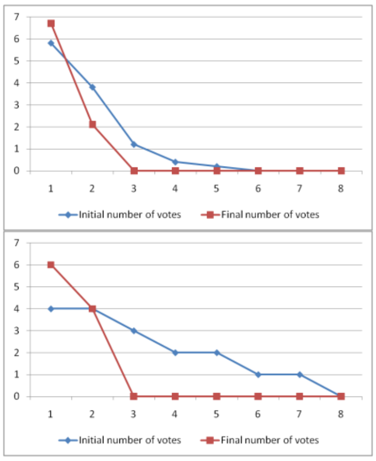


Figure 5: Number of votes for the *i*-th best alternative

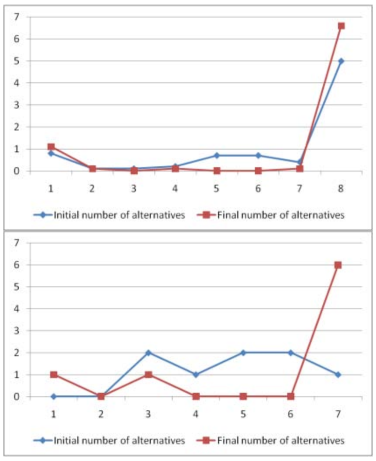


Figure 6: Number of alternatives ranked in the *k*-th place

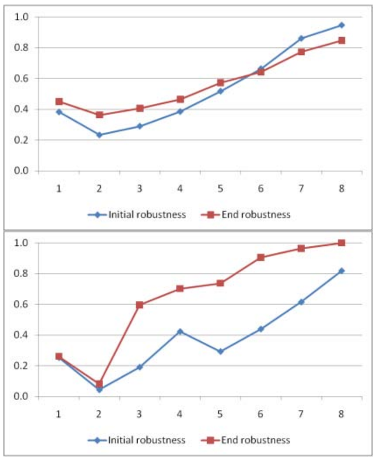


Figure 7: Robustness of the *i*-th best alternative

	Metric $M_{7.2}$	Metric $M_{7.4}$ (metric $M_{7.3}$)							
		Iter. 2	Iter. 3	Iter. 4	Iter. 5	Iter. 6	Iter. 7	Total	Average
DM_1	$n \cdot (m+18)+2$	2	2	2	2	2	2	12	2.000
DM_2	$n \cdot (m+1)+13$	26	2	2	26	2	2	60	10.000
DM_3	$n \cdot (m+4)+14$	2	2	2	2	2	2	12	2.000
DM_4	$n \cdot (m+18)+2$	2	57 (1)	2	2	2	2	67 (1)	11.167
DM_5	$n \cdot (m+18)+2$	2	2	26	2	2	2	36	6.000
DM_6	$n \cdot (m+18)+2$	2	2	2	2	26	2	36	6.000

Finally, the table above gives details about metrics $M_{7.2}$, $M_{7.3}$ and $M_{7.4}$ for each decision-maker and each iteration of the case study. All decision-makers have to provide $m \cdot n$ criteria-wise evaluations of alternatives and n values of the referential profile in the first iteration. Other parameters vary according to the form in which preferences are expressed. In each subsequent iteration, only two values have to be mandatory observed by the decision-maker: the overall consensus degree and the total individual agreement degree. An active decision-maker must additionally process m categories, robustness degrees and ranks for each choice. If an individual is active and has to manually adjust preferential parameters, newly derived values of parameters and violations of constraints are relevant for him as well. It should be noted that only one manual adjustment is required throughout the negotiation process.

6 CONCLUSION

Both research methods – the simulation study and the case study – produce comparable results. Their complementary application clearly and consistently exposes the characteristics of the evaluated aggregation-disaggregation approach to negotiations, which is proved to be efficient. Within the scope of further research work, a questionnaire survey will also be performed to obtain and analyse the feedback of decision-makers on the usefulness of such negotiation and group decision support systems.

References

[1] Bregar, A. Efficiency of problem localization in group decision-making. Proceedings of 10th International Symposium on Operational Research in Slovenia, 139–149, 2009.

[2] Bregar, A. Celovit model vrednotenja skupinskih odločitvenih metod in sistemov. Proceedings of 17th Symposium Dnevi Slovenske Informatike (DSI), 13 pp., 2010.

[3] Bregar, A. Metode na temelju prednostne relacije in njihova uporaba v postopkih sku-pinskega odločanja: študija primera. Proceedings of 18th DSI, 15 pp., 2011.

[4] Bregar, A., Györkös, J., Jurič, M. B. Interactive aggregation/disaggregation dichotomic sorting procedure for group decision analysis based on the threshold model. Informa-tica, 19 (2), 161–190, 2008.

[5] Damart, S., Dias, L. C., Mousseau, V. Supporting groups in sorting decisions: Metho-dology and use of a multi-criteria aggregation/disaggregation DSS. Decision Support Systems, 43 (4), 1464–1475, 2007.

[6] Dias, L. C., Clímaco, J. N. ELECTRE TRI for groups with imprecise information on parameter values. Group Decision and Negotiation (GDN), 9 (5), 355–377, 2000.

[7] El Fallah, S. A., Moraitis, P., Tsoukiàs, A. An aggregation-disaggregation approach for automated negotiation in multi-agent systems. Proceedings of ICSC Symposium on Multi-Agents and Mobile Agents, 7 pp., 2000.

[8] Espinasse, B., Picolet, G., Chouraqui, E. Negotiation support systems: A multi-criteria and multi-agent approach. European Journal of Operational Research (EJOR), 103 (2), 389–409, 1997.

[9] Jabeur, K., Martel, J.-M., Ben Khélifa, S. A distance-based collective preorder inte-grating the relative importance of the group's members. GDN, 13 (4), 327–349, 2004.

[10] Kilgour, D. M., Colin, E. Handbook of Group Decision and Negotiation. Springer, Dordrecht, 2010.

[11] Leyva-López, J. C., Fernández-González, E. A new method for group decision support based on the ELECTRE III methodology. EJOR, 148 (1), 14–27, 2003.

[12] Matsatsinis, N. F., Grigoroudis, E., Samaras, A. P. Aggregation and disaggregation of preferences for collective decision-making. GDN, 14 (3), 217–232, 2005.

[13] Matsatsinis, N. F., Samaras, A. P. MCDA and preference disaggregation in group deci-sion support systems. EJOR, 130 (2), 414–429, 2001.

[14] Mousseau, V., Slowinski, R., Zielniewicz, P. A user-oriented implementation of the ELECTRE TRI method. Computers & Operations Research, 27 (7–8), 757–777, 2000.

HOW TO USE THE 5WS & H TECHNIQUE TO DETERMINE THE WEIGHTS OF INTERACTING CRITERIA

Vesna Čančer
University of Maribor, Faculty of Economics and Business
Razlagova 14, SI-2000 Maribor, Slovenia
vesna.cancer@uni-mb.si

Abstract: This paper introduces the use of the 5Ws & H technique, for the establishing of the criteria weights. It adapts and completes the steps of this technique based on questions to establish the importance of criteria by the methods based on interval scale. It extends its use to the weighting step of MCDM where synergies and redundancies among criteria are considered by using the Choquet integral. The applicability of the proposed approach is also discussed and introduced via a practical case – the selection of the most appropriate blade.

Keywords: 5Ws & H technique, Choquet integral, SMART, SWING, synthesis, weighting.

1 INTRODUCTION

When solving problems with multi-criteria decision-making (MCDM) methods, decision makers are encouraged to follow one of the MCDM procedures. The frame procedure of MCDM for the group of methods based on assigning weights includes the following steps: problem definition, elimination of unacceptable alternatives, problem structuring, measuring local alternatives' values, criteria weighting, synthesis and ranking, and sensitivity analysis [4]. The procedure was well-verified in practice [3], but lacked the support of the problem definition techniques. To eliminate this deficiency, in terms of prescriptive approach, we adapted and completed the steps of the 5Ws & H technique, which is based on questions [2], to establish the weights of criteria and of the subsets of them. The discrete Choquet integral [5, 10] was used to consider interactions among criteria. The applicability of the proposed approach is introduced via a practical case – the selection of the most appropriate blade.

2 WEIGHTING AND AGGREGATION TOOLS IN MCDM

The most common aggregation tool that is used in MCDM is the weighted arithmetic mean. Under the assumption of independence among criteria, it requires the assignment of a weight to each criterion [9]. This step is usually carried out by decision makers and thus reflects their point of view on the multi-criteria decision problem [9]. Since, in practical applications, decision makers very often tell the relative importance of criteria directly with difficulty, the criteria's importance can be expressed by using the methods based on ordinal (e.g. SMARTER), interval (e.g. SWING and SMART) and the ratio scale (i.e. AHP) [1]. In this paper, special attention is given to the use of the methods for establishing the judgements on criteria's importance, based on the interval scale. In SMART, a decision maker is first asked to assign 10 points to the least important attribute change from the worst criterion level to its best level, and then to give points (≥ 10, but ≤ 100) to reflect the importance of the attribute change from the worst criterion level to the best level relative to the least important attribute range [6]. In SWING, a decision maker is asked first to assign 100 points to the most important attribute change from the worst criterion level to the best level, and then to assign points (≤ 100, but ≥ 10) to reflect the importance of the attribute change from the worst criterion level to the best level relative to the most important attribute change [6]. In SMART and SWING, the weight of the j^{th} criterion, w_j, is obtained by [6]:

$$w_j = \frac{t_j}{\sum_{j=1}^{m} t_j}, \tag{1}$$

where t_j corresponds to the points given to the j^{th} criterion, and m is the number of criteria. When the criteria are structured in two levels, the weight of the s^{th} attribute of the j^{th} criterion, w_{js}, is obtained in SMART and SWING by:

$$w_{js} = \frac{t_{js}}{\sum_{s=1}^{p_j} t_{js}}, \tag{2}$$

where t_{js} corresponds to the points given to the s^{th} attribute of the j^{th} criterion, and p_j is the number of the j^{th} criterion sub-criteria.

The local values of alternatives with respect to each criterion on the lowest level (attribute) can be measured by value functions, pair-wise comparisons or directly [1]. During synthesis, the additive model is usually used [1]. When the criteria are structured in two levels, the aggregate alternatives' values are obtained by [4]:

$$v(X_i) = \sum_{j=1}^{m} w_j \left(\sum_{s=1}^{p_j} w_{js} v_{js}(X_i) \right), \text{for each } i = 1, 2, \ldots, n \tag{3}$$

where $v(X_i)$ is the value of the i^{th} alternative, and $v_{js}(X_i)$ is the local value of the i^{th} alternative with respect to the s^{th} attribute of the j^{th} criterion.

If there is interaction among the criteria, decision makers usually return to the hierarchy and redefine the criteria. They can also use other models to obtain the aggregated alternatives' values; it has already been delineated how to complete the additive model into the multiplicative one [4]. Further, the concept of fuzzy measure has been introduced [10]: it is useful to substitute to the weight vector w a non-additive set function on K allowing to define a weight not only on each criterion, but also on each subset of criteria. A suitable aggregation operator, which generalizes the weighted arithmetic mean, is the discrete Choquet integral. Following [5, 10], this integral is viewed here as an m-variable aggregation function; let us adopt a function-like notation instead of the usual integral form, where the integrand is a set of m real values, denoted by $v = (v_1, \ldots, v_m) \in \Re^n$. The (discrete) Choquet integral of $v \in \Re^n$ with respect to w is defined by:

$$C_w(v) = \sum_{j=1}^{m} v_{(j)} \left[w(K_{(j)}) - w(K_{(j+1)}) \right], \tag{4}$$

where $(\cdot)$ is a permutation on K – the set of criteria, such that $v_{(1)} \le \ldots \le v_{(m)}$. Also, $K_{(j)} = \{(j), \ldots, (m)\}$.

3 THE CRITERIA WEIGHTING BASED ON QUESTIONS

In MCDM, by using the groups of methods based on assigning weights, it is assumed that decision makers are able to express their judgments about the criteria's importance. However, very often decision makers are not aware of the relationships among different criteria. Marichal and Roubens [11] have already emphasized that it is important to ask the decision maker the kind of good questions to determine the weights of interacting criteria from a reference set; since questions are not evident from [11], we propose the use of the problem definition method based on questions to determine the weights of interacting criteria.

According to Cook [2], the 5Ws & H technique is a structured method that examines a problem from multiple viewpoints. It is based on who, what, when, where, why and how questions. When the technique is used for problem definition, the process may be summarized as follows [2]. We state the problem starting with 'In what ways might ... ?' and write down the questions that are relevant to the problem. Participants answer the questions, examine responses to each question and use them to stimulate new problem definition. Any redefinitions suggested are written down so that one redefinition that best captures the problem we are trying to resolve is selected.

In this paper, we propose the following process of establishing the judgments about the criteria's importance by the 5Ws & H technique:

1. In what ways might the criteria weights be determined?
2. The questions regarding the considered problem are put and written down.
3. The questions are answered and the weights are determined and re-determined.

4 A PRACTICAL CASE: BLADE SELECTION

This section describes the process of establishing judgments about the criteria's importance by using the 5Ws & H technique in decision-making about blade selection. The MCDM model for the selection of the most suitable blade was built together with an IT company with the aim of presenting possible solutions to their current and potential customers: medium-sized and large companies. The blades that can be offered are described as alternatives in Table 2. The criteria structure is presented in Table 1.

The decision maker with appropriate knowledge for the problem definition techniques and for MCDM (in this case, the co-ordinator) asks (Q) and answers (A) the typical question of the first step of the 5Ws & H technique process:

Q: In what ways might the criteria weights be determined?

A: Directly, by using several methods based on the interval scale (e.g. SWING, SMART), ordinal (SMARTER) and ratio scale (AHP); individually, in groups; by assuming independence between two criteria, by considering interactions among multiple criteria.

In the second step, the co-ordinator asked, and in the third step, answered the questions regarding the responsibility and competency regarding expressing judgments about the criteria's importance. After the participants of the group for solving the problem were defined (project manager, seller, engineers in the considered IT company, responsible for pre-sales support, customers), they answered the questions, successively put by the co-ordinator. To determine the first level criteria's importance, the SWING method was used:

Q: Which criterion change from the worst to the best level is considered the most important?

A: The change from worst to best costs.

Q: With respect to this change importance, how many points less and how many points are given to other first-level criteria changes?

A: 20 points less, i.e. 80 points are given to the change from the worst to the best vision, and 40 points less (i.e. 60 points) are given to the technology change.

Similar questions were asked to determine the importance of the attributes of technology. To determine the weights of the vision attributes, the SMART method was used:

Q: Which criterion change from the worst to the best level is considered the least important?

A: The change from worst to best innovativeness.

Q: With respect to this change importance, how many points more and how many points are given to other vision attributes changes?

A: 10 points more, i.e. 20 points are given to the change from the worst to the best product development, and 15 points more (i.e. 25) are given to the market strategy change.

Similar questions were asked to determine the importance of the attributes of the costs. The weights in Table 1 were determined by considering the above-written answers, (1) for the weights of the first level criteria and (2) for the weights of the second level criteria.

Table 1: The criteria structure and the weights for the selection of blade.

First level criteria	Weights of the first level criteria	Second level criteria	Weights of the second level criteria
TECHNOLOGY	$w_1 = 0.250$	Chassis Blade number Connectivity Deployment Features	$w_{11} = 0.138$ $w_{12} = 0.345$ $w_{13} = 0.207$ $w_{14} = 0.103$ $w_{15} = 0.207$
COSTS	$w_2 = 0.417$	Energy efficiency Purchase price Management	$w_{21} = 0.444$ $w_{22} = 0.333$ $w_{23} = 0.222$
VISION	$w_3 = 0.333$	Market strategy Product development Innovativeness	$w_{31} = 0.455$ $w_{32} = 0.364$ $w_{33} = 0.182$

Table 2: Alternatives' data with respect to the attributes.

	Data type	Alternative 1	Alternative 2	Alternative 3	Measuring local alternatives' values
Chassis	Quantitative: number of choices	5	3	2	Value function, LB: 1, UB: 5
Blade number	Quantitative: number	14	16	10	Value function, LB: 6, UB: 16
Connectivity	Qualitative, verbal evaluation	Flexible	Limited	Flexible	Pair-wise comparisons
Deployment	Quantitative, MU: h	12	8	8	Value function, LB: 4, UB: 16
Features	Quantitative: number	3	3	1	Value function, LB: 1, UB: 4
Energy efficiency	Quantitative: in 1000 kWH	190	240	165	Value function, LB: 60, UB: 240
Purchase price	Quantitative, MU: 1000 €	100	140	130	Value function, LB: 60, UB: 140
Management	Quantitative, MU: 1000 €	400	400	600	Value function, LB:400, UB: 600
Market strategy	Qualitative, verbal evaluation	Good	Very good	Not enough	Pair-wise comparisons
Product development	Qualitative, verbal evaluation	Good	Very good	Good	Pair-wise comparisons
Innovative-ness	Qualitative, verbal evaluation	Medium	Medium	Low	Pair-wise comparisons

Symbols: MU – measurement unit, € – Euro, h – hour, kwH – kilo watt hour, LB – lower bound, UB – upper bound, Alternative 1 – IBM BladeCenter [8], Alternative 2 – HP BladeSystem c7000 Enclosure [7], Alternative 3 – Oracle's Sun Blade 6000 [12]; Sources: [7, 8, 12]; own experience of the observed company's pre-sales support engineers

Table 2 shows the alternatives' data with respect to the criteria of the lowest hierarchy level, together with the methods that are used to measure the local alternatives' values. The alternatives' values with respect to the higher level criteria and the aggregate alternatives' values obtained by the additive model (3) are presented in Table 3. They allow us to report that Alternative 1 is the most appropriate alternative.

Table 3: The alternatives' values, obtained with the additive model.

	Alternative 1	**Alternative 2**	**Alternative 3**
Value with respect to 'technology' v_1	0.675	0.650	0.330
Value with respect to 'costs' v_2	0.512	0.222	0.227
Value with respect to 'vision' v_3	0.244	0.601	0.156
Aggregate alternative's value v	0.463	0.455	0.229
Rank	1.	2.	3.

This is a complex MCDM process, where interactions among criteria should be considered; the co-ordinator found that the Choquet integral as an aggregation function has proven quite useful and convenient in this direction. Since engineers in the considered IT company, who are responsible for pre-sales support, can evaluate the synergies and redundancies between factors on the bases of their professional experience, detailed data from the principal, and the project goals directly, he asked them the following questions:

Q: Which synergies should be taken into consideration?

A: The customers' (especially top) managers that make the blade purchase decisions are interested in the interactions among higher level criteria, in this case among 'technology', 'costs' and 'vision'.

Q: Where are the synergies/redundancies in this model?

A: Synergy: between 'costs' and 'vision,' redundancy: between 'technology' and 'vision'.

Q: Why is there synergy between 'costs' and 'vision' and what does it mean?

A: Appropriate vision enables better cost controlling. In the concept of the Choquet integral this means: $w_{2,3} > w_2 + w_3$; $w_2 + w_3 = 0.75$ (Table 1), $w_{2,3} = 0.85$.

Q: Why is there redundancy between 'technology' and 'vision' and what does it mean?

A: Because the vision determines the technology. For the concept of the Choquet integral, this means that $w_{1,3} < w_1 + w_3$; note that $w_1 + w_3 = 0.583$ (see Table 1), $w_{1,3} = 0.45$.

Considering the above-written answers, the weights were re-determined. Table 4 presents the Choquet integrals for the selection of the most suitable blade, obtained by (4).

Table 4: The alternatives' values, obtained by considering interactions among criteria with the Choquet integral.

	Alternative 1	**Alternative 2**	**Alternative 3**
Choquet integral C	0.463	0.405	0.229
Rank	1.	2.	3.

Studying the results in Table 4 we can report that considering interactions among criteria did not change the final rank of alternatives. However, when comparing it with the values obtained by the additive model (Table 3), it can be concluded that redundancy between 'technology' and 'vision' decreased the value of the Choquet integral C of Alternative 2 (note that $v_2 < v_3 < v_1$). Because $v_3 < v_2 < v_1$ for alternatives 1 and 3, the above mentioned redundancy did not influence C of alternatives 1 and 3, and the synergy between 'costs' and 'vision' did not come into forefront. Although the aggregate values of alternatives 1 and 2, obtained by the additive model (3) (Table 3), are extremely sensitive to the changes of the weights of ‘costs’ and ‘vision’, the redundancy between ‘technology’ and ‘vision’ considerably decreases *C* of Alternative 2 and thus strengthen the decision-making basis.

5 DISCUSSION AND CONCLUSIONS

The problem definition techniques based on questions (W, 5Ws & H, Why, the 5 Whys) and visualisation (cognitive mapping, fishbone diagrams, mind mapping) are usually applied in problem definition – the first step of the frame procedure of MCDM, based on assigning weights, in order to find and describe a problem, relevant criteria and alternatives. However, we illustrated that in this paper adapted and completed steps of the 5Ws & H technique enable decision makers to consider several aspects regarding establishing the criteria weights. The described approach requires the co-ordinator that can be a member of the decision-making group, but has knowledge on both the considered problem definition (5Ws & H) technique and the computer supported MCDM methods based on assigning weights (e.g. SMART and SWING). It enables other participants to focus on professional aspects of the considered problem without knowing the particularities of weighting and aggregation procedures for interacting criteria. We recognize further application possibilities of the methods based on questions in measuring the local alternatives’ values.

References

[1] Belton, V., Stewart, T. J. (2002): *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers.

[2] Cook, P. (1998): *Best Practice Creativity*. Aldershot: Gower Publishing Limited.

[3] Čančer, V. (2008): *A Frame Procedure for Multi-criteria Decision-making: Formulation and Applications.* In: V. Boljunčić, L. Neralić and K. Šorić (eds.): KOI 2008 Proceedings. Pula, Zagreb: Croatian Operational Research Society.

[4] Čančer, V. (2009): *Considering Interactions in Multi-criteria Decision-making.* In: L. Zadnik Stirn, J. Žerovnik, S. Drobne and A. Lisec (eds.): SOR ’09 Proceedings. Ljubljana: Slovenian Society Informatika, Section for Operational Research.

[5] Grabisch, M. (1995): Fuzzy Integral in Multicriteria Decision Making. *Fuzzy Sets and Systems*, 69(3): 279-298.

[6] Helsinki University of Technology (2002): *Value Tree Analysis.* http://www.mcda.hut.fi/value_tree/theory, consulted in August 2011.

[7] Hewlett-Packard Development Company (2011): *HP BladeSystem c7000 Enclosure.* http:// h18004.www1.hp.com/products/quickspecs/12810_div/12810_div.pdf, consulted in June 2011.

[8] IBM Systems and Technology (2011): *IBM BladeCenter – Build smarter IT.* http://public.dhe.ibm.com/common/ssi/ecm/en/blb03002usen/BLB03002USEN.PDF, consulted in May 2011.

[9] Kojadinovic, I. (2004): Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functional. *European Journal of Operational Research*, 155(3): 741-751.

[10] Marichal, J. L (2000): An Axiomatic Approach of the Discrete Choquet Integral as a Tool to Aggregate Interacting Criteria. *IEEE Trans. Fuzzy Systems*, 8: 800-807

[11] Marichal, J. L., Roubens, M. (2000): Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, 124(3): 641-650.

[12] Oracle (2010): *Sun Blade 6000 Chassis.* http://www.oracle.com/us/products/servers-storage/servers/blades/033613.pdf, consulted in May 2011.

ECONOMIC CRITERIA IN DECISION-MAKING ON NUMBER OF FUNCTIONAL REGIONS: THE CASE OF SLOVENIA

Samo Drobne* and **Marija Bogataj****
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: samo.drobne@fgg.uni-lj.si
** University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož,
& MEDIFAS, Nova Gorica, Slovenia, e-mail: marija.bogataj@guest.arnes.si

Abstract: In this paper, we suggest the method for decision-making on number of functional regions. The method considers economic variables of the average monthly gross earnings in the functional region as well as the guidelines for the size of the regions. The method was tested in case study of Slovenia. There are also some recommendations to improve the method.

Keywords: functional region, commuting, decision-making, number of functional regions, Slovenia.

1 INTRODUCTION

The concept of regions is anchored deep in the history of Europe. Nowadays, the idea of regions is often connected with the integration of European Union (EU). However, different actors understand the very concept of a region quite differently. Administrative or statistic regions are defined by their borders and they are required to cover whole the respective territory and to be of comparable size. But, the functional regions of economy and/or society are product of interrelations, they are quite diverse in terms of their size and population, and they may overlap as well as not fully cover the territory [12].

The NUTS (Nomenclature of Territorial Units for Statistics) classification is a hierarchical system for dividing up the economic territory of the EU for the purpose of (a) the collection, development and harmonisation of EU regional statistics; (b) socio-economic analyses of the regions; (c) framing of EU regional policies. For the purpose of socio-economic analyses, three levels of regions have been established inside each EU member: (b1) major socio-economic regions at NUTS 1 level, (b2) basic regions for the application of regional policies at NUTS 2 level, and (b3) small regions for specific diagnoses at NUTS 3 level. For the purpose of framing of EU regional policies areas eligible under the other priority objectives have mainly been classified at the NUTS 3 level. The current NUTS classification lists 1303 regions at NUTS 3 level [10].

In Slovenia, there are twelve “statistical regions” at NUTS 3 level also called “development regions”. The first version of statistical regions is from the middle of seventies of the previous century. At that time, statistical regions were established for the purpose of regional planning and cooperation in various fields. The first regionalization of statistical regions was supported by exhaustive gravity analysis of labour markets, education areas and supply markets in twelve regional and their sub-regional centres – that is the reason, why Slovenian regions at NUTS 3 level are very stable [17].

Regions on NUTS 3 level are normally functional regions. A functional region is characterised by its agglomeration of activities and by its intra-regional transport infrastructure, facilitating a large mobility of people, products, and inputs within its interaction borders. The basic characteristic of a functional region is the integrated labour market, in which intra-regional commuting as well as intra-regional job search and search for labour demand is much more intensive than the inter-regional counterparts [11]. Consequently, the border of a labour market region is a good approximation of the border of a functional region [1,3,4,5,9,11,15].

There are several approaches and methods for delimitation of functional regions; for taxonomy and discussion about approaches see [11]. In this paper we analyse functional regions of Slovenia by commuting zone approach using Intramax method[1] and suggest model for decision-making on number of functional regions in the country.

2 MATERIALS AND METHODOLOGY

In 2009, there were a total of 390,500 labour commuters (persons in employment who commute) between 210 municipalities of total 812,315 labour commuters in Slovenia. Labour commuter is person in employment whose territorial unit of workplace is not the same as territorial unit of residence. The source of data was the Statistical Register of Employment (SRDAP), which was kept by Statistical Office of the Republic of Slovenia. SRDAP covers persons in paid employment and self-employed persons who are at least 15 years old and who have on the basis of the employment contract compulsory social insurance or are employed on the territory of the Republic of Slovenia. Employment can be permanent or temporary, full time or part time [18].

To analyse economic criteria in decision-making on number of regions, set of functional regions were modelled first using Intramax method. The method, which was introduced by Masser and Brown in [13] and improved in [14], carries out a regionalization of an interaction matrix. The objective of the Intramax procedure is to maximise the proportion within the group interaction at each stage of the grouping process, while taking account of the variations in the row and column totals of the matrix. In the grouping process, two areas (municipalities in our case) are grouped together for which the objective function T is maximised [2]:

$$\max T, \quad T = \frac{T_{ij}}{O_i \cdot D_j} + \frac{T_{ji}}{O_j \cdot D_i}, \tag{1}$$

where T_{ij} is the interaction between origin location i and destination location j, $O_i = \sum_j T_{ij}$ is the total of interactions originating from origin i, $D_j = \sum_i T_{ij}$ is total of interactions coming to destination j, and O_i and $D_j > 0$.

The Intramax analysis is a stepwise analysis. In each step two areas are grouped together and the interaction between the two municipalities becomes internal interaction for the new resulting area. This new area takes the place of the two parent areas at the next step of the analyses. So with N areas after $N-1$ steps all areas are grouped together into one area (region) and all interaction become internal.

In the analysis on number of regions, the Flowmap software [2], with implemented Intramax method, was used to delimitate functional regions of Slovenia. In Flowmap, the outcome of an Intramax analysis is a report in table form and a dendogram which municipalities are grouped and how. We modelled fourteen systems of two to fifteen functional regions based on commuting data between 210 municipalities of Slovenia in 2009. This set of functional regions was used to develop economic criteria on decision-making on number of regions in the country. Here we considered two economic criteria: (a) the EU guidelines for the size of region, and (b) economic homogeneity of regions.

[1] Functional regions of Slovenia defined by labour market approach are in [6,7,8].

There are EU guidelines for size of regions (population) on NUTS 3 level. The thresholds in the Tab. 1 are used as guidelines for establishing the regions, but they are not applied rigidly.

Table 1: Guidelines for establishing the regions at NUTS 3 level [10].

Level	***Minimum population (EU)***	***Maximum population (EU)***
NUTS 3	150,000	800,000

On the other side, the most useful and the most often used economic criterion in different regional development analysis is gross domestic product (GDP). Economic prosperity can be determined in three ways, all of which should, in principle, give the same result. There are the product (or output) approach, the income approach, and the expenditure approach. The income approach measures GDP by adding incomes that firms pay households for the factors of production they hire- wages for labour, interest for capital, rent for land and profits for entrepreneurship. Normally, GDP is measured only for regions at NUTS 3 level, or higher. There are no data for GDP at lower levels of regions. For that reason, we chose average monthly gross earnings per person in paid employment [19] in the municipality of destination as a measure of economic prosperity. The average monthly gross earnings were grouped according pre-modelled functional regions of Slovenia. This economic criterion was calculated per persons in employment as well as per capita in functional region.

Model for decision-making on number of regions in the country is based on the variation of average monthly gross earnings per capita in the (functional) region and variation of population in the (functional) region regarding the EU guidelines:

$$\begin{aligned} &\min_K f(K,a) \\ &f = a \cdot CV_{GEAR}(SI) + (1-a) \cdot CD_{POP}(EU) \end{aligned} \tag{2}$$

where K is the number of regions in the country, $CV_{GEAR}(SI)$ is coefficient of variation of average monthly gross earnings per capita in the Slovene functional region, $CD_{POP}(EU)$ is coefficient of deviation of population in the region regarding the EU guidelines, a is the weight for Slovenian criterion, respectively $1-a$ is weight for EU criterion, and

$$CV_{GEAR}(SI) = \frac{\sigma_{GEAR}}{\mu_{GEAR}} \tag{3}$$

$$CD_{POP}(EU) = \sqrt{\frac{1}{K}\sum_i^K D_i^2} \bigg/ \frac{(POP(EU)_{\min} + POP(EU)_{\max})}{2} \tag{4}$$

$$D_i^2 = \begin{cases} POP_i < POP(EU)_{\min} & \rightarrow \quad (POP_i - POP(EU)_{\min})^2 \\ POP_i > POP(EU)_{\max} & \rightarrow \quad (POP_i - POP(EU)_{\max})^2 \\ \quad otherwise & \rightarrow \quad 0 \end{cases} \tag{5}$$

where $POP(EU)_{\min}$ is the minimum population in region regarding the EU guidelines and $POP(EU)_{\max}$ is the maximum population in region regarding the EU guidelines.

Model (2) allows decision-makers to include domestic and/or EU preferences; i.e. weights for domestic (a) respectively EU criteria ($1-a$).

3 RESULTS

Using data on commuting between 210 municipalities in Slovenia in 2009 and Intramax procedure in Flowmap software, fourteen systems of two to fifteen functional regions were modelled. According the EU guidelines for regions at NUTS 3 level (see Tab. 1) the lowest suggested system is system of five functional regions (minimum of population in functional region is 199,011 and maximum is 821,703).

Besides EU recommendations on population in the regions, here suggested model for defining appropriate number of (functional) regions in the country considers in addition economic variable on average monthly gross earnings in the (functional) region. Results in Chart 1 show that the systems of three, four, nine and fourteen functional regions are the most interesting systems of functional regions considering solely analysed economic variable in the region (variation of economic variable has a local minimum).

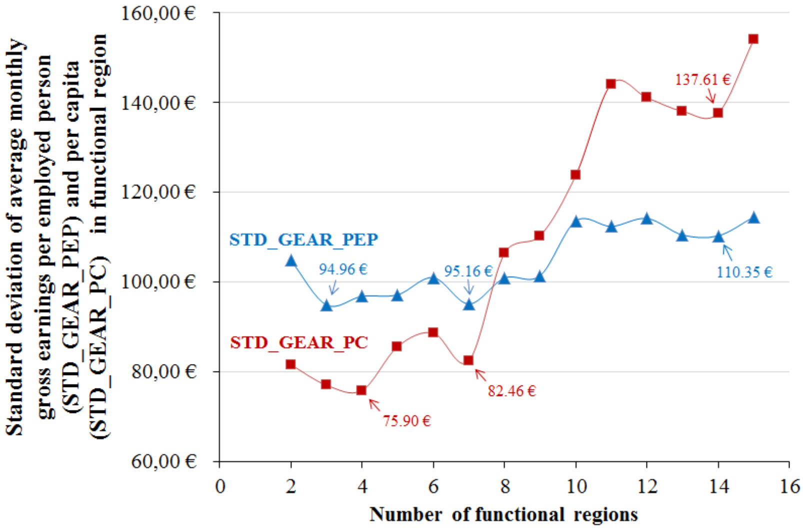


Chart 1: Standard deviation of average monthly gross earnings in fourteen systems of two to fifteen functional regions in Slovenia in 2009.

But, using the model (2) – i.e. considering both (Slovenian and EU) requirements – where the weight for Slovenian criterion a is changing from 0.1 to 0.9, there are two suggestions for the number of functional regions in Slovenia: five functional regions for $0.1 \le a \le 0.8$ and seven functional regions for $a = 0.9$. So, laying (great) stress on EU guidelines suggests us to consider the system of smaller number of (five) functional regions (Fig. 1 on the left), while forcing the most economically homogeneous regions (the more uniform distribution of average monthly gross earnings per capita in the functional regions) gives us the system of seven functional regions of Slovenia in 2009 (Fig. 1 on the right). Considering solely "local" economic variable and the fact that EU guidelines for regions are not applied rigidly for EU members the system of fourteen (functional) regions (Fig. 2) becomes also a candidate for the "right" number of functionally modelled regions. The system of fourteen functional regions is the most close to the current system of twelve statistical regions of Slovenia.

4 CONCLUSIONS

In the paper, we suggested the method for decision-making on number of functional regions in the country. Software implementation of algorithms for delimitation of functional regions enables modelling of many different systems of functional regions. So, the question about the "right" number of functional regions should not be ignored. Suggested model for decision-making on the number of functional regions considers the average monthly gross earnings per capita in the functional region as well as the guidelines for the size of the regions. The model for decision-making on number of functional regions could be improved by considering other more holistic economic parameters (i.e. GDP) or by including other functional criterion in decision-making (e.g. travelling costs in functional region as whole or to regional centre as future administrative centre).

In our case study of Slovenia, functional regions were delimitated by Intramax method considering flows of labour commuters between 210 municipalities in Slovenia in 2009. The results show that there were two (conditionally three) systems of functional regions in 2009. But, for more adequate results, functional regions should be studied at longer time horizon.

The complex territorial organization of most EU member’s political and administrative systems is rooted in history and tradition as well as in a strong political will. Most parts of the provincial structure (states) and of the district structure of administration have been already inherited from the past and reflect the administrative entities of different social systems. But, for various motivations, the creation of a middle layer of regional government or administration should be established in some new member states of the EU where no intermediate level, except state and municipality level, of territorial organisation is organised [6,8,16]. This is also the case in Slovenia. Here suggested model could help decision-makers to decide about new level of administrative regions.

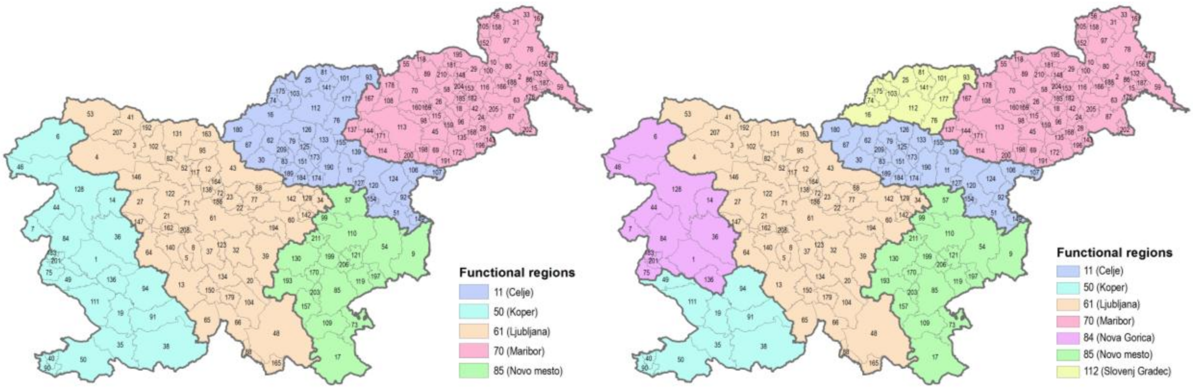


Figure 1: Five (left) and seven (right) functional regions of Slovenia in 2009.

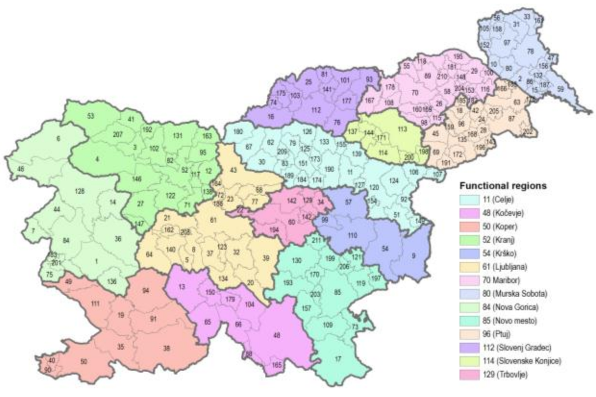


Figure 2: Fourteen functional regions of Slovenia in 2009.

References

[1] Andersen A. K. (2002). Are commuting areas relevant for the delimitation of administrative regions in Denmark?, Regional Studies, 36:833–844.

[2] Breukelman J., G. Brink, T. de Jong, H. Floor (2009). Manual Flowmap 7.3. Faculty of Geographical Sciences, Utrecht University, The Netherlands. (http://flowmap.geo.uu.nl; accessed: 15.8.2010).

[3] Casado-Di´Az J. M. (2000). Local labour market areas in Spain: A case study. Regional Studies, 34:843–856.

[4] Coombes M. G., Green A. E. and Openshaw S. (1986). An efficient algorithm to generate official statistical reporting areas: The case of the 1984 travel-to-work-areas revision in Britain. Journal of the Operational Research Society, 37:943–953.

[5] Cörvers F., M. Hensen and D. Bongaerts (2009). Delimitation and coherence of functional and administrative regions. Regional Studies, 43(1):19-31.

[6] Drobne, S., M. Konjar and A. Lisec (2009). Delimitation of Functional Regions Using Labour Market Approach. In: Zadnik Stirn L., J. Žerovnik, S. Drobne and A. Lisec (Eds.), Proceedings of SOR'09, 10th International Symposium on Operational Research in Slovenia. Ljubljana. Slovenian Society Informatika (SDI), Section for Operational Research (SOR): 417–425.

[7] Drobne, S., M. Konjar and A. Lisec (2010). Razmejitev funkcionalnih regij Slovenije na podlagi analize trga dela = Delimitation of functional regions of Slovenia based on labour market analysis. Geod. vestn. 54(3):481-500.

[8] Drobne, S., A. Lisec, M. Konjar, A. Zavodnik Lamovšek and A. Pogačnik (2009). Functional vs. Administrative regions: Case of Slovenia. In: Vujošević M. (Ed.), Thematic Conference Proceedings. Vol. 1. Belgrade. Institute of Architecture and Urban & Spatial Planning of Serbia: 395–416.

[9] Eurostat (1992). Commission Regulation amending annexes I, II and III to Regulation (EC) No1059/2003 of the European Parliament and of the Council on the establishment of a common classification of territorial units for statistics (NUTS). CPS 2006/60/1/EN, Eurostat, Luxembourg.

[10] Eurostat (2011). NUTS - Nomenclature of territorial units for statistics. (http://aesop2005.scix.net/cgi-bin/papers/Show?667; accessed: 1.5.2011).

[11] Karlsson C. and M. Olsson (2006). The identification of functional regions: theory, methods, and applications. Ann Reg Sci, 40:1–18.

[12] Maier K. (2005). New policy? New regions? New borders? AESOP 2005 Congress, 13-17 July 2005, Vienna, Austria, 9 pp. (http://aesop2005.scix.net/cgi-bin/papers/Show?667; accessed: 25.2.2011).

[13] Masser, I. and P. J. B. Brown (1975). Hierarchical aggregation procedures for interaction data. Environment and Planning A, 7(5): 509-523.

[14] Masser, I. and P.J.B. Brown (1977). Spatial representation and spatial Interaction. Papers of the Regional Science Association 38, 71-92.

[15] OECD (2002). Redefining territories – The functional regions. Organisation for Economic Co-operation and Development. Paris, France.

[16] Schrerrer W. (2006). Economic Aspects of a "Middle Layer" of Administration and Government: Some Experience from Austria, Uprava, 4(2-3):25-34.

[17] SORS (2011a). Administrative-territorial division = "Upravno-teritorialna razdelitev", Statistical Office of the Republic of Slovenia, Ljubljana. (http://www.stat.si › Splošno ›; accessed: 25.3.2011).

[18] SORS (2011b). Persons in employment (excluding farmers) by municipalities of residence and municipalities of workplace by sex, municipalities, Slovenia, annually. Statistical Office of the Republic of Slovenia. Ljubljana. (http://pxweb.stat.si/pxweb/Database/Municipalities/Municipalities.asp; accessed: 1.6.2010)

[19] SORS (2011c). Average monthly gross and net earnings per person in paid employment and persons in employment, 2010, municipalities, Slovenia. Statistical yearbook, Statistical Office of the Republic of Slovenia. Ljubljana. (http://www.stat.si/letopis/2010/31_10/31-08-10.htm; accessed: 25.2.2011)

TOWARDS CRITERIA SELECTION IN DEA BY CONJOINT ANALYSIS

Milena Djurovic, Gordana Savic, Marija Kuzmanovic and **Milan Martic**
University of Belgrade, Faculty of Organizational Sciences
Jove Ilica 154, Belgrade, Serbia
djurovicmilena86@gmail.com; goca@fon.rs; mari@fon.rs; milan@fon.rs

Abstract: Data envelopment analysis (DEA) is a method for evaluating the relative efficiency of decision making units (DMUs) with multiple inputs and outputs. DEA results depend heavily on the criteria included into analysis. Therefore the selection of adequate criteria is one of the most important phases in DEA application. In this paper, statistical technique, Conjoint analysis is suggested as supporting method for the selection of the criteria. Conjoint analysis determines the relative importance of each criterion, based on stakeholders' preference, which can be used as guide for selection of inputs and outputs for efficiency assessment of DMUs.

Keywords: Data envelopment analysis, criteria selection, stakeholder preferences, Conjoint analysis.

1 INTRODUCTION

Data envelopment analysis (DEA) is a linear programming-based procedure that measures the relative efficiencies of peer decision-making units (DMUs). This special mathematical technique was firstly introduced by Charnes, Cooper and Rhodes [4]. The DEA methodology has been widely applied in areas such as banking [12], energy, flexible manufacturing cells, highway maintenance, individual physician practice, and telecommunications.

For a given set of inputs and outputs, DEA produces a single comprehensive measure of performance (efficiency score) for each DMU. Obtained results rely heavily on the set of inputs and outputs that are used in the analysis. Therefore, one of the most important DEA phases is criteria selection. The effort increases significantly when the available data are growing. In the literature relatively little attention has been paid to how, in a real-world situation, these inputs and outputs should be chosen. Many authors treat the inputs and outputs used in their studies as simply „givens" and then go on to deal with the DEA methodology.

On the other hand, statistical methods, such as a regression and a correlation analysis, have been used in order to decrease the number of the criteria. Finally, criteria selection, as well as proper DEA model selection, may differ depends on particular case, the objectives and the purposes of the analysis. The aim of this paper is to show possibilities of using Conjoint statistical techniques as support tool for DEA inputs and outputs selection.

The paper is organized as follows: Section 2 describes DEA basics and implementation process, followed by literature survey concerning the criteria (inputs and outputs) selection. The Conjoint analysis, including the procedure for the determining the criteria importance is given in the Section 3. The proposed methodological framework of DEA criteria selection by Conjoint analysis is given in the Section 4. Finally, main conclusions are summarized in Section 5.

2 DATA ENVELOPMENT ANALYSIS

The creators of DEA [4], introduced the basic DEA CCR model as a new way to measure efficiency of DMUs, and since then a lot of variations of DEA models have been developed: the BCC model [2] which assumes variable return to scale, the additive model [1]which is non-radial, Banker and Morey [3] model which involves qualitative inputs and outputs, Golany and Roll [8] model in which input-output weights are restricted to certain ranges of values.

DEA empirically identifies the efficient frontier of a set of DMUs based on the input and output variables. Assume that there are *n* DMUs, and the *j*th DMU, produces *s* outputs ($y_{ij},..,y_{sj}$) by using *m* inputs ($x_{ij},..,x_{mj}$). The efficiency score of the observed DMU$_k$ is given as ratio of the virtual outputs (sum of weighted outputs) to the virtual inputs (sum of weighted inputs). The basic CCR ratio model is as follows:

$$\begin{aligned}
&\max \theta_k = \sum_{r=1}^{s} u_r y_{rk} \Big/ \sum_{i=1}^{m} v_i x_{ik} \\
&s.t \\
&\sum_{r=1}^{s} u_r y_{rj} \Big/ \sum_{i=1}^{m} v_i x_{ij} \leq 1,\ j = 1, \ldots, n \\
&u_r \geq 0,\ r = 1, \ldots, s,\ v_i \geq 0,\ i = 1, \ldots, m
\end{aligned} \tag{1}$$

where $u_r \geq 0$, are weights assigned to the *r*th outputs, $r = 1, \ldots, s$, and $v_i \geq 0$, are weights assigned to the *i*th inputs, $i = 1, \ldots, m$ in order assess DMU$_k$ as efficient as possible. This basic CCR DEA ratio model, which can easily be linearized, should be solved *n* times, once for each DMU$_k$. The index θ_k shows relative efficiency of DMU$_k$, obtained as maximum possible achievement in comparison with the other DMUs under the evaluation.

Emrouznejad and Witte [6] suggest a complete procedure of DEA efficiency assesment. It is a framework which can be further adapted and modified along the specific needs of the researcher. They suggested a COOPER-framework, which involves six interrelated phases:

(1) Concepts and objectives,
(2) On structuring data,
(3) Operational models,
(4) Performance comparison model,
(5) Evaluation, and
(6) Results and deployment.

The first two phases of the COOPER-framework correspond to defining the problem and understanding how decision making units operate. The last two phases correspond to summarization of the results and documentation of the project for non-DEA experts. Obviously, the phases are interrelated and affect each other.

The attention in this paper is paid to the phases 1, 2 and 3, since the objective definition in the model (1) obviously indicated that efficiency of DMU*k* is crucially related to the criteria selection. Jenkins and Anderson [10] claimed the greater the criteria, the less constrained are the model weights assigned to the criteria, and the less discerning are the DEA results. The number of criteria may be large, and it is not clear which one to choose. Even worse, different selections of criteria can lead to different efficiency evaluation results. It is obviously possible to consider all criteria for evaluation, but too many of them may lead to too many efficient units, and it gives rise to difficulties in distinguishing truly efficient units from inefficient ones. For this reason, the problem of selecting adequate criteria becomes an important issue for the improvement of discrimination power of DEA.

While it is advantageous to limit the number of variables, there is no consensus on how best to do this. Banker, Charnes and Cooper [2] suggested that the number of DMUs should be at least three times larger than the number of criteria. As noted by Golany and Roll [8], few studies give an overall view of DEA as an application procedure that must focus on the choice of data variables in addition to the methodology of DEA.

Several methods have been proposed that involve the analysis of correlation among the criteria, with the goal of choosing a set of criteria that are not highly correlated with each other. Unfortunately, studies have shown that these approaches yield results which are often inconsistent in the sense that removing criteria that are highly correlated with others can still have a large effect on the DEA results [15]. Morita and Haba [14] select criteria so as to distinguish between two groups based on external information, where a 2-level orthogonal layout experiment is utilized and optimal variables can be found statistically.

Ediridsinghe and Zhang [5] have proposed a generalized DEA approach to select criteria by maximizing the correlation between the DEA score and the external performance index. They utilize a two-step heuristic algorithm that combines random sampling and local search to find an optimal combination of inputs and outputs.

Morita and Avkiran [13] considered the criteria selection method based on discriminant analysis using external evaluation. They used a 3-level orthogonal layout experiment to find an appropriate combination of inputs and outputs, where experiments are independent of each other.

Lim [11] proposed a method for selection of better combinations of input-output factors. It is designed to select better combinations of input-output factors that are well suited for evaluating substantial performance of DMUs. Several selected DEA models with different combinations of input-output factors are evaluated, and the relationship between the computed efficiency scores and a single performance criterion of DMUs is investigated using decision tree. Based on the results of decision tree analysis, a relatively better DEA model can be chosen, which is expected to effectively assess the true performance of DMUs.

3 CONJOINT ANALYSIS

Conjoint analysis is a multivariate technique that can be used to understand how individual's preferences for products or services are developed [9]. Specifically, Conjoint analysis is used to gain insights into how customers value various product attributes based on their valuation of the complete product. Green and Krieger [7] pointed out the potential usefulness of Conjoint analysis to deal with some marketing problems, in particular to develop new multi-attribute products with optimal utility levels over other competitive products, to estimate market shares in alternative competitive scenarios, to benefit segmentation, and to design promotion strategies, among other uses.

Implementation of Conjoint analysis can be simply described as follows. Researchers at first develop a set of alternative products (real or hypothetical) in terms of bundles of quantitative and qualitative attributes through fractional factorial designs. These products, referred to as profiles, are then presented to the customers during the survey. The customers are asked to rank order or rate these alternatives, or choose the best one. Because the products are represented in terms of bundles of attributes at mixed "good" and "bad" levels, the customers have to evaluate the total utility from all of the attribute levels simultaneously to make their judgments. Based on these judgments, the researchers can estimate the part-worths for the attribute levels by assuming certain composition rules. The manner that respondents combine the part-worths of attribute levels in total utility of product can be explained by these roles. The simplest and most commonly used model is the linear additive model. This model assumes that the overall utility derived from any combination of attributes of a given good or service is obtained from the sum of the separate part-worths of the attributes. Thus, respondent i's predicted utility for profile j can be specified as follows:

$$U_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{L_k} \beta_{ikl} x_{jkl} + \varepsilon_{ij}, \quad i = 1,...,I, \quad j = 1,...,J \tag{2}$$

where I is the number of respondents; J is the number of profiles; K is the number of attributes; L_k is the number of levels of attribute k. β_{ikl} is respondent i's utility with respect to level l of attribute k. x_{jkl} is such a (0,1) variable that it equals 1 if profile j has attribute k at level l, otherwise it equals 0. ε_{ij} is a stochastic error term.

The parameters β_{ikl}, also known as part-worth utilities, can be used to establish a number of things. Firstly, the value of these coefficients indicates the amount of any effect that an attribute has on overall utility of the profiles– the larger the coefficient, the greater the impact. Secondly, part-worths can be used for preference-based segmentation. Namely, given that part worth utilities are calculated at the individual level, if preference heterogeneity is present, the researcher can find it. Respondents who place similar value to the various attribute levels will be grouped together into a segment. Thirdly, part-worths can be used to calculate the relative importance of each attribute, also known as an importance value.

Importance values are calculated by taking the utility range for each attribute separately, and then dividing it by the sum of the utility ranges for all of the factors:

$$FI_{ik} = \frac{\max\{\beta_{ik1}, \beta_{ik2}, ...\beta_{ikL_k}\} - \min\{\beta_{ik1}, \beta_{ik2}, ...\beta_{ikL_k}\}}{\sum_{k=1}^{K}\left(\max\{\beta_{ik1}, \beta_{ik2}, ...\beta_{ikL_k}\} - \min\{\beta_{ik1}, \beta_{ik2}, ...\beta_{ikL_k}\}\right)}, \quad i = 1,...,I, k = 1,...,K \tag{3}$$

where FI_{ik} is the relative importance that ith respondent assigned to the factor k. The results are then averaged to include all of the respondents:

$$FI_k = \sum_{i=1}^{I} FI_{ik} \Big/ I, \quad k = 1,...,K. \tag{4}$$

If the market is characterized by heterogeneous customer preferences, it is possible to determine the importance of each attribute for each of isolated market segments.

4 DEA CRITERIA SELECTION USING CONJOINT ANALYSIS

Here is a framework designed to help the selection of the criteria relevant to the analysis and to obtain results that best reflect the research objectives. The adequate criteria selection is especially significant in the case of non-profit sector as well as for the studies that include a large number of categorical (discrete, intangible) variables.

The main phases of proposed framework are shown in Figure 1. In the first phase, the research objectives and stakeholders should be defined. In the second phase, Conjoint analysis should be conducted. Based on the objectives defined in the first stage, selection of the key attributes and their levels are performed. After the data are collected, Conjoint parameters estimation should be done, as described in Section 4. In the real-world applications, the parameters obtained from Conjoint analysis shows respondents' preferences to the particular criteria.

One of the important objectives of Conjoint analysis is to determine what combination of a limited number of criteria is most influential on respondent choice or decision making. Particularly, FI_k, $k = 1,.., K$, represents the importance of each criterion, which may be starting point for DEA criteria selection (see Fig. 1).

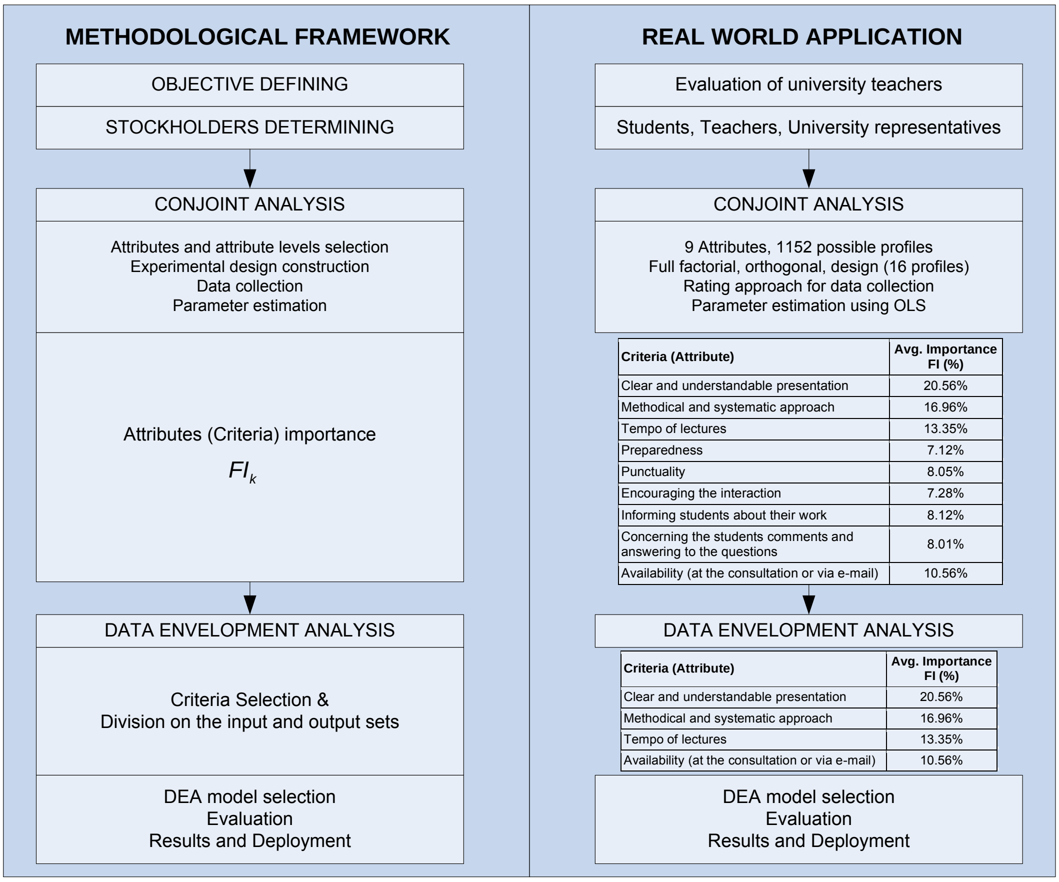


Figure 1: Model of criteria selection in DEA using Conjoint data

The illustrative example is given in the right part of the Fig. 1. The study is motivated by teachers' evaluation on the University of Belgrade, Serbia. The survey, usually, carries out twice a year and results are meant to represent students' standpoint. But the final estimation of each teacher is calculated as average of mean values of 9 criteria given as attributes set up by the University representatives. A survey, partially presented here, is conducted with 98 undergraduate students. The results obtained by Conjoint analysis shows that some criteria, such as *Clear and understandable presentation* is far more significant than the others. Four criteria, shown in DATA ENVELOPMENT ANALYSIS rectangle, are distinguished as the most important from the students' point of view. These criteria are going to be used as DEA outputs, since marks are desirable to be as higher as possible.

5 CONCLUSION

The aim of each entity is to provide the most reliable, useful and inexpensive business analysis. It can be DEA which can help the management to decrease cost, to make processes easier and to focusing on the key business competencies. DEA is an effective tool for evaluating and managing operational performance in a wide variety of settings. Since DEA gives different indexes of efficiencies with different combination of criteria, the selection of inputs and outputs is one of the most important steps in DEA.

A large number of criteria require a great effort to obtain efficiency index of each DMU. DEA efficiency index is relative measure, depends on the number of DMUs and number and structure of criteria included into the analysis. The criteria reduction has been usually done by statistical methods, such as regression and correlation analysis. This paper has suggested using Conjoint analysis as supporting tool for more realistic criteria selection. The stakeholders' preferences obtained by Conjoint analysis represents starting point for making the most suitable combination of criteria used in next phase of DEA efficiency measurement.

The described framework provides better criteria selection which is well suited to the stakeholders and allows selection of different criteria combination suited to the different objectives and the number of DMUs. This paper also provides some interesting and promising lines for further research.

References

[1] Ahn, T., Charnes, A., Cooper, W. W. 1988. Efficiency Characterizations in Different DEA Models. Socio-Economic Planning Sciences, 22 (6), 253-257.

[2] Banker, R., Charnes, A., Cooper, W. W. 1984. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science, 30 (9), 1078-1092.

[3] Banker, R. D., Morey, R. C. 1986. The Use of Categorical Variables in Data Envelopment Analysis. Management Science, 32 (12), 1613-1627.

[4] Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring efficiency of decision making units. European Journal of Operations Research 2, 429-44.

[5] Edirisinghe, N.C.P., Zhang, X. 2007. Generalized DEA model of fundamental analysis and its application to portfolio optimization. Journal of Banking & Finance, 31, 3311–3335.

[6] Emrouznejad, A., Witte, K. 2010. COOPER-framework: A unified process for non-parametric projects. European Journal of Operational Research, 207, 1573–1586

[7] Green, P.E., Krieger, A.M. 1991. Segmenting markets with conjoint analysis. Journal of Marketing, 55, 20–31.

[8] Golany, B., Roll, Y. 1989. An application procedure for DEA. OMEGA International Journal of Management Science 17 (3), 237–250.

[9] Hair, J. F., Anderson, R. E., Tathan, R. L., and Black, W. C. 1995. Multivariate Data Analysis, Englewood Cliffs, NJ: Prentice Hall.

[10] Jenkins, L., Anderson, M., 2003. A multivariate statistical approach to reducing the number of variables in data envelopment analysis. European Journal of Operational Research 147, 51–61.

[11] Lim, S., 2008. A Decision Tree-Based Method for Selection of Input-Output Factors in DEA. Proceedings of the 2008 International Conference on Machine Learning; Models, Technologies and Applications, MLMTA'08, July 14-17, 2008, Las Vegas, Nevada, USA, 762-769.

[12] Luo, X. 2003. Evaluating the Profitability and Marketability Efficiency of Large Banks - an Application of Data Envelopment Analysis. Journal of Business Research, 56 (8), 627-635.

[13] Morita, H., Avkiran, K. N. 2009. Selecting inputs and outputs in data envelopment analysis by designing statistical experiments. Journal of the Operations Research, Society of Japan, Vol. 52, No.2, 163-173

[14] Morita, H., Haba, Y. 2005. Variable selection in data envelopment analysis based on external information. Proceedings of the eighth Czech-Japan Seminar on Data Analysis and Decision Making under Uncertainty, 181–187.

[15] Nunamaker, T.R., 1985. Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation. Managerial and Decision Economics 6 (1), 50–58.

INTERVAL COMPARISON MATRICES IN GROUP AHP

Petra Grošelj and **Lidija Zadnik Stirn**
University of Ljubljana, Biotechnical Faculty
Jamnikarjeva 101, 1000 Ljubljana, Slovenia
petra.groselj@bf.uni-lj.si, lidija.zadnik@bf.uni-lj.si

Abstract: Analytic hierarchy process is a well-known approach for handling multi-criteria decision making problems. In group decision making the main difficulty is aggregating individual preferences into joint weights, so that the weights satisfy all decision makers as much as possible. One possibility, when the group cannot reach a consensus on a single judgment is to express it with an interval. In the paper a new way of constructing the interval comparison matrix from individual judgments is proposed. The problem of generating weights from interval comparison matrices is discussed and a numerical example from the field of natural resources management is provided.

Keywords: multiple criteria decision making; group decision making; analytic hierarchy process; interval judgments; management of natural resources

1 INTRODUCTION

Analytic hierarchy process (AHP) is a well-known approach for handling multi-criteria decision making problems. It is based on pairwise comparisons. The 1-9 ratio scale is used for expressing the strength of preference between the compared objects. Since one decision maker is limited by his/her knowledge, experiences and perspective, group decision making is often applied inside the AHP model. In group decision making the main difficulty is aggregating individual preferences into joint weights, so that the weights satisfy all decision makers as much as possible. Many approaches have been applied, including aggregating individual judgments or priorities [1] and aggregating the base of data envelopment analysis concepts [2].

In one decision maker’s case intervals can be used instead of the crisp values employed in pairwise comparisons. The exact values sometimes cannot express the subjectivity and the lack of information of decision maker, or the complexity and uncertainty of the real world decision problems. Interval judgments are more natural in such cases. Another possibility for interval judgments arises in group decision making when the group cannot reach a consensus on a single judgment and expresses it with an interval [3]. Several analysts have examined the problem of generating weights from interval comparison matrices recently. A derived priority vector can be a vector of interval weights, the representative vector from the assurance region or a weakly efficient vector of crisp weights in the case when assurance region is empty [4].

In the paper we discuss the problem of combining individual judgments into group interval judgment and provide a new concept of aggregation. For deriving the interval weights from interval comparison matrix we employed the concept of Liu [5], which is based on constructing two crisp matrices from the interval matrix and deriving weights from them.

2 THE CONSTRUCTION OF GROUP INTERVAL MATRIX

Let m be the number of decision makers included in the process of evaluating n criteria (or alternatives) according to the element on the next higher level. Let $A^{(k)} = \left(a_{ij}^{(k)}\right)_{n\times n}$, $k=1,\dots,n$ be their comparison matrices. Group intervals can be constructed using minimum and maximum judgments for the endpoints of the intervals [6]. So the outstanding judgments determine the group intervals. Despite eliminating outliers, the intermediate values do not impact the endpoint of the intervals. We suggest another approach to overcome this drawback. The lower bound of the interval should be influenced by all values that are lower or equal to the median. Since it is the lower bound the influence should not be equal for all values. The degree of influence should be greater for smaller values and smaller for the values that are closer to the median. Similarly, the upper bound of the interval should be influenced by all values that are greater than or equal to the median. The simplest way for mathematical record of such approach is employing the Ordered Weighted Geometric (OWG) operator [7].

Definition 1: An OWG operator of dimension m is a mapping $F:\mathbb{R}^m \to \mathbb{R}$, that has associated a weighting vector $W=\left(w_1,...,w_m\right)$ having the properties:

$$w_i \in [0,1],\ \sum_{i=1}^{m} w_i = 1 \text{ and such that: } F\left(a_1,...,a_m\right)=\prod_{i=1}^{m} c_i^{w_i} \tag{1}$$

where c_i is the ith largest value from the set $\left\{a_1,...,a_m\right\}$.

We selected the OWG operator since it preserves reciprocity. Different vectors W assign different weights to the values $a_1,...,a_m$. We assume that all decision makers are equally important.

We define two vectors $W_L=\left(w_1^L,...,w_m^L\right)$ and $W_U=\left(w_1^U,...,w_m^U\right)$ for the lower and upper bounds of the intervals, respectively. The description is made separately for even or odd number of decision makers. If m is an odd number, then $\frac{m+1}{2}$ is the median of numbers $1,2,...,m$ and $s_{\frac{m+1}{2}}=\frac{(m+1)(m+3)}{8}$ is the sum of numbers from 1 to $\frac{m+1}{2}$. Then we employ the judgments of decision makers that are smaller than or equal to the median to influence the lower bound of the group interval. The judgments of decision makers that are greater than or equal to the median influence the upper bound of the group interval:

$$W_L^{odd}=\left(\underbrace{0,...,0}_{\frac{m-1}{2}},\frac{1}{s_{\frac{m+1}{2}}},\frac{2}{s_{\frac{m+1}{2}}},...,\frac{\frac{m-1}{2}}{s_{\frac{m+1}{2}}},\frac{\frac{m+1}{2}}{s_{\frac{m+1}{2}}}\right) \text{ and } W_U^{odd}=\left(\frac{\frac{m+1}{2}}{s_{\frac{m+1}{2}}},\frac{\frac{m-1}{2}}{s_{\frac{m+1}{2}}},...,\frac{2}{s_{\frac{m+1}{2}}},\frac{1}{s_{\frac{m+1}{2}}},\underbrace{0,...,0}_{\frac{m-1}{2}}\right). \tag{2}$$

If n is an even number, then median of numbers $1,2,...,m$ is not an integer and $s_{\frac{m}{2}}=\frac{m(m+2)}{8}$ is the sum of numbers from 1 to $\frac{n}{2}$, which are smaller than median. Then

$$W_L^{even}=\left(\underbrace{0,...,0}_{\frac{m}{2}},\frac{1}{s_{\frac{m}{2}}},\frac{2}{s_{\frac{m}{2}}},...,\frac{\frac{m-2}{2}}{s_{\frac{m}{2}}},\frac{\frac{m}{2}}{s_{\frac{m}{2}}}\right) \text{ and } W_U^{even}=\left(\frac{\frac{m}{2}}{s_{\frac{m}{2}}},\frac{\frac{m-2}{2}}{s_{\frac{m}{2}}},...,\frac{2}{s_{\frac{m}{2}}},\frac{1}{s_{\frac{m}{2}}},\underbrace{0,...,0}_{\frac{m}{2}}\right). \tag{3}$$

Then the aggregated interval group matrix A^{group} is defined as

$$A^{group}=\begin{bmatrix} 1 & \left[\prod_{k=1}^{m}\left(c_{12}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{12}^{(k)}\right)^{w_k^U}\right] & \cdots & \left[\prod_{k=1}^{m}\left(c_{1n}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{1n}^{(k)}\right)^{w_k^U}\right] \\ \left[\prod_{k=1}^{m}\left(c_{21}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{21}^{(k)}\right)^{w_k^U}\right] & 1 & \cdots & \left[\prod_{k=1}^{m}\left(c_{2n}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{2n}^{(k)}\right)^{w_k^U}\right] \\ \vdots & \vdots & \ddots & \vdots \\ \left[\prod_{k=1}^{m}\left(c_{n1}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{n1}^{(k)}\right)^{w_k^U}\right] & \left[\prod_{k=1}^{m}\left(c_{n2}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{n2}^{(k)}\right)^{w_k^U}\right] & \cdots & 1 \end{bmatrix} \tag{4}$$

where $c_{ij}^{(k)}$ is the kth largest value from the set $\left\{a_{ij}^1,...,a_{ij}^m\right\}$.

3 DERIVING INTERVAL WEIGHTS FROM INTERVAL COMPARISON MATRIX

For deriving interval weights from interval comparison matrix A^{group} we use the approach of separating A^{group} into two crisp comparison matrices $A_L^{group}=\left(a_{ij}^L\right)$ and $A_U^{group}=\left(a_{ij}^U\right)$ [5].

Let $\left[l_{ij},u_{ij}\right]:=\left[\prod_{k=1}^{m}\left(c_{ij}^{(k)}\right)^{w_k^L},\prod_{k=1}^{m}\left(c_{ij}^{(k)}\right)^{w_k^U}\right]$ for $i,j=1,...,n$. Then

$$a_{ij}^L=\begin{cases} l_{ij}, & i<j \\ 1, & i=j, \\ u_{ij}, & i>j \end{cases} \qquad a_{ij}^U=\begin{cases} u_{ij}, & i<j \\ 1, & i=j. \\ l_{ij}, & i>j \end{cases} \tag{5}$$

Matrices A_L^{group} and A_U^{group} are reciprocal comparison matrices, since the OWG operator preserves reciprocity.

The weights can be obtained from A_L^{group} and A_U^{group} in many ways, which are suitable for crisp comparison matrices. We employed the eigenvector method [8], where the priority vector, derived from a comparison matrix A, is the eigenvector belonging to the maximal eigenvalue of matrix A. This is the most commonly used method for deriving weights in AHP. The results are the vectors $\omega_{A_L}=\left(\omega_1^{A_L},...,\omega_n^{A_L}\right)$ and $\omega_{A_U}=\left(\omega_1^{A_U},...,\omega_n^{A_U}\right)$ for matrices A_L^{group} and A_U^{group}, respectively. The interval weights belonging to A^{group} are defined as

$$\omega_i=\left[\omega_i^L,\omega_i^U\right]=\left[\min\left\{\omega_i^{A_L},\omega_i^{A_U}\right\},\max\left\{\omega_i^{A_L},\omega_i^{A_U}\right\}\right]. \tag{6}$$

For ranking interval weights the matrix of degrees of preference could be used:

$$P=\begin{bmatrix} - & p_{12} & \cdots & p_{1n} \\ p_{21} & - & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & - \end{bmatrix} \tag{7}$$

In recent years the possibility- degree formula for p_{ij} has been used several times [9-12]:

$$p_{ij}=P\left(\omega_i>\omega_j\right)=\frac{\max\left\{0,\omega_i^U-\omega_j^L\right\}-\max\left\{0,\omega_i^L-\omega_j^U\right\}}{(\omega_i^U-\omega_i^L)+(\omega_j^U-\omega_j^L)},\ i,j=1,...,n,\ i\neq j \tag{8}$$

The preference ranking order is provided using row-column elimination method [10].

4 CASE STUDY

Natura 2000 is a European network of ecologically significant areas of nature, as specified on the basis of the EU Bird and Habitat Directives. The Slovenian Decree of special protected areas [13] directed that the Natura 2000 sites will be managed only throughout sectorial management plans. The agricultural priorities are outlined in the Rural

Development Programme of the Republic of Slovenia 2007 – 2013 [14] and the objectives are divided on four axes:

1. Axis 1 – improving the competitiveness of the agricultural and forestry sector
   The activities under the first axis should support modernization and innovations and raise the qualification and competitive position. They should contribute to improved employment possibilities, increased productivity, and added value in agriculture and forestry.
2. Axis 2 – improving the environment and rural areas
   The activities under the second axis should contribute to environmental and water resource protection, conservation of natural resources, and implementation of nature friendly technologies in agriculture and forestry. They should provide sustainable development of rural areas and ensure a favorable biodiversity status and the preservation of habitats in the Natura 2000 sites.
3. Axis 3 – the quality of life in the rural areas and diversification of the rural economy
   The activities under the third axis promote entrepreneurship and raise the quality of life in rural areas through enhanced employment opportunities, rural economic development, and natural and cultural heritage conservation.
4. Axis 4 – LEADER initiative
   LEADER is a bottom - up method of delivering support for rural development through implementing local development strategies. The activities under the fourth axis should stimulate the cooperation and connection of local action groups (LAS).

To assure the best results for the Natura 2000 sites we should rank these four aims and seek a balance between them. The weighting depends on the view of the objectives which differ among stakeholders. With an objective of incorporating different perspectives, we identified three main stakeholders for the Natura 2000 sites: representatives of environmental protection, representatives of farmers, and the government. The pairwise comparisons of the four objectives and results are represented by matrices A, B and C for environmental, farmer, and government views, respectively.

$$A=\begin{bmatrix}1 & \frac{1}{4} & \frac{1}{3} & 2\\ 4 & 1 & 2 & 3\\ 3 & \frac{1}{2} & 1 & 2\\ \frac{1}{2} & \frac{1}{3} & \frac{1}{2} & 1\end{bmatrix},\ B=\begin{bmatrix}1 & 3 & 1 & 1\\ \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{2}\\ 1 & 3 & 1 & 2\\ 1 & 2 & \frac{1}{2} & 1\end{bmatrix},\ C=\begin{bmatrix}1 & \frac{1}{2} & 3 & 8\\ 2 & 1 & 4 & 6\\ \frac{1}{3} & \frac{1}{4} & 1 & 3\\ \frac{1}{8} & \frac{1}{6} & \frac{1}{3} & 1\end{bmatrix}$$

The priorities of the three stakeholders, gained by the eigenvector method are presented in Table 1.

Table 1: The priorities and the ranking of the four axes for three stakeholders.

	A environmantalist		B farmer		C government	
	priorities	ranks	priorities	ranks	priorities	ranks
axis 1	0.1397	3	0.3015	2	0.3372	2
axis 2	0.4647	1	0.1100	4	0.4832	1
axis 3	0.2799	2	0.3584	1	0.1265	3
axis 4	0.1156	4	0.2301	3	0.0531	4

The ranking differs between the stakeholders. The environmentalist prefers axis 2, which is the most favors nature protection. Farmers favor rural economic development, which is reflect by axis 3 and 1, which are the highest evaluated. The government weights indicate that it is focused on sharing funds for particular objectives.

The comparison matrices *A*, *B* and *C* are aggregated in the A^{group} (4) according to the OWG operator (1). The associated lower and upper weighted vectors (2) are defined as $W_L = \left(0, \frac{1}{3}, \frac{2}{3}\right)$ and $W_U = \left(\frac{2}{3}, \frac{1}{3}, 0\right)$. The intervals in the matrix A^{group} (9) are presented on four decimals.

$$A^{group} = \begin{bmatrix} 1 & [0.3150, 1.6510] & [0.4807, 2.0801] & [1.2599, 5.0397] \\ [0.6057, 3.1748] & 1 & [0.6057, 3.1748] & [0.9086, 4.7622] \\ [0.4807, 2.0801] & [0.3150, 1.6510] & 1 & [2.0000, 2.6207] \\ [0.1984, 0.7937] & [0.2100, 1.1006] & [0.3816, 0.5000] & 1 \end{bmatrix} \quad (9)$$

The interval weights arising from the A^{group} are composed from the eigenvectors belonging to the maximal eigenvalues of matrices A_L^{group} and A_U^{group} by the equation (6):

$$\omega = \begin{bmatrix} [0.158, 0.414] \\ [0.281, 0.355] \\ [0.161, 0.368] \\ [0.071, 0.193] \end{bmatrix} \quad (10)$$

Ranking of interval weights (10) has been done over the matrix of degrees of preference (7)

$$P = \begin{bmatrix} - & 0.403 & 0.547 & 0.909 \\ 0.597 & - & 0.692 & 1 \\ 0.453 & 0.308 & - & 0.902 \\ 0.091 & 0 & 0.098 & - \end{bmatrix}$$, which presents the ranking $\omega_2 \overset{59.7\%}{\succ} \omega_1 \overset{54.7\%}{\succ} \omega_3 \overset{90.2\%}{\succ} \omega_4$.

The final ranking sets axis 2 – the protection of nature - as the most important aim in managing Nature 2000 sites. This objective is expected to be the most vital. The second most significant aim is axis 1. Improving the competitiveness of the agricultural and forestry sectors will indirectly contribute to improved environmental, water, and air quality through new technologies and renewable energy sources [14]. Axis 3 is third in importance, but close to axis 2 and should upgrade, complement, and refine the effects of axes 1 and 2 [14]. Axis 4 is ranked last but should be included in implementing the other three axes.

5 CONCLUSIONS

In the paper we discussed the problem of aggregating individuals' judgments into group interval judgments, deriving interval weights from the interval comparison matrix, and the ranking of interval weights. The case study indicates that the group approach with interval matrices could be appropriate and contributes to managing group decision problems from different areas.

Additional issues must be addressed however:

- We assumed that all decision makers are equally important for our analysis. It will be interesting to construct a vector *W* in the OWG operator (1) that considers weights reflecting the importance of the various decision makers.
- The concept of acceptable consistency which is important in AHP should be included.

- Finally, if every decision maker in the group provides an interval comparison matrix, their aggregation is even more complicated as in the case of crisp matrices. Some approaches have already been offered for this issue [15-16] but here still a lot of space remains for new ideas.

References

[1] Forman, E. and Peniwati, K., 1998. Aggregating individual judgments and priorities with the analytic hierarchy process. European Journal of Operational Research, 108(1), pp. 165-169.

[2] Grošelj, P., Pezdevšek Malovrh, Š. and Zadnik Stirn, L., 2011. Methods based on data envelopment analysis for deriving group priorities in analytic hierarchy process. Central European Journal of Operations Research, In press, pp. 1-18.

[3] Arbel, A. and Vargas, L., 2007. Interval judgments and Euclidean centers. Mathematical and Computer Modelling, 46(7-8), pp. 976-984.

[4] Conde, E. and de la Paz Rivera Pérez, M., 2010. A linear optimization problem to derive relative weights using an interval judgement matrix. European Journal of Operational Research, 201(2), pp. 537-544.

[5] Liu, F., 2009. Acceptable consistency analysis of interval reciprocal comparison matrices. Fuzzy Sets and Systems, 160(18), pp. 2686-2700.

[6] Chandran, B., Golden, B. and Wasil, E., 2005. Linear programming models for estimating weights in the analytic hierarchy process. Computers & Operations Research, 32(9), pp. 2235-2254.

[7] Chiclana, F., Herrera, F. and Herrera-viedma, E., 2000. The Ordered Weighted Geometric Operator: Properties and Application. Proc. 8th Conf. Inform. Processing and Management of Uncertainty in Knowledgebased Systems (IPMU), pp. 985-991.

[8] Saaty, T. L., 2006 Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process Second Edition Second Printing ed RWS Publications, Pittsburgh.

[9] Facchinetti, G., Ricci, R. G. and Muzzioli, S., 1998. Note on ranking fuzzy triangular numbers. International Journal of Intelligent Systems, 13(7), pp. 613-622.

[10] Wang, Y.-M., Yang, J.-B. and Xu, D.-L., 2005. A two-stage logarithmic goal programming method for generating weights from interval comparison matrices. Fuzzy Sets and Systems, 152(3), pp. 475-498.

[11] Xu, Z. and Chen, J., 2008. Some models for deriving the priority weights from interval fuzzy preference relations. European Journal of Operational Research, 184(1), pp. 266-280.

[12] Xu, Z. S. and Da, Q. L., 2002. The uncertain OWA operator. International Journal of Intelligent Systems, 17(6), pp. 569-575.

[13] UL. (2004) Uredba o posebnih varstvenih območjih - območjih Natura 2000 (Slovenian Decree of special protected areas).

[14] PRP. (2007) Program razvoja podeželja Republike Slovenije za obdobje 2007 - 2013 (Rural Development Programme of the Republic of Slovenia 2007 – 2013).

[15] Entani, T. and Inuiguchi, M. (2010) Group Decisions in Interval AHP Based on Interval Regression Analysis. In Huynh, V.-N., Nakamori, Y., Lawry, J. and Inuiguchi, M. (eds), Integrated Uncertainty Management and Applications. Springer Berlin / Heidelberg.

[16] Yang, W., Ge, Y. and Xiong, B., 2010. Group Decision Support Systems with Probability Density Functions for Deriving Priorities from Consistent Interval Judgments. Journal of Computational Information Systems, 6(14), pp. 4811-4818.

A GOAL PROGRAMMING APPROACH TO RANKING BANKS

Višnja Vojvodić Rosenzweig
Ekonomski fakultet u Zagrebu
Kennedyjev trg 6, 10000 Zagreb
Phone: ++385 1 2383 333; E-mail: vvojvodic@efzg.hr

Hrvoje Volarević
Zagrebačka škola ekonomije i managementa
Jordanovac 110, 10000 Zagreb
Phone: ++385 1 2354 151; E-mail: hrvoje.volarevic@zsem.hr

Mario Varović
Zagrebačka škola ekonomije i managementa
Jordanovac 110, 10000 Zagreb
Phone: ++385 1 2354 151; E-mail: mario.varovic@zsem.hr

Abstract: Ranking of commercial banks based on seven proposed criteria is performed by using goal programming, in which the goal of every bank is the best business performance (evaluated with multiple criteria), and which is represented by a Score. The Score is obtained by calculating weights as a solution of a goal programming problem. Profitability indicators are the most important indicators for the five observed Croatian banks. Other indicators, for credit risk and productivity, are far less important for the final ranking of the chosen banks.

Key words: Commercial banks, Multi-criteria ranking, Goal programming, Business performance.

1. INTRODUCTION

Banks play an extremely important role in each country's economy, particularly in countries with a rather less developed financial system, as is the case with the Republic of Croatia. The banking sector in the Republic of Croatia consists of thirty banks that are mostly owned by foreign proprietors, generally by Italian, Austrian, French, and Hungarian banks. The dominant position, based on their total assets and the size of equity, is occupied by two largest Croatian Banks, Zagrebačka banka d.d. and Privredna banka Zagreb d.d. In addition to these, the top ten Croatian banks also include Erste & Steiermarkische bank d.d., Raiffeisenbank Austria d.d., Hypo-Alpe-Adria-bank d.d., Societe Generale - Splitska banka d.d., Hrvatska poštanska banka d.d., OTP banka d.d., Volksbank d.d., and Podravska banka d.d. The scope of this study encompasses the following five banks: Erste & Steiermarkische bank (ERSTE), Raiffeisenbank Austria (RBA), Hypo-Alpe-Adria-bank (HYPO), Hrvatska poštanska banka (HPB) and Podravska banka (POBA).

These banks were chosen primarily because of their comparability with regard to the criteria of total assets and size of equity, as well as for the online availability of their annual reports with financial statements for the year 2010. Moreover, because of the fact that only one out of the five - HPB bank has domestic (Croatian) ownership, these five banks represent a representative sample for the Croatian banking sector. The two largest banks that participate in over 50% of the Croatian banking sector are excluded from analysis since their results would not be comparable with the financial position of the other banks studied. Particular emphasis will be put on the interpretation of the results relating to the HPB Bank, since it is the only large bank in Croatia owned by domestic capital, i.e. mainly a state-owned bank. The results of the analyses will imply certain conclusions and recommendations for the purpose of repositioning the HPB bank, but also other banks covered in the study, on the Croatian banking market.

A mathematical multicriteria decision making model will be used, that will consist of seven individual criteria classified into three basic groups - profitability, credit risk, and productivity. Multicriteria business performance of each bank will be evaluated using a score calculated as the weighted sum of relative values of individual indicators. There is an assumption that each bank goal is the maximum score that they wish to obtain. The score is dependent on the weights assigned to individual indicators. The deviation from the goal will be measured using two distance functions. The formulated mathematical model uses goal programming to determine the weights and the score for each bank. This approach is used in paper [6]; however, in that paper the goal of each bank is the score closest to the performance of all indicators, which will not be the case here.

The rest of this paper is presented in the following manner. All seven criteria are presented in the second section, followed by formulation of the multicriteria optimalisation model in the third. The approach to solving this kind of a model is illustrated in the fourth section on the basis of the examples that include five banks and seven selected attributes (criteria). The closing considerations are presented in the final section of this paper.

2. SELECTION OF CRITERIA

Ranking of commercial banks is a classic problem of multicriteria decision-making. In the first place, it is necessary to select the criteria on the basis of the ranking of the banks in a descending order (from the best to the worst). In this paper seven individual criteria have been chosen, categorized in three fundamental groups (profitability, credit risk, and productivity) as follows:

1. Return on average assets – ROAA represents one of the most well-known indicators of profitability that is often used not only in the banking sector, but also in the real sector. The value of this indicator is obtained from the next relation:

X_1 = Return on average assets (ROAA) = profit before taxation / average assets of the bank (1)

Profit before taxation can be found in the Income statement (P&L), while the average assets of the bank are calculated as the arithmetical mean of the balance sheet's positions on the asset side for two consecutive business years (in this case for the years 2009 and 2010). The obtained values are expressed as percentages, and are desirable to be as high as possible for each bank.

2. Return on average equity – ROAE also represents a well-known profitability indicator, as well as Return on average assets. The value of this indicator is obtained as follows:

X_2 = Return on average equity (ROAE) = profit after taxation / average equity of the bank (2)

Profit after taxation is the final entry of the Income statement, while the average equity of the bank is calculated in the same way as the average assets of the bank (arithmetical mean of the balance sheet's positions of the equity for the two sequential business years). The obtained values are also expressed as percentages, and are desirable to be as high as possible for each bank.

3. Income from interest bearing assets and expenses on interest bearing liabilities represents a specific indicator of profitability that is solely applied to the banking sector. The value of this indicator is obtained as follows:

X_3 = Income from interest bearing assets and expenses on interest bearing liabilities = (interest income / average interest bearing assets) / (interest expenses / average interest bearing liabilities) (3)

Interest income and interest expenses represent the initial positions in the Income statement of every business bank because they define the financial result that is derived from basic banking activity - receiving deposits and lending loans. Interest bearing assets are the total of all positions on the asset side of the balance sheet that represent the ground for calculating active interest, by which banks' income is generated. On the other hand, interest bearing liabilities are the total of all positions on the liability side of the balance sheet as the ground for calculating passive interest that make banks' expenditures. The obtained values are expressed as absolute values and it is desirable that the obtained results of this ratio be as high as possible in order to confirm the profitability of banks dealings.

4. Coverage represents the indicator commonly used in banks for credit risk evaluation. The value of this indicator is obtained as follows:

X_4 = Coverage = (total of investments impairment + provisions) / (total of investments + contingent liabilities) (4)

The numerator of the ratio consists of the total of investments impairment and provisions, where the impairment stands for the cumulative of all recognized losses for bad and doubtful loans that are not expected to be repaid, that is reimbursed, while the term provisions refers to the balance sheet position on the liability side that is recognized in the banks expenses as future observed and estimated liabilities (for example provisions for legal actions, that is lawsuits filed against the bank). The denominator of the ratio consists of the total of investments comprised divided by the total of all balance sheet positions on the asset side of the bank that represent the basis for generating income, and the other part of the denominator relates to contingent liabilities that are, as a rule, booked on the off-balance sheet, and consist of given guarantees and open letters of credit as typical banking affairs. The obtained values are expressed as percentages, and it is desirable that the obtained results of this ratio should be as high as possible, which implies that the bank management is aware of possible credit risk in business activities and of the necessity for its anticipation.

5. Quality of investments represents an indicator that pertains to the credit risk assessment, as well as coverage, because it assesses the percentage of bank investments that can be reimbursed. The value of this indicator is obtained as follows:

X_5 = Quality of investments = (1 – (total of investments impairment / total of investments)) (5)

The equation listed above puts in ratio two positions from the asset side of the bank's balance sheet. The obtained values are expressed as percentages and their maximum value is 100%, which means that all the bank's investments can be repaid and that there is no need for investment impairment. Taking into account the existing risk when making credit investments, this situation should not be expected to be realistic.

6. Assets per employee is a typical banking indicator that belongs to the category of productivity indicators because it represents the ratio of the realized output (total of assets, i.e. total bank’s property) against actors in bank business operations (which means all bank's employees). The value of this indicator is obtained as follows:

X_6 = Assets per employee = total assets / total number of employees (6)

The values in this equation are obtained from the balance sheet and the notes accompanying financial statements (information about the number of employees). The obtained values are expressed as absolute values, i.e. money units, and are desirable to be as high as possible.

7. Interest income per employee represents the banking indicator that also belongs to the category of productivity indicators. The value of this indicator is obtained as follows:

X_7 = Interest income per employee = Interest income / total number of employees (7)

The numerator of the ratio is obtained from the Income statement, while the denominator consists of the number of employees that can be found in the notes accompanying financial statements. The obtained values are also expressed as absolute values, i.e. money units, and are desirable to be as high as possible, just as with all the previous indicators.

Based on the former formulas, the calculated values of all seven individual criteria (X_1,...,X_7) for the five selected banks, and all the obtained results are presented in the following decision-making table (Tab. 1):

Table 1. The values of seven individual indicators (X_1, X_2, X_3, X_4, X_5, X_6 and X_7), categorized into three basic groups (profitability, credit risk, and productivity) for the five selected banks (ERSTE, HPB, HYPO, POBA and RBA).

BANK:	PROFITABILITY:			CREDIT RISK:		PRODUCTIVITY:	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1. ERSTE	1,52%	10,55%	2,26	4,37%	95,65%	26,17	1,51
2. HPB	0,40%	5,55%	2,05	5,44%	94,10%	14,61	0,81
3. HYPO	0,72%	3,56%	1,77	5,71%	93,90%	22,82	1,24
4. POBA	0,58%	3,52%	2,20	5,58%	94,53%	9,11	0,54
5. RBA	1,13%	6,77%	2,17	2,85%	96,97%	17,44	0,97

All obtained results of individual indicators are positively directed, but the benefit criteria are not displayed in the same measurement units. Therefore the next step is the transformation of the positively directed criteria values. The percentage transformation is used here as it leads to proportional changes in the results. The obtained results are presented in Table 2.

Table 2. The transformed values of seven individual criteria (X_1, X_2, X_3, X_4, X_5, X_6 and X_7) as part of the three basic groups (profitability, credit risk, and productivity) for the five selected banks (ERSTE, HPB, HYPO, POBA and RBA).

BANK:	PROFITABILITY:			CREDIT RISK:		PRODUCTIVITY:	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1. ERSTE	0,3508	0,3522	0,2163	0,1825	0,2013	0,2903	0,2979
2. HPB	0,0912	0,1854	0,1964	0,2271	0,1980	0,1621	0,1597
3. HYPO	0,1663	0,1189	0,1692	0,2383	0,1976	0,2531	0,2436
4. POBA	0,1326	0,1176	0,2108	0,2329	0,1989	0,1011	0,1071
5. RBA	0,2591	0,2259	0,2072	0,1191	0,2041	0,1934	0,1918

3. MULTICRITERIA PROBLEM AND GOAL PROGRAMMING

The weighted sum model is the most frequently used approach for the estimation of multicriteria performance of specific alternatives that are also used in this paper. To each bank i we assign score S_i based on the values of individual indicators (attributes) and weights assigned to them. The weights w_j of indicators j determine the score and by varying different weight different scores can be obtained for the same bank. Since the score of the alternative is its multicriteria value, it is assumed here that the goal of each bank is the maximum value of the score. In that sense the goal programming problem will be formulated. The notations in the model are as follows:

i - Bank, $i = 1,\ldots,n$.
j – Indicator (Attribute), $j = 1,\ldots,p$.
w_j *–Weight of Attribute j,* $j = 1,\ldots,p$.
x_{ij} *– Value of Indicator j of Alternative i.*
S_i *- Score Alternative i,* $S_i = w_1 x_{i1} + \ldots + w_p x_{ip}$.

As it was mentioned earlier, the goal for every bank i is the highest score, and therefore it is valid to define:

$$g_i = \max \{S_i (w): w_1 + \ldots + w_p = 1,\ w_1,\ldots,w_p \geq 0\} \tag{8}$$

If $d = (d_1,\ldots,d_n)$ represents a vector whose components d_i are deviations from components g_i of the goal $g = (g_1,\ldots,g_n)$, and S is vector $S = S(w) = (S_1,\ldots,S_n)$, the problem **(GP)** that we are solving is as follows:

$$\textbf{(GP)} \qquad \text{Min } \|g-S(w)\|_\alpha \tag{9}$$

With limitations:

$$S(w) + d = g,\ \ d \geq 0$$

$$w_1 + \ldots + w_p = 1$$

$$w_1,\ldots,w_p \geq 0$$

The solution of the problem depends on the selection of the norm i.e. on the values of the weights (w_j) of the goal programming problem (**GP**).

4. IMPLEMENTATION

The problem is solved for the five selected banks and the seven individual indicators. In this paper, the norm suggested by Dinckelbach and Isermann is used, as the first one:

$$\| g\text{-}S(w) \|_\alpha = \| g\text{-}S(w) \|_\infty + (1/\alpha)\| g\text{-}S(w) \|_1,\ \alpha \geq 1 \qquad (10)$$

The problem is solved for $\alpha = 1$, 10 and 100. For all mentioned values of parameter α, the same solution is obtained. The following weights for every individual criterion are obtained:

$w_1 = 0.3951$, $w_2 = 0.2235$, $w_3 = 0$, $w_4 = 0.3783$, $w_5 = 0$, $w_6 = 0$ and $w_7 = 0.032$.

The banks scores are (S_1 – ERSTE, S_2 – HPB, S_3 – HYPO, S_4 – POBA, S_5 – RBA):

$S_1 = 0.2855$, $S_2 = 0.1655$, $S_3 = 0.1855$, $S_4 = 0.1665$, $S_5 = 0.2001$.

Apart from using the Dinckelbach and Isermann's norm, the problem is also solved using the Euclid's norm in which the sum of square deviations is the smallest. The following weights are obtained for every individual criterion:

$w_1 = 0.24$, $w_2 = 0.22$, $w_3 = 0.19$, $w_4 = 0.22$, $w_5 = 0$, $w_6 = 0$ and $w_7 = 0.13$.

The banks' scores are (S_1 – ERSTE, S_2 – HPB, S_3 – HYPO, S_4 – POBA, S_5 – RBA):

$S_1 = 0.28$, $S_2 = 0.17$, $S_3 = 0.19$, $S_4 = 0.16$, $S_5 = 0.20$.

The results are rounded up to two decimal points, unlike the previous problem, since this is a square programming problem.

The final ranking list of the five selected banks for both norms we used is as follows:

I. Dinckelbach and Isermann's norm:	II. Euclid's norm:
ERSTE (S_1)	ERSTE (S_1)
RBA (S_5)	RBA (S_5)
HYPO (S_3)	HYPO (S_3)
POBA (S_4)	HPB (S_2)
HPB (S_2)	POBA (S_4)

As one can see from the obtained results, the score (S_i) of every bank is approximately the same regardless of the norm used in the model, and the ranking is approximately the same in both cases. The only difference in the ranking is between the two banks with the lowest rank (HPB and POBA); their rank changes according to the norm used.

Furthermore, in both cases the largest weights are assigned to profitability indicators (over 60%) while the weight of the fifth indicator equals zero because all the banks have approximately the same values of that indicator (quality of investments). Moreover, the weight of the sixth indicator (assets per employee) equals zero because its values are approximately the same as the values of the seventh indicator from the list of indicators (interest income per employee).

The first place of the ranking list is taken by a bank with moderate risk in business activities (ERSTE), while the bank with the highest risk in business activities (RBA) sits in the second place

On the other hand, HPB has small risk and small productivity, and therefore has small profitability, which puts the bank in the last or next to the last place in the total ranking (it changes places with POBA depending of the norm used). HYPO bank in both observed cases firmly holds the third position.

5. CONCLUSION

The commercial bank ranking problem can be efficiently solved with goal programming. The first step is to determine the criteria in advance, as the basis for executing multicriteria ranking and find the best business performance of the selected banks accordingly. The second step consists of using a goal programming mathematical model, in which the decision maker has the choice of using different norms. Two norms (Dinckelbach and Isermann, and Euklid's norm) are used in this paper, and the obtained results, weights, and scores are approximately the same in both cases. The obtained results for the five proposed banks suggest that the most important indicators in the model are profitability indicators, whose weights prevail in relation to the remaining two groups of indicators – credit risk and productivity – that have far less importance for the final bank ranking. This conclusion exclusively applies to the banking sector in the Republic of Croatia, while results might be different for some other countries and their banking markets [6].

Having analyzed the obtained score values for every bank selected in the model, it is beyond question that the two banks with the best score (ERSTE and RBA) have the adequate ratio for accomplished profitability and productivity, related to embedded risk in the business process. On the other hand, the same cannot be said for HPB and POBA that achieve just the opposite results, while HYPO is somewhere in between, which means there is room for improvement. HPB bank needs to improve its productivity and increase embedded risk in the business process. In that way, the bank ought to strengthen its market share in the Croatian banking sector, which would eventually lead to its repositioning regarding other banks. An alternative solution for HPB bank, as the only large bank in Croatia owned by domestic capital, would be referring to the possible recapitalization from its strategic partner, which should lead to necessary restructuring of its current business activity.

References

[1] Atrill, P., McLaney, E., 2006, Accounting and Finance for Non-Specialists, 5th edition, Prentice Hall, Harlow, England.

[2] Blocher, E., J., Chen, K., H., Lin, T., W., 2002, Cost Management: A Strategic Emphasis, McGraw-Hill/Irwin, New York, USA.

[3] Ehrgott, M., Klamroth, K., Schwehm, C., 2004, An MCDM approach to portfolio optimization, European Journal of Operational Research, Vol. 155, pp. 752-770.

[4] Feroz, E., H., Kim, S., Raab, R., L., 2003, Financial Statement Analysis: A Data Envelopment Analysis Approach, Journal of the Operational Research Society, Vol. 54, pp. 48-58.

[5] Gallizo, J., L., Jimenez, F., Salvador, M., 2003, Evaluating the effects of financial ratio adjustment in European financial statements, European Accounting Review, Vol. 12(2), pp. 357-377.

[6] Garcia, F., Guijarro F., Moya I., 2010, Ranking Spanish savings banks: A multicriteria approach, Mathematical and Computer Modeling, Vol.52, pp. 1058-106.

[7] Horngren, C., T., Oliver, M., S., 2010, Managerial Accounting, Upper Saddle River, New Jersey, Pearson Prentice Hall.

[8] Ng, W., L., 2007, An Efficient and simple model for multiple criteria supplier selection problem, European Journal of Operational Research, doi:10.1016/j.ejor.2007.01.018.

[9] Sawaragi, Y., Nakayama, H., Tanino, T., 1985, Theory of Multiobjective Optimization, Academic Press, Inc., Orlando, USA.

[10] Triantaphyyllou, E., 2000, Multi-Criteria Decision Making Methods: A Comparative Study, Kluwer Academic Publishers.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section VI:

Econometric Models and Statistics

SECTORIAL GROWTH DRIVERS OF WOOD PROCESSING AND FURNITURE MANUFACTURING IN CROATIA

Martina Basarac
Economic Research Division
Croatian Academy of Sciences and Arts
Strossmayerov trg 2, 10000 Zagreb, Croatia
mbasarac@hazu.hr

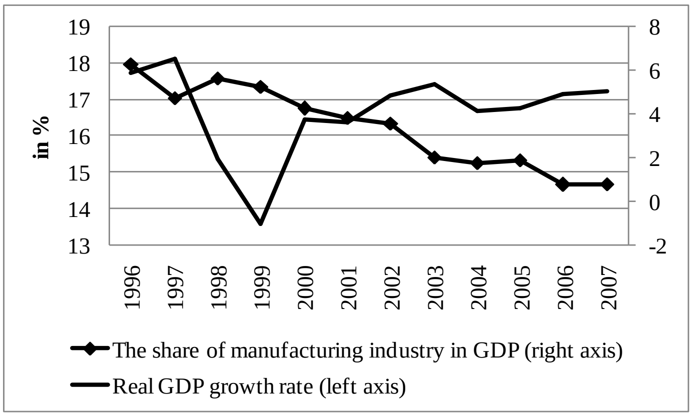
Abstract: Reducing the share of manufacturing industry in gross domestic product, even in the years of higher rates of economic growth, leads to a decrease in competitiveness of Croatian economy. Therefore, this paper investigates the trends in the wood processing industry and furniture manufacturing, as one of the oldest industries in Croatia. Using the econometric models, we estimate which macroeconomic indicators influence sectorial production. According to the multiple linear regression analysis results, one can conclude that production is affected by both real gross domestic product growth rates and unit labour costs. Real export values also play an important role.

Keywords: wood processing, manufacture of furniture, regression analysis, Croatia.

1 INTRODUCTION

As we slowly emerge from the first global recession since World War II, governments and businesses share an overarching aim – to steer their economies toward increasing competitiveness and growth [3]. In this effort, long-term economic growth and development of the country depend, among others, on the competitive performance of individual industries. The efficacy and competitiveness of these industries in turn depend on the relative wealth (and hence cost) of resources, as well as the actions and possibilities to use them in a seminal and sustainable way. Therefore, analysing specific sectors is the key to understanding competitiveness and growth.

Within this context, the issue of manufacturing production, as an important source of attractive jobs and export revenue, has frequently been addressed in the analysis of the Croatian economy. Nonetheless, numerous questions relating to the forest-based industries have not been adequately explored empirically (with the exception of [4] and [5]). The issue of scarce empirical analysis is even more pronounced since reducing the share of gross value added of manufacturing industry in gross domestic product (GDP), even in the years of higher rates of economy growth, leads to a decrease in Croatian economy competitiveness (see Figure 1).



Source: Croatian Bureau of Statistics – CBS (at the end of 2010 CBS released revised annual GDP data for period from 1995 to 2007).

Figure 1: The share of gross value added of manufacturing industry in gross domestic product, in period 1996-2007.

In that sense, a detail analysis of wood processing industry and furniture manufacturing, rather than looking at the aggregate macroeconomic level, will reveal notable insights. The study highlights the patterns and trends of the observed Croatian industries and shows the importance of these sectors, their major characteristics and level of international competitiveness. Furthermore, in order to address its driving factors, the current study on sectorial growth drivers aims to identify the key determinants of Croatian wood-based industries. Namely, many determinants, e.g. macroeconomic stability, affect industrial sectors and vary considerably between them, defining the environment within which industries operate. Therefore, it is important to focus on industry performance of output, because an industry analysis can contribute to the understanding of forces underlying competitiveness.

The following regression analysis fills the gap in literature on the determinants of industrial production and shows the importance of the forest-based industries in Croatian economy. Using the econometric models, we estimate which macroeconomic indicators (e.g. economic activity, unit labour costs and international trade) influence sectorial production and to what extent. The results show that these models explain the dependent variable quite well.

The remainder of this paper is organized as follows. In the next section we briefly discuss relevant characteristics of selected industries, addressing some important issues and general aspects of the wood-based industries. Third section provides a structural overview of the sectors. Fourth section gives a description of all variables used in models, as well as the analysis and interpretation of the estimated models. Finally, section five concludes.

2 WOOD PROCESSING INDUSTRY AND FURNITURE MANUFACTURING IN CROATIA

Almost 48% of Croatian territory is covered by woods and forests. As one of the oldest industry in Croatia, the wood processing is labour-intensive, low-technology sector dominated by small and medium-sized companies. Although it is a relatively small sector of the Croatian economy, it is a large employer. Furthermore, wood processing industry is the significant consumer of forestry products and supplies its products mainly to the furniture and the construction industry. On the other hand, as demonstrated by the advanced European Union countries, modern furniture industry is becoming more of a capital-intensive sector.

In general, wood processing and manufacture of furniture have always been significant export-oriented parts of Croatian economy. These sectors have been developed on high quality of forest raw material, long wood-processing tradition and good quality of human resources. In spite of tradition and prerequisites such as available infrastructure, long-term principle of sustainable management, labour and raw potential, past years show negative economic trends in certain macroeconomic indicators. These sectors should be competitive, profitable and internationally important, with high degree of post-processing products (especially in furniture), and with high share of value added. However, although the crisis started in the financial sector, its impact on the ‘real economy’ has now materialised; it has spread throughout the whole economy as all sectors are interconnected. Yet, this industry has been showing signs of deteriorating competitiveness even before the crisis started. In this respect, many governments are tempted to focus on emerging, innovative sectors as the key to their economies’ future competitiveness. Boosting the competitiveness of such sectors is not sufficient to sustain economy-wide growth in large, diversified economies [3]. Therein is the potential for enhancing the competitiveness of forest-based industries.

This study covers forest-based activities regarding the Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials (refer to Section C – Manufacturing, Division 16) and the Manufacture of furniture (refer to Section C – Manufacturing, Division 31). The activities are classified according to NACE Rev. 2, the statistical classification of economic activities which has been in force since 2007 in Croatia.

3 SECTORIAL PERFORMANCES

The evolution of Croatian production indices for wood and wood products manufacturing on the one hand, and furniture on the other, were relatively similar in the ten years through until 2007. For both of these activities, output growth was relatively strong in the period between 2000 and 2007 (albeit with a temporary fall in the output of furniture in 2005). The growth in output over the ten year period through until 2010 averaged 1.9% per annum for wood and wood products and 3.0 % per annum for furniture.

The workforce trends in wood processing register more than 19.000 workers in both industrial sectors, accompanied by increasing labour productivity and decreasing unit labour costs. However, in the nineties as well as before the global crisis, that number was about 28.000 workers (in 1997). Namely, the global financial crisis continued to significantly impact the Croatian forest sector in 2009 and 2010.

Furthermore, global demand shrank tremendously, including the demand for Croatia's forest products exports. As exports are important to its economy, Croatia's forest sector faced great challenges. According to data, we had the highest coverage of import by export for wood products in 2000 (204.6%). But, in 2007, 2008 and 2009 exports of most forest products have been seriously affected by the economic downturn (133.6, 129.0 and 134.8%).

Moreover, in the past few years, total exports exceeded imports owing to the increased export of raw wood and wood products, while the import of furniture still exceeds exports. In 2010, the most important export markets for wood products were Italy, Slovenia and Egypt to which more than 50% of product value has been exported. Although the Italy remained the largest export market for Croatian wood products (accounting for a 36% share of Croatian exports) in 2010, this share recorded a sharp decrease relative to the corresponding value for the year 1997 (i.e. 57%). On the other hand, the most important export markets for furniture products were Germany, Italy and Slovakia (in 2010).

The indicators presented encompass key dimensions of industrial performance and the relevant characteristics of wood processing and manufacture of furniture in Croatia.

4 EMPIRICAL ANALYSES

Since there has been conducted only a few research on macroeconomic drivers of sector growth, we adjust the list of variables to be considered to include those that will be relevant for a wood processing and manufacture of furniture growth study. The general idea is to concentrate on three measures – economic activity, unit labour costs and international trade. Furthermore, we expect to find that two variables (real GDP growth rate and export) boost industrial production in wood processing and manufacture of furniture, while an increase of unit labour costs (as a proxy for cost competitiveness) is expected to have negative impact on production.

4.1 Data and Methodology

Data used in the analysis encompass the period from the first quarter 2000 to the second quarter 2010, providing altogether 42 observations. The source of the data and thus the construction of the variables are based on official data as published by the Croatian Bureau of Statistics [1].

In a small open economy in transition, such as Croatia, performance in international trade plays an important role. The value of exports was deflated using the exchange rate. Data for unit labour costs were obtained by multiplying the average gross wage and the number of persons employed in observed industries, and then divided by industrial production. Economic activity is proxied by real GDP growth rate. All variables are expressed in indices, 2000=100 (except the real GDP growth rate).

The described approach enables us to investigate the microeconomic dynamics behind growth in observed sectors and to analyze the impact of different macroeconomic variables on industrial production. All series are seasonally adjusted and expressed in logarithms (except real GDP growth rate). Before the regression was specified, in order to avoid spurious regression, all of the time series were tested for the presence of the unit root. If the mean and variance are constant over time, then the series is stationary. On the other hand, if the mean and variance change over time, the series is non-stationary and it should be transformed to stationary ones by taking the first difference. A series that has stationary first differences is I(1) or integrated of order 1. In order to analyze the observed data, an augmented Dickey-Fuller test [2] (ADF test) is applied. Table 1 presents the results of the ADF test on the presence of the unit root. The results of ADF test in levels and first differences, suggest that all series are I(1). This means that the transformation of the original series by using first differences in the model is sufficient to obtain stationary series.

Table 1: Test Values for ADF test, in levels and in differences.

Name of the variable	**In levels/in first differences**	**ADF test statistic**	
		Intercept	**Trend & intercept**
LIND_WOOD	in levels	-1.631649(1)	-3.427235(0)
DLIND_WOOD	in first differences	-9.290167(0)*	-9.220420(0)*
LIND_FUR	in levels	-2.465237(2)	-3.751371(2)
DLIND_FUR	in first differences	-4.421063(0)*	-5.038612(7)*
GDP_GR	in levels	-1.208661(0)	-1.898058(0)
DGDP_GR	in first differences	-6.119240(0)*	-6.083650(0)*
LULC_WOOD	in levels	-3.480046(0)	-3.844119(0)
DLULC_WOOD	in first differences	-7.006564(0)*	-6.898132(0)*
LULC_FUR	in levels	-3.328738(2)	-3.479316(2)
DULC_FUR	in first differences	-6.059656(0)*	-5.955493(0)*
LEX_WOOD	in levels	-1.093496(0)	-1.441408(0)
DLEX_WOOD	in first differences	-6.069704(0)*	-5.975776(0)*
LEX_FUR	in levels	-1.811743(0)	-0.890117(0)
DLEX_FUR	in first differences	-5.244475(0)*	-5.533666(0)*

Note: IND_WOOD, IND_FUR – value of the industrial production of wood processing and manufacture of furniture; GDP_GR – value of real GDP growth rate; ULC_WOOD, ULC_FUR – unit labour costs of wood processing and manufacture of furniture; EX_WOOD, EX_FUR – value of export of wood processing and manufacture of furniture. L and D denote natural logarithm and first differences respectively. Numbers in the brackets are the lag length (automatic based on SIC, MAXLAG=9). *Null hypothesis on the existence of unit root rejected at the 1 percent significance level.
Source: Author's calculations.

4.2 Estimation Results

The next step is a multiple linear regression analysis. We estimate two regressions, with the industrial production index for both industries as dependant variables. As explanatory variables we use real GDP growth rate, and industry specific data on unit labour costs and real exports. The results of the regression analysis are presented in Table 2 and Table 3.

Table 2: Results of a time series regression for Division 16.

Dependent Variable: DLIND_WOOD
Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DGDP_GR	0.011167	0.005156	2.165.966	0.0368**
DLULC_WOOD	-0.772018	0.129237	-5.973.667	0.0000*
DLEX_WOOD	0.324378	0.144652	2.242.471	0.0310**
C	0.005394	0.008258	0.653138	0.5177

R-squared = 0.567393; Adjusted R-squared = 0.532317
F-statistic = 1.617598; Prob(F-statistic) = 0.000001

Breusch-Godfrey Serial Correlation LM Test:
F-statistic = 1.684259; Prob. F(3,34) = 0.1888
Obs*R-squared = 5.304714; Prob. Chi-Square(2) = 0.1508

Heteroskedasticity Test: Breusch-Pagan-Godfrey
F-statistic = 0.768821; Prob. F(1,38) = 0.5188
Obs*R-squared =2.405838; Prob. Chi-Square(1) = 0.4925

Jarque-Bera Test = 1.831706; Probability = 0.400175

Variance Inflation Factor = 2.31 **Tolerance** = 0.43

Note: *p-value less than 0.01; **p-value less than 0.05; ***p-value less than 0.1.
Source: Author's calculations.

Table 3: Results of a time series regression for Division 31.

Dependent Variable: DLIND_FUR
Included observations: 41 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DGDP_GR	0.003711	0.002177	1.705.201	0.0965***
DLULC_FUR	-0.299460	0.056020	-5.345.625	0.0000*
DLEX_FUR	0.060623	0.039675	1.527.981	0.1350
C	0.007238	0.003618	2.000.849	0.0528***

R-squared = 0.469493; Adjusted R-squared = 0.426479
F-statistic = 1.091486; Prob(F-statistic) = 0.000028

Breusch-Godfrey Serial Correlation LM Test:
F-statistic = 0.988891; Prob. F(2,35) = 0.3821
Obs*R-squared = 2.192913; Prob. Chi-Square(2) = 0.3341

Heteroskedasticity Test: Breusch-Pagan-Godfrey
F-statistic = 0.181084; Prob. F(1,38) = 0.9086
Obs*R-squared =0.593271; Prob. Chi-Square(1) = 0.8980

Jarque-Bera Test = 0.567908; Probability = 0.752801

Variance Inflation Factor = 1.88 **Tolerance** = 0.53

Note: *p-value less than 0.01; **p-value less than 0.05; ***p-value less than 0.1.
Source: Author's calculations.

Both models satisfy all diagnostic tests (autocorrelation, heteroskedasticity, normality and, multicollinearity), which are shown below the tables. Ramsey’s RESET Test shows that there is no evidence of specification error in any of the models. Furthermore, the CUSUM and CUSUM of Squares tests clearly indicate stability in both equations during the observed period (because the cumulative sum of the recursive residuals and the cumulative sum of squares are within the 5% significance lines). The results of the regression analysis indicate that real GDP growth rate and unit labour costs are statistically significant in both models, and have the expected signs. More specifically, the coefficient -0.772 (in Table 2) means that: on average, holding DGDP_GR and DLEX_WOOD fixed, an increase of one index point of unit labour costs is predicted to decrease industrial production by 0.772 index points. Hence, an increase in the unit labour costs has a negative impact on production in both industries while an increase in real GDP growth rate has a positive and statistically significant impact on production in wood processing and manufacture of furniture. Furthermore, real exports has statistically significant and positive coefficient only in wood processing industry, suggesting that an increase in export leads to the increase in output. This is in line with basic characteristics of analysed industries.

5 CONCLUDING REMARKS

Strong and healthy forest-based industries with high levels of export competitiveness are essential in order to fully exploit the Croatia's potential for growth and to enhance and sustain its overall economic development. The economic importance of the domestic wood industry will increase in the future, with its primary influence in both its export orientation and its existing raw material potential.

As industry-specific results are of interest, the analysis is conducted for each industry separately. Regression on macroeconomic variables presents similar stories: for the wood processing and furniture manufacturing, real GDP growth rates and unit labour costs play an important role. For wood processing industry model, real export is also statistically significant and has theoretically plausible sign. The models proved their adequacy in terms of various diagnostic tests. This empirical investigation leads to the conclusion that the main possibilities for the Croatian wood industry to maintain and enhance its competitiveness lays in export-oriented production and lower unit labour costs.

References

[1] Central Bureau of Statistics, Republic of Croatia, First releases and Statistical Reports, http://www.dzs.hr.

[2] Dickey, D. A., Fuller, W. A., 1979. Distributions of the estimators for autoregressive time series with unit root. Journal of the American Statistical Association, 74, pp. 427-431.

[3] McKinsey Global Institute, 2010. How to compete and grow: A sector guide to policy. McKinsey and Company, 53 p.

[4] Motik, D., Pirc, A., 2009. Comparison of wood products production and consumption in the Republic of Croatia. In: Bičanić, K. (ed.), Competitiveness of wood processing and furniture manufacturing, Šibenik, WoodEMA, International Association, pp. 27-32.

[5] Tkalec, M., Vizek, M., 2010. The Impact of Macroeconomic Policies on Manufacturing Production in Croatia. Economic Trends and Economic Policy, Vol. 19, No. 121, pp. 61-92.

ARIMA MODELS AND THE BOX – JENKINS APPROACH IN ANALYSING AND FORECASTING VARIABLES IN FIELD OF SUSTAINABLE DEVELOPMENT – THE CASE OF CROATIA

Mirjana Čižmešija and **Jelena Knezović**
University of Zagreb, Faculty of Economics and Business
Trg J. F. Kennedyja 6
10000 Zagreb, Croatia
mcizmesija@efzg.hr
jelenaknezovic@yahoo.com

Abstract: The purpose of the paper is applying the Box-Jenkins approach in developing an appropriate ARIMA model with the aim to analyze and forecast the total energy consumption in Croatia as a most important variable in the field of sustainable development. The analysis was conducted for the period since 1992 up to 2008 and forecasts values were determined up to 2014. There are two ARIMA models selected as representative: ARIMA (1,1,0) and ARIMA (1,1,1).

Keywords: sustainable development, total energy consumption, Box-Jenkins approach, ARIMA model, ADF test.

1 INTORODUCTION

Macroeconomic forecasting is a vital element of the total economic policy in a certain country or a region. The variables of interest usually are: gross domestic product, unemployment rate, industrial production etc. In this paper, the justification and motivation for analysing and forecasting total energy consumption in Croatia is its relevance in recent years, its impact on economic development and its significant role in the formation of macroeconomic policies in the state. The importance of this impact was recognized even fifty years ago, because of increasing scarcity of energy. During this time a growing need has been detected of developing a suitable forecasting model that would be able to predict energy consumption trends in the country.

2 THE CONCEPT OF SUSTAINABLE DEVELOPMENT

The concept of sustainable development becomes a relevant issue during the 1970s. At that time people began to realize that it is very difficult to have a healthy society and a growing economy in a world with so much poverty and environment degradation. In accordance with this direction, one should develop a model of economic development that will not be as detrimental for environment and social development. There exist different understandings of the sole concept of sustainable development; therefore it is very difficult to determine its unique definition. A large number of such definitions are based on the report of Burtland (1987), which says that sustainability means "meeting the needs of present generations without compromising the ability of future generations to meet their own needs". In the context of such an understanding, Goddland and Ledec (1984) comprehend the concept of sustainable development as a model of social and structural-economic transformation that displays economic and social benefits of now living people, without compromising the benefits of future generations. On the other hand, Ress (1988), as well as Robbinson and Tinker (1995), see sustainability in the context of merging economic, social and ecological systems, emphasizing the importance of limited ecological capacity. Despite the differences in perception of the concept of sustainable development, in general we can say that it is a process towards achieving a balance between economic, social and environmental requirements to ensure "meeting the needs of present generations without compromising the ability of future generations to meet their own needs".

Croatia has adopted the "Resolution on Environmental Protection” already in 1972. In 1992, after the World conference in Rio de Janeiro, where the Declaration was adopted, Croatia elect for sustainable development. But only since 2000, the topic of sustainable development becomes a current issue of public and economic interest. In February 2009, pursuant to article 44 paragraph 4 of the Environmental Protection Act (*Official Gazette*, No. 110/07) the Strategy for Sustainable Development of the Republic of Croatia has been adopted in the Croatian Parliament. It is a document that focuses on the long-term Croatian economic and social development and environmental protection to ensure sustainable development. In order to raise awareness of sustainable development, in this paper the variable of interest is the total energy consumption in Croatia. In the past few decades, the role of energy in economic growth is becoming increasingly important and a very common scientific topic of many authors. It is very clear that the role of energy in economic development should be given more attention. Since October 2009, when the Croatian Parliament adopted the Strategy of Energy Development, Croatia has a strategic document that relates to energy development. In accordance with this, the Strategy is not only important in terms of energy but also in terms of political and socially important documents because the energy situation in one country "spills over" to other very importance areas.

3 BOX – JENKINS (ARIMA) METHODOLOGY AND DATASET

The Box-Jenkins methodology is an iterative approach of identifying, fitting and checking ARIMA models with time series data (Hanke and Wichern, 2009). The chosen model can be used for forecasting. Forecasts follow directly from the form of the fitted model. In this paper the dataset of total energy consumption in Croatia is used. The yearly data cover the period from 1992 to 2008 and are taken from the Croatian Bureau of Statistics (http://www.dzs.hr)[1].

Phenomena in nature do not behave deterministically; there is indeed a wide variety of different influences acting on the observed variables. In this case we should use the appropriate analytical model that expresses the correlation of the time series with itself, lagged by 1,2, or more periods. In such model, values of the observed series with a shift in time take the role of independent variables. It is necessary to find an appropriate analytical expression (model) that can express the dependence of the current value of the phenomena of its lagged values (Šošić, 2006). Consequently, an economic phenomenon can be defined as a stochastic process $\{Y_t,\ t = 0, \pm 1, \pm 2, ...\}$.

Time series of economic and energy variables often have nonstationarity problem that can be resolved through appropriate procedures.[2]

One of the reasons why the time series of these variables have the characteristics of nonstationarity may be constant changes in legal and technical principles and rules which certainly affect economic relations that have implications for changes in the time series of variables from this area. ARIMA(p,d,q) models are used to analyze processes with nonstationary components, in other words, they are used to model the processes that contain a periodic variation in time. The mentioned models are often used to describe the dynamics of a large number of economic variables. Precisely because of these reasons they are also suitable for the analysis of total energy consumption in Croatia.

[1] The data used in paper are the newest available.
[2] Differentiation of time series of original value.

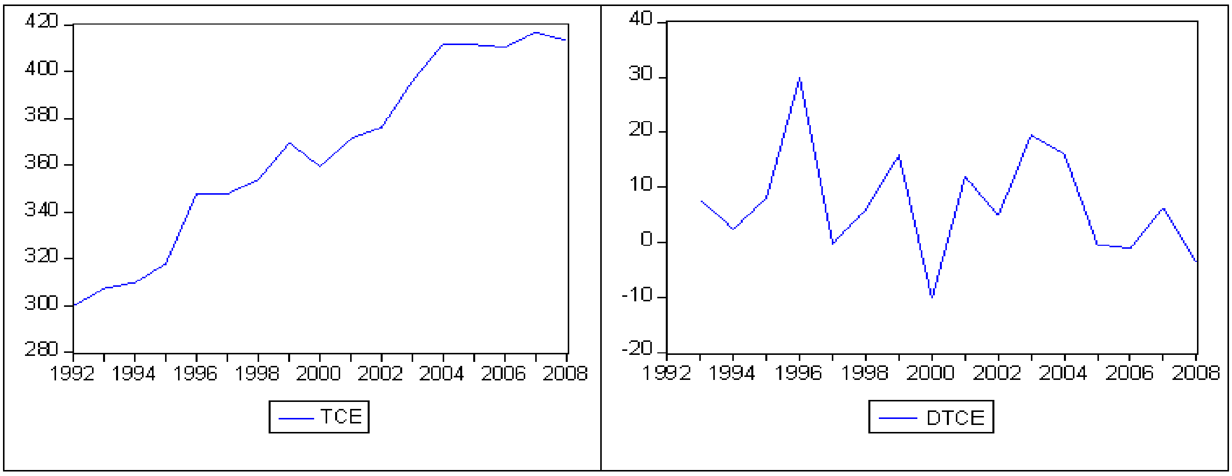
ARIMA (p,d,q) can be expressed as follows:

$$\phi(B)(1-B)^d Y_t = \theta(B)\varepsilon_t, \tag{1}$$

where $\phi(B)$ indicates an autoregressive polynomial of order p, and $\theta(B)$ represents the moving average polynomial of order q, assuming that zero points of these polynomials all lie outside the unit circle and the polynomials have no common zero point. d is a positive number and indicates the order of differencing (Bahovec and Erjavec, 2009). In other words, d numerically represents how many times a time series is differentiated to eventually become stationary.

4 EMPIRICAL RESULTS

This paper first presents the results of unit root tests, Dickey-Fuller and Augmented Dickey-Fuller tests which are used to identify the order of integration for the variable total energy consumption (TCE) in Croatia. After unit root tests result, further follows an ARIMA model in the analysis of mentioned variable using Box-Jenkins approach to model selection and forecasting selected model. The forecast follows directly from the form of the fitted model (Hanke and Wichern, 2009).



Source: http://www.dzs.hr and authors´ calculation

Figure 1: Total energy consumption (TCE) and first differences of total energy consumption (DTCE) in Croatia (expressed in petajoule)

The time series shown in Figure 1 indicates the presence of an upward trend in total energy consumption in Croatia for the observed period. This means that the total energy consumption in Croatia (TCE) for the observed period has the characteristics of nonstationarity.[3]

If the series is nonstationary, it can often be converted to a stationary series by differencing. That is, the original series is replaced by a series of differences (Hanke and Wichern, 2009). To eliminate the nonstationarity of the time series, the first differences (DTCE) were calculated. A series of first differences is tested and it has been found to be

[3] Same conclusion is made by analyzing the correlogram, and based on the results of Ljung-Box test. Results are available upon request to the authors.

stationary[4]. It is concluded that the time series of total energy consumption in Croatia is integrated order of 1, $I(1)$.

Stationarity was also noticed from the Sample autocorrelation function (SACF) and Sample partial autocorrelation function (SPACF) for time series D(TCE)[5]. Both functions have a tendency to decrease. In the identification phase, the initial models were chosen: ARIMA(1,1,0) and ARIMA(1,1,1) with estimated parameters, as shown in tables 1 and 2.

Table 1. Estimation of ARIMA(1,1,0) model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	452,4773	77,73031	5,821119	0
AR(1)	0,920802	0,063305	14,54558	0
R-squared	0,937936	Mean dependent var		370,2412
Adjusted R-squared	0,933503	S.D. dependent var		37,7843
S.E. of regression	9,743446	Akaike info criterion		7,507535
Sum squared resid	1329,086	Schwarz criterion		7,604109
Log likelihood	-58,06028	F-statistic		211,5739
Durbin-Watson stat	2,398034	Prob(F-statistic)		0

Source: Authors' calculations (in EViews 5.0)

ARIMA(1,1,0) with the backward shift operator can be expressed as:

$$(1-0{,}92B)(1-B)Y_t = \varepsilon_t \tag{2}$$

p-value for the AR parameter is 0,0000 which leads to the conclusion that the AR parameter is significant in the observed model. It is known that the $\mathrm{AR}(p)$ process has the property of invertibility by definition; therefore it is not necessary to examine the conditions for satisfying this property. $\mathrm{AR}(p)$ process is stationary if it has $\mathrm{MA}(q)$ representation, or an $\mathrm{AR}(1)$ process is stationary if $|\phi|<1$. In table 1 the AR parameter is $\phi = 0{,}92$ which is less than 1. This means that the ARIMA model satisfies the property of stationarity.

ARIMA(1,1,1) with the backward shift operator can be expressed as:

$$(1-0{,}935B)(1-B)Y_t = (1-0{,}940B)\varepsilon_t \tag{3}$$

[4] *p*-value is less than the significance level $\alpha = 0{,}05$. ADF test results are available upon request to the authors.

[5] Results are available upon request to the authors.

Table 2. Estimation of ARIMA(1,1,1) model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	487,1824	46,27584	10,52779	0.0000
AR(1)	0,935301	0,025053	37,33249	0.0000
MA(1)	-0,940203	0,056029	-16,78054	0.0000
R-squared	0,959462	Mean dependent var		370,2412
Adjusted R-squared	0,953225	S.D. dependent var		37,7843
S.E. of regression	8,171812	Akaike info criterion		7,206619
Sum squared resid	868,1206	Schwarz criterion		7,351479
Log likelihood	-54,65295	F-statistic		153,842
Durbin-Watson stat	1,623505	Prob(F-statistic)		0

Source: Authors' calculations (in EViews 5.0)

Empirical level of significance (p-value) for all three parameters is equal to 0.0000 which is less than any conventional theoretical significance level. As values of AR and MA parameters are less than 1, this means that the ARIMA(1,1,1) model has the property of both stationarity and invertibility. At the phase of diagnostic checking the following criteria are used: adjusted R square ($\overline{R}^2$), residual sum of squares (RSS) and information criteria. Since the model is used for forecasting purposes, to select the most appropriate model, measures of predictive efficiency can also be used. Forecasting errors that are often compared are: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE) and Mean Absolute Error (MAE).

Table 3. Comparison of criteria for ARIMA(1,1,0) and ARIMA(1,1,1) models

Criteria	ARIMA(1,1,0)	ARIMA(1,1,1)
$\overline{R}^2$	0,933503	0,953225
SR	1329,086	868,1206
AIC	7,507535	7,351479
$RMSE$	9,320695	9,598116
$MAPE$	2,242391	2,192870
MAE	8,164889	7,836630

Source: Authors' calculations (in EViews 5.0)

In table 3 we can see a comparison of measures of representativeness that were used in the evaluation of selected models. Between those two models, ARIMA(1,1,0) and ARIMA(1,1,1), there are very small differences when we take into account the adjusted R square and Akaike information criterion ($\overline{R}^2$) is slightly larger and AIC is slightly smaller for ARIMA(1,1,1) than for ARIMA(1,1,0). If we compare the predictive efficiency of models, MAPE and MAE are smaller for ARIMA(1,1,1) model, while RMSE is slightly greater than the RMSE for the ARIMA(1,1,0) model[6].

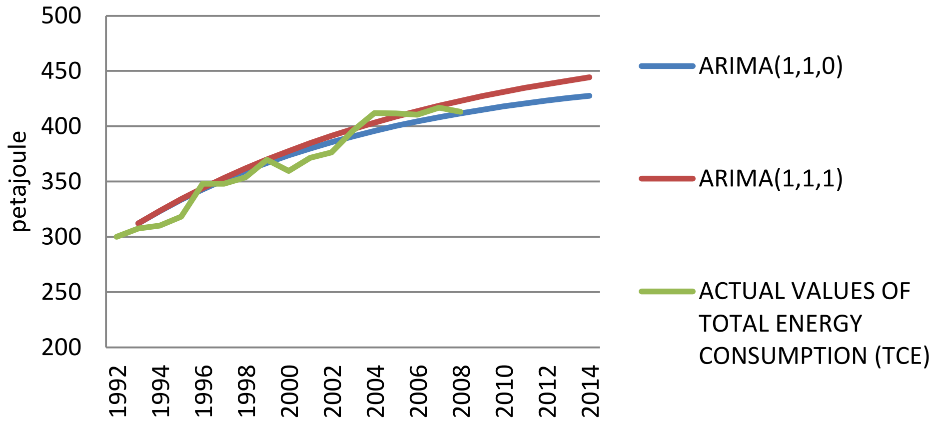
After comparing the two potentially "good" models with the selected criteria, it is difficult to draw conclusions on the selection of the final model. Although, looking at the relations between values of selected criteria, more of them are in favour of the ARIMA(1,1,1) model. But differences are almost minimal. As the art of selecting the model

[6] In the selected models, there is no residual autocorrelation up to the fourteenth lag.

depends on experience and the ability of analysts, we decided to employ both models in forecasting future values of total energy consumption in Croatia.

5 FORECASTING AND CONCLUDING REMARKS

Using the selected models, prognostic values for the next 6 years were found (see figure 2.)



Source: Authors' calculations

Figure 2. Actual and prognostic values of time series variable total energy consumption in Croatia for the models ARIMA (1,1,0) and ARIMA (1,1,1)

Although the two selected forecasting models are almost equally representative, planning and forecasting, besides the quantitative part, also contains the judgmental approach. So in conditions in which the Croatian economy came (high energy prices and insufficient orientation to renewable energy sources and prediction of the recession conditions in the next three years) it is expected for the ARIMA(1,1,0) to be more appropriate forecasting model than the model ARIMA(1,1,1) because it predicts a slightly lower level of total energy consumption in Croatia up to 2014.

References

[1] Bahovec, V., and Erjavec, N., (2009), Uvod u ekonometrijsku analizu, Zagreb, Element

[2] Henke, J. E., Wichern and Dean W., (2009), Business forecasting, Pearson Prentice Hall

[3] Ledec, G., and Goodland, R. J., A., (1984), The role of environmental management in sustainable development, New York

[4] Robinson, J., and Tinker, J.,(1995), Reconciling ecological, economic and social imperatives: Toward an analytical framework, SDRI Discussion Papers Series 1995-1, Sustainable Development Institute, Vancouver, Canada

[5] Rees, W., E., (1988), Sustainable development and how to achieve it, University of British Columbia

[6] Šošić, I. (2006), Primijenjena statistika, Zagreb, Školska knjiga

REGRESSION-NEURO-FUZZY APPROACH TO ANALYSE DISTANCE FUNCTION IN INTERNAL INTER-REGIONAL MIGRATIONS IN EU COUNTRIES

Samo Drobne*, Marija Bogataj, Danijela Tuljak Suban***** and **Urška Železnik******
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: samo.drobne@fgg.uni-lj.si
** University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož,
& MEDIFAS, Nova Gorica, Slovenia, e-mail: marija.bogataj@guest.arnes.si
*** University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož,
e-mail: danijela.tuljak@fpp.uni-lj.si
**** Dolenja vas 30, 1410 Zagorje ob Savi, e-mail: u.zeleznik@gmail.com

Abstract: In this paper, we analyse the distance regression and neuro-fuzzy distance functions in internal inter-regional migration flows of EU and candidate countries. The significant impact of time distance on internal migration flows is presented. To analyse the perception of distance by analysed countries, we introduced neuro-fuzzy modelling approach.

Keywords: migration, distance function, regression analysis, fuzzy logic, neuro-fuzzy, NUTS 2, EU.

1 INTRODUCTION

Migrations, and periodical commuting, are regarded as the most important factors influencing the demographic and socio-economic compositions of regions at different levels. For the general process of regional changes, an understanding of inter-regional migrations (and daily or weekly commuting) is vital. As Codwallader (1992) and many other researchers (e.g. Anjomani, (2002) and Chun (1996)) have pointed out, policy-makers have become increasingly aware of the role of migrations as migration of the human resources for any production or services and in the context of any other socio-economic issues, especially as regional growth. The growth of regions relates closely to population growth, which is mostly a result of migrations.[1] In this paper, we analyse the distance function in inter-regional migration for twenty EU and candidate countries in 2006 using regression-neuro-fuzzy approach. From Tab. 1 it is evident that the most active population in migration were in Turkey and United Kingdom, but Poland and Slovakia were the less active countries.

2 MODEL OF INTERNAL INTER-REGIONAL MIGRATION IN EU COUNTRIES

Data for internal inter-region migrations in analysed EU and candidate countries in 2006 were obtained from EUROSTAT (2010). For those countries which did not reported for internal inter-regional flows for analysed year of 2006, geometric mean of data for 2004, 2005 and 2007 was calculated. In the regression analysis of the parameters in the gravity model, time distance function $\Psi(d(t)_{ij})$ was analysed where $d(t)_{ij}$ describes time spending distance by car between regional centres on NUTS 2 level.[2] Time distance function $\Psi(d(t)_{ij})$ was later considered by neuro-fuzzy reasoning. Data on road network in 2005 were obtained from JRC-IPTS (2010). Using data on roads, we constructed network models,

[1] 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 than any migration (Nijkamp, 1987).

[2] The NUTS (Nomenclature of Territorial Units for Statistics) classification is a hierarchical system for dividing up the economic territory of the EU. Regions at NUTS 2 level are used for the application of regional policies.

which were the basis for calculation of optimal (shortest) time-spending distances between regional centres on NUTS 2 level.

Table 1: Population and number of internal inter-regional migrants on NUTS 2 level in EU countries in 2006 (source: (EUROSTAT, 2010) and own calculation; * Slovenia is analysed for NUTS 3 regions).

ID	Country	Country code	Number of NUTS 2 regions	Population (in 1000)	Migrants	
					Number	Percentage
1	Austria	AT	9	8282	89,677	1.08%
2	Belgium	BE	11	10,548	147,469	1.40%
3	Bulgaria	BG	6	7699	49,999	0.65%
4	Croatia	HR	3	4442	17,684	0.40%
5	Czech Republic	CZ	7	10,269	73,985	0.72%
6	Denmark	DK	5	5437	107,518	1.98%
7	Finland	FI	5	5266	75,573	1.44%
8	France	FR	26	63,438	731,312	1.15%
9	Germany	DE	39	82,376	1,437,174	1.74%
10	Hungary	HU	7	10,071	166,713	1.66%
11	Italy	IT	21	58,942	335,643	0.57%
12	The Netherlands	NL	12	16,346	265,057	1.62%
13	Poland	PL	16	38,141	113,348	0.30%
14	Romania	RO	8	21,588	89,454	0.41%
15	Slovakia	SK	4	5391	16,110	0.03%
16	Slovenia*	SI	12	2007	8499	0.42%
17	Spain	ES	19	44,116	533,128	1.21%
18	Sweden	SE	8	9081	164,755	1.81%
19	Turkey	TR	26	71,553	2,099,033	2.93%
20	United Kingdom	UK	37	60,596	1,729,963	2.85%

2.1 Regression model

Gravity models, based on classical physics, posit that the flow of migrants between two nodes will be proportional to the population at both nodes, and inversely proportional to the distance between them (Sen and Smith, 1998). In our previous work for case studies of Slovenia (Bogataj, Drobne and Bogataj, 1995; Bogataj and Drobne, 2005; Drobne, Bogataj and Bogataj, 2008) we proved that there is correlation between gross migration and daily commuting. For this reason, we used time-spending distance between regional centres to analyse the impact of distance on migration flows.[3] The gravity model that has been analysed for twenty countries is

$$M_{ij} = \alpha \cdot P_i^{\beta_1} \cdot P_j^{\beta_2} \cdot \Psi(d(t)_{ij}) = \alpha \cdot P_i^{\beta_1} \cdot P_j^{\beta_2} \cdot d(t)_{ij}^{\gamma} \qquad (1)$$

where we denote with i the region of origin and with j the region of destination, M_{ij} is the number of inter-regional migrants, P_i and P_j are the populations in the region of origin respectively destination, $d(t)_{ij}$ is travel time between the region of origin and region of destination, α is the intercept, and β_1, β_2, γ are the powers. Intercept and the powers were estimated in the regression analysis. Tab. 2 shows the results of the inter-regional migrating gravity model (1) between regions on NUTS 2 level by analysed country for 2006. The most reliable gravity models were estimated for those countries where adjusted R^2 is high and (intercept and) powers are significant; those countries are: Austria, Denmark, France, Germany, Hungary, Italy, The Netherlands, Poland, Spain, Sweden, Turkey and United Kingdom. We got estimators of less reliable models for Belgium, Bulgaria, Croatia, Czech

[3] We assume that in internal inter-regional migrations barriers like institutional and language barriers do not exist or do not play significant role.

Republic, Finland, Romania and Slovakia. Note that regression analysis for Slovenia was done for NUTS 3 regions (while there were only two NUTS 2 regions in Slovenia) and that there are only three regions on NUTS 2 level in Croatia.

Table 2: Powers (exponents) of the inter-regional migrating gravity model (1) between regions on NUTS 2 level by country in 2006 (parentheses denote intercept and/or powers which significance is bad, P-value>0.1; in the most of other cases the significance is very good, 0.01<P-values; * denotes that regression analysis for Slovenia was done for regions at NUTS 3 level).

ID	*Country*	*Country code*	*Number of relations*	*Adj. R^2 (%)*	*Intercept α*	*β_1 [P_i]*	*β_2 [P_j]*	*γ [$d(t)_{ij}$]*
1	Austria	AT	72	82.9	0.01	0.99	0.96	-1.41
2	Belgium	BE	110	34.8	(0.09)	0.77	0.56	-2.32
3	Bulgaria	BG	30	39.3	1E-03	0.69	1.38	(-0.49)
4	Croatia	HR	6	69.4	(3E-08)	(0.37)	3.13	(-0.25)
5	Czech Republic	CZ	56	67.0	(0.32	1.03	(0.26)	-1.39
6	Denmark	DK	20	83.3	1E-03	1.12	1.24	-1.62
7	Finland	FI	20	92.7	1E-05	1.43	1.47	(-0.18)
8	France	FR	462	79.4	3E-03	0.93	0.87	-0.91
9	Germany	DE	1482	72.7	8E-04	0.94	1.05	-1.41
10	Hungary	HU	42	76.5	2E-04	1.22	1.33	-1.53
11	Italy	IT	418	83.9	1E-04	1.04	0.99	-0.38
12	The Netherlands	NL	132	85.0	(0.38)	0.65	0.64	-2.03
13	Poland	PL	240	71.0	2E-03	0.71	1.13	-1.49
14	Romania	RO	56	47.5	8E+04	(0.31)	(-0.65)	-1.05
15	Slovakia	SK	12	54.6	(992.41)	(0.54)	(-0.31)	-1.47
16	Slovenia*	SI	132	72.7	0.05	0.68	0.83	-2.38
17	Spain	ES	342	79.9	4E-03	0.84	0.84	-0.19
18	Sweden	SE	56	75.7	(2.92)	0.48	0.66	-0.61
19	Turkey	TR	650	63.3	9E-04	1.10	0.95	-0.59
20	United Kingdom	UK	1332	70.2	0.16	0.72	0.59	-1.04

The powers in Tab. 2 and on Fig. 1 shows that travel distance between regions has significant influence on migration flows (the power is high and the estimator is significant) in Austria, Belgium, Czech Republic, Denmark, France, Germany, Hungary, The Netherlands, Poland, Romania, Slovakia and United Kingdom (conditionally in Sweden and Turkey). Travel distance doesn't play significant role (powers are low) in decision on migration in Italy and Spain. For Bulgaria, Croatia and Finland estimators of regression coefficient for travel distance are not significant. Comparing the results on the map (Fig. 1), one can find more significant impact of time distance on migration flows for countries in the centre of Europe. The differentiation of the impact of travel distance on migration flows derived by regression analysis led us to investigate the role of travel distance by countries deeper using neuro-fuzzy approach.

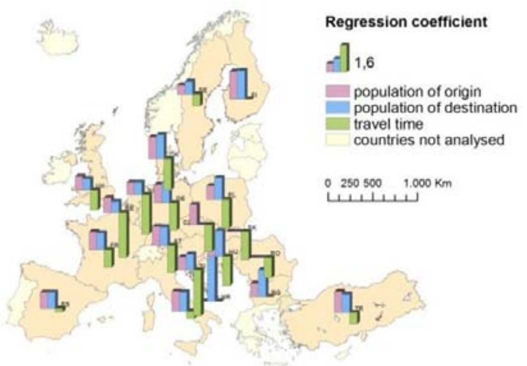


Figure 1: Statistical significant regression coefficients of internal inter-regional migration flows on NUTS 2 level in analysed EU countries in 2006.

2.2 Neuro-fuzzy approach

Fuzzy logic is used as a complementary tool to determinate the distance perception in decision on internal inter-regional migration between regions on NUTS 2 level in Europe. The approach is described with simple "if-then" relations based on input linguistic variables and output linguistic variable all described with three terms (*low*, *medium*, *high*). This permits to find quickly a good solution, but in our case there was a limitation: the knowledge which allows the formulation of fuzzy system was deep hidden in data sets. So we decided use neuro-fuzzy approach to derive the structure of the model from the data sets of internal inter-regional migration flows. Neural nets can learn from data sets while fuzzy logic solutions are easy to verify and optimize (Rajasekaran and Vijayalakshmi Pai, 2004).

The structure of the fuzzy system was defined by the program FuzzyTECH 5.72d[4] that permeated us to improve the outcome of the model with the experience of the authors and the Neuro Fuzzy software tool. The algorithm included the following steps (von Altrock, 1995): (1) Creating a Fuzzy Logic System, (2) Obtain Training Data, (3) Define the Neuro Fuzzy Learning, (4) Training Phase, (5) Optimization and Verification.

On the base of the experience acquired in this field (Bogataj, Tuljak Suban and Drobne, 2011; Bogataj, Drobne and Tuljak Suban, 2011) we defined a fuzzy inference structure consisted of three components (Shing and Jang, 1993): (1) the rule block, which was composed of 27 rules, (2) the database or dictionary that defined the membership functions of three input linguistic variables (time_hours $\lfloor d(t)_{ij} \rfloor$, pop_orig $\lfloor P_i \rfloor$, pop_dest $\lfloor P_j \rfloor$) and of the output linguistic variable (mig $\lfloor M_{ij} \rfloor$) that are used in the fuzzy rules, and (3) the reasoning mechanism, which performed the inference procedure upon the rules and computed a plausible output. The explanation of input and output linguistic variables is: *time_hours* is time distance (travel time) $d(t)_{ij}$ between regions on NUTS 2 level in the country in hours, *pop_orig* is the population in the region of origin P_i, *pop_dest* is the population in the region of destination P_j, and *mig* is the number of inter-regional migration flow in the country M_{ij}. All accessible information from a designer's experience on rules formulation and about training data quality were included in the fuzzy inference system. In the next step, we selected the parts of the system that the Neuro Fuzzy Module may modify. This is a major advantage of a neural net, since it lets to control the learning process. We opened for learning only the membership functions since the rules had been defined on the base of previous experiences (Bogataj, Drobne and Tuljak Suban, 2011). The learning process was done in random sequence on the data sets. The effects of corrections were measured by the estimation of R^2_{REG} and R^2_{FUZZ}, where R^2_{REG} is R^2 derived in regression analysis and R^2_{FUZZ} is R^2 derived in neuro-fuzzy analysis.

The results of neuro-fuzzy approach are comparable with those obtained by the regression analyse. Even more, in some simulations where the size of training sets was appropriate (~50 data) and included "good" information, the fuzzy model was better than the regression one; i.e. $R^2_{REG}(\text{Czech Republic}) = 0.6702$ and $R^2_{FUZZ}(\text{Czech Republic}) = 0.6931$, $R^2_{REG}(\text{Romania}) = 0.4752$ and $R^2_{FUZZ}(\text{Romania}) = 0.5164$, $R^2_{REG}(\text{Sweden}) = 0.7569$ and $R^2_{FUZZ}(\text{Sweden}) = 0.7570$, etc. Additionally, the results of neuro-fuzzy model allow us to analyse the perception of distance – i.e. to determine the points of discontinuity of the

[4] FuzzyTECH 5.72d is the copyright of INFORM GmbH and Inform Software Corporation.

distance function – in internal inter-regional migrations. Tab. 3 shows the perception of distance for internal inter-regional migrants by analysed countries in EU in 2006. In the presented fuzzy model the range of time distance functions $d(t)_{ij}$ is uniform in all simulations, $d(t)_{ij} \in [0.28, 60.28]$. Minimum and maximum $d(t)_{ij}$ are time distances between the nearest, respectively the farthest, analysed regions on NUTS 2 level in the country, and low, medium and high $d(t)_{ij}$ are the breakpoints of distance function in the country, where the value of the terms membership functions become different from one. In these breakpoints the perception of distance changes.

Table 3: The perception of time distance in inter-regional migration and minimum and maximum time distances between regions on NUTS 2 level in EU and candidate countries (in hours).

ID	***Country***	***Country code***	***Number of NUTS 2 regions***	***Minimum*** $d(t)_{ij}$	***Maximum*** $d(t)_{ij}$	***Low*** $d(t)_{ij}$	***Medium*** $d(t)_{ij}$	***High*** $d(t)_{ij}$
1	Austria	AT	9	0.8	5.92	0.79	3.66	4.42
2	Belgium	BE	11	0.37	2.29	0.94	1.48	11.50
3	Bulgaria	BG	6	1.88	5.36	1.17	3.60	14.74
4	Croatia	HR	3	2.53	3.57	0.76	4.05	13.98
5	Czech Republic	CZ	7	0.31	4.83	0.93	4.59	14.74
6	Denmark	DK	5	1.3	4.23	3.29	6.04	13.89
7	Finland	FI	5	5.34	19.93	10.75	10.79	10.81
8	France	FR	26	1.08	17	3.80	18.17	18.32
9	Germany	DE	39	0.59	7.44	3.06	23.72	24.73
10	Hungary	HU	7	1.62	5.28	3.21	3.57	13.98
11	Italy	IT	21	1.24	15.49	0.90	3.82	5.90
12	The Netherlands	NL	12	0.76	3.35	1.35	7.96	14.15
13	Poland	PL	16	1.6	9.92	3.08	13.29	14.51
14	Romania	RO	8	0.4	9.06	1.77	6.54	14.12
15	Slovakia	SK	4	1.77	4.79	0.75	2.28	14.13
16	Slovenia	SI	2	N/A	N/A	N/A	N/A	N/A
17	Spain	ES	19	1.6	60.28	2.06	5.39	6.07
18	Sweden	SE	8	1.61	21.35	1.02	7.52	19.20
19	Turkey	TR	26	2.32	20.91	4.48	8.58	27.04
20	United Kingdom	UK	37	0.28	11.36	0.76	5.33	11.46

From Tab. 3 we can see that the perception of migration distance vary a lot across analysed countries. The longer distances are more acceptable in Germany and Turkey, where the same distance is assumed to be low while in other countries is taken to be medium or high. For Finland distance is not the meter of concern. The perception of low time distance between regions of migration varies from Slovakia, Croatia and United Kingdom with the lowest perception to the Finland with the highest perception of low time distance. But, the lowest perception of high time distance for migrations is in Austria, Italy and Spain, and the highest (internal inter-regional migrants are the most tolerant for the distance) is in Turkey and Germany.

4 CONCLUSIONS

In this paper, we analysed the distance function in internal inter-regional migration flows in twenty EU and candidate countries using regression-neuro-fuzzy approach. We found out more significant impact of time distance on migration flows for the countries in the centre of Europe than for the countries on the border of the analysed area – this is typical for Bulgaria, Croatia, Italy and Spain. Using neuro-fuzzy modelling approach, we analysed the distance function for migrants in the analysed countries as step function of low, medium and high value. The results show that the perception of migration distance varies a lot across analysed countries – but, in general, internal inter-regional migrants are the less tolerant for the migration distance in Austria, Italy and Spain and the most tolerant migrants are in Turkey and Germany. Regression-neuro-fuzzy approach to analyse migration flows was used for the year of 2006 (according to the most recent availability of data in EUROSTAT and ESPON databases). But, for more reliable results, here suggested methodology can be tested for data on wider time horizon. In the analysis we assumed that in internal inter-regional migrations barriers like institutional and language barriers did not play significant role.

References

[1] Anjomani A. (2002). Regional growth and interstate migration, Social – Economic Planning Science, Vol. 36, pp. 239-265.

[2] Bogataj M. and S. Drobne (2005). Does the improvement of roads increase the daily commuting? Numerical analysis of Slovenian interregional flows. In: Zadnik Stirn L., M. Indihar Štemberger, L. Ferbar and S. Drobne (Eds.). Selected Decision Support Models for Production and Public Policy Problems. Slovenian Society Informatika, Ljubljana. 185-206.

[3] Bogataj M., D. Tuljak Suban and S. Drobne (2011). Samo. Regression-fuzzy approach to land valuation. Cent. Eur. j. oper. res., 13 pp., http://www.springerlink.com/content/r08r73t2611831j2/fulltext.pdf, doi: 10.1007/s10100-010-0188-x; accessed: 25.3.2011)

[4] Bogataj M., S. Drobne and D. Tuljak Suban (2011). A fuzzy approach to forecasting the attractiveness of regions for human resources. 5th International Conference on Industrial Engineering and Industrial Management, XV Congreso de Ingeniería de Organización. Cartagena, Spain, 7.-9. Septembre, 2011. To be appeared.

[5] Bogataj M., S. Drobne and L. Bogataj (1995): The influence of investment and fiscal policy on growth of spatial structure. Suvremeni promet, Zagreb. 15(5):239–245.

[6] Cadwallader M. (1992). Migration and residential mobility: Macro and micro approaches. University of Wisconsin Press, Wisconsin.

[7] Chun J. (1996). Interregional migration and development. Ashgate Publishing Company, Brookfield, Vermont.

[8] De Beer J., N. Van der Gaag, R. Van der Erf, R. Bauer, H. Fassmann, D. Kupiszewska, M. Kupiszewski, P. Rees, P. Boden, A. Dennett, M. Jasińska, J. Stillwell, P. Wohland, A. De Jong, M. Ter Veer, J. Roto, L. Van Well, F. Heins, C. Bonifazi and G. Gesano (2010). DEMIFER, Demographic and migratory flows affecting European regions and cities. Applied Research 2013/1/3. Final Report. Version 30/09/2010, ESPON.

[9] Drobne S., Bogataj M. and L. Bogataj (2008): Spatial interactions influenced by European corridors and the shift of the Schengen border regime. In Boljunčić V., L. Neralić and K. Šorić (Eds.). The 11th International Conference on Operational Research, Pula, 27th-29th September, 2006. KOI 2006: proceedings. Croatian Operational Research Society, Zagreb. 185-198.

[10] EUROSTAT (2010). Regional migration statistics. Luxemburg. (http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database; accessed: 15.8.2010)

[11] JRC-IPTS (2010). Network for base scenario (year 2005). European Commission, Joint Research Centre, Institute for Prospective Technological Studies. (http://energy.jrc.ec.europa.eu/transtools/FTP.html; accessed: 15.8.2010)

[12] Rajasekaran S. and G. A. Vijayalakshmi Pai (2004). Neural networks, fuzzy logic and genetic algorithms: Synthesis and applications. PHI Learning Pvt. Ltd. New Delhi.

[13] Sen, A. and T. E. Smith (1998). Gravity models of spatial interaction behaviour. Springer-Verlag, Heidelberg.

[14] Shing J. and R. Jang (1993). ANFIS : Adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man and Cybernetics, 23(3)665-685.

[15] von Altrock C. (1995). Fuzzy logic and neurofuzzy applications explained. Prentice-Hall, Inc. Upper Saddle River, USA.

MULTIVARIATE REGRESSION ANALYSIS OF PERSONAL CONSUMPTION IN CROATIA

Ksenija Dumičić, Anita Čeh Časni and **Irena Palić**
Department of Statistics, Faculty of Economics and Business, University of Zagreb,
Trg J.F. Kennedy 6, HR-10000 Zagreb, Croatia
{kdumicic,aceh,ipalic}@efzg.hr

Abstract: The paper discusses short run determinants of personal consumption in Croatia which is analyzed in traditional way using multiple regression model. The results of the analysis showed that wages and interest rates on housing loans play the most important role in determining personal consumption. Besides wages and interest rates, regression model results reveal that in Croatia total loans to households and housing price are also important for explaining personal consumption, but to lesser degree than wages and interest rates on housing loans.

Keywords: personal consumption, income, housing markets variables, total loans to households, multiple regression analysis

1 INTRODUCTION

From the beginning of development of economic thought, scientists have studied the theory of consumption and the theory of growth. Household decision about consumption is a microeconomic question, since it refers to the decisions of individuals. Consumption decisions are important for short-term macroeconomic analysis because of its impact on aggregate demand. Moreover, decisions of households can be the causes of economic shocks. Also, the marginal propensity to consume determines the fiscal policy multiplier [18]. Furthermore, consumption decisions are essential for long-term macroeconomic analysis because of its role in economic growth [11]. The main scientific problem outlined in this research is how housing market variables, namely: house price, interest rate on housing credits and housing loans, combined with income and total loans to households affect personal consumption in Croatia. The research hypothesis is that housing market variables together with income and loans to households impact personal consumption in Croatia. Besides determinants which have been elaborated in economic theory for years, this paper includes housing market variables in analyzing personal consumption. Since household consumption in Croatia in 2010 accounts for about 60% of gross domestic product [6], it is obvious that consumption has a significant impact on economic activity. Therefore, understanding the determinants of households' consumption represents important macroeconomic policy issue.

Keynes consumption theory defines personal consumption as a function of current income including autonomous consumption:

$$C = c_0 + c_1(Y - T), \tag{1}$$

where C represents current consumption, Y is current income and T denotes taxes. Therefore, Y-T represents disposable income. Moreover, c_0 is autonomous consumption, and c_1 is marginal propensity to consume. Keynes states that other determinants of consumption cannot be covered qualitatively; therefore they are not included in consumption function [16]. Despite the fact that the classical economists before Keynes thought that a higher interest rate leads to a reduction in consumption, Keynes did not think that the interest rate can theoretically affect the savings. However, he held that the effect of short-term interest rates on consumption is secondary and relatively unimportant [4]. According to Fisher`s theory based on consumer preferences, change in real interest rates can reduce or increase personal consumption. The increase in real interest rates increases the demand in the second period, but in the first period could theoretically lead to a reduction as well as to an increase in consumption [18]. Regarding consumption theories, it is important to note that decision about lifelong consumption is complex, since the individual should take into consideration income, investment strategies, macroeconomic trends and risks over a longer period of time. It is difficult for the average individual in this situation to be perfectly rational and in that sense to maximize ones utility [3].

Concerning determinants of personal consumption, income and total loans to households are included into the model among variables that affect household consumption and saving decisions, which is in line with economic theory and relevant empirical research (see, for example, [5], [10], [12], [14], [15]).

Besides income and total loans to households, changes in house prices, interest rates on housing credits and housing loans are also included in the analysis in order to quantify the effect of housing market variables on personal consumption.

Personal consumption of a household comprises natural and financial consumption of products and services used for meeting existential needs of household members. Basic characteristics of household consumption and household income in Croatia are provided in Household Budget Survey. The Survey is last conducted in 2009 using random sample of households'. In 2009, consumption expenditures by household amounted on average 76,188 kuna. The largest share of consumption expenditures belongs to the food and non-alcoholic beverages, more precisely 32.06%. This expenditure group is followed by housing and energy consumption, which represents 14.39% of total household consumption. Transport accounts for 10.98%, followed by other goods and services (8.45%), clothing and footwear (7.29%), recreation and culture (5.99%), and expenditures for furnishing, household equipment and routine maintenance as well as communication with a share of about 5%. All other groups in the consumption expenditures had a share of less than 5% and the lowest share in consumption expenditures was recorded in education (1.00%) and health (2.77%) [6].

For the purpose of the empirical analysis in this research, personal consumption time series is taken from quarterly gross domestic product (GDP) estimate. Quarterly GDP data are shown by using basic expenditure categories and activity sections: final consumption expenditure which is divided into household consumption and government consumption, gross fixed capital formation, changes in inventories and net exports of goods and services. Real personal consumption in Croatia shown in Figure 1 follows growth trend from first quarter 2001 to first quarter 2008. In first quarter 2008 it reaches its peak and afterwards it continuously decreases up to second quarter 2009 as a result of global financial crisis. After one quarter of slight recovery, it diminishes again up to first quarter 2010.

Financial crisis can have a lasting notable effect on economic activity when they reflect the unwinding of financial imbalances that funded real sector imbalances [9]. Clearly, housing developments are intertwined with and integral to the crisis that has gripped financial markets since August 2007 and then escalated to a near complete paralysis of credit flows in late 2008 [9]. There is substantial amount of contributing factors and lesson to be learned from that crisis, far too many to review in this study (for more details see, for example [1], [13]).

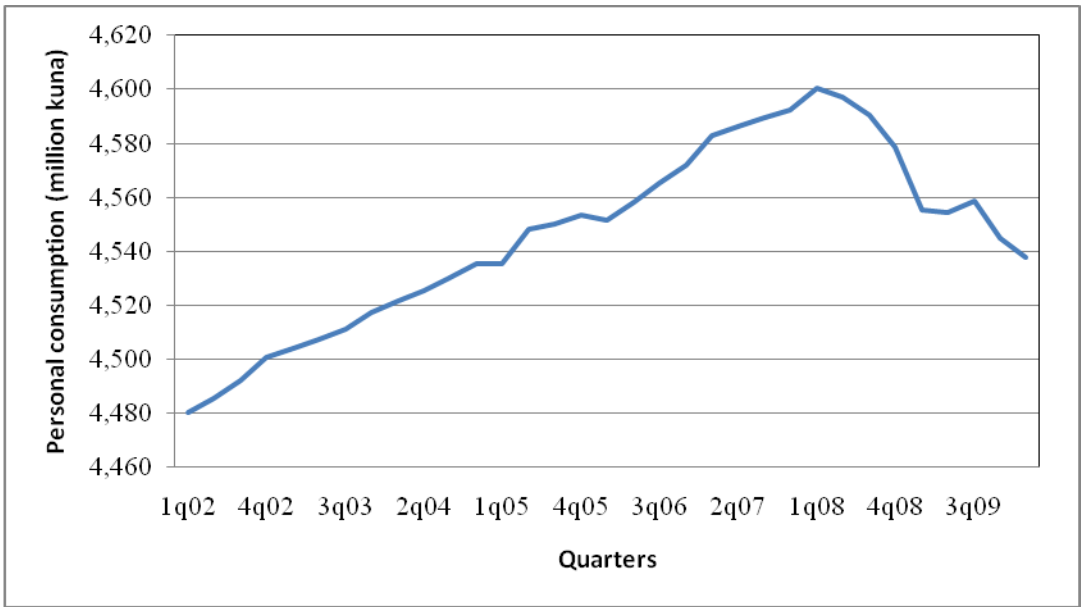


Figure 1: Real personal consumption in Croatia for the period Q1 2002 to Q12010 in million kuna (Source: Croatian Bureau of Statistics)

2 DATA AND METHODS

According to availability of data and existing literature that deals with modeling personal consumption, for the purpose of empirical analysis in this paper the following variables were chosen: household consumption, wage, house price, interest rate on housing credits, (total) loans to households and housing loans [7]. The analysis is conducted using quarterly data ranging from first quarter 2002 to fourth quarter 2010 of original time series. All series were deflated using Consumer price index and first differences were calculated in order to make variables stationary. In table 1 data description and sources are given.

Table 1: Data description and sources

Variable	Description	Source
Household consumption (PC)	Millions of national currency, chain-linked volumes, reference year 2000	Statistical Office of the European Communities (Eurostat)
Wage (W)	Average net nominal wages of the employees in legal entities, in national currency	Croatian Central Bureau of Statistics
House price (HP)	Average market prices per square meter of dwellings, in national currency, total for Croatia	Real Estate Exchange database
CPI deflator	Derived using Consumer price index	International Financial Statistics
Interest rate on housing credits (R)	Weighted averages of monthly interest rates in % on annual basis	Croatian national bank
Total loans to households (TC)	Millions of national currency on annual basis	Croatian national bank
Housing loans (HC)	Millions of national currency on annual basis	Croatian national bank

Analysis was conducted using statistical software EViews 7 (Econometric Views developed by Quantitative Micro Software (QMS)) and Statistica 9 (a suite of analytics software products and solutions developed by Stat Soft Inc.). Although in relevant literature on modeling personal consumption error correction model is usually used [16], in this paper Multivariate regression analysis using Ordinary Least Square Method is conducted due to unavailability of adequate length of relevant time series.

Multiple regression analysis makes possible to discriminate between the effects of the explanatory variables, allowing for the fact that they may be correlated. The regression coefficient of each independent variable provides an estimate of its influence on dependent variable, controlling for the effects of all the other independent variables. So, estimators are unbiased if the model is correctly specified and the assumptions relating to the regression model are fulfilled [8]. This is shown in section 3, namely the empirical analysis of personal consumption in Croatia. More detailed explanation of multiple regression models can be found in [2], [17], [19].

3 EMPIRICAL ANALYSIS OF PERSONAL CONSUMPTION IN CROATIA

In order to investigate which factors have the largest influence on personal consumption in Croatia, multiple regression model with four independent variables was estimated, namely: wage (which was proxy for income), interest rate on housing credits, house price and total loans to households. The aim of the analysis was that all the variables of interest were statistically significant on 5%, there for housing loans were excluded from the analysis.

The following multiple regression model of personal consumption was estimated:

$$\Delta PC = \underset{\substack{(468.9306)\\ [-0.4142]}}{-194.2305} + \underset{\substack{(6.6954)\\ [10.5439]}}{70.59549}\Delta W - \underset{\substack{(295.6287)\\ [-5.7467]}}{1698.879}\Delta R - \underset{\substack{(0.6610)\\ [-1.7541]}}{1.159511}\Delta HP - \underset{\substack{(0.2329)\\ [-1.8339]}}{0.427214}\Delta TC \tag{2}$$

$AdjR^2=0.8124 \quad F=33.4756 \quad DW=2.3237,$

where numbers in parentheses denote standard error of estimate and numbers in brackets denote t-statistics.

According to estimated regression model, all the coefficients have meaningful economic interpretation, since their signs are in line with economic theory and relevant research. Namely, when change in wages rises for 1000 kuna, with all the other variables held constant, the regression value of change in personal consumption rises for 70.60 millions of kuna, whereas the coefficient next to first independent variable is statistically significant on any reasonable significance level. Further, if change in interest rate on housing credits rises for 1 percentage point, with all the other variables held constant, according to the regression model, change in personal consumption decreases for 1698.88 millions of kuna. Coefficient next to second independent variable is also statistically significant on any reasonable significance level. Also, if change in house price rises for 1000 kuna, with all the other variables held constant, the regression value of change in personal consumption decreases for 1.16 thousand kuna. Coefficient next to third independent variable is statistically significant at 4.5% significance level. Finally, if change in total loans to households would rise for 1 000 000 kuna, with all the other variables held constant, then the regression value of change in personal consumption would decrease for 0.427 millions of kuna with the coefficient being statistically significant on 3.5% significance level.

According to the adjusted coefficient of determination 81.24% of total variation in personal consumption is explained by this model. Ljung-Box test of autocorrelation of higher order resulted in not rejecting the null hypotheses of no autocorrelation to order k on every usual significance level. The Jarque-Bera test of residuals resulted in not rejecting the null hypothesis that error terms are normally distributed on every reasonable significance level. Furthermore, the error terms were tested for serial correlation using Breusch-Godfrey serial correlation LM test which also resulted in not rejecting the null hypothesis of no autocorrelation of order 4 on every usual significance level. The model was also tested for heteroscedasticity using Breusch-Pagan LM test and White heteroscedasticity test and both tests resulted in accepting the null hypothesis on homoscedasticity of variance on every usual significance level.

Since the model was satisfactory according to conducted regression diagnostics, the parameter stability tests were applied. According to standardized cumulative recursive residual (CUSUM) and standardized cumulative recursive residual of squares (CUSUMSQ) tests of residuals on structural stability of the tested model, regression equation was stable since test statistics did not cross the 5% significance border lines. Results of CUSUMSQ test on 5% significance level is showed in Figure 2.

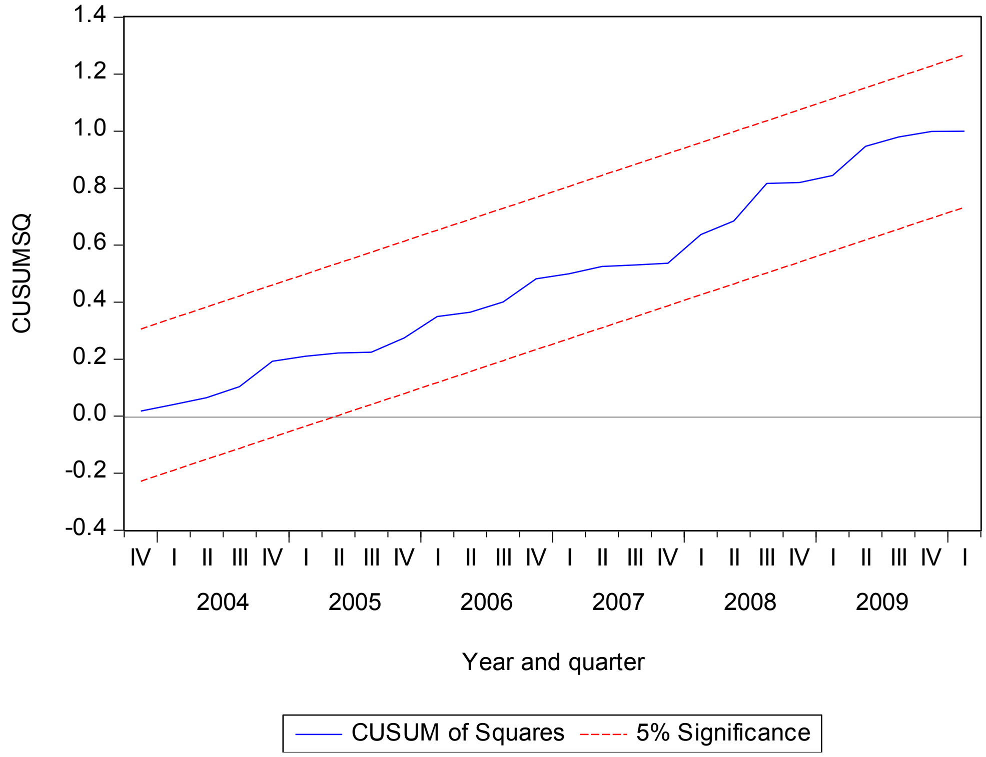


Figure 2: The results of CUSUMSQ test of residuals (*EViews 7*)

Conducted multivariate regression analysis resulted with unbiased estimators of personal consumption since all the model assumptions were fulfilled and the parameters of the model showed stable. The signs of all the regression coefficients are in line with the economic theory. Empirical analysis conducted in this paper suggests that wages and interest rates on housing credits have greatest influence on personal consumption in Croatia, while total loans to households and housing price have less, but still significant influence on personal consumption.

4 CONCLUSION

The aim of this article is to analyze the short run determinants of personal consumption in Croatia. Multiple regression analysis showed that wages and interest rates on housing loans play the most important role in determining personal consumption. Besides wages and interest rates, regression model results reveal that in Croatia total loans to households and housing price are also important for explaining personal consumption, but to lesser degree than wages and interest rates on housing loans.

However, this paper leaves space for further research. First of all, additional housing market explanatory variables should be included in econometric analysis. Also, a different approach to modeling personal consumption could also be used to analyze the short-run determinants of personal consumption in Croatia for instance vector autoregression framework could be applied.

References:

[1] Acharya, V. V., Schnabl, P., 2009. How banks played the leverage game. In: Acharya, V.V., Richardson, M. (Eds), Restoring Financial Stability: How to Repair a broken System, Wiley

[2] Bahovec, V. i Erjavec, N.,2009. Uvod u ekonometrijsku analizu, Element, Zagreb

[3] Bernheim, B. D., Scholz, J. K., 1993. Private saving and public policy, Tax Policy and the Economy 7

[4] Čibarić, I., 2010. Analiza odrednica private štednje u Republici Hrvatskoj primjenom modela vektorske autoregresije, Postgraduate specialist paper, Faculty of Economics and Business Zagreb

[5] Corbo, V., Schmidt-Hebbel, K.S., 1991. Public Policies and saving in Developing Countries, Policy, Research and External Affairs Working Papers, World Bank

[6] Croatian Bureau of Statistics, www.dzs.hr (1st March 2011)

[7] Croatian National Bank, www.hnb.hr (1st March 2011)

[8] Dougherty, C., 2006. Introduction to Econometrics, 3rd ed., Oxford University Press

[9] Duca, J. V., Muellbauer, J., Murphy, A., Housing markets and the financial crisis of 2007-2009. Lessons for the future, Journal of Financial Stability 6 (2010), 203-217

[10] Dumičić, K., Čibarić, I., 2010. The Analysis of Household Savings in Republic of Croatia Using Cointegration Approach, The Business Review, Cambridge, Vol. 15, No. 1, pp. 220-226. ISSN: 1553-5827.

[11] Gartner, M., 2006. Macroeconomics, 2nd ed, Prentice Hall: Financial Times, Harlow etc.

[12] Hussein, K.A., Thirlwall, A.P., 1999. Explaining differences in the domestic savings ratio across countries: A panel data study, The Journal of Development Studies 36

[13] IMF, 2009.Global financial stability report: responding to the financial crisis and measuring systemic risks, April 2009, http//www.imf.org/external/pubs/ft/gfsr/2009/01/index.htm

[14] Jappelli T., Pagano, M.,1998. The Determinants of Saving: Lessons from Italy, Centre for Studies in Economics and Finance Working paper, Department of Economics, Università di Salerno (Italy)

[15] Loayza, N., Schmidt-Hebbel, K., Serven, L.. 2000. What Drives Private Saving Across the World?,Review of Economics and Statistics 82(2)

[16] Lovrinčević, Ž., 2000. Analiza agregatne potrošnje u Hrvatskoj, PhD thesis, Faculty of Economics and Business Zagreb

[17] Maddala, G. S., 2006. Introduction to Econometrics. 3rd ed., Wiley

[18] Mankiw, N.G., 1997. Macroeconomics, 3rd Ed., New York: Worth Publishers

[19] Wooldridge, J. M. ,2008. Introductory Econometrics: A modern approach, 4th ed., South-Western College Pub

IMPACT OF APPLIED ACCEPTANCE SAMPLING PLAN ON DECISIONS IN QUALITY MANAGEMENT

Ksenija Dumicic and **Berislav Zmuk**
Department of Statistics, Faculty of Economics and Business, University of Zagreb
Trg J.F. Kennedy 6, HR-10000 Zagreb, Croatia
{kdumicic, bzmuk}@efzg.hr

Abstract: The goal of the research presented in this paper is to recognize whether an acceptance sampling plan applied, single or double, may influence quality costs, and a quality manager's decision about accepting or rejecting an inspected lot. At the same AQL and LTPD, using reasonable sample sizes, both plans allow the same probabilities of the lot fraction defective, with no statistically significant difference. With a lack of the lot quality, intentional manipulations by decision makers using a certain acceptance sampling plan based on reasonable sample sizes are not possible.

Keywords: acceptance sampling plan, lot, alpha risk, beta risk, OC curve, AQL, LTPD

1 INTRODUCTION

Products should usually be inspected to ensure that they have an appropriate level of quality which is determined by certain quality standards. This inspection would be unnecessary if the process itself would ensure achieving those quality standards. In general, there are three approaches to lot inspection [4]. The first approach implies acceptance of a lot without inspection. This approach is recommended when a supplier's process provides no defective units, and is used when no economic justification to make an inspection exists. The second approach includes a 100% inspection. Inspecting every unit in a lot and removing all defective units, if any would be found, would be recommended when defective units might result in a considerably high failure costs for the buyers, or when it is known that a supplier's processes do not meet a certain level of quality standards. The third approach implies an acceptance sampling in which samples of units in every lot are inspected.

In this paper, the focus is on acceptance sampling, which is not providing any direct form of quality control, but which is used for inspecting a sample from each lot. Acceptance sampling is used: when testing of units is destructive; when costs of the 100% inspection are very high and the inspection requires too much time for conducting; when there are too many units in a lot which can therefore result in inspection errors in the 100% inspection; when the supplier has an excellent quality history and there is no need to make a 100% inspection [5]. Even if all lots are of the same quality, it could happen that, because of sampling, some lots will be accepted and some will be rejected. For this reason, it must be observed that there are risks of accepting "bad" lots and rejecting "good" lots. The risk of rejecting "good" lots is known as alpha or supplier's risk and the risk of accepting "bad" lots is known as beta or buyer's risk. A major classification of acceptance sampling is by variables and attributes [5]. Here the variables are quality characteristics that are measured on a numerical scale. Attributes are quality characteristics that can obtain only two values implying that an inspection leads to a binary result (e.g. a good-bad lot). There are different types of acceptance plans to choose from: single, double, multiple, sequential and skip lot sampling plans [6].

The aim of this paper is to investigate how usage of different acceptance sampling plans could lead to different conclusions and decisions about a lot's quality. Both single and double sampling plans and three different cases are considered. Parameters for the single sampling plan are held constant, while parameters for the double sampling plan are changed. Inference is based on a statistical test for the difference in proportions. When the test shows that the difference is statistically significant, the decision maker could lower quality costs easily by choosing one of the observed sampling plans. The paper is emphasizing the probability of this fraud and misleading of customers.

2 DATA, METHODS AND ANALYSIS

2.1 Single sampling plan

A single sampling plan is the basic plan for conducting inspections of lots. In this plan n units for inspecting are randomly chosen from the lot. This paper analyses only the single sampling plan for attributes, so, in this case, an acceptance limit or the number of c defective units which can be tolerated must be set. That means, if the number of defective units in the chosen sample of n units is equal or smaller than c, the lot will be accepted. Otherwise, the lot will be rejected. The way of reaching a conclusion is shown in Figure 1.

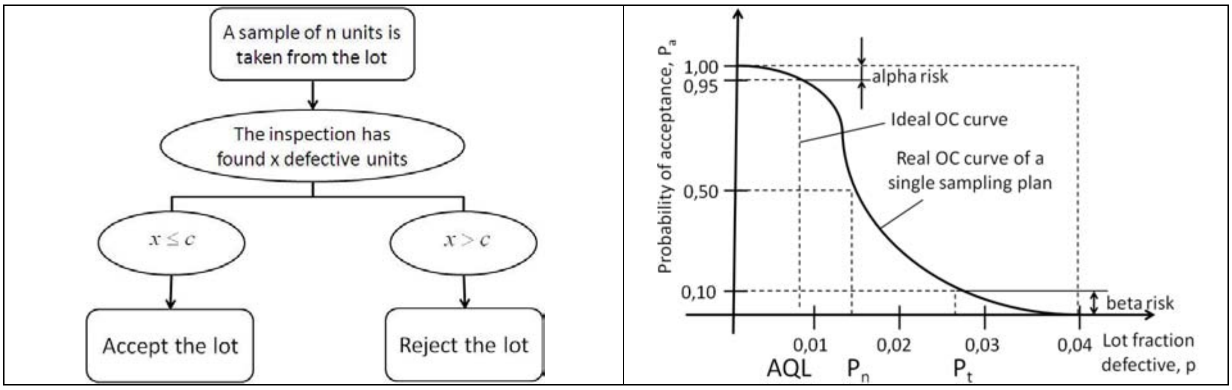


Figure 1: Decision Making and the OC Curve of a Single Sampling Plan

An important measure of performance of an acceptance sampling plan is the operating characteristic (OC) curve. The general shape of the OC curve for a single sampling plan is shown in Figure 1 [3]. An OC curve shows the discriminatory power of the sampling plan. This means that the OC curve shows the probability that a lot, submitted with a certain fraction defective, will be either accepted, or rejected [5]. The lot fraction of defective units representing the proportion (or probability) of defective units that are made in a certain process and denoted as p is shown on the x- axis. The probability of acceptance of a certain lot is denoted as P_a and it is shown on the y-axis. If the lot size N is large enough to be declared as infinite, the distribution of the number of defectives x in a random sample of n units will be binomial with parameters n and p. In this case, the probability of observing exactly x defectives is:

$$P\{x\,defectives\} = f(x) = n! \big/ [x!(n-x)!]\, p^x (1-p)^{n-x} \tag{1}$$

From that, the probability of acceptance or probability that x is less than or equal to c is:

$$P_a = P\{x \le c\} = \sum_{x=0}^{c} n! \big/ [x!(n-x)!]\, p^x (1-p)^{n-x} \tag{2}$$

The OC curve contains some specific points. The first specific point is the acceptable quality level (AQL) and represents the poorest or minimum level of quality of supplier's process that the buyer would consider to be acceptable as a process average [5]. The decision of accepting or rejecting the observed lot is made through the AQL. Even if the supplier's process has a smaller lot fraction defective than it is defined by the AQL, the probability of rejecting a lot still exists because the chosen sample can show that there is a higher lot fraction defective than it is allowed by AQL. This probability is called supplier's risk or alpha risk. On other side, buyer's risk or beta risk is the probability that the buyer will accept a lot of poor quality. The lot tolerance percent defective (LTPD) is the minimum quality level that the buyer is willing to accept, and in Figure 1 it is marked as P_t. Between the AQL and P_t points there is an area of indifference. The point P_n shows the neutral quality level which means that in this point there is probability of 0.5 that the lot will be accepted and probability of 0.5 that it will be rejected. Dumičić, Bahovec and Kurnoga Živadinović [1] have observed the movements of the OC curve if one of the parameters is changed and others are kept at the same level.

2.2 Double sampling plan

A double sampling plan implies conducting inspections of lots in which, when it is necessary, two samples of units from the lot are taken. The procedure of a double sampling plan is shown in figure 2.

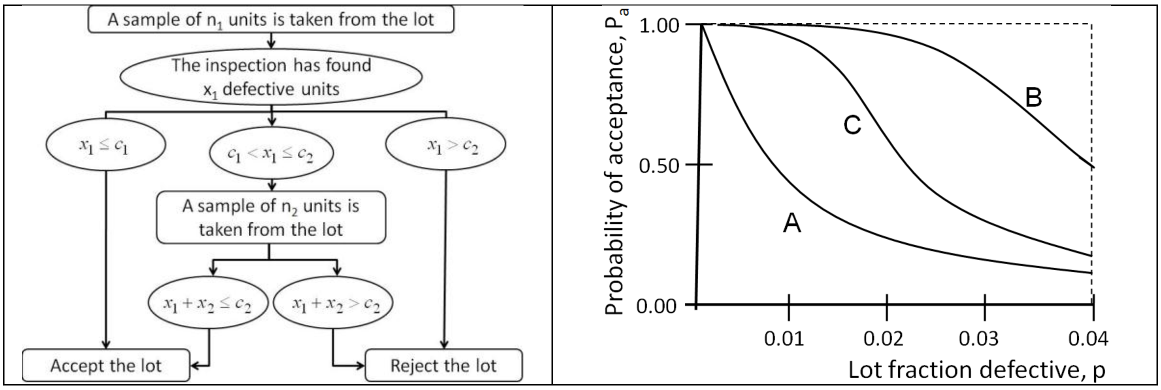


Figure 2: Decision Making and the OC Curve of a Double Sampling Plan

The first step of conducting the sampling plan is to form a sample of n_1 units from the observed lot. Usually, that sample is smaller than the sample which would be used if a single sampling plan was performed. This is the main advantage of the double over the simple sampling plan, the advantage is obvious only if this first sample leads directly to the conclusion to accept or reject the lot. The lot will be accepted if the number of defective units (x_1), found in the first sample (n_1), is smaller than the acceptance level for the first sample (c_1). The lot will be rejected if the number of defective units (x_1), found in the first sample (n_1), is bigger than the acceptance level for both samples (c_2). It is assumed that the rejection number on the first sample (r_1) is equal to the rejection number for both samples (r_2) and it is equal to c_2+1. If the number of defective units (x_1), found in the first sample (n_1), is bigger than the acceptance level for the first sample (c_1), but not bigger than the acceptance level for both samples (c_2), the second sample from the lot must be chosen. The second sample contains n_2 units and is usually equal to the first sample or twice the first [7]. If the overall number of defective units in both samples together (x_1+x_2) is found to be bigger than the acceptance level for both samples (c_2), then the lot will be rejected.

The OC curve for a double sampling plan is shown in Figure 2 [2]. It is more complex to construct than the one for a single sampling plan. The curve A in Figure 2 shows the probability of acceptance on the first sample, and the curve B describes the probability of not rejecting the lot on the first sample. The difference between the two curves is the probability of taking a second sample. The final OC curve is the curve C which describes the probability of acceptance for the double sampling plan. If the lot size is relative large in comparison to the sample size, the binomial distribution may be used. The probability of acceptance for the double sampling plan can be calculated as sum probabilities of acceptance on the first and second samples:

$$P_a = P_a^I + P_a^{II} = P\{x_1 \le c_1\} + \sum_{i=c_1+1}^{c_2} P\{x_1 = i\} P\{x_2 \le c_2 - i\} \qquad (3)$$

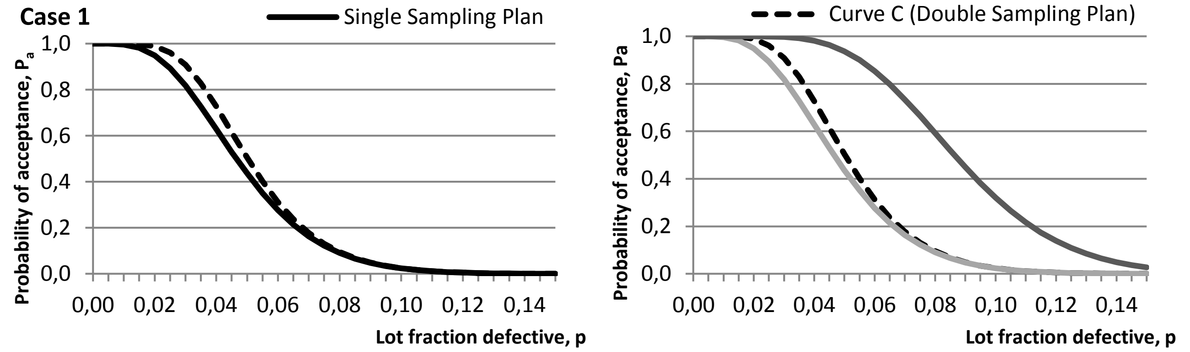
The probability of acceptance on the first sample (P_a^I) includes probability for $x_1 \le c_1$ and the probability of acceptance on the second sample includes probability of different ways in which the second sample can be obtained. A second sample is chosen only if $c_1 \le x_1 \le c_2$.

2.3 Example

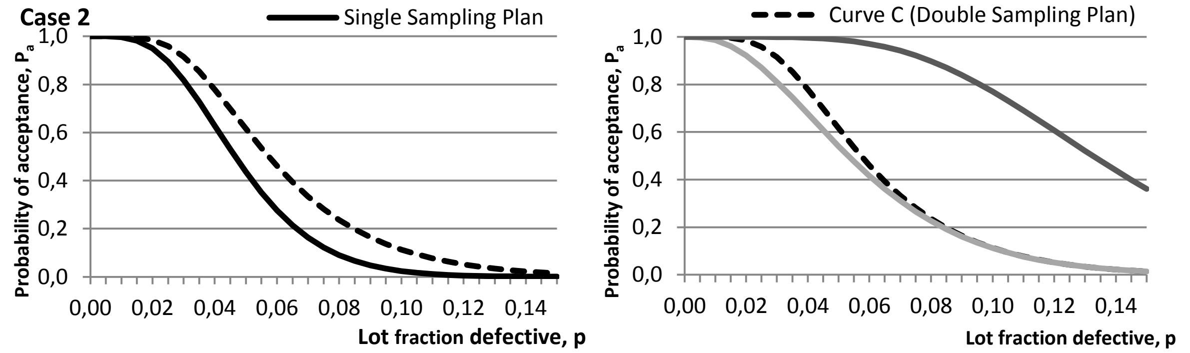
An example in which probabilities of acceptance and rejection of a lot are going to be compared using a single and a double sampling plan is presented. In other words, probabilities of alpha and beta risks in a single versus a double sampling plan will be compared. The comparison shows when any significant difference appears. Probabilities of alpha and beta risks are shown on the OC curves, and the calculations are made in Excel 2007. A conclusion about existence of any significant difference between the compared probabilities of alpha and beta risks in a single versus a double sampling plan at the same level of AQL and LTPD is reached using the test for differences in the two proportions.

For the purposes of this example, AQL is set at 0.05. This means, the buyer assumes that the supplier's process is making 5% bad products. On the other side, the buyer is establishing LTPD at 0.10 or 10%. The supplier is sending a lot which contains 10,000 units of a certain product. The buyer considers using a single or a double sampling plan for conducting inspection of the lot. A random sample of n=100 units is selected to conduct the inspection of the lot using a single sampling plan. As an acceptance number or the maximum allowed number of observed defective units in the lot c=4 is determined. These settings are held constant throughout the whole example. In the *Case 1*, values defining the double sampling plan are n_1=100, c_1=4, n_2=100, and c_2=8. It could be noticed that n=n_1 and c=c_1, so that the OC curve of the single sampling plan is equal to the curve A of the double sampling plan. Because it is not common that n=n_1, since this neutralizes the main advantage of performing a double sampling plan, in the *Case 2* the size of the first sample is n_1=50 units. Hence the other values are c_1=2, n_2=100, and c_2=6. In the *Case 3* the size of samples is the same or n_1=n_2=50 because usually the second sample size is the same or double size of the first sample. The acceptance levels are c_1=2 and c_2=4.

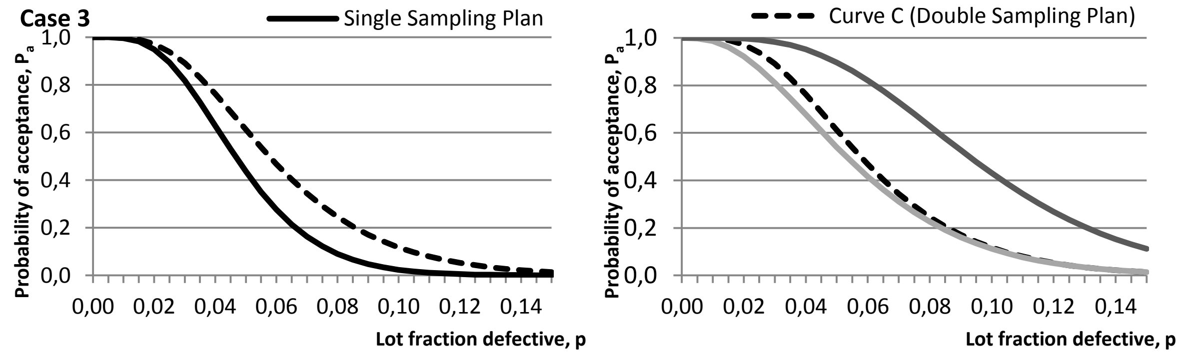
The OC curves of the single and the double sampling plans for all the three cases are shown in Figure 3. On the left side, the OC curves of the single and the double sampling plans are compared, and on the right side, curves A, B and C for the double sampling plan are shown. If the sample size is n=100, the maximum allowed number of the observed defective units in the lot c=4, and the probability of acceptance P_a or AQL=0.05, then, for the single sampling plan, the lot fraction defective is p=0.019906 or the lots are 1.99% defective. So, when 100 lots from a process that manufactures 1.99% defective products are submitted to this sampling plan, it will be expected to accept 95 of the lots and reject 5 of them. In this case the alpha risk or the risk of rejecting a good lot is 0.05. With the given sample size, the allowed observed defective units in the lot, and with the probability of acceptance P_a or LTPD=0.10, the probability of the lot fraction defective for the single sampling plan is p=0.078347. This means that when 100 lots from a process that manufactures 7.83% defective products are submitted to this sampling plan, it will be expected to accept 90 of the lots and reject 10 of them. In this case the beta risk or the risk of accepting a bad lot is 0.10. Because the parameters of the single sampling plan are kept constant through all cases, the probabilities of the lot fraction defective remained at the same values. Double sampling plans are used, first, with varying sample sizes, and after that with a different maximum allowed number of observed defective units in the lot, which are kept proportional to sample sizes. It follows, for the observed probabilities of acceptance, the lot fraction defectives are different in each case. It must be emphasized that it is assumed that the second sample must be selected to make the decision.



	Lot fraction defective, p		Sample size (n; n_1+n_2)
	AQL = 0.05	LTPD=0.10	
Single sampling plan	0,0199	0,0783	100
Double sampling plan	0,0264	0,0793	200



	Lot fraction defective, p		Sample size (n; n_1+n_2)
	AQL = 0.05	LTPD=0.10	
Single sampling plan	0,0199	0,0783	100
Double sampling plan	0,0262	0,1032	150



	Lot fraction defective, p		Sample size (n; n_1+n_2)
	AQL = 0.05	LTPD=0.10	
Single sampling plan	0,0199	0,0783	100
Double sampling plan	0,0235	0,1041	100

Figure 3: OC Curves of the Single and the Double Sampling Plans

Tests for differences between two proportions were conducted. The first proportion was taken from a single sampling plan, and the second from a double sampling plan. In fact, proportions are the lot fraction defectives taken from both sampling plans for AQL=0.05 or alpha risk of 5%, and then for LTPD=0.10 or beta risk of 10%. The null hypothesis states that the lot fraction defective is equal in a single and a double sampling plan at the same levels of risks. The alternative hypothesis assumes that there is a statistically significant difference between the lot fraction defectives for the same level of risks between both sampling plans. In Table 1 results of the tests conducted are shown.

Table 1: Test for differences in the two proportions

	Probability of acceptance, P_a	Lot fraction defective, *p*		Emp. Z	P-value
		Single sampling plan	Double sampling plan		
Case 1	AQL=0.05	0.019906	0.026351	-0.3424	0.7320
	LTPD=0.10	0.078347	0.079267	-0.0279	0.9778
Case 2	AQL=0.05	0.019906	0.026155	-0.3185	0.7501
	LTPD=0.10	0.078347	0.103244	-0.6631	0.5073
Case 3	AQL=0.05	0.019906	0.023493	-0.1741	0.8618
	LTPD=0.10	0.078347	0.104101	-0.6325	0.5271

According to Table 1, all p-values are high, so at all common levels of significance it can be concluded that there is not enough evidence to reject the null hypothesis and this means that there is not any significant difference in the lot fraction defective between a single and a double sampling plan at given levels of alpha and beta risks for the given example.

3 CONCLUSIONS

Using three examples and a test for the difference in proportions, this paper analyses the probability of acceptance of lots when using a single or a double acceptance sampling plan. All three cases have proved that using moderate sample sizes (50 to 200) and the given values AQL=0.05 and LTPD=0.10 results in no statistically significant differences between the probability of acceptance of lots between the two observed acceptance sampling plans. The named analysis leads to the conclusion that both observed acceptance sampling plans result in the same probability of lot acceptance, i.e. in the same expected products quality, thus implying that there is no possibility of manipulation by choosing one of the two acceptance sampling plans.

It can be concluded that using statistical tests, random samples of reasonable size and applying criteria of statistical significance, there is no possibility of unnoticed manipulation with the quality level of products by choosing a single or a double acceptance sampling plan. In other words, applying different acceptance sampling plans in the conditions of the same quality level will not result in different conclusions, and therefore neither in different decisions by quality managers regarding rejection or acceptance of a lot. In further studies, we suggest using non-proportional values of the maximum allowed number of the observed defective units in proportion to the sample size as well as conducting an analysis with the hypothesis that it is not necessary to choose a second sample in a double acceptance sampling plan.

References

[1] Dumičić, K., Bahovec, V., Kurnoga Živadinović, N., 2006. Analysing the Shape of an OC Curve for an Acceptance Sampling Plan: A Quality Management Tool. WSEAS Transactions on Business and Economics, Issue 3, Volume 3, March 2006, pp. 169-177.

[2] Grant, E.L., Leavenworth, R.S., 1980. Statistical Quality Control. 5th Ed. New York: McGraw-Hill.

[3] Kondić, Ž., 2004. Kvaliteta i metode poboljšanja. Čakovec: Zrinski.

[4] Montgomery, D. C., 2009. Statistical Quality Control: A Modern Introduction, 6th Ed. New York: John Wiley & Sons.

[5] Montgomery, D. C., Jennings, C. L., Pfund, M. E., 2011. Managing, Controlling, and Improving Quality. New York: John Wiley & Sons.

[6] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/pmc/section2/pmc22.htm, 14.6.2011.

[7] Wadsworth, H. M., Stephens, K. S., Godfrey, A. B., 2002. Modern Methods for Quality Control and Improvement. 2nd Ed. New York: John Wiley & Sons.

PRINCIPAL COMPONENT ANALYSIS FOR AUTHORSHIP ATTRIBUTION

Amir Jamak, Alen Savatić and **Mehmet Can**
International University of Sarajevo, Faculty of Engineering and Natural Sciences
Hrasnićka Cesta 15, 71000 Sarajevo, Bosnia and Herzegovina
amir.jamak@bhtelecom.ba , alen@savatic.net , mcan@ius.edu.ba

Abstract: A common problem in statistical pattern recognition is that of feature selection or feature extraction. Feature selection refers to a process whereby a data space is transformed into a feature space that, in theory, has exactly the same dimension as the original data space. However, the transformation is designed in such a way that the data set may be represented by a reduced number of "effective" features and yet retain most of the intrinsic information content of the data; in other words, the data set undergoes a dimensionality reduction.
In this paper the data collected by counting words and characters in around a thousand paragraphs of each sample book underwent a principal component analysis performed using neural networks. Then first of the principal components is used to distinguished the books authored by a certain author.

Keywords: principal components, authorship attribution, stylometry, text categorization, function words, classification task, stylistic features, syntactic characteristics.

1 INTRODUCTION

Of all text categorization problems, that of authorship attribution is probably the oldest; however, it is also possibly the least well organized, and its history is marred with the mishandling of statistical techniques. And yet, it still promises to provide useful applications in spheres as diverse as law, security, and education.

The origin of non-traditional authorship attribution, or stylometry, is often said to be Augustus de Morgan's suggestion in 1851 that certain authors of the Bible might be distinguishable from one another if one used longer words (Holmes 1998). In 1887, Mendenhall began investigating this hypothesis, searching for a characteristic difference in the distribution of different-sized words in writings of different languages and presentation styles. In 1901, he turned his methods to Shakespeare, Bacon and Marlowe, and found that while Shakespeare and Marlowe were nearly indistinguishable, they were both significantly and consistently different from Bacon (Williams 1975). The difference was mainly observed in the relative frequency of three- and four-letter words: Shakespeare used more four - letter words and Bacon more three-letter words.

Authorship studies also began independently around the same time in Russia; it seems, with Morozov proposing a model for measuring style that garnered the interest of A. Markov (Kukurushkina et al. 2002). In the West, it took 30 years or so for Mendenhall's studies to be resumed by other linguists. George Zipf examined word frequencies and determined not a stylometric but a universal law of language, Zipf's Law: that the statistical rank of a word varies inversely to its frequency. G. Udny Yule devised a feature known as "Yule's characteristic K," which estimated 'vocabulary richness' by comparing word frequencies to that expected by a Poisson distribution, but like Mendenhall's word lengths, this too was later found to be an unreliable marker of style (Holmes 1998). In fact, most of the measurements proposed in this period proved unhelpful: among others, researchers tried average sentence length, number of syllables per word, and other estimates of vocabulary richness such as Simpson's D index and a simple type/token ratio (a ratio of the number of unique words, or types, to the number of total words, or tokens) (Juola et. al. 2006).

A breakthrough was needed, and it came in 1963 with Mosteller and Wallace's study on the Federalist Papers. In 1787 and 1788, John Jay, Alexander Hamilton and James Madison collectively wrote 85 newspaper essays supporting the ratification of the constitution. Published under the pseudonym "Publius," the authors later revealed which of the Federalist Papers they had written; however, while authorship of 67 were undisputed, 12 were claimed by both Hamilton and Madison. Mosteller and Wallace hoped to characterize each author's style through their choice of function words, such as "to," "by," and so forth. Function words are regarded as good markers of style because they are (assumed to be) unconsciously generated and independent of semantics (meaning, or what the author is trying to convey). That is, an author may have a preference for modes of expression (for instance, the active vs. the passive voice) that emphasize certain function words, and the same broad set of function words will be used regardless of the topic at hand.

Despite the fact that Hamilton and Madison have otherwise very similar styles—nearly identical sentence length distributions, as noted by (Juola 2006)—Mosteller and Wallace found sharp differences in their preference for different function words: for instance, the word "upon" appears 3.24 times per 1000 words in Hamilton, and just 0.23 times in Madison (quoted in Holmes 1998). Adjusting these frequencies with a Bayesian model, they showed that Madison had most likely written all 12 disputed papers. Traditional scholarship had already long come to the same conclusion, but Mosteller and Wallace's conclusion was independent, and thus a great achievement of the then quite exploratory field of stylometry. The Federalist Papers problem is still regarded as a very difficult test case, and as an unofficial benchmark it has been used to test most methods of authorship attribution developed since then (see, for instance, Kjell 1994, Holmes & Forsyth 1995, Bosch & Smith 1998, and Fung 2003).

2 PROBLEM DEFINITION

In this paper an application of principal component analysis is presented. The authorship attribution is considered as a classification task (Chaski, C. 2001 , 2005). Texts studied are literary works of three Bosnian writers, Ivo Andrić (1892-1975) , M. Meša Selimović (1910-1982), and Derviš Sušić (1925–1990). Feature selected to describe texts are lexical and syntactical components that show promising results when used as writer invariants because they are used rather subconsciously and reflect the individual writing style which is difficult to be copied. Principal components of data elicited from texts possess generalization properties that allow for the required high accuracy of classification (Hayes 2008).

2.1 Texts Used

In research texts of two famous Bosnian writers, Ivo Andrić, M. Meša Selimović, and Derviš Sušić are used. Their novels provide the corpora which are wide enough to make sure that characteristic features found based on the training data can be treated as representative of other parts of the texts and this generalized knowledge can be used to classify the test data according to their respective authors.

Obviously literary texts can greatly vary in length; what is more, all stylistic features can be influenced not only by different timelines within which the text is written but also by its genre. The first of these issues is easily dealt with by dividing long texts, such as novels, into some number of smaller parts of approximately the same size.

Described approach gives additional advantage in classification tasks as even in case of some incorrect classification results of these parts the whole text can still be properly attributed to some author by based the final decision on the majority of outcomes instead of all individual decisions for all samples.

Whether the genre of a novel is reflected in lexical and syntactic characteristics of it is the question yet to be answered. If the influence is significant, then lexical and syntactic features cannot be used as the writer invariant as unreliable.

Hence all together we have selected thousands of paragraphs from "Na Drini Ćuprija, Znakovi Pored Puta, Prokleta Avlija " by Ivo Andrić, "Derviš i Smrt, Tvrdjava" by M. Meša Selimović, and “Pobune” by Derviš Sušić.

2.2 Feature Selection

Establishing features that work as effective discriminators of texts under study is one of critical issues in research on authorship analysis which are lexical. In this research five textual descriptors are used, numbers of characters, words, sentences, commas, and conjecture “and”, in Bosnian “i”, and other characteristics in paragraphs. Means and variances of the textual descriptors for the texts Ivo Andrić: Na Drini Ćuprija, and M. Meša Selimović: Derviš i Smrt are shown in Table 1.

Table 1: Paragraph averages and variances of the textual descriptors used in this research

	Ivo Andrić: Na Drini Ćuprija		M. Meša Selimović: Derviš	
Textual descriptors	Mean	Variance	Mean	Variance
Sentence length	84.331	2090.92	58.710	2053.855
Word length	2.157	2.877	2.155	3.460
Word count	79.208	5861.724	60.362	4756.432
Sentence count	4.395	16.886	5.012	29.411
Comma count	6.432	45.95	7.130	87.211
dots count	0.052	0.135	0.002	0.002
i count	5.375	35.072	2.235	9.659
ili count	0.250	0.514	0.302	0.688
je count	2.798	11.991	2.552	11.531
se count	1.852	4.823	1.615	4.478
pa count	0.140	0.216	0.098	0.133
da count	1.935	6.853	2.262	9.613
ne count	0.637	1.695	0.968	2.718
kao poput count	0.662	1.106	0.480	1.007
Total		8080.760		6970.200

As it is seen, there is statistical difference between the usage of textual descriptors, for instance, Ivo Andrić prefers longer paragraphs. In average Ivo Andrić ‘s paragraphs contain 79 words with variance 5861.7, while Meša Selimović’s average is 62 with variance 4756.4.

In the next chapter the pattern captured by principal components will be displayed.

3 PRINCIPAL COMPONENT ANALYSIS

The methods of Mosteller and Wallace have proved as enduring as the problem they investigated: they were only modestly altered when Burrows described his method of stylometric analysis in a series of papers published in the late 1980s and early 1990s (Holmes 1998; see, for instance, Burrows 1992). The Burrows method essentially involves computing the frequency of each of a list of function words (larger than that of Mosteller and Wallace), and performing principle component analysis (PCA) to find the linear combination of variables that best accounts for the variations in the data. Rather than analyze this result statistically, the transformed data are simply plotted (a two-dimensional plot of the first two principal components) and inspected visually for trends, which occur as clusters of points (Holmes 1998). (Later, cluster analysis would accomplish this step.)

This simple but effective method continues to be used today, partly because of the ease with which the results are communicated and interpreted. For example, Binongo used this method to study the problem of the authorship of L. Frank Baum's last book, which historians had long suspected of being mostly the work of Baum's successor, Ruth P. Thompson (Binongo 2003). He confirmed this suspicion independently, demonstrating that Thompson was much more prone to use position words such as "up," "down," "over," and "back," than Baum. This was not demonstrated using complex statistical techniques; rather, function word frequencies were tallied, the authors' tallies compared, PCA used to reduce the dimensionality of the data, and the resulting plots inspected: the two authors' works form obvious clusters. Similar procedures can be found in (Holmes & Forsyth 1995, Holmes et al. 2001, and Peng & Hentgartner 2002).

4 PRINCIPAL COMPONENTS OF SAMPLE TEXTS

Next, random samples of 400 data are chosen from data sets for the textual descriptors for the texts authored by Ivo Andrić: Na Drini Ćuprija, and M. Meša Selimović: Derviš i Smrt, X_{ivo}, and X_{mesa}, and for other four books. These are all 400×14 matrices. Their covariance matrices C_{ivo}, C_{mesa} are 14×14 matrices. The information in the covariance matrices are used to define a set of new variables $P_{ivo} = X_{ivo}.C_{ivo}$, and $P_{mesa} = X_{mesa}.C_{mesa}$ as a linear combination of the original variables in the data matrices**.** The new variables are derived in a decreasing order of importance. The first column of $P_{(\ )}$ is called first principal component and accounts for as much much as possible of the variation in the original data. The second column is called second principal component and accounts for another, but smaller portion of the variation, and so on.

If there are p variables, to cover all of the variation in the original data, one needs p components, but often much of the variation is covered by a smaller number of components. Thus PCA has as its goals the interpretation of the variation and data reduction.

Variances and percentage variances covered by fourteen principal components of the textual descriptors for the sample texts Ivo Andrić: Na Drini Ćuprija, and M. Meša Selimović: Derviš i Smrt are shown in Table 2.

Table 2: Variances and percentage variances covered by fourteen principal components of the textual descriptors used in this research.

	Ivo Andrić: Na Drini Ćuprija		M. Meša Selimović: Derviš	
P. Comp.	Variance	% Variance covered	Variance	% Variance covered
1	7447.15422	75.6006325	5374.75843	77.1105536
2	2376.67024	24.1270381	1561.30450	22.3997146
3	8.130187	0.0825345	14.211600	0.2038909
4	5.310098	0.0539061	6.152396	0.0882672
5	3.199071	0.0324757	3.335845	0.0478587
6	2.811245	0.0285387	2.884413	0.0413821
7	2.152849	0.0218549	2.027011	0.0290811
8	1.569122	0.0159291	1.644081	0.0235873
9	1.345059	0.0136545	1.530064	0.0219515
10	0.830111	0.0084270	1.078908	0.0154789
11	0.777950	0.0078975	0.700177	0.0100453
12	0.477576	0.0048482	0.451615	0.0064792
13	0.148686	0.0015094	0.116703	0.0016743
14	0.074267	0.0007539	0.002465	0.0000354
	8080.76	100	6970.20	100

Table 2 reveals that the first two principal components cover more than %99 of variances of principal components.

In Figure 1 below, one first principal component of each of samples from Cuprija na Drina and Derviš i Smrt data are displayed.

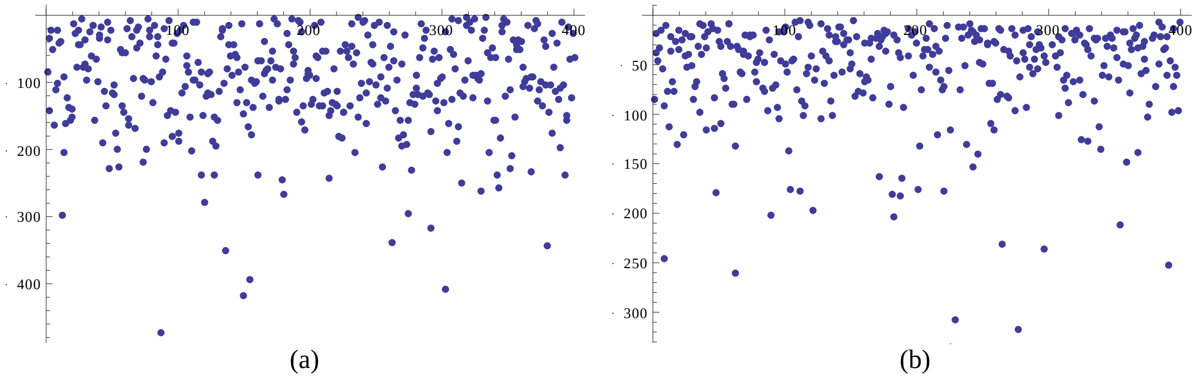


Figure 1: First principal components of samples from Cuprija na Drina (a) and Derviš i Smrt (b) data.

These figures are similar, and do not seem to be used as writeprints of authors. It is the same for the second principal components. To search for a writerprint, we transform this information into the frequency domain. A common range for the contents of these two vectors is the interval $[-500, 0]$. We divide this interval into 25 bins of equal length of 20

$$\{[-500, -480), [-480, -460), \ldots, [-40, -20), [-20, 0]\}$$

and count the numbers of entries of first component vectors in these bins. Figure 2 displays the data in Figure 1 in frequency domain.

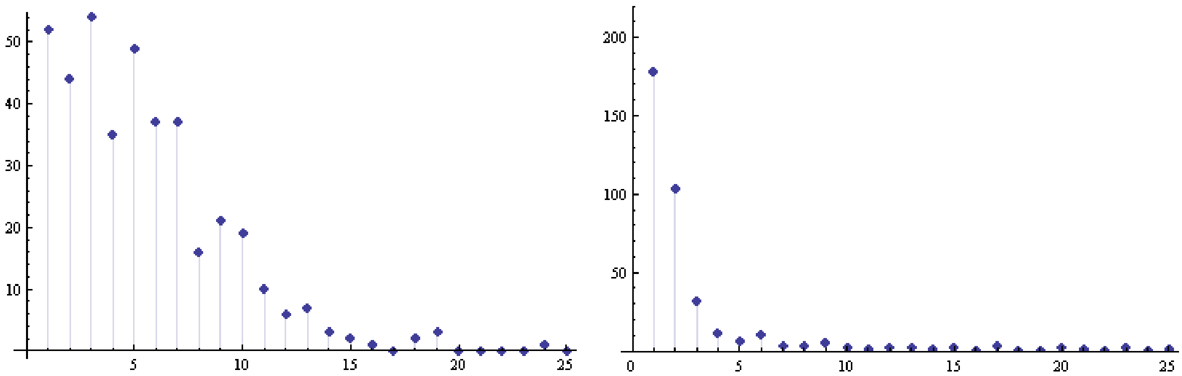


Figure 2: Frequencies of elements of first principal component vectors of random samples from Cuprija na Drina (a) and Derviš i Smrt (b) data in 25 bins.

It is seen that the writeprints of the two authors are distinguishable. To see whether the captured features remains similar through random samplings from data sets, we sketch together the frequencies of ten different samples in figure 3.

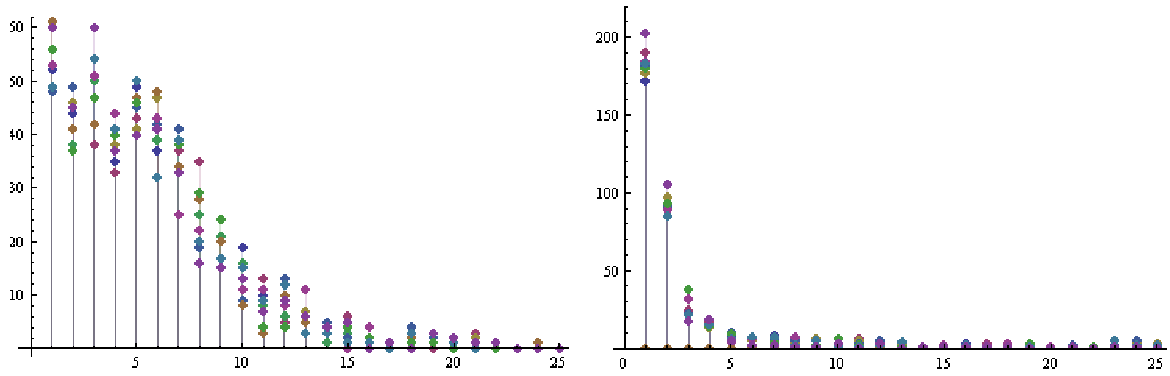
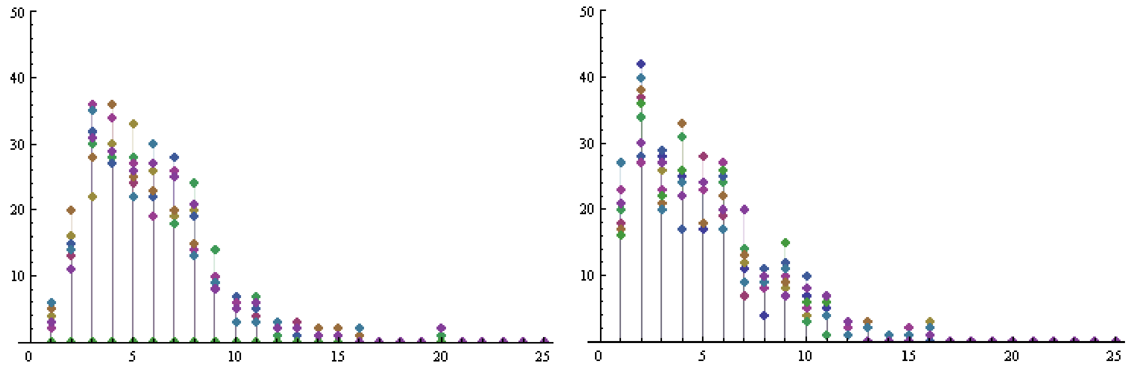


Figure 3: Frequencies of elements of first principal component vectors of ten random samples from Cuprija na Drina (a) and Derviš i Smrt (b) data in 25 bins.

To check whether these patterns are characteristic for other books of the two authors, two more books of Ivo Andrić; Znakovi Pored Puta, and Proklet Avlija, and one other book of Meša Selimović; Tvrdjeva, as well as Pobune authored by a third novelist Derviš Sušić are investigated.

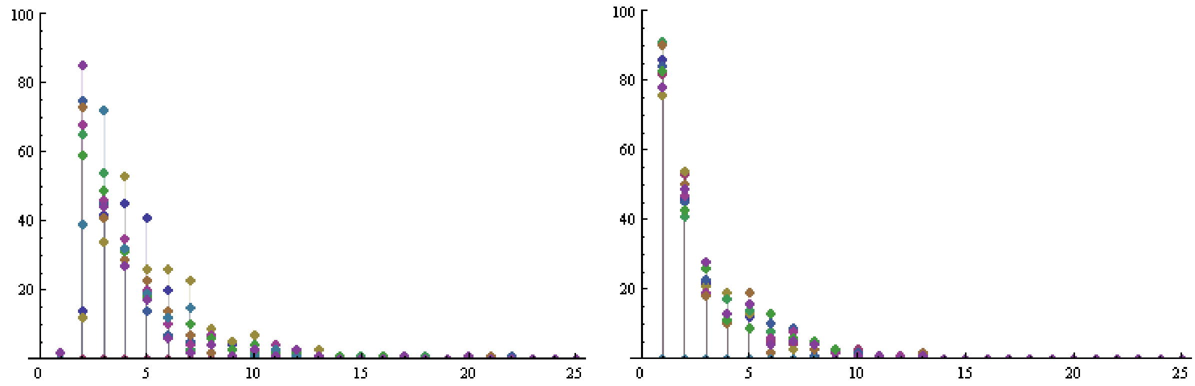
The comparison of the frequencies in the first principal components of the three books authored by Ivo Andrić: Cuprija na Drina, Znakovi Pored Puta, Proklet Avlija are shown in Figure 3 below. The writing print of Ivo Andrić is the lower peaks – less than 70 – at the lowermost values of the principal components.



a) b)

Figure 3: Frequencies of data in the first principal components of the two other books authored by Ivo Andrić: Znakovi Pored Puta a), and Proklet Avlija b).

The first principal components of the another book authored by Meša Selimović; Tvrdjeva displayed in Figure 4a, a third author's text Pobune (Sušić 1966) in Figure 4a. The writing print of Meša Selimović is revealed as twice higher peaks compared to the corresponding Ivo Andrić peaks, and differs significantly from pattern for Derviš Sušić.



a) b)

Figure 4: Frequencies of data in the first principal components of the book authored by Meša Selimović; Tvrdjeva a), and third author's text Pobune b).

5 CONCLUSIONS

The research described in this paper concerning author identification analysis shows that the method of principal component analysis (PCA) is an efficient a tool. Yet conclusions as to the choice of textual descriptors used as features for recognition process, based only on results presented in the previous sections and leading to some arbitrary statement that syntactic attributes are more effective in authorship attribution, would be much too hasty and premature. Undeniably true in the studied example, it would have to be verified against much wider corpora as for other writers other features could give better results.

Thus a series of future experiments should include artificial neural networks -based methodology to wider range of authors, definition of new sets of textual descriptors, and test for other types and structures of neural networks, and search the possibility of inheritance through translation into other languages.

Once a method for finding write prints, it is not difficult to deal with the author attribution problems, simply by the use of perceptrons of artificial neural networks. Indeed in a series of articles, the authors of this article, with a group of researcher at the International University of Sarajevo follow this path (Can, Jamak, Savatić 2011, Savatić, Can, Jamak 2011, Can, Hadžiabdić, Demir 2011, Selman, Turan, Kuşakçı, 2011)

References

[1] Andrić, I. 1981. Na Drini Ćuprija, Svjetlost, Sarajevo.

[2] Andrić, I. 1989. Znakovi Pored Puta,, Svjetlost, Sarajevo.

[3] Andrić,I. 1980. Prokleta Avlija, Svjetlost, Sarajevo.

[4] Binongo, J. 2003. Who wrote the 15th book of Oz? An application of multivariate analysis to authorship attribution. Chance 16(2): 9–17.

[5] Bosch, R., and J. Smith. 1998. Separating hyperplanes and the authorship of the disputed federalist papers. American Mathematical Monthly 105(7): 601–8.

[6] Burrows, J. 1992. Not unless you ask nicely: The interpretative nexus between analysis and information. Literary and Linguistic Computing 7(2): 91–109.

[7] Can, M., A. Jamak, and A. Savatić. 2011. Teaching Neural Networks to Detect the Authors of Texts Using Lexical Descriptors, 9th International Conference on Knowledge, Economy & Management Proceedings,ISBN: pp. 3607-3624.

[8] Can M., K.K. Hadžiabdić, N. M. Demir. 2011. Teaching Neural Networks to Detect the Authors of Texts, Using Lexical Descriptors, 9th International Conference on Knowledge, Economy & Management Proceedings, pp. 1393-1402.

[9] Chaski, C. 2001. Empirical evaluations of language-based author identification techniques. Journal of Forensic Linguistics. 8(1): 1–65.

[10] Chaski, C. 2005. Who’s at the keyboard? Authorship attribution in digital evidence investigations. International Journal of Digital Evidence 4(1).

[11] Fung, G. 2003. The disputed Federalist Papers: SVM feature selection using concave minimization. Proceedings of the 2003 Conference on Diversity in Computing. 42–6.

[12] Hayes, J. F. 2008, Authorship Attribution: A Principal Component and Linear Discriminant Analysis of theConsistent Programmer Hypothesis, I. J. Comput. Appl. 15, No. 2, 79-99 (2008).

[13] Holmes, D. 1998. The evolution of stylometry in humanities scholarship. Literary and Linguistic Computing 13(3): 111–7.

[14] Holmes, D., and R. Forsyth. 1995. The Federalist revisited: New directions in authorship attribution. Literary and Linguistic Computing 10(2): 112–27.

[15] Holmes, D., L. Gordon, and C. Wilson. 2001. A widow and her soldier: Stylometry and the American Civil War. Literary and Linguistic Computing 16(4): 403–20.

[16] Juola, P. 2006. Authorship attribution. Foundations and Trends in Information Retrieval 1(3): 233–334.

[17] Juola, P., J. Sofko, and P. Brennan. 2006. A prototype for authorship attribution studies. Literary and Linguistic Computing 21: 169–78.

[18] Kjell, B. 1994. Authorship determination using letter pair frequency features with neural network classifiers. Literary and Linguistic Computing 9(2): 119–24.

[19] Kolman, B., and D. R. Hill. 2004. Elementary Linear Algebra, Pearson, New Jersey.

[20] Kukushkina, O., A. Polikarpov, and D. Khmelev. 2002. Using literal and grammatical statistics for authorship attribution. Problemy Peredachi Informatsii 37(2).

[21] Peng, R., and N. Hengartner. 2002. Quantitative analysis of literary styles. The American Statistician 56(3): 175–85.

[22] Savatić, A., A. Jamak, and M.Can. 2011. Detecting the Authors of Texts by Boosting Neural Network Committee Machines, Proceedings of the 2nd International Scientific and professional Conference of Graphic Technology and Design, 9-11 June 2011, Kselajak, BiH, pp. 223-232.

[23] Selimović, M. M.1966 , Derviš i smrt, Svjetlost, Sarajevo.

[24] Selimović, M. M. 1970. Tvrdjava, Svjetlost, Sarajevo.

[25] Selman S., K. Turan, and A. O. Kuşakçı. 2011. Distingtion of the Authors of Texts Using Multilayered Feedforward Neural Networks, 9th International Conference on Knowledge, Economy & Management Proceedings, pp. 1419-1429.

[26] Sušić, D. 1966. Pobune, Veselin Masleša, Sarajevo.

[27] Williams, C. 1975. Mendenhall’s studies of word-length distribution in the works of Shakespeare and Bacon. Biometrika 62(1): 207–12.

The 11th International Symposium on
Operational Research in Slovenia

SOR '11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Section VII:

Production and Inventory

VOLUME DISCOUNTS IN MULTIPRODUCT SUPPLIER SELECTION PROBLEM

Zoran Babić[*] and **Tunjo Perić**[**]
[*]Faculty of Economics, Split, Matice Hrvatske 31, E-mail: babic@efst.hr
[**]Bakeries Sunce, Rakitje, Rakitska cesta 98, 10437 Bestovje, E-mail: tunjo.peric1@zg.t-com.hr

Abstract: Supplier selection is a multi-criteria decision making problem that is affected by quantitative and qualitative conflicting factors. For these problems, if suppliers offer quantity or money volume discount and a buyer wants to buy multiple products as well, the problem becomes more complicated.
This paper presents only the first part of the problem, namely vendor selection problem with volume discounts in the multiproduct case. The problem is solved with one objective only, minimizing the total purchasing costs and leaving the multi-criteria problems for the further research. The paper deals with a concrete problem of flour purchase by a company that manufactures bakery products.

Keywords: vendor selection, volume discounts, multiproduct.

1 INTRODUCTION

The problem of vendor (supplier) selection and determination of material quantities supplied is the key element in the purchasing process in manufacturing which is one of the most important activities in supply chain. If all the selected vendors are able to meet the buyer's requirements completely, then the selection process becomes easier and is based only on the selection of the most suitable vendor in terms of purchasing costs, product quality, and vendor reliability. Nevertheless, practice shows that it is not good to rely on one vendor only. Therefore the management of the purchasing company generally enters into contracts with several vendors. Moreover, even if one supplier dominates the others, the buyer might choose to give business to all of them. There are more strategic reasons for such decision (multi-sourcing). For example, a buyer might wish to prevent any supplier to become a monopolist and buyer may award contracts to several suppliers to keep them solvent and thereby encourage their continued presence in future contract competitions. Their number usually ranges from two to five for each sort of material. Also, there are cases when no vendor can meet the buyer's demand, or will not do it in order to protect his own business interests.

In this paper we will discuss the supplier selection problem when there are some limitations on suppliers' capacity, quality and so on, where no supplier can satisfy the buyer's total requirements and the buyer needs to purchase some of the needed material from one supplier and some from another to compensate for the shortage of capacity or low quality of the first supplier.

Identifying suppliers with the lowest item price in a given industry becomes a major challenge for purchasing managers, especially when suppliers offer multiple products and volume-based discount pricing schedules. In this environment, the supplier induces the buyer into making large purchases by offering discounts on the total value of sales volume, not on the quantity or variety of products purchased over a given period of time.

In traditional quantity discount pricing schedules, price breaks are function of the order quantity existed for each product, irrespective of the total magnitude of business the buyer contracts with the supplier over a given period of time. When every vendor offers the variety of products, vendors are finding it more meaningful to give discounts on the total value of multiproduct orders placed by a given buyer.

Vendor selection is an important issue dealt with by numerous researchers. Great efforts are made to define appropriate models for vendor selection and determination of supply quotas from the selected vendors and to apply the adequate methods to solve such models. Of course, vendor selection is a multi-criteria problem by nature. In this paper this problem is solved with one objective only, minimizing the total purchasing costs and leaving the multi-criteria problems for some further researches. We will first present the methodology of vendor selection and determination of supply quotas. Then the proposed methodology will be tested on the concrete example of vendor selection by a bakery.

2 MULTIPRODUCT VENDOR SELECTION MODEL WITH VOLUME DISCOUNTS

Consider a procurement situation in which we have m items P_i to be purchased $i \in I = \{1,2,...,m\}$ from n vendors V_j, $j \in J = \{1,2,...,n\}$ that provide different levels of item price, product quality, reliability performance, and supply capacity for each item they sell. Also, depending on the buyer's total purchase value, vendor V_j offers a business volume discount having r discount brackets, $r \in R_j = \{1,2,...,m_j\}$.

In order to formulate the multiproduct vendor selection model with volume discounts, the following notations are defined:

J_i = set of vendors offering item P_i , $J_i \subseteq J$,

I_j = set of items offered by vendor V_j , $I_j \subseteq I$,

D_i = units of item i demanded by the buyer,

c_{ij} = unit price of item P_i quoted by vendor V_j ,

Q_{ij} = maximum quantity of item P_i that may be purchased from vendor V_j ,

u_{jr} = upper cutoff point of discount bracket r for vendor V_j ,

u_{jr}* = slightly less than u_{jr} ,

d_{jr} = discount coefficient associated with bracket r of vendor j's cost function.

Decisions variables are:

x_{ij} = units of item P_i to purchase from vendor V_j ,

v_{jr} = volume of business awarded (*money amount*) to vendor V_j in discount bracket r. Observe that $v_{jr} > 0$ only if the money amount of purchase made from vendor j falls within bracket r of its cost function, otherwise it is zero.

y_{jr} = 0 or 1; 1 - if the volume of business awarded to vendor V_j falls on segment r of its cost function, 0 - otherwise.

The mathematical programming formulation of the model is as follows:

$$\min Z_1 = \sum_{j \in J} \sum_{r \in R_j} \left(1 - d_{jr}\right) \cdot v_{jr} \tag{1}$$

Equation (1) minimizes the total purchase cost where d_{jr} are the percent of discount.

Constraints:

$$\sum_{j \in J_i} x_{ij} = D_i \ , \ i \in I \tag{2}$$

Constraints (2) represent the condition that the total demand of each item will be satisfied.

$$x_{ij} \le Q_{ij}, \ i \in I, \ j \in J_i \tag{3}$$

Constraints (3) ensure that the total number of items procured by each supplier is within the production and shipping capacity of that supplier.

$$\sum_{i \in I_j} c_{ij} x_{ij} = \sum_{r \in R_j} v_{jr} \ , \ j \in J \tag{4}$$

Constraints (4) determine the money amount of business awarded to vendor *j*.

$$v_{jr} \le u_{jr} * \cdot y_{jr},\ j \in J,\ r = 1,\cdots,m_j - 1 \qquad (5)$$

$$v_{jm_j} \le u_{jm_j} \cdot y_{jm_j},\ j \in J \qquad (6)$$

$$v_{j,r+1} \ge u_{jr} \cdot y_{j,r+1},\ j \in J,\ r = 1,\cdots,R_j - 1 \qquad (7)$$

Constraints (5), (6) and (7) link the purchase of the item with the business volume discount to the appropriate segment of the discount pricing schedule for each vendor.

$$\sum_{r \in R_j} y_{jr} = 1, \quad j \in J \qquad (8)$$

Constraints (8) ensure that only one discount brackets for each vendor's volume of business will apply.

$$y_{jr} \in \{0,1\},\ v_{jr} \ge 0,\ j \in J,\ r \in R_j \qquad (9)$$

$$x_{ij} \ge 0,\ i \in I,\ j \in J_i \qquad (10)$$

Constraints (9) and (10) ensure integrality and non-negativity on the decision variables.

In the model we may have some additional constraints, for example if the buyer wants to reduce a number of suppliers, i.e. is willing to do business to a maximum of *K* suppliers. This constraint requires replacing constraint (8) by the following set of equations:

$$\sum_{r \in R_j} y_{jr} \le 1, \quad j \in J \qquad (8a)$$

$$\sum_{j \in J} \sum_{r \in R_j} y_{jr} \le K \qquad (8b)$$

These optional constraints ultimately affect the type and number of vendors selected, their respective order quantities, as well as the total cost, quality and delivery outcomes of the procurement process.

The other additional constraint might be the business volume constraint with which the buyer limits its volume of business with supplier *j* to a maximum money value U_j. This constraint is expressed as follows:

$$\sum_{r \in R_j} v_{jr} \le U_j,\ j \in J \qquad (8c)$$

In the next section we will show how the model (1) - (10) can be applied in a vendor selection problem for the different types of flour for an industrial bakery.

3 ILLUSTRATIVE EXAMPLE

In this section, we illustrate how the proposed model can be applied in determining the order quotas from different vendors in an industrial bakery. It is to be noted that in production of bread and bakery products the purchase of flour is contracted for the period of one year, from harvest to harvest. After the harvest flour producers have the information on the available wheat quantity, price and quality, which allows them to define the price, quality and quantity of flour they can supply in the subsequent one-year period.

In the one-year period the bakery plans to consume 4000 tons of wheat flour Type 550 (P_1), 1500 tons of wheat flour Type 850 (P_2), 500 tons of wheat flour Type 1100 (P_3) and 1000 tons of rye flour Type 1150 (P_4), or $D_1 = 4000$, $D_2 = 1500$, $D_3 = 500$ and $D_4 = 1000$.

The company contacts 4 potential flour suppliers and defines the upper limit of flour supplied by a single vendor and for every type of flour as it can be seen in matrix Q:

$$
\mathrm{Q} = [\, Q_{ij}\,] = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} V_1 & V_2 & V_3 & V_4 \end{array} \\ \begin{bmatrix} 2000 & 2000 & 2000 & 2000 \\ 1000 & 1000 & 1000 & 1000 \\ 500 & 500 & 500 & 500 \\ 500 & / & 500 & 500 \end{bmatrix} \end{array}
$$

It can be seen that second vendor (V_2) does not offer the flour P_4, so we have $J_4 = \{1,3,4\}$ and $J_1 = J_2 = J_3 = \{1,2,3,4\}$ as the sets of vendors that offer item P_i.

Similarly, sets of items that are offered by vendor V_j are $I_1 = I_3 = I_4 = \{1,2,3,4\}$ and $I_2 = \{1,2,3\}$.

The proposed prices for each vendor and every type of flour are given in the subsequent matrix (C_{ij}):

$$
\mathrm{C} = [\, c_{ij}\,] = \begin{array}{c} \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{array} \begin{array}{c} \begin{array}{cccc} V_1 & V_2 & V_3 & V_4 \end{array} \\ \begin{bmatrix} 220.10 & 207.60 & 207.10 & 230.80 \\ 200.20 & 189.12 & 192.14 & 208.16 \\ 190.11 & 182.14 & 185.20 & 190.40 \\ 420.25 & / & 440.70 & 430.80 \end{bmatrix} \end{array}
$$

Depending on the buyer's total purchase value, vendors offer a business volume discount having r discount brackets, $r \in R_j = \{1,2,\ldots,m_j\}$, and it can be seen in the Table 1.

Table 1: Vendor's discount brackets

Vendor 1	Money value (000 Euros)	M < 100	100 ≤ M < 400	400 ≤ M ≤ 900
	Discount (%)	5	8	10
Vendor 2	Money value (000 Euros)	M < 100	100 ≤ M < 400	400 ≤ M ≤ 650
	Discount (%)	/	5	8
Vendor 3	Money value (000 Euros)	M < 150	150 ≤ M < 500	500 ≤ M ≤ 900
	Discount (%)	/	4	6
Vendor 4	Money value (000 Euros)	M < 100	100 ≤ M < 300	300 ≤ M ≤ 900
	Discount (%)	3	5	10

Maximum money value for each supplier in the last column is obtained as the money amount if the vendor reaches its upper limit for the all types of flour.

Finally the model for purchasing the flour for this bakery is:

$$
\min Z_1 = \sum_{j \in J} \sum_{r \in R_j} \left(1 - d_{jr}\right) \cdot v_{jr} = 0.95v_{11} + 0.92v_{12} + 0.90v_{13} + v_{21} + 0.95v_{22} + 0.92v_{23} +
$$

$$
+ v_{31} + 0.96v_{32} + 0.94v_{33} + 0.97v_{41} + 0.95v_{42} + 0.90v_{43} \tag{1}
$$

$$
\begin{aligned}
& x_{11} + x_{12} + x_{13} + x_{14} = 4000 \\
& x_{21} + x_{22} + x_{23} + x_{24} = 1500 \\
& x_{31} + x_{32} + x_{33} + x_{34} = 500 \\
& x_{41} + \quad x_{43} + x_{44} = 1000
\end{aligned} \tag{2}
$$

$$
x_{ij} \le Q_{ij}, \; i \in I, \; j \in J_i \tag{3}
$$

As we have fifteen variables x_{ij} in relation (3) there are fifteen constraints which can be included in the model as the upper bound for these variables, for instance $0 \le x_{11} \le 2000$, $0 \le x_{21} \le 1000$, and so on.

Relation (4) $\sum_{i \in I_j} c_{ij} x_{ij} = \sum_{r \in R_j} v_{jr}, \quad j \in J$ determines the money amount of business awarded to vendor j, and we have these four constraints:

$$\begin{aligned}
&220.10x_{11} + 200.20x_{21} + 190.11x_{31} + 420.25x_{41} = v_{11} + v_{12} + v_{13} \\
&207.60x_{12} + 189.12x_{22} + 182.14x_{32} = v_{21} + v_{22} + v_{23} \\
&207.10x_{13} + 192.14x_{23} + 185.20x_{33} + 440.70x_{43} = v_{31} + v_{32} + v_{33} \\
&230.80x_{14} + 208.16x_{24} + 199.40x_{34} + 430.80x_{44} = v_{41} + v_{42} + v_{43}
\end{aligned} \tag{4}$$

Constraints (5), (6) and (7) link the purchase of the item with the business volume discount to the appropriate segment of the discount pricing schedule for each vendor. There are twenty of these equations and the first three of them are:

$$v_{11} \le 99999 \cdot y_{11}, \; v_{12} \ge 100000 \cdot y_{12}, \; v_{12} \le 399999 \cdot y_{12}.$$

Finally we have constraints (8) that ensure that only one discount brackets for each vendor's volume of business will apply, and integrality and non-negativity constraints (9) and (10):

$$y_{11} + y_{12} + y_{13} = 1, \; y_{21} + y_{22} + y_{23} = 1, \; y_{31} + y_{32} + y_{33} = 1, \; y_{41} + y_{42} + y_{43} = 1 \tag{8}$$

$$y_{jr} \in \{0,1\}, v_{jr} \ge 0, \; j \in J, r \in R_j \tag{9}$$

$$x_{ij} \ge 0, i \in I, j \in J_i \tag{10}$$

Our model has 39 variables, (15 x_{ij} , 12 v_{jr} and 12 y_{jr}), and 32 constraints. Solving this model with mixed integer linear programming software (WINQSB) we obtained the results that are shown in the table 2.

Table 2: Model results

04-20-2011 10:32:52	Decision Variable	Solution Value	Unit Cost or Profit C(j)	Total Contribution	Reduced Cost	Basis Status
1	x11	0	0	0	11,5000	at bound
2	x12	2.000,0000	0	0	0	basic
3	x13	2.000,0000	0	0	0	basic
4	x14	0	0	0	2,9643	at bound
5	x21	0	0	0	9,2536	at bound
6	x22	1.000,0000	0	0	0	basic
7	x23	446,5493	0	0	0	basic
8	x24	53,4507	0	0	0	basic
9	x31	0	0	0	7,3324	at bound
10	x32	131,5260	0	0	0	basic
11	x33	0	0	0	1,0432	at bound
12	x34	368,4740	0	0	0	basic
13	x41	500,0000	0	0	0	basic
14	x43	0	0	0	0	basic
15	x44	500,0000	0	0	0	basic
16	v11	0	0,9500	0	0,0300	at bound
17	v12	210.125,0000	0,9200	193.315,0000	0	basic
18	v13	0	0,9000	0	0	basic
19	v21	0	1,0000	0	0,0800	at bound
20	v22	0	0,9500	0	0	basic
21	v23	628.276,1000	0,9200	578.014,1000	0	basic
22	v31	0	1,0000	0	0,0896	at bound
23	v32	0	0,9600	0	0,0496	at bound
24	v33	500.000,0000	0,9400	470.000,0000	0	basic
25	v41	0	0,9700	0	0,1296	at bound
26	v42	0	0,9500	0	0,1096	at bound
27	v43	300.000,0000	0,9000	270.000,0000	0	basic
28	y11	0	0	0	6.574,1860	at bound
29	y12	1,0000	0	0	18.000,0300	at bound
30	y13	0	0	0	0	basic
31	y21	0	0	0	-18.400,0000	at bound
32	y22	0	0	0	2.999,9850	at bound
33	y23	1,0000	0	0	0	basic
34	y31	0	0	0	0	at bound
35	y32	0	0	0	0	at bound
36	y33	1,0000	0	0	0	basic
37	y41	0	0	0	-34.696,0100	at bound
38	y42	0	0	0	-34.696,0100	at bound
39	y43	1,0000	0	0	0	basic
	Objective	Function	(Min.) =	1.511.329,0000		

From the table 2 we can see that first vendor delivers only the fourth type of flour in the quantity of 2000 t ($x_{11} = x_{21} = x_{31} = 0$, $x_{41} = 2000$).

As the unit price for this vendor and that type of flour is $c_{41} = 420.25$ that means that the buyer's cost for that purchase is $v_{12} = 210125$ Euros and we are in the second price level with the discount of 8% (see Table 1). Of course $v_{11} = v_{13} = 0$, and also zero-one variable $y_{12} = 1$.

For the second vendor we have $x_{12} = 2000$, $x_{22} = 1000$, $x_{32} = 131.526$, $x_{42} = 0$ (second vendor does not have the fourth type of flour) and because the total amount of money for these quantities is 628276.1 (> 400000) we are in the third price bracket and have the maximum discount of 8%.

The third vendor delivers 2000 t of the first type of flour ($x_{13} = 2000$), 446.55 t of the second type ($x_{23} = 446.55$), and does not deliver any flour of the third and fourth type ($x_{33} = x_{43} = 0$). Value for these shipments is 500000 Euros and we are again in the third price bracket for this vendor and have the maximum discount of 6%.

For the fourth vendor we have $x_{14} = 0$, $x_{24} = 53.45$, $x_{34} = 368.47$ and $x_{44} = 500$ with total cost for this vendor of $v_{43} = 300000$ Euros, and we again have the maximum discount of 10%.

The optimal value of the objective function, the total purchase cost, is 1511329 Euros.

4 CONCLUSION

Since orders from external suppliers present a significant item for the majority of firms, vendor selection problem has a decisive influence upon firm competitiveness. Vendor 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 goals that a customer wants to achieve when selecting a supplier. For these reasons this model needs to be considered as the multi-criteria problem, but that is left for further researches.

The developed model, verified in this paper on the real case study, has a general value because it can be successfully used in solving similar practical problems dependent on numerous qualitative and quantitative criteria.

References:

[1] Beil D. R. (2010): *Supplier selection*, Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran, John Wiley & Sons, Inc., eorms0852.

[2] Chaudhry, S.S., F.G. Forst and J. L. Zydiak (1993): *Vendor selection with price breaks*, European Journal of Operational Research, 70, 52-66.

[3] Dahel N-E. (2003): *Vendor selection and order quantity allocation in volume discount environments*, Supply Chain Management: An International Journal, 8/4, 335–342.

[4] Jurun E., N.T. Plazibat, Z.Babić (1999): *Supplier Selection Problem in City of Split Kindergartens*, Proceedings of the 5th International Symposium on Operational Research, Preddvor , Slovenia, October 1999, p. 99-104.

[5] T. Perić, Z. Babić (2007): *A Decision System for Vendor Selection Problem*, Proceedings of the 9th International Symposium on Operational Research, Nova Gorica, Slovenia, p. 191-196.

[6] T. Perić, Z. Babić (2011): *Quantity Discounts in Supplier Selection Problem by Use of Fuzzy Multi-criteria Programming*, Croatian Operational Research Review, Vol. 2, 49-58.

[7] Wang, T-Y., Y-H. Yang (2009): *A fuzzy model for supplier selection in quantity discount environments*, Expert Systems with Applications 36, 12179–12187.

[8] Weber, C.A., J. Current (1993): *A multiobjective approach to vendor selection*, European Journal of Operational Research 68, 173-184.

[9] Xia, W., Z. Wu (2007): *Supplier selection with multiple criteria in volume discount environments*, Omega 35, 494 – 504.

PREDICTING MANUFACTURING DUE DATE

T. Berlec[1]**, P. Potočnik**[2]**, E. Govekar**[2] and **M. Starbek**[1]
[1]Laboratory for Manufacturing Systems and Production Process Planning, Faculty of Mechanical Engineering, University of Ljubljana, Slovenia
[2]Laboratory for Synergetics, Faculty of Mechanical Engineering, University of Ljubljana, Slovenia

Abstract: One of the largest problems of today's time are too long lead times of orders.
To make a bid just on the basis of experience of employees is very risky. Therefore we propose a procedure by which – on the basis of actual lead times of orders processed in the company's workplaces in the past – expected lead times of planned (and indirectly – production) orders can be predicted.
Using the proposed procedure, the sales department can make a prediction for the customer about delivery time of the planned order.
A case study is presented: lead time of order for the "tool for manufacturing the filter housing" was predicted; the tool is manufactured in the Slovenian company ETI d.d..

Keywords: prediction, lead time, operating order, empirical distribution, percentile

1 INTRODUCTION

Companies on the global market offer similar or the same products at comparable price and quality. The main difference between these companies is in the predicted order development time and in observance of the deadlines for delivery time.

That is why a company cannot afford to make estimates based on personal experience. The bids may be based on wrong delivery time which can cause that the company does not receive the order.

Development of information and communication technologies (ICT) simplifies many business-related tasks [1]. Every company that wants to be competitive on the global market needs a suitable enterprise resource planning system – ERP system. There are several ERP systems available on the market [2] and it is the task of each company to select the optimal system [3].

The article will present how the data stored in the ERP system can be used for calculation of lead times of orders (and indirectly: lead times of manufacturing orders) on the basis of theory developed at the IFA Hannover [4, 5]. Furthermore, the calculation of percentiles of manufacturing order lead times will be shown, which allows the calculation of the confidence interval. The purpose of this paper is therefore to propose a procedure for predicting lead times of manufacturing orders on the basis of past gathered data on actual lead times.

2 METHOD FOR PREDICTING MANUFACTURING ORDER LEAD TIMES

When talking about "an order", it is necessary to distinguish between operational, manufacturing, assembly and production order [4], as presented in Figure 1.

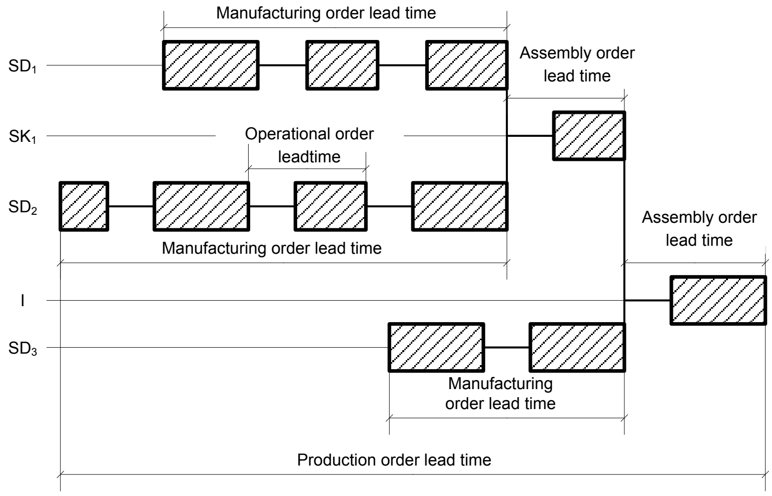


Figure 1: Order lead times [4]

When designing a procedure for predicting manufacturing order lead times, it will be assumed that the company uses an ERP system, whose database contains data about past operational and assembly orders in company workplaces. Any ERP system should therefore provide data on: production and assembly-order code, manufacturing-order code, operational-order code, type and sequence of operations on manufacturing and assembly orders, IDs of workplaces where operational orders have been carried out, actual execution times of operational orders, date of completing a particular operational or assembly order in the previous workplace and date of finishing a particular operational or assembly order in the observed workplace.

ERP system output data should be available in Microsoft Excel format (.xls).

Based on previous research on problems of determination of lead times let us to the conclusion, that the proposed procedure for predicting order lead times consists of the following steps:

Step 1: Determining actual lead times of already processed operational orders in the company's workplaces

Vectors of actual lead times of operational orders, executed in the past in all company workplaces, will present the basis for predicting expected lead times of the planned new production orders.

Step 2: Forming assembly structure of the planned production order and technology routings for parts and components of planed order

The assembly structure and technology routings of realization of component parts and assemblies of manufacturing orders are made on the basis of available documentation of the planned production order.

Step 3: Random sampling and summing of vector element values of actual lead times of operational orders of individual manufacturing or assembly order

Figure 2 presents the principle of random sampling and summing of vector element values of operational order actual lead times in the past of planned manufacturing and assembly orders.

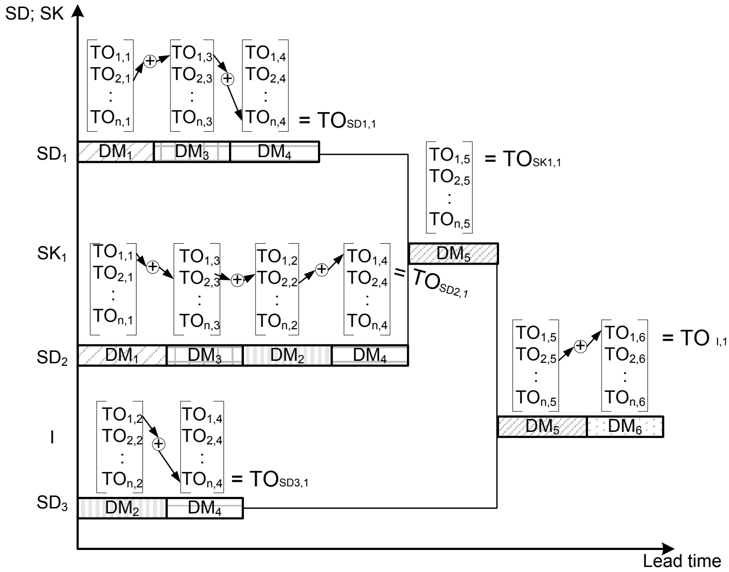


Figure 2: Random sampling and summing of vector element values of actual operational order lead times achieved in the past

$TO_{SD1,1}$ … lead time of component part SD1, got after first iteration
$TO_{SK1,1}$ … lead time of assembly SK1, got after first iteration
$TO_{I,1}$ … lead time of product I, got after first iteration

Figure 2 shows a schematic presentation of random sampling and addition of lead times values achieved in the past from workplace vectors, defined by technology and assembly routings for manufacturing parts and assembly of components.

It is necessary to select the number of iterations for random sampling of lead times of manufacturing and assembly orders of planned production order (computer supported). The number of iterations is affected by the order type – by increasing its complexity, the number of iterations should be increased, too.

Step 4: Setting up vectors of expected lead times of manufacturing and assembly orders of planned order

Results of step 3 allow setup of vectors of the expected lead times of the manufacturing and assembly orders of planned production order.

The number of elements in individual vector depends on the number of performed iterations k. The number of required iterations can be established on the basis of tests, as it is necessary to assure a stable process, which cannot be achieved by a small number of iterations. A criterion for an adequate number of iterations is that multiple use of the procedure must yield comparable results, by which the stability (convergence) of the procedure is achieved.

Step 5: Definition of vector of expected lead times of the planned order

In order to define the vector of expected lead times of the planned production order V_I, it is necessary to transform the Gantt chart of production order into an activity network diagram of order and enter into it the vectors of expected lead times of manufacturing and assembly orders of planned order (Figure 3).

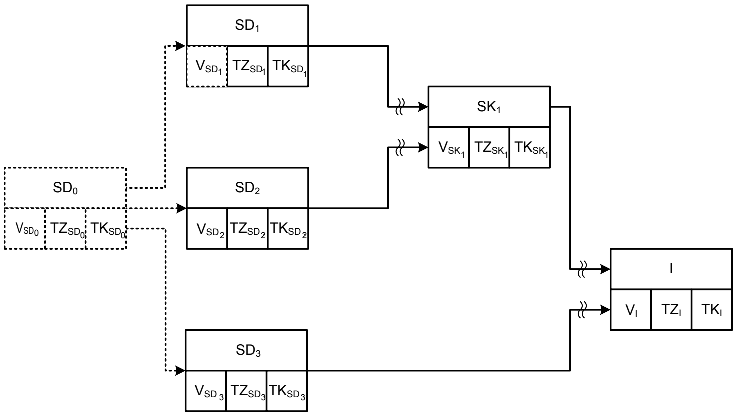


Figure 3: Activity network diagram of the planned (new) order

The calculation is carried out as follows:

- For sequential operations, individual randomly sampled lead times from vectors of sequentially listed workplaces are summed up. The result of each iteration is stored into a new vector, which represents the sum for one component or part.
- For parallel operations it is necessary to collect randomly sampled lead times from vectors of parallel workplaces, and then find a maximum lead time for each parallel path. Thus obtained results in each iteration should be stored in the common vector of maximum times of parallel paths, as the critical path in the activity network diagram is always the path with the longest required lead time for realization the manufacturing order.

Calculation has to be repeated for the selected number of iterations.

Thus obtained expected lead times of the planned order will represent empirical distribution of lead times of the planned order.

Step 6: Predicting delivery time of the planned order

Step 5 in predicting order lead time has lead us to the vector of expected lead times of order, i.e. to a certain distribution of lead times.

In real life, however, an exact value of lead time for delivery of order is required.

The most probable delivery lead time for the planned order can be estimated by using median [6], which means that there is a 50% probability that the actual delivery time will be shorter, and 50% probability that it will be longer than stated. As the 50% risk is usually unacceptable for the client, so the estimated value should be stated for a wider confidence interval.

For instance, 90% confidence interval is defined as 90% probability that the order will be delivered before the stated time. Therefore, maximum delivery time that can be guaranteed to the customer with 90% reliability, corresponds to the 90th percentile of empirical distribution of prediction of the planned order.

Based on our tests, we propose that the 90th percentile be used as a standard. In this way it is possible to state with 90% confidence that the order will be completed within the expected time.

If the company wants to achieve even higher reliability, it can use an even higher percentile (for example 99th) – and thus minimize the risk.

Naturally, the choice of the percentile may depend on importance of the order and the customer – the more important the customer, or the more important the order, the higher is the interest of the company to get a particular order.

In the proposed procedure for predicting manufacturing order lead times, in addition to MS Excel, the MATLAB software will be used [7], which allows execution of mathematical operations and graphical presentation of results.

3 TESTING THE PROCEDURE FOR PREDICTING ORDER LEAD TIMES

The procedure for predicting order lead times was tested in the tool shop of ETI d.d. company from Izlake, Slovenia. It produces tools for transforming and cutting, tools for injection moulding of thermoplastic and duroplastic materials, jet and press machines for duroplastic materials, press machines for ceramic materials, and automated assembly appliances.

This tool shop speciality is design and manufacturing of high-quality tools for injection moulding of thermoplastic and duroplastic materials. Thanks to its many years of experience in making tools for ETI company, the tool shop started producing tools and appliances for external customers in the following fields: automotive industry, household appliances, medical technology, electrical engineering and electronics and illumination.

Tool shop uses Largo ERP system, developed by Perftech d.o.o. company from Bled [8], Slovenia. Because of their way of production (tools are made for known customers and each tool is unique) it is very difficult to precisely predict duration of production – but this is essential data for making bids and winning orders.

Company management decided to test the suitability of the proposed procedure for predicting lead times of orders in a case study of determining lead time of order for the "tool for manufacturing filter housing # 705429".

Six steps of the procedure for predicting manufacturing order lead time for the “tool for manufacturing filter housing # 705429 were carried out.

For the experiment, the Largo ERP system data were used in the period from 12 December 2002 till 22 August 2005. During that time, 22,850 manufacturing orders were processed in the production, with 57,951 operational orders on 35 workplaces.

It was necessary to export data from Largo ERP system to MS Excel format. From the database the following data were exported: order number, arrival date, departure date, manufacturing time, and sequence of operational orders.

After 10,000 iterations the final vector of expected order lead time for the planned order for the tool # 705429 with distribution presented in Figure 4 was obtained.

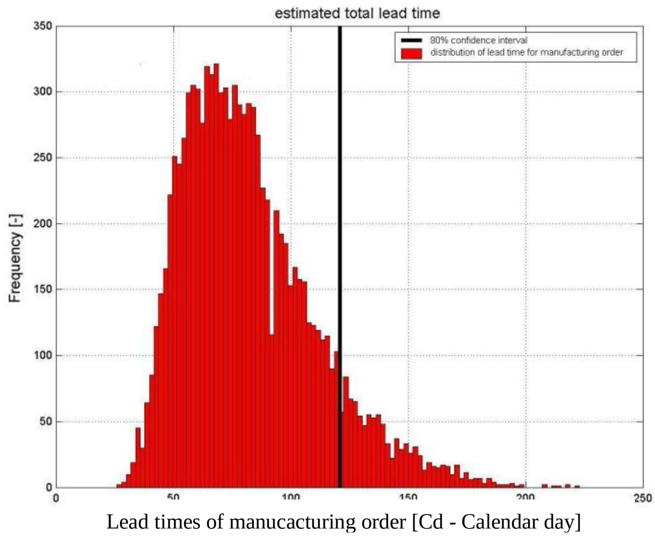


Figure 4: Distribution of lead times for manufacturing order # 705429

However, no customer is interested in a lead-time vector (or distribution of lead times), so as the first approximate delivery time value, the median of this vector is used and for this order it is:

$$TO_{med} = 77 \text{ Cd.}$$

TO_{med} … median lead time of the tool production order # 705429
Cd ... calendar day

Expected lead time therefore equals to the 50th percentile of the vector of expected lead times.

Confidence intervals are defined according to the selected confidence levels.

If a very high confidence level is required, the 90th percentile is used. So, for the order # 705429 it can be stated with 90% confidence level that this order will be produced within

$$TO_{90\%} = 120 \text{ Cd.}$$

$TO_{90\%}$ … lead time at 90^{th} percentile

The company has to decide itself, which risk level it is ready to accept when signing the contract with the customer.

4 CONCLUSION

Due to ever fiercer competition of companies on domestic and foreign markets, and due to transition from the market of sellers to the market of customers, the companies have to continuously reduce delivery times of the order.

The article proposes a procedure for predicting manufacturing order lead times on the basis of past actual lead times of operational orders.. The use of the proposed procedure allows:

- prediction of lead time needed for delivery of any new order,
- variation of delivery-time calculations on the basis of acceptable risk level by selecting the confidence interval.

The basis for calculation of delivery time is median, while 90th percentile may represent the upper (safe) limit, which can be offered to the customer as the latest delivery lead time. On the basis of its experience and its willingness to risk, the company can choose different confidence interval

The procedure for predicting order lead times has been tested in a case study of predicting lead times for producing a tool for manufacturing a filter casing in the tool shop of the ETI d.d. company from Izlake, Slovenia. Case study was performed on the basis of data, defined in calendar days collected in three years in the data base of ERP system Largo.

By using this procedure of predicting manufacturing order lead times, sales department can make a well-defined bid for the customer in a short time without the sales person needing many years of experience – (s)he only needs well-defined technology routings, while the company management provides the confidence interval.

In the future it is planned that the proposed procedure will be improved by taking into account the sequence of operations required to complete an order, the influence of the number of operations per order, the influence of manufacturing time, etc.

5 ACKNOWLEDGEMENTS

We would like to thank to the tool shop of ETI d.d. company for disposal with the data from their ERP system, for their help. We would also like to thank to the Slovenian Ministry of Higher Education, Science and Technology for their financial aid during development of this method.

References

[1] Leem C.S., Suh J.W.: Techniques in integrated development and implementation of enterprise information systems, 2005, Intelligent knowledge-based systems, Business and technology in the new millennium, Vol. 2, Information technology, Kluwer Academic Publishers, 3 – 26.

[2] Scherer E.: ERP – Projekte auf dem Prüfstand der Praxis, 2005, ERP Management, GITO mbH Verlag, 34-37

[3] M. Starbek, J. Grum: Selection and implementation of a PPC system, Production planning & control, 2000, Vol. 11, pp.765-774

[4] Wiendahl H.P.: Load-Oriented Manufacturing Control, 1995, Springer Verlag, Berlin, 37-199.

[5] Nyhuis P., Wiendahl H.P.: Logistische Kennlinien, 1999, Springer Verlag, Berlin, 81-94.

[6] Rice John A.: Mathematical Statistics and Data Analysis, Second Edition, International Thomson Publishing, California, 1995

[7] The MathWorks, Inc.: Getting Started with MATLAB, ver. 6, 2002

[8] http://www.perftech.si/index.php?page=vsebina&page_id=40, checked on 9 June 2011

A SIMPLE DISCRETE APPROXIMATION FOR THE RENEWAL FUNCTION

Alenka Brezavšček
University of Maribor, Faculty of Organizational Sciences
Kidričeva cesta 55a, SI-4000 Kranj, Slovenia
alenka.brezavscek@fov.uni-mb.si

Abstract: In the paper we propose a simple discrete approximation for calculating the renewal function which is an important characteristic in the areas of reliability, maintenance and spare components inventory planning. The approximation enables easy and relatively accurate calculation of the renewal function irrespective of the type of the probability density function of component failure times. It is especially applicable to the peak-shaped functions (i.e. components are subject to wear-out failures) when the analytical solution for the renewal function hardly ever exists. To prove the usefulness of the approximation some numerical results are given.

Keywords: probability density function of component failure times, random failures, wear-out failures, renewal function, approximation

1 INTRODUCTION

The renewal function $H(t)$ plays an important role in areas of reliability, maintenance and spare part inventory planning (see e.g. [8], [2],[19]). Calculation of $H(t)$ depends of the type of the probability density function of inter-renewal times. In the reliability applications this function is equal to the probability density function of component failure times $f(t)$. However, it is not possible to obtain $H(t)$ analytically for any type of the function $f(t)$. For most peak-shaped distribution families the closed form of $H(t)$ is not available. Many different approaches have been developed in the literature to approximate or numerical compute the renewal function considering the peak-shaped underlying distributions (e.g. [20], [5], [17], [7], [1], [8], [9], [15], [4] and [13]). Unfortunately, simple approximate formulas for $H(t)$ are obtained for a limited range of functions $f(t)$ only.

In this paper we develop a simple discrete approximation which enables easy calculation of the renewal function $H(t)$ irrespective of the type of the probability density function $f(t)$ of the component failure times. The approximation proposed is especially useful for the peak-shaped density functions when the analytical solution of the renewal function hardly ever exists.

2 PRELIMINARIES

2.1 Probability density functions of component failure times

Technical systems such as industrial systems consist of a number of components of different types. During the system operation its components fails. Time to failure of a component of a given type is a random variable distributed according to the probability density function $f(t)$. The behaviour of the function $f(t)$ depends on the region of the component lifetime. From the reliability theory it is known that the lifetime of a component can be roughly divided into three regions: the region of early failures, the region of random failures (also called the region of the normal operation) and the region of wear-out failures shown in Fig. 1. Early failures are usually eliminated by screening or burn-in tests before the components are put into operation. The operating period of the component thus comprises only the region of random failures and the region of wear-out failures. Random failures are mainly due to inherent slow-acting defects in the component, or to sudden excessive loading. The main failure mechanism in the wear-out region is deterioration of the component materials. The region of the random failures is characterized by monotonically decreasing function $f(t)$. In the region of wear-out failures the distribution of component failure times follows a peak shaped curve. In Fig. 1, general form of the function $f(t)$ during the component lifetime is presented. Besides, the instantaneous failure rate $\lambda(t)$ is also shown. It can be seen from Fig. 1 that during the region of random failures $\lambda(t)$ has approximately constant value λ, called simply the failure rate. During the wear-out region, instantaneous failure rate $\lambda(t)$ increases with time. The behaviour of $\lambda(t)$ indicates that during the random failures region only component replacements on failure are reasonable, while in the wear-out region also preventive replacements make sense.

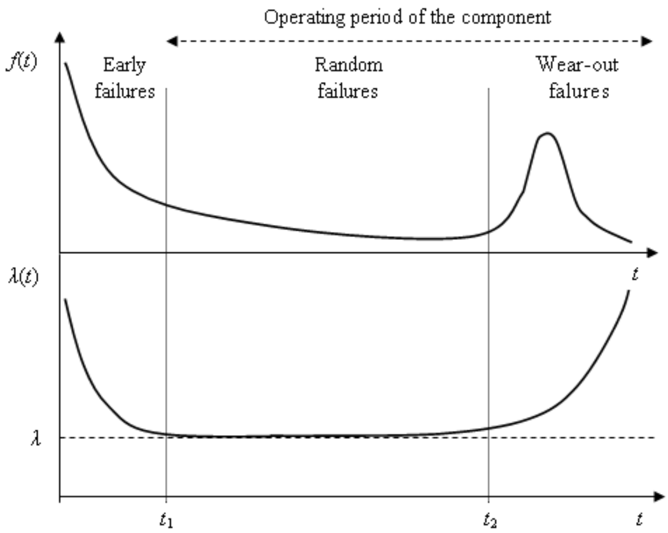


Figure 1: General form of component failure functions: the probability density function $f(t)$ of failure times, and the instantaneous failure rate $\lambda(t)$

An appropriate mathematical model for $f(t)$ in the region of random failures is exponential probability density function with the parameter λ. In the region of wear-out failures, times to failure of a component are distributed around a peak value, specific for a given deterioration mechanism due to material aging. In this region the normal, lognormal, Weibull or Gamma density functions are frequently used to describe the function $f(t)$.

2.2 The renewal process and the renewal function

In the case of system failure, a failed component should be replaced by a new one to ensure the continuity of the system operation. If the replacement time is negligible (e.g. a spare component is available in the inventory), the process of successive corrective replacements of a particular component can be described by an ordinary renewal process.

A realization of the renewal process consists of a series of point events (renewals) occurring singly in time and completely randomly. The renewal process is ordinary when all inter-renewal times (in our case times to failure) are independent identically distributed random variables, all with probability density function $f(t)$ (e.g. [6]).

The essential characteristic of an ordinary renewal process is the renewal function $H(t)$ defined as the expected number of renewals of a single component in the interval (0,t):

$$H(t) = E\left[N(t)\right]$$

where $N(t)$ represents the number of component renewals in the interval (0,t).

The renewal function $H(t)$ can be calculated according to the equation:

$$H(t) = \sum_{r=1}^{\infty} F_r(t) \tag{1}$$

with $F_0(t) = 1$. The symbol $F_r(t)$ in equation (1) denotes the r-fold convolution of the cumulative distribution function $F(t) = \int_0^t f(x)dx$ with itself. For an arbitrary time t a simple solution of (1) is obtainable for some specific types of $f(t)$ only.

Theoretically, the calculation of the renewal function $H(t)$ according to (1) can be simplified using Laplace transformations because there is a simple relation between the Laplace transform of a function and the Laplace transform of its integral. The Laplace transform $H^*(s)$ of the renewal function $H(t)$ is given by the expression:

$$H^*(s) = \frac{1}{s}\frac{f^*(s)}{1 - f^*(s)} \tag{2}$$

where $f^*(s)$ denotes the Laplace transform of the function $f(t)$. The renewal function $H(t)$ is than obtained by the inversion of the expression (2):

$$H(t) = L^{-1}H^*(s) \tag{3}$$

Another presentation of the renewal function is provided by the so-called renewal equation (e.g. [21]):

$$H(t) = F(t) + \int_0^t H(t-x)\, f(x)dx \tag{4}$$

2.3 Calculation of the renewal function for different types of $f(t)$

It is evident that the calculation of $H(t)$ depends on the type of the probability density function $f(t)$ of component failure times. Regarding the region of the component lifetime, the function $f(t)$ can be described either by the exponential or by one of the peak-shaped density functions (see Section 2.1).

Exponential $f(t)$

If the function $f(t)$ is exponential with the parameter λ the analytical solution for the renewal function $H(t)$ can be simply derived using the equation (3): $H(t) = \lambda t$.

Normal $f(t)$

The exact inversion of (2) is not obtainable, and the analytical solution for $H(t)$ do not exists. However, the numerical calculation of $H(t)$ according to (1) is rather simple because the r-fold convolution of the normal distribution function $F(t)$ with the parameters μ and σ is also a normal distribution function with the parameters $r\mu$ and $\sigma\sqrt{r}$.

Lognormal $f(t)$

The exact inversion of (2) is not obtainable, so the analytical solution for $H(t)$ do not exists. Unfortunately, the numerical calculation is not trivial because the closed form of $F_r(t)$ is not directly available. Some approximate formulas are proposed in the literature but calculations are quite tedious (see e.g. [3], [18] and [14]).

Weibull $f(t)$

If the shape parameter β is equal to 1, the Weibull density function becomes exponential, and the calculation of the renewal function is trivial. When $\beta > 1$, the Weibull density function exhibits a peak-shaped form. In such a case the closed form of $F_r(t)$ is not available, and the numerical calculation of $H(t)$ is quite tedious (see e.g. [12]). A comprehensive overview of different numerical calculations of the Weibull renewal function is given in the book [16].

Gamma $f(t)$

If the shape parameter β is equal to 1, the Gamma density function becomes exponential and the calculation of the renewal function is trivial. Similar as the Weibull density function, the Gamma density function exhibits a peak-shaped form when $\beta > 1$. Analytical solution for $H(t)$ exists for some integer values of β only (e.g, $\beta = 2$ or $\beta = 3$). For example, if $\beta = 2$ the renewal function is

$$H(t) = \frac{1}{4}\left(e^{-\frac{2t}{\eta}} + \frac{2t}{\eta} - 1\right)$$

where the symbol η denotes the scale parameter of the Gamma probability density function. The numerical calculation of $H(t)$ according to (1) is easy for any value of parameter β because the r-fold convolution of the Gamma distribution function with parameters β and η is also a Gamma distribution function with parameters $r\beta$ and η.

We can conclude that the analytical solution for $H(t)$ is trivial when $f(t)$ is exponential (i.e. components operate in the period of random failures) while the closed form of $H(t)$ is hardly ever available when the underlying density function is a peak-shaped (i.e. the components operate in the period of wear-out failures). The derivation of an approximate solution of $H(t)$ for a peak-shaped $f(t)$ has attracted attention of many authors, but simple formulas are obtained for a limited range of functions $f(t)$ only. More general approximate solution for $H(t)$ would be most desirable.

3 THE APPROXIMATE SOLUTION FOR THE RENEWAL FUNCTION

We use the recursive integral equation (4) to derive a discrete approximate formula for estimating the value of the renewal function in time T, $H(T)$. We assume that component renewals can occur only in time points i, i=1,2, …, with the probability p_i. Our discrete approximation of (4) reads as follows:

$$H(T) = \sum_{i=1}^{t} p_i + \sum_{i=1}^{t} H(T-i)\, p_i \qquad T=1, 2, \ldots \qquad (5)$$
$$H(0) = 0$$

The probability p_i in (5) is approximated by the integral $p_i = \int_{i-1}^{i} f(t)dt$, i=1,2, …, $\sum_{i=1}^{\infty} p_i = 1$.

This approximation implies that discrete and continuous distribution functions for inter-renewal times are equal since:

$$\sum_{i=1}^{T} p_i = \sum_{i=1}^{T} \int_{i-1}^{i} f(t)dt = \int_{0}^{T} f(t)dt = F(T)$$

Equation (5) can be then rewritten in the following way:

$$H(T) = F(T) + \sum_{i=1}^{T} H(T-i) \int_{i-1}^{i} f(t)dt \qquad T=1, 2, \ldots \qquad (6)$$
$$H(0) = 0$$

Our discrete approximation of $H(T)$ involves the assumption that the renewal of a component, which can actually occur anywhere within the interval $(i-1, i)$, i=1,2,…,T, between two successive time points, occurs exactly at the time point i. Consequently, the value $H(T)$ calculated according to (6) is usually too low. The error introduced in such a way can be diminished by choosing sufficiently short intervals $(i-1, i)$, i=1,2,…,T.

Because the mathematical operations in (6) are simple, the renewal function can be calculated easily, regardless of the type of the function $f(t)$. Similar approximations, derived in a different way, have been proposed in [11], [10] and [22].

4 NUMERICAL RESULTS

To prove the usefulness of the approximation (6) some numerical results are given (Tab. 1). In the calculations, the normal, lognormal, Weibull and Gamma probability density functions are considered. In reliability studies these density functions are widely used to describe the function $f(t)$ for components subject to wear-out.

We can see from Tab. 1 that the approximate solution (6) is applicable to all four probability density functions considered. Although all these density functions are of peak-shaped form, the results cannot be directly compared due to specific properties of a particular density function.

In the case of normal and Gamma density functions we can compare the values of the renewal function calculated according to our approximation (6) with the values calculated according to (1). We can conclude that our approximation gives relatively accurate results even for comparatively large values of time t.

Table 1: Results of using the approximation of the renewal function for the peak-shaped probability density functions widely used in the reliability applications

	t	0.25μ	0.5μ	0.75μ	μ	2μ	3μ
Normal pdf $\mu = 45 \quad \sigma = 15$	$H(t)$ calculated according to (1)	0.01233	0.06755	0.23068	0.51722	1.54166	2.55363
	$H(t)$ calculated according to (6)	0.01122	0.06674	0.22941	0.51392	1.53687	2.53418
Lognormal pdf $\mu' = 5 \quad \sigma' = 0.3$ $\Downarrow$ $\mu \approx 155 \quad \sigma \approx 48$	$H(t)$ calculated according to (1)	-	-	-	-	-	-
	$H(t)$ Calculated according to (6)	4.20E-06	0.01600	0.20572	0.55834	1.54237	2.53594
Weiblull pdf $\beta = 3.5 \ \eta = 200$ $\Downarrow$ $\mu \approx 180 \ \sigma \approx 57$	$H(t)$ calculated according to (1)	-	-	-	-	-	-
	$H(t)$ calculated according to (6)	0.00539	0.05939	0.22488	0.51039	1.53147	2.53710
Gamma pdf $\beta = 2.9 \quad \eta = 50$ $\Downarrow$ $\mu = 145 \ \sigma \approx 85$	$H(t)$ calculated according to (1)	0.04395	0.20226	0.42552	0.67093	1.67240	2.67241
	$H(t)$ calculated according to (6)	0.04322	0.20129	0.42478	0.66968	1.66778	2.66436

5 CONCLUSION

The renewal function is an important characteristic, needed in the areas of reliability analyses, maintenance optimization and spare components inventory planning. Its calculation depends on the behaviour of the probability density function of failure times during the component operating period. In the period of random failures, the calculation of the renewal function is trivial because the probability density function of component failure times is described by the exponential density function. In the period of wear-out failures, the probability density function for component failure times exhibits a peak-shaped form. Calculation of the renewal function in such a case is in general difficult because the analytical solutions hardly ever exist. The derivation of an approximate solution of the renewal function for a peak-shaped failure time distribution has attracted attention of many authors, but simple approximate formulas are obtained for a limited range of functions only.

In the paper, we proposed a simple discrete approximation for estimating the value of the renewal function. The approximation is derived from the so called renewal equation. Due to simple mathematical operations involved it enables easy and relatively accurate calculation of the renewal function for any type of the probability density function of component failure times. It is especially applicable to the peak-shaped density functions when the analytical solutions hardly ever exist. To prove the usefulness of the approximation some numerical results considering normal, lognormal, Weibull and Gamma probability density function are given.

The approximation proposed is useful everywhere in the reliability analysis and the maintenance policy optimization where the renewal function needs to be evaluated.

References

[1] Ayhan H., Limon-Robles, J. and Wortman M.A., 1999. An approach for computing tight numerical bounds on renewal functions, IEEE Transactions on Reliability 48, 182-188.

[2] Barlow, R.E., Proschan, F., and Hunter, L.C. 1996. Mathematical theory of reliability, SIAM, 1996.

[3] Barouch, E. & Kaufman, G.M. (1976). On Sums of Lognormal Random Variables. Working paper, Alfred P. Sloan School of Management, Cambridge, Massachusetts, available on http://dspace.mit.edu/bitstream/handle/1721.1/48703/onsumsoflognorma00baro.pdf (10.06.2011).

[4] Bebbington, M., Davydov, Y. and Zitikis, R., 2007. Estimating the renewal function when the second moment is infinite', Stochastic Models 23, 27 – 48.

[5] Chaudhry, M.L., 1995. On computations of the mean and variance of the number of renewals: a unified approach, The Journal of the Operational Research Society 46, 1352-1364.

[6] Cox, D.R., 1970. Renewal Theory, Methuen.

[7] Garg, A. and Kalagnanam, J.R., 1998. Approximations for the renewal function, IEEE Transactions on Reliability 47, 66-72.

[8] Gertsbakh, I., 2000. Reliability Theory, With applications to Preventive Maintenance, Springer Verglag.

[9] Hu, X., 2006. Approximation of partial distribution in renewal function calculation, Computational Statistics & Data Analysis 50, 1615-1624.

[10] Jardine, A.K.S. and Tsang, A.H. C., 2006. Maintenance, Replacement, and Reliability: Theory and Applications, CRC/Taylor & Francis, Boca Raton.

[11] Jardine, A.K.S., 1973. Maintenance, Replacement, and Reliability, Pitman, London.

[12] Jiang, R. (2008). A Gamma–normal series truncation approximation for computing the Weibull renewal function , Reliability Engineering & System Safety, 93 (4): 616-626.

[13] Jiang, R. (2010). A simple approximation for the renewal function with an increasing failure rate Reliability Engineering & System Safety, (95) (9): Pages 963-969.

[14] Lam, C.-L.J. & Le-Ngoc, T. (2006). Estimation of typical sum of lognormal random variables using log shifted gamma approximation, IEEE Communications Letters, 10 (4): 234 – 235.

[15] Politis, K. and Koutras, M.V., 2006. Some new bounds for the renewal function, Probability in the Engineering and Informational Sciences 20, 231 – 250.

[16] Rinne, H. (2009), The Weibull Distribution: A Handbook, CRC Press, Taylor & Francis Group.

[17] Robinson, N.I., 1997. Renewal functions as series, Stochastic Models 13, 577 – 604.

[18] Romeo, M., Da Costa, V. & Bardou, F. (2003). Broad distribution effects in sums of lognormal random variables, The European Physical Journal B - Condensed Matter and Complex Systems, 32 (4): 513-525.

[19] Sheikh A.K., and M. Younas, 1985. Renewal Models in Reliability Engineering, (Edited by P.E. Deopker), ASME.

[20] Smeitink, E. and Dekker, R., 1990. A simple approximation to the renewal function, IEEE Transactions on Reliability 39, 71-75.

[21] Tijms, H.C., 2003. A First Course in Stochastic Models, John Wiley & Sons, Chichester.

[22] van Noortwijk, J.M. and van der Weide, J.A.M., 2008. Applications to continuous-time processes of computational techniques for discrete-time renewal processes, Reliability Engineering & System Safety 93 (12), 1853-1860.

INDUSTRIAL CONFIDENCE INDICATOR AND MANUFACTURING INDUSTRY IN THE EURO AREA

Mirjana Čižmešija, Maja Copak and **Nataša Kurnoga Živadinović**
University of Zagreb - Faculty of Economics & Business
Trg J. F. Kennedyja 6
10000 Zagreb, Croatia
mcizmesija@efzg.hr, majacopak@hotmail.com, nkurnoga@efzg.hr

Abstract: Business surveys are continuous qualitative economic surveys based on managerial perceptions and expectations in four sectors of national (and regional) economies. Surveys are conducted on a sample of economic agents at the specified time intervals. The aim of this paper is to determine the relationship between industrial confidence indicator (ICI) and the industrial production. Based on regression analyses it was concluded that in accordance of ICI changes, we can predict the direction of changes in industrial production, because business survey results are available before the official statistics is published.

Keywords: Business Survey, Industrial Confidence Indicator (ICI), Regression analysis, Euro area

1 INTRODUCTION

Business surveys are relatively new studies that began in 1949 in Germany and are now conducted in almost all EU members and potential accession countries. By decision of the European Commission, of 15th November 1961, The Joint harmonised EU Programme of Business and Consumer surveys was launched. It is managed by DG ECFIN - Directorate General for Economic and Financial Affairs.

Before the sole empirical analyses, it is necessary to seasonally adjust the data. Because respondents' attitudes are affected by various circumstances, such as weather conditions, strikes, elections, changes in exchange rates, public holidays (consumer fever) etc. Such variations in managers' and consumers' perceptions should be eliminated when comparing two or more consecutive months. Hence it is necessary to conduct seasonal adjustment procedure [6]. There are many seasonal adjustment methods, whereas the most important are: DAINTIES, Tramo&Seats, ASA II, X-12 ARIMA model etc.

For each of the four surveyed sectors confidence indicators are calculating: an indicator of confidence in industry (Industrial Confidence Indicator, ICI), which will be explained later, an indicator of confidence in retail trade (Retail Trade Confidence Indicator, RTCI), an indicator of confidence in services sector (Services Confidence Indicator, SCI) and an indicator of confidence in construction (Construction Confidence Indicator, BCI). Each confidence indicator is calculated as the simple arithmetic mean of (seasonally adjusted) balances of answers to specific questions in each individual survey.

The confidence indicator in industry (ICI), according to the EU methodology, is defined as the arithmetic average of seasonally adjusted balances (in percentage points) of the three variables: production expectations, order books and stocks of finished products (the last with inverted sign).

$$ICI = \frac{B_P + B_O + (-B_S)}{3} \qquad (1)$$

where ICI stands for Industrial Confidence Indicator, B_P - seasonally adjusted response balance[1] to the question about expected production. B_O - seasonally adjusted balance of responses to the question about order books , B_S - seasonally adjusted response balance to the question about stocks of finished products used with inverted sign [3].

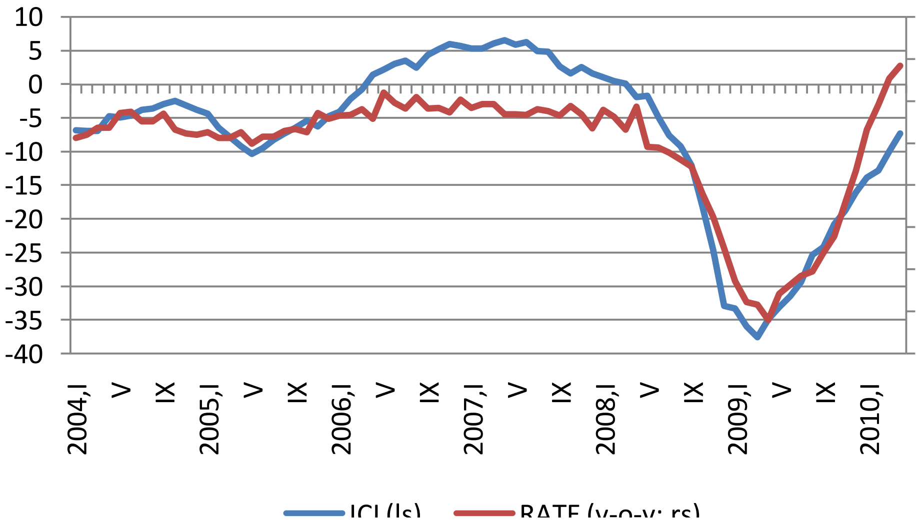


Figure 1: ICI and growth rate of industrial production (y-o-y) in the Euro area

The advantage of Business surveys is that research results are quickly available and published before the results of official statistics are published, and complement them. The disadvantage is the subjectivity of the respondents, because the answers are based on their ideas and attitudes.

The aim of this paper is to determine the relationship and analytical expression of relationship between confidence indicator and appropriate macroeconomic variable in the Euro area and determine whether, and how, business surveys can be applied in forecasting business cycles. In order to analyse these relationship, a regression models ware estimated.

2 DATA AND METHODS

Business surveys are qualitative economic studies based on managers' estimates and expectations in four sectors of national economies.

In practice, the questionnaires are being sent to firms every month. Managers' answers are subjective, based on their opinions and attitudes. The data are obtained by mail, fax, telephone and electronic media. An increasing importance is being given to the electronic data interchange - EDI. A problem that may occur in business surveys is a small response rate, which affects the quality of research results. The questionnaire usually includes four types of questions [5]: those referring to actual and expected directions of change of certain micro-variables endogenous to the firm, question asking the entrepreneurs to judge the actual level of certain endogenous variables in relation to some ideal (normal) level, questions regarding the actual or expected direction of change and the evaluation of the level of variables that are usually exogenous to the entrepreneur's decisions and the last type: quantitative questions.

[1] Balance is a difference between positive and negative answers percentages (more detail in European Economy, 1997).

The business survey results for the European Union and for the Euro area (based on the Joint Harmonised EU Programme of Business and Consumer Surveys) are available on the website of the European Commission, which is regularly updated. They published results of business surveys and consumers surveys for the Euro area (13 EU countries), all countries together (EU27) and for each country separately.

Business surveys are qualitative in character and therefore questions used in questionnaires are mainly qualitative. Questions are closed type, and answers are usually given according to a three-option ordinal scale: "increase" (P, +), "remain unchanged" (E, =) and „decrease" (M, -).

There are various methods of quantifying the results of these studies and deriving composite indicators with forecasting properties. Namely, based on changes in indicators, the direction of change of the reference macroeconomic series can be correctly predicted for some months ahead [1, 2, 4].

Since the manufacturing industry is dominated by the structure of the economy (both national and regional economy), and the Euro area is a "representative portion" of the European Union, the research was conducted specifically for the Euro area manufacturing industry.

The analysis is based on a data set comprising Industrial confidence indicator collected from Directorate General for Economic and Financial Affairs of the EU (DG ECFIN) published on http://ec.europa.eu/economy_finance and of the industrial production, obtained from the European Central Bank (ECB), (published on www.ecb.int) for the period from January 2004 until April 2010.

Industrial production is expressed (in accordance with the EU methodology) as the growth rate of industrial production (y-o-y; compared to the same period of the previous year).

3 EMPIRICAL RESULTS

The empirical (econometric) analysis is carried out on 76 pairs of values of the rate of growth of industrial production and the ICI (for the period from January 2004 until April 2010).

In order to analyse the relationship between Euro area business survey composite indicators and industrial production, we estimated the following regression model. The regression diagnostics was conducted to examine whether baseline assumptions of the model was fulfilled. White`s test was conducted and concluded that there is no problem of heteroscedasticity. However, Durbin-Watson statistics[2] showed that there was a problem of autocorrelation. Regression analysis results are given in Table 1. It appears that there exists a problem of positive residual autocorrelation, $d < d_L$ (0,318572 < 1,598).

[2] Critical values for the Durbin-Watson test for autocorrelation for n=76, k=1, α =0,05 are: d_L = 1,598 (the lower bound), d_U =1,632 598 (the upper bound).

Table 1. Estimated model parameters and Durbin-Watson test size

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3,593694	0,388245	9,256256	0
ICI	0,556904	0,028126	19,80062	0
R-squared	0,841224	Mean dependent var		-0,467312
Adjusted R-squared	0,839078	S.D. dependent var		7,163981
S.E. of regression	2,873836	Akaike info criterion		4,975136
Sum squared resid	611,1612	Schwarz criterion		5,036471
Log likelihood	-187,0552	Hannan-Quinn criter.		4,999649
F-statistic	392,0646	Durbin-Watson stat		0,318572
Prob(F-statistic)	0			

Source: Authors' calculations, EViews 7.0.

Given the presence of the problem of autocorrelation, the following regression model with AR(1) errors was estimated.

$$RATE_t = \beta_0 + \beta_1 ICI_t + u_t, \qquad u_t = \rho \cdot u_{t-1} + \varepsilon_t \tag{2}$$

Table 2 provides results of the analysis. The results show that the problem of autocorrelation is resolved because $d > d_u$, (2,166421> 1,632).

Table 2. Regression model with AR(1) errors

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4,29277	3,302386	1,2999	0,1978
ICI	0,42199	0,088129	4,788299	0
AR(1)	0,935091	0,061862	15,11567	0
R-squared	0,95243	Mean dependent var		-0,481647
Adjusted R-squared	0,951108	S.D. dependent var		7,211127
S.E. of regression	1,594483	Akaike info criterion		3,810155
Sum squared resid	183,0512	Schwarz criterion		3,902854
Log likelihood	-139,8808	Hannan-Quinn criter.		3,847168
F-statistic	720,7772	Durbin-Watson stat		2,166421
Prob(F-statistic)	0			

Source: Authors' calculations, EViews 7.0.

The estimated results presented in Table 2 show that ICI and AR(1) process are statistically significant for the growth rate of industrial production in Euro area (in the same period) at the significant level 0,05. It means that on the basis of ICI we can predict changes in industrial production (in the same month, but ICI is published around 15 days before the official statistics of industrial production is published).

4 CONCLUSION

Business surveys are qualitative studies conducted with the aim of obtaining, timely information about the condition of the national economy and to predict its future movements. The results of business surveys are used by economic experts, government institutions, banks and other entities. They only complement the data of official statistics.

In this paper regression analyses are conducted, based on business surveys results in the manufacturing industry and the growth rate of industrial production in Euro area. We conducted regression diagnostics, and solved the problem of autocorrelation.

Based on the econometric analysis it was concluded that ICI can, with fairly high confidence, predict the changes in industrial production (for the same month). Main advantage of business surveys (it means a forecasting power) is that the results of the research are quickly available and published before the results of official statistics, complementing them. Main disadvantage is the subjectivity of respondents. Business surveys constitute an important part of economic cycle analysis, especially for the tracking and forecasting the direction of changes in referent economic series.

References

[1] Bahovec, V., Cizmesija, M. and Kurnoga – Zivadinovic, N., 2008. The forecasts of some macroeconomic variables for Croatia using the information provided by business surveys, Proceedings of the 11th International Conference on Operational Research KOI 2006, Boljuncic, V. (editor), Pula, Croatia, September 27 - 29, 2006: 77 – 86.

[2] Claveria,O., Pons, E. and Ramos, R., 2005. Business and Consumer Expectations and Macroeconomic Forecasts, Proceedings of Conference on Survey data in Economics – Methodology and Applications– Cess IFO, Munich: 1-20.

[3] Čižmešija, M., 2008. Konjunkturni testovi Europske Unije i Hrvatske. Zagreb. Biblioteka Privrednog vjesnika.

[4] Čižmešija, M., N. Erjavec and V. Bahovec., 2010. EU Business and Consumer Survey Indicators and Croatian Economy, Zagreb international review of economics & business, XIII (2010.), 2; 15-25.

[5] European Economy, European Commission, Directorate-General for Economic and Financial Affairs 1997, The joint harmonised EU programme of business and consumer surveys

[6] European Economy, European Commission, Directorate-General for Economic and Financial Affairs 2007, The joint harmonised EU programme of business and consumer surveys, User guide.

[7] European Commission, http://ec.europa.eu/economy_finance, Accessed February 10, 2011

[8] European Central Bank - ECB, www.ecb.int Accessed February 10, 2011

LAYOUT OF COMPANY'S FUNCTION UNITS

Stanislav Gorenc[1], **Mitja Hrast**[2], **Tomaž Berlec**[3] and **Marko Starbek**[3]
[1] PLASTA proizvodnja in trgovina d.o.o., Šentrupert, Slovenia
[2] Kovinski izdelki Hrast Milan s.p., Breginj, Slovenia
[3] Faculty of Mechanical Engineering, Aškerčeva ulica 6, Ljubljana, Slovenia

Abstract: Production globalisation, increased dynamics and cost pressure force companies into permanent innovative adaptation of production structures – company's function units to changed conditions. A company will be competitive on the world market only if capable of a quick adaptation to market requirements and of producing a product within set deadline, in required quality and at a market acceptable price.
To be capable of meeting the requirements of either the market or customers a company must be capable of reacting to the changed production programme by an adequate layout of the company's function units, i. e. optimization of layout of company's function units.
The article will illustrate a method of optimum layout of company's function units emphasizing the design of a company's functional diagram and dimensioning of equipment, employees and surfaces of company's function units.

Keywords: layout, function unit, planning, dimensioning, capacities, work surface.

1 INTRODUCTION

Having reviewed published works on the problems of optimum layout of company's function units [1,2,3,4] and having carried out a research, we came to a conclusion that the layout of company's function units should undergo the methods shown in Figure 1.

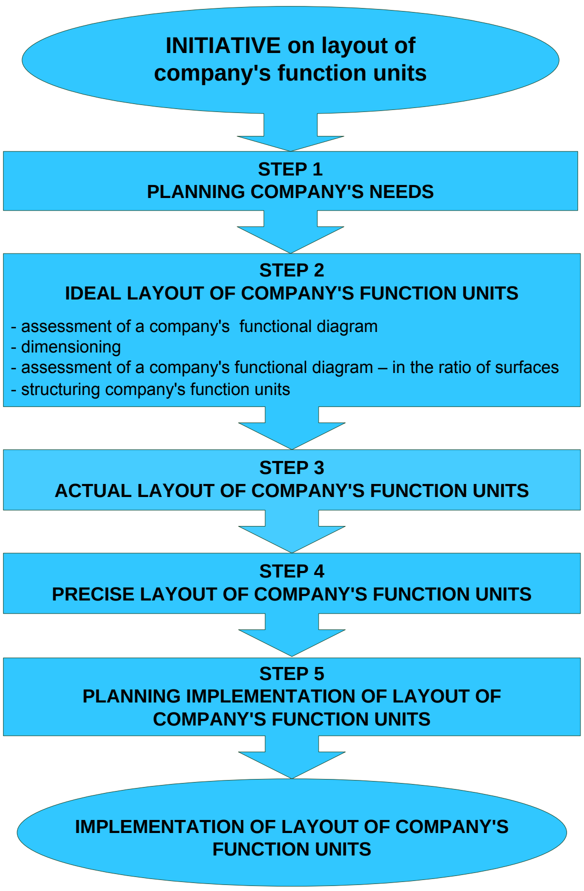


Figure 1: Method of layout of company's function units

Figure 1 shows that a method of layout of company's function units begins with a step of planning company goals. Then follow a step of ideal layout of function units, a step of actual layout, a step of fine layout of function units and finally a step of planning the implementation of layout of company's function units.

In the continuation of the article we will pay special attention to the implementation of step two, i. e. ideal layout of company's function units with an emphasis on determining the functional diagram of the company and dimensioning of equipment and employees needed for the realization of the production programme and needed surfaces.

2 IDENTIFICATION OF A COMPANY'S FUNCTIONAL DIAGRAM

A company's functional diagram shows the needed company's function units that are required for the implementation of the company's production programme and a planned link between function units based on material flow. In short, a company's functional diagram provides the data on the type and number of company's function units (departments, jobs), on relations of function units with respect to material flow and on production development logic.

A company's functional diagram is reached progressively by the implementation of the production programme analysis and design of a work flow diagram [1,6].

A company's production programme and designed product structures allow a calculation of secondary needs for elements of products (method of linear equations and matrix method).

Technological methods of production and assembly of products included in the production programme allow a creation of a work flow diagram. The work flow diagram indicates the required number of production / assembling operations and the sequence of operations for each element of the production programme (product, assembly, component).

The work flow diagram can be the basis for a functional diagram that shows the place where production and assembling operations are performed and the anticipated material flow.

Figure 2 shows an example of a well-established functional diagram of a company.

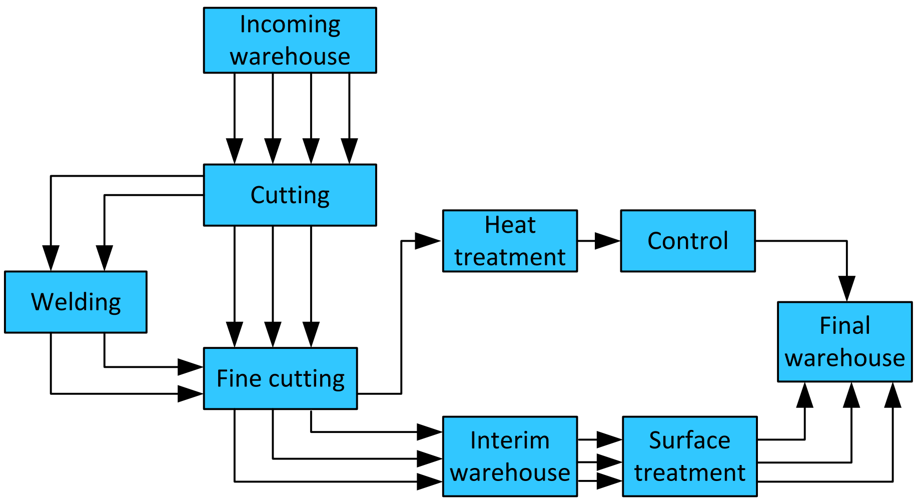


Figure 2: Company's functional diagram

3 DIMENSIONING OF EQUIPMENT, EMPLOYEES AND AREAS OF COMPANY'S FUNCTION UNITS

A basis for dimensioning equipment, employees and areas of company's function units is a production programme of a company [3]. The company's production programme dictates the needed capacities of equipment, employees and surfaces of company's function units.

3.1 Dimensioning of equipment

Equipment generally designate machines, plants, devices, measuring equipment and tools.
Dimensioning of equipment comprises both identification of type of equipment (qualitative capacities) and the quantity of equipment (quantitative capacities). The type of equipment is determined by a department of production preparation on the basis of technological requirements of a workpiece in question. The quantity of equipment or the need for equipment generally represents a ratio between the needed and available capacities of equipment:

$$\mathrm{QNE_j} = \frac{\mathrm{NC_j}}{\mathrm{AC_j}} \tag{1}$$

QNE_j – quantity of needed equipment of type *j*
NC_j – needed capacities of equipment of type *j* [min/year]
AC_j – available capacities of equipment of type *j* [min/year]

The needed capacities of equipment in individual production are of type *j*:

$$\mathrm{NC_j^*} = \sum_{i=1}^{n} \mathrm{Tp_{ij}} + \sum_{i=1}^{n} (\mathrm{m_{ij}} \cdot \mathrm{t_{eij}}) \tag{2}$$

NC_j* – needed capacities of equipment of type *j* – in individual production [min/year]
Tp_{ij} – time of preparation of a product *i* on a equipment *j* [min/year]
m_{ij} – produced quantity of a product *i* on a equipment *j* [pc/year]
te_{ij} – time per unit of the product *i* on a equipment *j* [min/year]
n – number of product varieties

The available capacities of equipment in individual production are of type *j*:

$$\mathrm{AC_j} = \mathrm{NWD_j} \cdot \mathrm{TAE_j} \cdot \mathrm{NS_j} \cdot \eta_j \tag{3}$$

AC_j – available capacities of equipment of type *j* [min/year]
NWD_j – number of working days of availability of equipment of type *j* [Dd/year]
TAE_j – available time of use of equipment of type *j* per working day and shift [min/Dd × shift]
$ŠI_j$ – number of shifts of equipment of type *j* [shifts/Dd]
η_j – hour utilisation rate of equipment of type *j* (η_j = 0.7 – 0.9)

If the equations (2) and (3) are inserted into the equation (1), the amount of needed equipment of type *j* in individual production equals to:

$$\mathrm{ANE_j^*} = \frac{\sum_{i=1}^{n} \mathrm{Tp_{ij}} + \sum_{i=1}^{n} \mathrm{m_{ij}} \cdot \mathrm{t_{eij}}}{\mathrm{NWD_j} \cdot \mathrm{TAE_j} \cdot \mathrm{NS_j} \cdot \eta_j} \tag{4}$$

ANE_j* – amount of needed equipment of type *j* – individual production

Series production calls for capacities of equipment of type *j*:

$$NC_j^{**} = \sum_{i=1}^{n}(Tp_{ij} \cdot NS_{ij}) + \sum_{i=1}^{n}(m_{ij} \cdot t_{eij}) \tag{5}$$

NC_j** – needed capacity of equipment of type *j* – in series production [min/year]
Tp_{ij} – preparation time of a series of product *i* on a equipment *j* [min/series]
NS_{ij} – number of series of product *i* on a equipment *j* – per year [series/year]
m_{ij} – produced quantity of product *i* on a equipment *j* [pc/year]
te_{ij} – time per unit of product *i* on a equipment *j* [min/pc]
n – number of product varieties

Available capacities of equipment of type *j* in series production equal the available capacities of equipment in individual production – equation (3).

If the equations (5) and (3) are inserted into the equation (1) the amount of needed equipment of type *j* in series production amounts to:

$$ANE_j^{**} = \frac{\sum_{i=1}^{n}(Tp_{ij} \cdot NS_{ij}) + \sum_{i=1}^{n} m_{ij} \cdot t_{eij}}{NWD_j \cdot TAE_j \cdot NS_j \cdot \eta_j} \tag{6}$$

ANE_j** – amount of needed equipment of type *j* – series production

3.2 Dimensioning of employees

Dimensioning of employees includes both defining the type of employees (qualitative needs) and the number of employees (quantitative needs). The number of employees or the need for employees is generally represented by a ratio between the total needed time for the implementation of all activities performed by the employees and the available working time of one person:

$$NP_j \geq \frac{t_{ne\,j}}{t_{av\,j}} \tag{7}$$

NP_j – number of persons of a professional group *j*
$T_{ne\,j}$ – total needed time for the implementation of all activities of the employees of a professional group *j* [Nh/year]
$t_{av\,j}$ – available working time for the implementation of all activities of the employees of a professional group *j* [Nh/year]

3.3 Dimensioning of surfaces of company's function units

The results of dimensioning of equipment and employees are the basis for dimensioning the areas of company's function units (workshops, workplaces).

Several methods [1,5] can be used to determine the needed areas of company's function units, however, the method of functional determination of workshops' surfaces is most often used. This method is based on the results of a research performed in small- and middle-size companies of engine construction.

The surface of a company's function unit area is determined by the equation:

$$S_{fu} = S_p + S_s + S_{tr} + S_{add} \tag{8}$$

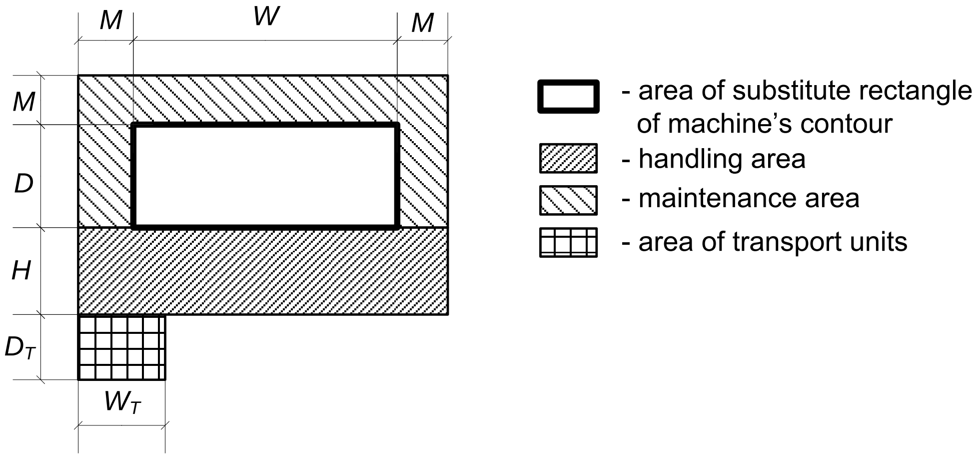
S_{fu} – surface of function unit's area [m^2]
S_p – surface for production [m^2]
S_s – surface of an intermediate storage facility [m^2]
S_{tr} – transport surface [m^2]
S_{add} – additional surface [m^2]

The surface for the production S_p of a function unit represents a sum of all working areas of function unit's workplaces:

$$\mathrm{S_p} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{S_{wa\,j}} \tag{9}$$

S_p – surface for production [m^2]
$S_{wa\,i}$ – work area of the i^{th} workplace of the function unit [m^2]
i – number of workplaces of a function unit

The work area e.g. of a machine workplace $P_{dm\,i}$ is determined by the method of substitute surfaces, wherein the surfaces of a substitute rectangle of a machine contour are added to the surface for machine handling and maintenance and to the surface of location of a transport unit (palette, box), which is shown in Figure 3.



H – handling area supplement (1 m)
M – maintenance supplement (0.4 m)
W – width of a machine
D – depth of a machine
W_T – width of a transport unit
D_T – depth of a transport unit

Figure 3: Structure of areas of the i^{th} machine workplace

The area of the i^{th} standing workplace assessed by the method of substitute areas is:

$$\mathrm{S_{wa\,i}} = (W + 2 \cdot \mathrm{M}) \cdot (D + \mathrm{H} + \mathrm{M}) + W_T \cdot \mathrm{D}_T \tag{10}$$

After the dimensioning of equipment, employees and function unit's areas a company's functional diagram can be designed in the ratio of areas. Thereafter follows structuring of the company's function units (material flow analysis, identification of production mode, design of ideal layout of company's function units).

4 CONCLUSION

The layout of company's function units has a considerable influence on the material flow on the relation incoming warehouse – production processes – outgoing warehouse.

The layout of company's function units must allow keeping the costs of internal material transport at the minimum level.

The article describes a method of layout of company's function units with the emphasis on the manner how a company's functional diagram is designed, how the needed equipment, employees and areas of company's function units are dimensioned.

Further research will focus on the problem of structuring company's function units and creating a proposal of several variants of ideal layout of company's function units (presentation and evaluation of material flow structure, assessment of production mode, design of variants of ideal layout of function units).

Figure 4 shows the example of using theoretical principles for preparing ideal layout of blown film extrusion and bag-making company.

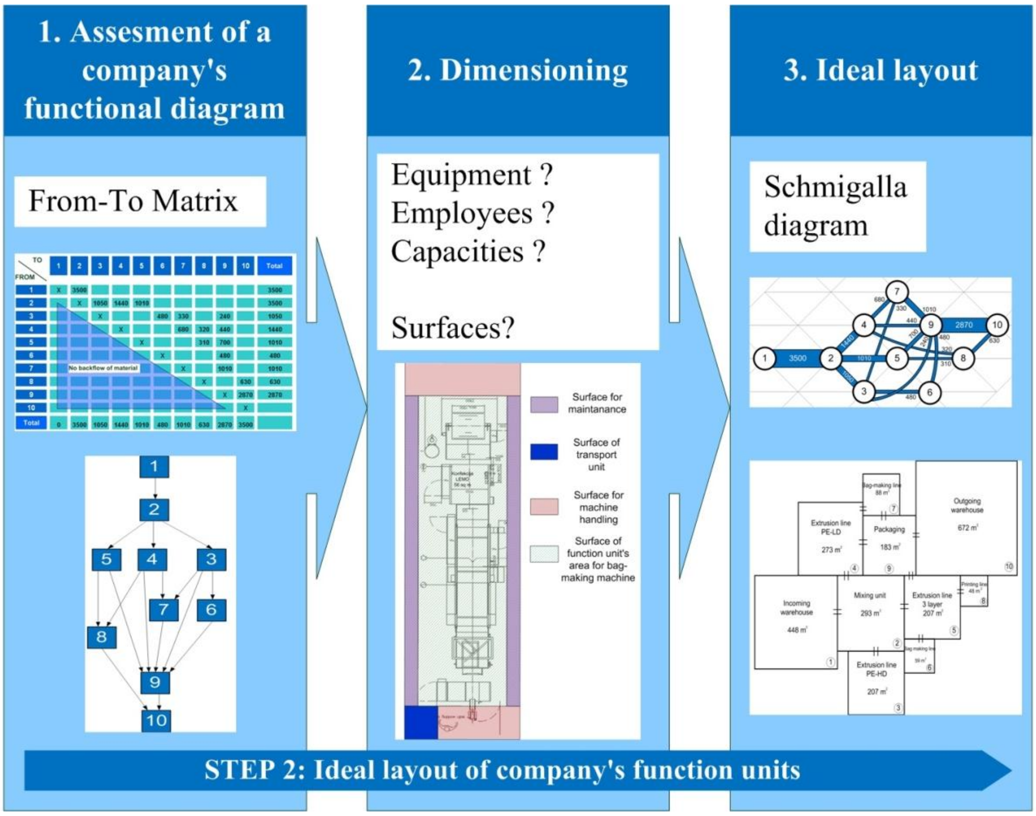


Figure 4: Example of building ideal layout of blown film extrusion and bag making company

References

[1] Bela Aggteleky: Fabrikplanung, Band 1,2,3, Carl Hanser Verlag München Wien, 1981

[2] Siegmar R. Frey: Plant Layout, Carl Hanser Verlag München Wien, 1975

[3] Claus G. Grundig: Fabrikplanung, Carl Hanser Verlag München Wien, 2009

[4] M. Schenk, S. Wirth: Fabrikplanung und Fabrikbetrieb, Springer Berlin, 2004

[5] H. Schmigalla: Fabrikplanung, Carl Hanser Verlag. 1995

[6] H.P. Wiendahl, J. Reichardt, P. Nyhuis: Handbuch Fabrikplanung, Carl Hanser Verlag München Wien, 2009

[7] Francis L.R., McGinnis F., White A.J.: Facility Layout and Location, Prentice-Hall, Inc., Engliwood Cliffs, New Jersey, 1991

VIRTUAL FACTORY – THE REALITY OF THE PRESENT

Mitja Hrast[1], **Stanislav Gorenc**[2], **Tomaž Berlec**[3] and **Marko Starbek**[3]
[1] Kovinski izdelki Hrast Milan s.p., Breginj, Slovenia
[2] PLASTA proizvodnja in trgovina d.o.o., Šentrupert, Slovenia
[3] Fakulteta za strojništvo, Aškerčeva ulica 6, Ljubljana, Slovenia

Abstract : This paper presents guidelines for building a network of distributed production units – i.e. virtual factory, which allows quick response to market demands and ensures coordinated cooperation of partner enterprises in a process of new product realisation. The described organisational aspects of a virtual factory form the foundation for beginning of virtual factory creation, on the basis of enterprise networks and partnership between enterprises that will be included in a virtual factory. Theoretical assumptions on the possibilities of organising a group of enterprises to a virtual factory are given, as well as advantages and disadvantages of such linking.

Keywords: virtual factory, enterprise network, partnership, information and communication infrastructure

1 INTRODUCTION

Development of information and communication technologies (ICT), endeavours of small and medium-sized enterprises (SMEs) for better competitiveness when producing more complex products and endeavours for better use of capacities lead to a new concept of integration of enterprises to the so-called virtual factory that consists of distributed production units of partner enterprises linked by ICT.

In today's globally-oriented production environment, the virtual factory (with its distributed production units) is becoming an important temporary organisational form of linked enterprises, which allows fast response to market opportunities even in unpredictable market changes, as well as an efficient distribution and use of the available resources of the virtual factory partners.

Today's production environment requires deeper cooperation with customer, sub-contractors and competitors; it also requires establishment of such an organisation type that can manage the changes, as well as uncertainty and complexity of the business environment by guiding the personnel and managing the information *[3]*.

Virtual factory consists of geographically dispersed production units of linked enterprises and sub-contractors *[7]*.

The basis for creation of a virtual factory is a suitable enterprise network and established ICT structure, which facilitates communication and information exchange between partners. Simple access to databases allows the virtual factory partners to reduce product realisation time *[1]*.

2 ENTERPRISE NETWORK – THE BASIS FOR CREATION OF VIRTUAL FACTORY

Enterprise network integrates distributed production units and (if needed) incorporates them in the particular product realisation process. Usually this is a temporary integration of enterprises, which dissolves after reaching the goal. Enterprises integrate or join to take advantage of better research and development opportunities, create greater financial resources or a wider distribution network. Enterprises included in the virtual factory share common development, common technology achievements, total costs, total profit and common market.

Networks serve both for informal exchange of information and for joint implementation of innovation projects, and for pilot applications and market launches *[6]*.

2.1 Enterprise partnerships

Selection of partners is one of the most important activities when creating a virtual factory. Criteria for the selection of partners should be based primarily on competence of partners (technology, costs, speed and flexibility, quality of products and services, information technology), location of partner enterprises and supply-chain efficiency.

Virtual factory is a group of partner enterprises that are temporarily involved in the product realisation process. The management structure must allow for easy inclusion and exclusion of partners, and independent operation of partners while they are part of the virtual factory.

Figure 1 shows a structure of the virtual factory, which consists of the leading enterprise and three partner enterprises.

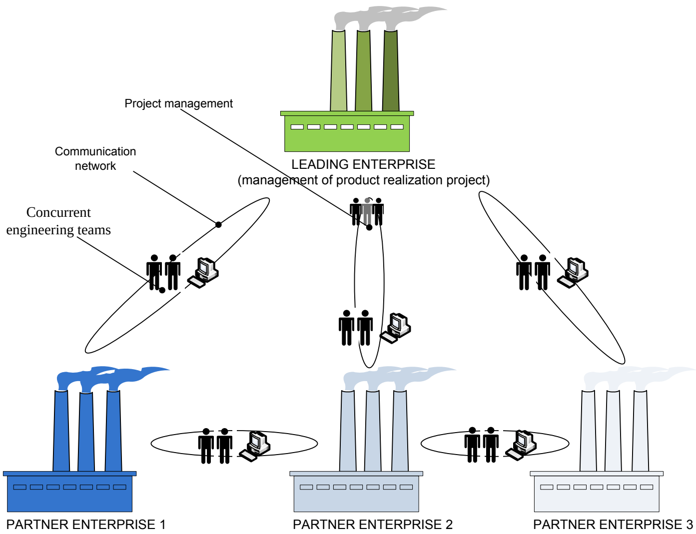


Figure 1: Virtual factory with the leading enterprise and three partner enterprises

There are different types of virtual factories, differing from each other mainly by the level of integration, duration of collaboration, type and volume of shared information, and division of tasks *[2], [9]*.

Regarding the enterprise partnership, three types of virtual factories exist:

- virtual factory with static partnership,
- virtual factory with agent partnership,
- virtual factory with decentralised partnership.

2.1.1 Virtual factory with static enterprise partnership

The static partnership consists of permanent partners with known division of competences and responsibilities, and the course of execution of the order is known in advance. Figure 2 shows a design of a virtual factory with static partnership.

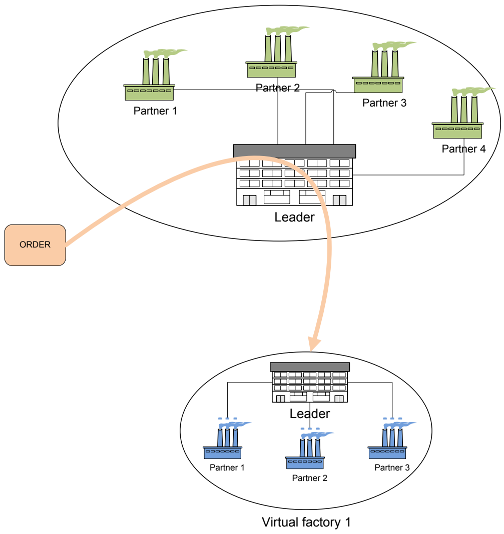


Figure 2: Virtual factory with static partnership

The costs of building a virtual factory of enterprises that are in static partnership are small, because partnership exactly defines division of services between partners, as well as risk- and profit sharing.

2.1.2 Virtual factory with agent partnership between enterprises

Agent partnership is a connection of equal partners, while management of the partnership and creation of a virtual factory is performed by an agent working under the agreed interests of all partners.

Figure 3 shows a design of a virtual factory with agent partnership between enterprises.

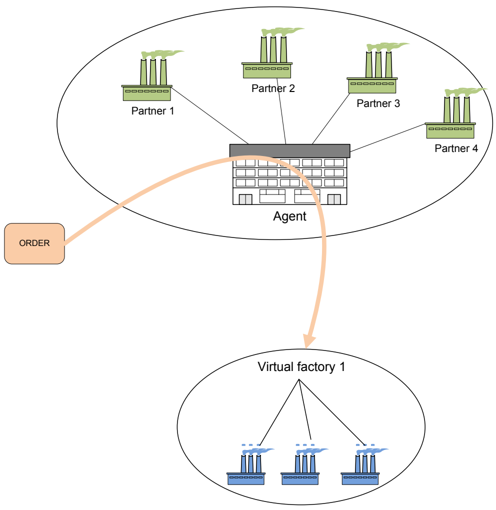


Figure 3: Virtual factory with agent partnership between enterprises

In agent partnership between enterprises the buyer contacts the agent who is responsible for clarifying the services related to the order offered and the construction of a virtual factory. The agent finds the best partners on the basis of the availability of the competences of partners and specific customer requirements.

2.1.3 Virtual factory with decentralised partnership between enterprises

In a decentralized partnership between enterprises, the task of creating a virtual factory is taken by the partner enterprise, which has accepted the order – the initiative partner.

Figure 4 shows the creation of a virtual factory in a decentralised partnership between enterprises *[8]*.

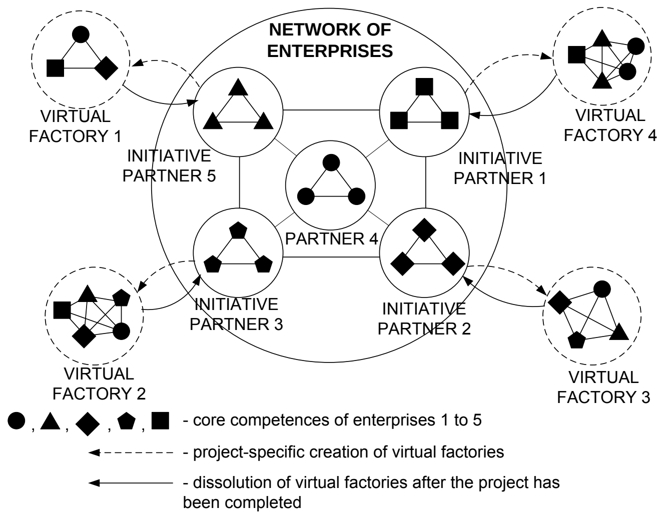


Figure 4: Virtual factory with decentralised partnership between enterprises

Inactive links represent a network of enterprises from which two or more virtual factories are created, composed of the core competences of individual enterprises. Initiative partner must provide a breakdown of the buyer's order to partial orders, description of partial orders and selection of appropriate partner enterprises.

3 CREATION OF A VIRTUAL FACTORY

The virtual factory is a manufacturing and technically organized virtual enterprise. Virtual factory is built on the basis of a network of enterprises (i.e. cooperation links between enterprises). Production network of enterprises is a set of links between two or more enterprises.

The lifetime of a virtual factory *[4]* consists of establishment phase, operation phase, development phase and the dissolution of the virtual factory, as shown in *Figure 5*.

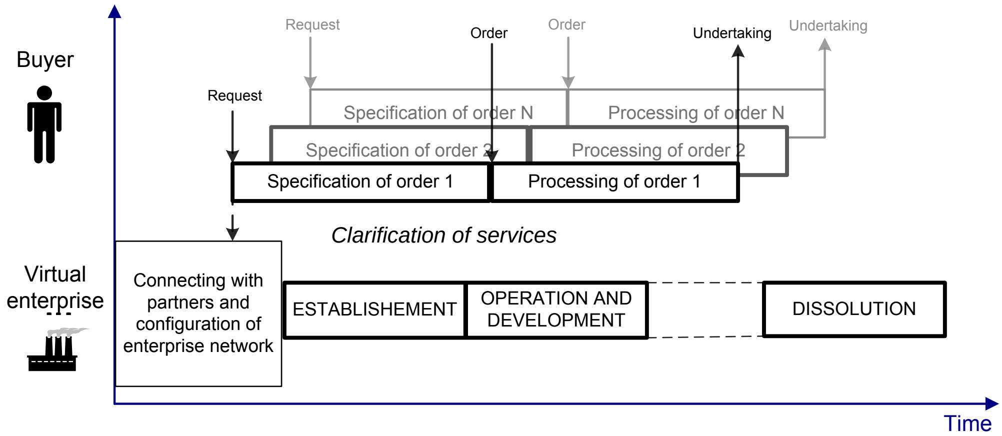


Figure 5: The lifetime phases of a virtual factory

Contrary to the classical enterprise production, the virtual factory with its distributed units has its production distributed according to competences and needs. Production in a virtual group of enterprises leads to a more complex organisation and it offers new challenges in the design of virtual factories *[1]*. Due to the nature and variability of the environment, traditional product realisation systems fail in this case, therefore it is necessary to find appropriate combination of managing the product realisation process by including project management and concurrent product realisation *[5]*.

Design of product or service realisation process is affected by the number of partners involved, the structure of partners and their flexibility, product complexity, project duration and organisation of work.

Management of a virtual factory creation project (during realisation of a product or service) begins with a breakdown of a product to segments, covered by enterprise's own resources or by resources of potential partners involved in the virtual factory. Such a creation of the virtual factory is built with a view to establish appropriate organisational guidelines in forming the product realisation process.

Creation of the virtual factory consists of the following phases:

Phase 1: Customer enquiry
Phase 2: Forwarding parts of tender to partner enterprises
Phase 3: Searching for proper tenders
Phase 4: Tenders that meet the criteria of enterprises
Phase 5: Preparation and dispatch of a bid
Phase 6: Preparation and dispatch of the bid to the customer
Phase 7: Sending the order
Phase 8: Preparation and dispatch of the order

The objective of building a virtual factory is to build such a virtual factory that will be able to respond quickly and organise itself in response to the changing market demands.

Communication, collaboration and sharing of data between the enterprises in a virtual factory are done via electronic data exchange on the basis of the agreed semantics and syntax.

After registering in the system and upon authorisation and identification, the user is granted the access to the data. The level of data access is agreed when a partner enters the virtual factory. The shared data allow rapid product realisation, but they do not contain the organisational knowledge of partners.

Information and communication infrastructure has completely different meaning in a virtual factory than in a traditional enterprise. In a traditional enterprise the ICT increases the efficiency of the enterprise, while in the virtual factory the ICT is an indispensable element that allows for the construction and operation of the virtual factory.

Figure 6 shows the ICT scheme in a virtual factory.

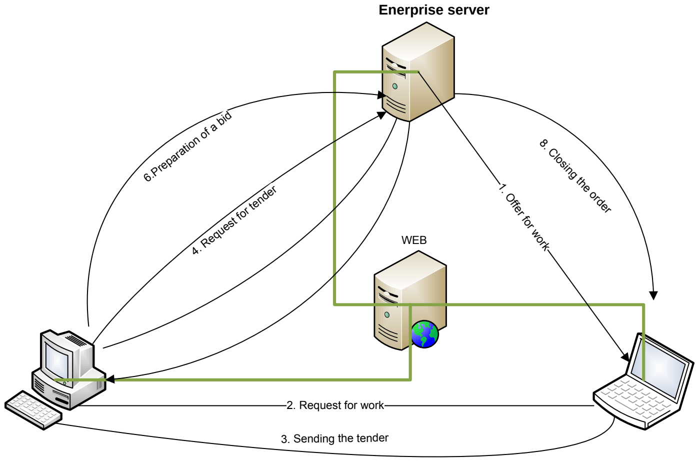


Figure 6: ICT infrastructure of a virtual factory

4 CONCLUSION

The development of information and communication technology and the changing circumstances of the business environment that require products with a high degree of complexity, combined with short delivery times and ever smaller series, force enterprises to unite. In such business environment it is essential for enterprises to establish more in-depth collaboration with customers, sub-contractors and competitors.

Enterprises are required to be very flexible and to be able to realise new product quickly. This is a major problem for SMEs, because they do not have sufficient knowledge, resources and technology to be able to quickly adapt to market demands and to be competitive.

Temporary long-term or short-term business relationship with subcontractors helps to improve flexibility and responsiveness of linked enterprises; however, the integration of activities in linked enterprises is complex. The latest trends in business associations lead to a new business concept: the virtual factory.

Support for managing the virtual factory is provided by the ICT. Links between different partners of a virtual factory, based on a shared technical knowledge, competences, resources and costs, allow better response to business opportunities.

Responsiveness in a virtual factory is achieved by integration of business partners, personnel and technology in a suitable entity on the basis of the ICT and adaptive organisational structures, and concurrent product realisation teams, supported by highly qualified, trained and motivated personnel.

Further research will focus on selection criteria for determining the type of partner links in virtual factory enterprises, and inventory of business process for design and management of the established virtual factory.

References

[1] Abair, R. A.,: Agile manufacturing: *This is not just repacking of material requirements planning and just-in-time*, American Society for Mechanical Engineers, design Engineering Devision (Publication) DE (2), 1995, pp. 19-198

[2] Camarinha-Mathos, L. M. et al., 1999, The virtual enterprise concept, Infrastructures for Virtual Enterprises, Kluwer Academic Publishers, Boston, pp. 3–14.

[3] Camarinha-Mathos, L.M.: Execution systems for distribited business precesses in a virtual enterprise, Future Generation Computer Systems 17, Elsevier, 2001, pp. 1009-1021

[4] Davidow, W., Malone, M., 1995, L’Entreprise a I’age du Virtuel, MAXIMA, Collection: Institut du Management d’EDF et de GDF, Paris.

[5] Duhovnik, J., Žargi, U., Kušar, J., Starbek, M., 2009, Project-driven concurrent product development. Concurr. eng. res. appl.,Vol. 17, No 3, pp. 225-236

[6] Goldman, S., e tal.: Competitors and Virtual Organizations, New York Organisations, New York, Van Nostrand Reinhold, 1995

[7] Gunasekaran, A.: Agile manufacturing: A framework: A framework for research and development, International Journal of Production Economics, 62, 1999, pp. 87-105

[8] Kušar, J., Brezovar, A., Berlec, T., Starbek, M., 2002, Concurrent Product Development in Virtual Distributed Companies, DET021st CIRP (UK) International Seminar on Digital Enterprise Technology, Durham, UK.

[9] Marko Starbek: Organizacijske strukture podjetij; Študijsko gradivo, Ljubljana, pp.16-38

[10] Prasad, B.: Concurrent Engineering Fundamentals, Volume I. Integrated Product and Process Organization, New Jersey. Prentice Hall PTR, pp. 216–276.

REVERSE LOGISTICS FACILITY LOCATION IN EXTENDED MRP THEORY

Danijel Kovačić and **Marija Bogataj**
Department of Mathematics and Operations Research, Faculty of Economics, University of Ljubljana,
Kardeljeva ploščad 17, SI-1000 Ljubljana, Slovenia
danijel.kovacic@ef.uni-lj.si, marija.bogataj@guest.arnes.si

Abstract: Grubbström's well-developed MRP Theory has been mostly used in modeling production processes. Global supply chains also contain distribution, consumption and recycling processes. For this reason theory was recently further extended to incorporate all kinds of activities. Such extended model, which consists of four main subsystems, is closed and can be used for several detailed analyses. Since the importance of reverse logistics is increasing, this paper will focus on parameters which determine Net Present Value of the whole system, depending on geographical location of recycling facilities. We will show how lead times, transportation costs, setup costs and price of labor and energy of individual location contribute to overall NPV of the system.

Keywords: MRP Theory, input-output analysis, reverse logistics, Net Present Value (NPV), environmental tax, transportation costs, labor price, energy

1 INTRODUCTION

Grubbström's MRP Theory, inspired by Orlicky [11], has been well researched and developed in many ways during last two decades, focusing mainly on multistage assembly systems [3], [5]. One of the Theory's most important ingredients is input-output analysis of Leontief [10]. Recently it has been further extended in such way that besides production (nr. 1) it also consists of distribution (nr. 2), consumption (nr. 3) and recycling (nr. 4) subsystems [4]. Those last three arborescent subsystems are successors of production (multistage assembly system), and create closed model where some items are circulating and some are exported as waste and replaced with newly purchased items on the market at the beginning of the next cycle [6], [7], [8] and [8]. Relation between waste and new items in such model is balanced with return rate α. Many factors influence optimal return rate besides lead times, which are one of the basic properties when calculating Net Present Value (NPV) using the MRP Theory. Transportation costs of used items coming into recycling, together with corresponding transportation costs of recovered items, which are going back to production, and waste items going to landfills, can significantly influence the NPV of the whole system **Error! Reference source not found.**, [1], [2]. Another important factor which should be considered is taxation connected with waste elements and pollution. Higher environmental taxes will stimulate higher return rates since recycling would have more rational impact on NPV compared to disposal of used items. Labor price, which covers quantity of needed work together with the price of workforce, and setup costs, which cover other physical costs (machinery, buildings, land price, etc.) also drastically influence NPV of the whole system **Error! Reference source not found.**.

Since many works about the theory are available, only the most important ones which are crucial for understanding of this paper are listed in this introduction. Reader can find detailed information in previously mentioned papers and many other works written by Grubbström et al.

The purpose of this paper is to further extend general model of all material flows in a cyclical system [8], [9], including production functions with labor and energy as inputs. Previously presented model and equations will be further developed in such way that it will be possible to evaluate every individual potential location of recycling facility, which is only one link in a closed chain of all activity cells of four sub-systems.

2 FURTHER EXTENSION OF THE MODEL

Individual activity cells of each of the four sub-systems are geographically dispersed. In the global world, components of the final product are usually assembled at different manufactures which can be far away from each other. The final products are distributed via distribution centers to unlimited number of micro locations. At the end of the lifecycle, final products enter recycling sub-system. In most cases, we may assume that consumption is the main source of items for recycling. However, some elements (components or final products) can enter recycling sub-system from any activity cell as a result of damage, insufficient quality level, etc. Consumption depends on demographic factors and is thus geographically not evenly distributed. Most of the consumption will usually be concentrated in the geographical centers, such as major cities and industrial centers. These urban and industrial centers are inappropriate for setting up recycling plants. Usually there will be several potential areas for locating a recycling plant around consumption centers, which are distinguished by technological, economic and legal-formal factors (Figure 1). Optimally chosen location may be geographically relatively far away from the center of consumption in case of attractive production factors such as technological development, low setup costs, labor force or energy costs or attractive legislation (which can be interpreted as low environmental taxation in our model).

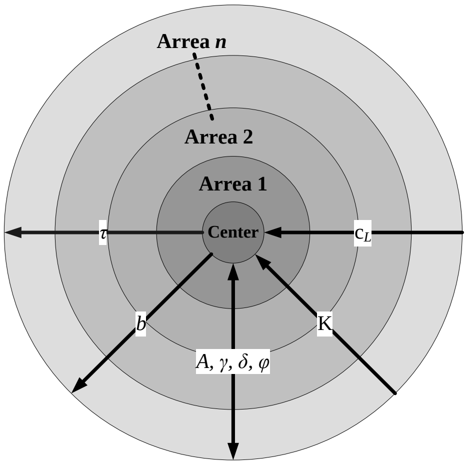


Figure 1: Parameters determining optimal geographic location of the recycling facility and their expected way of increasing/decreasing through areas 1, 2, …, *n* located around the center of consumption.

Optimal location of recycling activity cell will be determined by lead times and transportation costs on the following routes:
1) Consumption (collection centers) → recycling plant,
2) Recycling plant → next recycling plant (when disassembly activities are dislocated)
3) Recycling plant → landfills
4) Recycling plant → production plant (where recycled element is reused in a new cycle).

The more the recycling activity cell is away from the center of origin of waste items, higher lead times τ and transportation costs b can be expected. From this perspective locations close to the center of origin of waste items and production plants are more economical.

On the other hand, another important factor for optimization of the location is the price of labor force c_L. Consumption centers usually represent economic centers, where prices are higher, which is related to the level of household income and consequently determines the price of labor force. If recycling plant is built away from the economic centers (or even inside countries with lower labor costs), such location will be attractive in terms of labor cost, but it will usually result in a higher lead times and transportation costs.

Similar assumption can be made for setup costs K. Prices of real estates are declining with the distance from centers of economic activities. More remote locations will be more economically acceptable, but will result in higher lead times and transportation costs.

Price of the energy c_E and efficiency of disassembly processes are another important factors which should be considered when optimal location of recycling plant has to be chosen. Figure 2 presents inputs and outputs of recycling process which include labor force and energy.

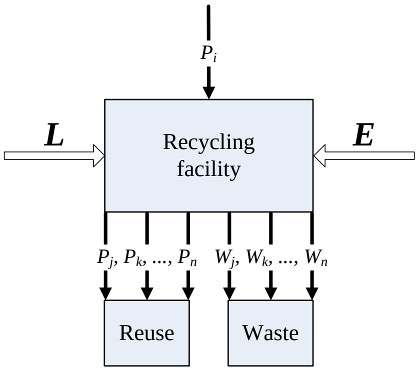


Figure 2: Used element P_i is entering recycling facility where it is disassembled using labor force L and energy E. The result of recycling activities are elements P_j, P_k, …, P_n, which can be reused in next production cycle of P_i, and waste elements W_j, W_k, …, W_n, which end their life-cycle and should be disposed of.

Generalized input and output matrices $\breve{\mathbf{H}}(s)$ and $\breve{\mathbf{G}}(s)$ of extended MRP model, which contain technical coefficients together with lead times, can be written as:

$$\breve{\mathbf{H}}(s)=\begin{bmatrix}\breve{\mathbf{H}}_{11}(s) & \breve{\mathbf{H}}_{12}(s) & 0 & \breve{\mathbf{H}}_{14}(s)\\ 0 & 0 & \breve{\mathbf{H}}_{23}(s) & \breve{\mathbf{H}}_{24}(s)\\ 0 & 0 & 0 & \breve{\mathbf{H}}_{34}(s)\\ 0 & 0 & 0 & 0\end{bmatrix} \qquad \breve{\mathbf{G}}(s)=\begin{bmatrix}\breve{\mathbf{G}}_{11}(s) & 0 & 0 & \breve{\mathbf{G}}_{14}(s)\\ 0 & \breve{\mathbf{G}}_{22}(s) & 0 & 0\\ 0 & 0 & \breve{\mathbf{G}}_{33}(s) & 0\\ 0 & 0 & 0 & \breve{\mathbf{G}}_{44}(s)\end{bmatrix}$$

Generalized input matrix $\breve{\mathbf{H}}(s)$ consists of several sub-matrices $\breve{\mathbf{H}}_{ij}(s)$, where every $\breve{\mathbf{H}}_{ij}(s)$ relates to items required by sub-system i because of processes in sub-system j. Sub-matrix $\breve{\mathbf{G}}_{ij}(s)$ relates to outputs of sub-system k as requirements of processes in sub-system l. Every sub-matrix $\breve{\mathbf{G}}_{ij}(s)$ always coincides with $\breve{\mathbf{H}}_{ij}(s)$ in dimension. Detailed description of every individual coefficient h_{kl}^{ij} and g_{kl}^{ij} which appear in generalized sub-matrices $\breve{\mathbf{H}}_{ij}(s)$ and $\breve{\mathbf{G}}_{ij}(s)$, can be found in [6], [8]. Obviously, coefficients g_{kl}^{14} and $g_{(k-2l)l}^{44}$, found in sub-matrices $\breve{\mathbf{G}}_{14}(s)$ and $\breve{\mathbf{G}}_{44}(s)$, are defining recycling activities. g_{kl}^{14} and $g_{(k-2l)l}^{44}$ can be expressed as: $g_{kl}^{14}=h_{kl}^{14}*\alpha_{kl}$ and $g_{(k-2l)l}^{44}=h_{kl}^{11}*(1-\alpha_{kl})$ where α_{kl} shows return rate of elementary item k being recycled from component l. Final return rate α_{kl} depends on all partial coefficients α and β, where β is defining individual material flows between two activity cells. Final return rate α_{kl} was formulated in the following way:

$$\alpha_{kl} = \sum_{n'=l}^{3} \Big(\prod_{i=k,l,l-1,\ldots n',4}^{n',4} \beta_i \alpha_i \Big)_{n'} \tag{1}$$

where i shows sequence of activity cells from the lowest to the highest level and n' defines exiting activity cell into recycling sub-system.

Production functions should be used for proper observation of recycling processes at every individual location [4]. In classical Cobb-Douglas production function:

$$Q = AL^{\gamma} K^{\delta} \tag{2}$$

which is now used for reverse logistics purposes, output Q can be replaced with $Q = \alpha \hat{P}$, where $\hat{P}$ is batch size of elements being recycled. The final products (outputs) of recycling processes are recovered elementary items, having some value. In order to produce certain quantity of recovered elementary items, sufficient quantity of used items should enter into recycling process (inputs). Consequently, more inputs (used items from production, distribution, consumption) result in higher quantity of outputs (recovered elementary items). This means that used items play vital role as inputs of the production function and since their quantity determines the output of the recycling sub-process, they can be treated as input capital K. Input capital of recycling sub-system is then a batch of used input elements: $K = \hat{P}$. Further, if we add energy as an input [12], extended Cobb-Douglas production function can be written in the following way:

$$\alpha \hat{P} = AL^{\gamma} \hat{P}^{\delta} E^{\varphi} \tag{3}$$

with standard assumptions concerning parameters $A > 0$, $0 < \gamma, \delta, \varphi \leq 1$ and $0 < \gamma + \delta + \varphi \leq 1$, giving decreasing returns to scale. Using (1) and (3), final return rate α_{kl} can now be written as:

$$\alpha_{kl} = \sum_{n'=l}^{3} \Big(\prod_{i=k,l,l-1,\ldots n',4}^{n',4} \beta_i A_i \hat{L}_i^{\gamma} \hat{P}_i^{\delta-1} \hat{E}_i^{\varphi} \Big)_{n'} \tag{4}$$

Overall Net Present Value of the whole system, containing of four sub-systems, can now be calculated as:

$$\mathrm{NPV} = \mathbf{p}\big(\breve{\mathbf{G}}(\rho) - \breve{\mathbf{H}}(\rho)\big)\tilde{\mathbf{P}}(\rho) - \hat{\mathbf{K}}\tilde{\mathbf{v}}(\rho) - \mathbf{U}^{\mathrm{T}}\big(\tilde{\mathbf{\Pi}}_{\mathrm{G}}(\rho) + \tilde{\mathbf{\Pi}}_{\mathrm{H}}(\rho)\big)\tilde{\mathbf{P}}(\rho) - \mathbf{c}_{\mathbf{L}}\tilde{\mathbf{L}}(\rho) - \mathbf{c}_{\mathbf{E}}\tilde{\mathbf{E}}(\rho) \tag{5}$$

were $\mathbf{p}$ presents price vector, $\tilde{\mathbf{P}}(\rho)$ production plan, $\tilde{\mathbf{v}}(\rho)$ given timing, $\hat{\mathbf{K}}$ is vector of setup costs, $\tilde{\mathbf{\Pi}}_{\mathrm{G}}(\rho)$ and $\tilde{\mathbf{\Pi}}_{\mathrm{H}}(\rho)$ are corresponding input-output matrices containing transportation costs between two activity cells, $\mathbf{c}_{\mathrm{L}}$ and $\mathbf{c}_{\mathbf{E}}$ are prices of labor and energy per unit, $\tilde{\mathbf{L}}(\rho)$ and $\tilde{\mathbf{E}}(\rho)$ are labor and energy plans and $\mathbf{U}^{\mathrm{T}}$ is row vector of unit values.

3 CALCULATIONS

In this chapter some further numerical calculations are presented. Calculations are further development of numerical example found in [8], [9]. Tables 1, 2 and 3 present technology parameters of activity cells located at individual geographical locations which serve as outflow of elements into recycling sub-system.

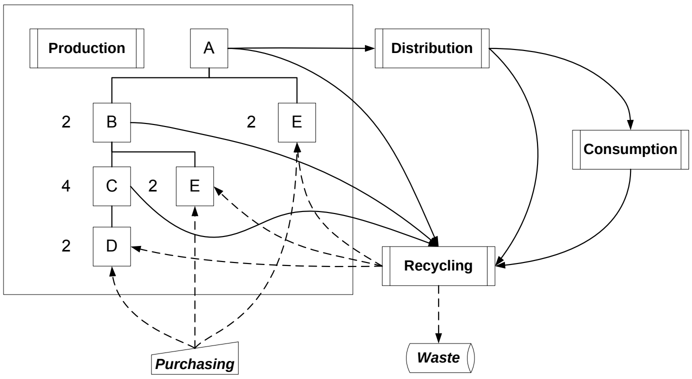


Figure 2: Graphical presentation of numerical example, using bill of materials.

Table 1: Technology parameters of activity cells which serve as exit point from production to recycling.

Elementary item	Sequence		A	$\hat{L}$	γ	$\hat{P}$	δ	$\hat{E}$	φ
D	From C	D, C	0.58	160000	0.31	12500	0.22	260000	0.29
D	From B	D, C, B	0.41	110000	0.34	12500	0.18	190000	0.31
D	From B	C, B	0.35	80000	0.27	2500	0.23	140000	0.28
D	From A	D, C, B, A	0.39	120000	0.29	12500	0.27	180000	0.24
D	From A	C, B, A	0.43	70000	0.28	2500	0.22	150000	0.25
D	From A	B, A	0.56	210000	0.23	1000	0.09	390000	0.26
E	From B	E, B	0.53	60000	0.34	2500	0.23	160000	0.19
E	From A	E, A	0.54	220000	0.24	1000	0.23	170000	0.23
E	From A	E, B, A	0.46	180000	0.21	2500	0.28	310000	0.24
E	From A	B, A	0.57	230000	0.21	1000	0.14	410000	0.25

Table 2: Technology parameters of activity cells which serve as exit point from distribution to recycling.

Elementary item	Sequence		A	$\hat{L}$	γ	$\hat{P}$	δ	$\hat{E}$	φ
D	From A	D, C, B, A	0.38	120000	0.29	12500	0.27	180000	0.24
D	From A	C, B, A	0.41	70000	0.28	2500	0.22	150000	0.25
D	From A	B, A	0.52	210000	0.23	1000	0.09	390000	0.26
E	From A	E, A	0.51	220000	0.24	1000	0.23	170000	0.23
E	From A	E, B, A	0.44	180000	0.21	2500	0.28	310000	0.24
E	From A	B, A	0.54	230000	0.21	1000	0.14	410000	0.25

Table 3: Technology parameters of activity cells which serve as exit point from consumption to recycling.

Elementary item	Sequence		A	$\hat{L}$	γ	$\hat{P}$	δ	$\hat{E}$	φ
D	From A	D, C, B, A	0.37	120000	0.29	12500	0.27	180000	0.24
D	From A	C, B, A	0.39	70000	0.28	2500	0.22	150000	0.25
D	From A	B, A	0.51	210000	0.23	1000	0.09	390000	0.26
E	From A	E, A	0.49	220000	0.24	1000	0.23	170000	0.23
E	From A	E, B, A	0.43	180000	0.21	2500	0.28	310000	0.24
E	From A	B, A	0.53	230000	0.21	1000	0.14	410000	0.25

Using above parameters and equation (4) final return rates α_{kl} of elementary elements D and E can be calculated (Table 4).

Table 4: Final return rates α_{kl} of elementary elements D and E.

n'	*l3*	*l2*	*l1*	2	3	Sum
α_{k4l3}	0.1128	0.0316	0.0059	0.0041	0.0148	0.1693
α_{k5l2}		0.1053	0.0352	0.0255	0.0980	0.2640
α_{k5l1}			0.1617	0.1221	0.4694	0.7532

Finally, using equation (5) Net Present Value of whole system can be calculated:

$$\begin{aligned}\mathrm{NPV} &= \mathbf{p}\left(\breve{\mathbf{G}}(\rho)-\breve{\mathbf{H}}(\rho)\right)\tilde{\mathbf{P}}(\rho)-\hat{\mathbf{K}}\tilde{\mathbf{v}}(\rho)-\mathbf{U}^{\mathrm{T}}\left(\tilde{\mathbf{\Pi}}_{\mathrm{G}}(\rho)+\tilde{\mathbf{\Pi}}_{\mathrm{H}}(\rho)\right)\tilde{\mathbf{P}}(\rho)-\mathbf{c}_{\mathbf{L}}\tilde{\mathbf{L}}(\rho)-\mathbf{c}_{\mathbf{E}}\tilde{\mathbf{E}}(\rho)= \\ &= 2076084.10-101950.19-2140360.58-276507.43-325910.13= \\ &= \text{-768.644,22}\end{aligned}$$

As we can see NPV of the system is negative so it is making loss. It would be possible to increase the total NPV in several ways. For example, we can see that transportation costs are relatively high. There is a high probability that locations of activity cells are not optimal. It might be also possible to increase production part of NPV with a change of recycling activity cells location. Repositioning them to geographic locations with better technology parameters could increase return rates and consequently NPV of the whole system. Such analysis can be a subject of further researches.

References

[1] Bogataj, M., Grubbström, R.W., 2011. On the Representation of Timing for Different Structures within MRP Theory. International Journal of Production Economics, Doi: 10.1016/j.ijpe.2011.04.016

[2] Bogataj, M., Grubbström, R.W., Bogataj, L., 2010. Efficient location of industrial activity cells in a global supply chain. International Journal of Production Economics, Doi: 10.1016/j.ijpe.2010.09.035.

[3] Grubbström, R.W., 2007. Transform Methodology Applied to Some Inventory Problems. Zeitschrift für Betriebswirtschaft, 77(3), pp. 297–324.

[4] Grubbström, R.W., Bogataj, M., Bogataj, L., 2007. A compact representation of distribution and reverse logistics in the value chain. Ekonomska fakulteta, Ljubljana.

[5] Grubbström, R.W., Tang, O., 2000. An Overview of Input-Output Analysis Applied to Production-Inventory Systems. Economic Systems Review, 12, pp. 3-25.

[6] Kovačić, D., Bogataj, L., 2009. Multistage reverse logistics of assembly systems in extended MRP theory. In: Zadnik Stirn, L., Žerovnik, J., Drobne, S., Lisec, A. (eds.), Proceedings of the 10th International Symposium on Operational Research SOR '09 in Slovenia, Nova Gorica, September 23-25, 2009, Slovenian Society Informatika, Section for Operational Research, Ljubljana, pp. 207-218.

[7] Kovačić, D., Bogataj, L., 2009. The reverse logistics in a virtual supply chain. 9th ISIR Summer School on Changing Paradigm for Inventory Management in a Supply Chain Context, 25 - 29 August, 2009, book of abstracts, Karol Adamiecki University of Economics, Katowice, pp. 11-12.

[8] Kovačić, D. , Bogataj, L., 2011. Multistage reverse logistics of assembly systems in extended MRP Theory consisting of all material flows. Central European Journal of Operations Research, DOI: 10.1007/s10100-010-0168-1.

[9] Kovačić, D., Bogataj, M., 2011. Lead times and repeated reverse logistics in extended MRP Theory. 10th ISIR Summer School on Emerging Trends in Inventory Modeling for Service and Manufacturing, 22 – 26 August, 2011, Istanbul, to appear.

[10] Leontief, W., 1966. Input-output economics. New York, Oxford: Oxford University Press.

[11] Orlicky, J.A., 1975. Material Requirements Planning. McGraw-Hill, New York.

[12] Wei, T., 2006. Impact of energy efficiency gains on output and energy use with Cobb-Douglas production function. Energy Policy, 35, pp. 2023-2030.

RISK ASSESSMENT AND MANAGEMENT IN THE MINING INDUSTRY

Marija Kuzmanović[1], **Biljana Panić**[1], **Mirko Vujošević**[1] and **Slobodan Vujić**[2]
[1]University of Belgrade, Faculty of Organizational Sciences
Jove Ilića 154, 11000 Belgrade, Serbia
mari@fon.bg.ac.rs
[2]University of Belgrade, Faculty for Mining and Geology
Djušina 10, 11000 Belgrade, Serbia
vujic@rgf.bg.ac.rs

Abstract: Mining operations represent an activity with plenty of decision problems involving risk and uncertainty. Mining project managers frequently face important decisions regarding the best allocation of scarce resources among mining ventures that are characterized by substantial financial risk and uncertainty. The challenge of sustainable development requires the mining industry to adopt proactive risk management approaches that recognize, integrate and implement the three pillars of social, environmental and economic sustainability. This paper deals with some aspects of risk management in the mining industry, with special emphasis on methods of risk assessment.

Key words: Risk, mining industry, risk assessment, qualitative and quantitative methods.

1 INTRODUCTION

The nature of mining can often present a range of uncertainties—on the extent of environmental impacts, social benefits, economic outcomes, geologic conditions and even political risks. Stakeholders will have different perceptions of uncertainty and the various aspects of mining. Thus, the challenge of sustainable development presents a variety of risks for the mining industry. These need to be considered in light of social, environmental and economic risks to all stakeholders affected by mining – local communities, investors and shareholders, governments, mining companies and so on. These pillars of sustainability are pivotal in understanding risks and the inter-relationships between them.

In applying risk management across the minerals industry, well-credentialed risk management frameworks need to be applied for all aspects of the life cycle, including mining, processing and downstream stewardship of minerals and metals products. Such an approach provides, wherever possible, a transparent risk management approach more likely to have the wide support of stakeholders.

To manage risk effectively, uncertainty and unpredictability must be recognised and, where possible, key information gaps need to be filled to reduce uncertainty. Many techniques, ranging from simple qualitative methods to advanced quantitative methods, are available to help identify and analyze hazards. This paper deals with some aspects of risk management in the mining industry, with special emphasis on methods of risk assessment.

2 MINING PROJECT RISKS

Risk is the probability that a hazard will turn into a disaster. Risk is defined as the combination of the probability (or likelihood) and consequence of an event (or outcome or result of exposure) [4]. This gives rise to the widely used concept of risk:

$$\text{Risk} = \text{Probability} \times \text{Consequence} \tag{1}$$

Risks present in two ways: strategic and operational. *Strategic risks* are those which affect business survival, strategic goals or the long-term sustainability of an operation. Strategic risks also relate to the interdependencies between an operation's activities and the broader business environment. *Operational risks* affect more systematic aspects of a process or operation. They are those risks that can be readily identified as having one or more types of impact and which affect an expected outcome. Different operations and their activity areas will face any number of operational risks. These types of risks are an integral and unavoidable component of business. Both strategic and operational risks need to be recorded and communicated to appropriate levels throughout the organization so that decision makers can effectively allocate resources.

Mining operations face many types of risks including workplace health and safety, environmental, public health and safety, regulatory, production, reputation and financial risk [3]. *Health risks* are those which affect people's health through chronic exposure leading to illness. *Environmental risk* may be defined in two ways: firstly, and more commonly, in terms of the impact of exploration, mining or mineral processing activities on the environment; secondly, in terms of environmental factors which may present a risk to the sustainability of the operation. *Community health risks* address potential impacts from an operation's activities that may affect the health of the local community. This type of risk generally relates to mining processing operations' emissions to air, water or land.

Changes to regulatory, legislative or compliance regimes can pose risks that are among the most challenging for the mining industry. *Production risk* must be managed to control and sustain operational activities or to benefit from an identified opportunity. These risks are identified in areas of the process that impact production volume or product quality and the cost and revenue of the business. These risks are mainly economic in nature, but may be closely associated with non-economic risks. *Risk to reputation* is in some respects a flow-on consequence from most of the other risk categories. *Financial impact or economic consequences* should be evaluated for all risk types identified for an operation's activities. Financial risks need to be assessed as they relate to capital expenditure, schedule, operating cost, production and revenue, all of which have the potential to affect the profitability and net present value of the operation. *Risks associated with closure* and post-closure phases in the mine life cycle cover both economic and non-economic consequence types. These risks are long term in nature. The expectations of the local community, government, landowners, neighbouring property owners and NGOs need to be taken into account.

The environmental and economic risks of mining are generally well identified and managed, but social risks remain a more challenging area for the minerals industry. Social risk can manifest in a variety of ways - through community development, workforce issues and so on. The relationships between social, environmental and economic risks are often not clearly defined or easy to clarify – yet they must be incorporated into risk management to ensure the minerals industry contributes strongly to sustainable development

3 RISK MANAGEMENT PROCESS

3.1 Risk management elements

The main elements of risk management are: (1) communicate and consult, (2) establish the context, (3) risk assessment, (6) treat risks, (7) monitor and review (Fig. 1).

Historically, risk management approaches have focused on the technical aspects of risk management. Contemporary risk approaches now place more emphasis on communication processes at each stage of risk management. It is important for risk practitioners and managers to fully appreciate the relationship between effective risk management, risk communication and the technical risk assessment process.

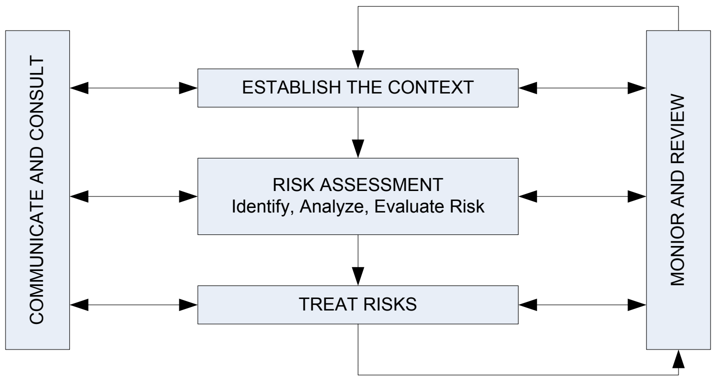


Figure 1: Risk management process

3.2 Applying risk processes to all stages of mine life

Project risks must be identified and managed at all stages of a mine life cycle. The principal stages of a mine development are: (1) exploration and discovery; (2) concept, order of magnitude, pre-feasibility, feasibility, design and project approval; (3) construction and commissioning of mine and mineral processing facilities; (4) operation and production; (5) closure, decommissioning and rehabilitation; (6) post-closure monitoring.

Each stage presents significant challenges for the mining company from a risk management perspective. It is essential that a mining/processing operation evaluates its technical risks wherever possible. Many mining companies today employ a risk process to help identify and weigh operational risks, before allocating resources to address priority risks. Listening to the questions and concerns a broad range of stakeholders will help the mining company define the scope of the risk management process.

If potentially affected stakeholders have concerns that are not adequately addressed by the mining company, then relationships will be damaged—often for the long term. Once trust is lost, it may be very difficult to regain. Where a significant risk has been identified, key questions that must be asked by the assigned risk manager are [3]:

- Which stakeholders could be affected or are likely to have an interest in the risk?
- Which stakeholder groups could perceive the risk differently?
- What are the stakeholder expectations and understandings in relation to the risk?
- What are the legal requirements?
- What technical solutions to risk minimization can be employed?
- What risk communication strategies should be applied for the various stakeholders?

These questions are applicable to all stages in the mine life cycle. Stakeholders are more likely to accept that significant risks are well managed if they are defined, communicated, understood and satisfactorily addressed early in the mine life cycle. This can only occur if those stakeholders are identified in the first place. Some types of stakeholders have an interest in the mining project throughout all stages (landowners or neighbouring property owners); some other stakeholder groups may only have an interest for one or two stages of the mine life. For example, employees and their families, suppliers and customers have a direct interest in the mine during its operating life.

Effectively applying risk management principles early lays the foundation for good relationships throughout the whole mine life cycle. There are many examples of relationships being damaged at the exploration/discovery stage or during mine feasibility. This creates difficulties for stakeholder relationships that can carry through to the construction, operational and closure phases of mining and may require significant additional management effort, delay project start-up or adversely affect the life of the mine. As technical solutions to risks are planned and implemented, the effectiveness of these solutions should be canvassed among stakeholders in order to build and retain confidence in the risk management process.

4 RISK ASSESSMENT: IDENTIFICATION, ANALYSIS AND EVALUATION

Risk assessment is process of analysis to identify and measure risks from natural hazards that affect people, property and the environment. This process can also encompass the assessment of available resources to address the risks. Recently, a considerable expansion has occurred in the strategic business importance placed on risk assessment and on the level of effort put into *risk identification, analysis* and *ongoing management* [2]. Consequently, a range of risk management methods and expertise is available to the mining industry.

The aim of the *risk identification process* is to understand all the key risk events that are relevant to a project, activity, or other situational context; define their cause-and-effect relationships; identify the nature and extent of all potential consequences (for example, financial, environmental, social, economic, safety); and understand their likelihood for occurring. The objective of *risk analysis* is to produce outputs that can be used to evaluate the nature and distribution of risk, and to develop appropriate strategies to manage the risk.

Many techniques, ranging from simple qualitative methods to advanced quantitative methods, are available to help identify and analyze hazards. Some of the more commonly used techniques include [1]: preliminary hazard analysis (PHA), failure modes and effects analysis (FMEA), hazard and operability studies (HAZOP), fault-tree analysis (FTA), and event-tree analysis (ETA). Considerations in analyzing risk include: (1) Investigating the frequency of particular types of disasters (often versus seldom); (2) Determining the degree of predictability of the disaster; (3) Analyzing the speed of onset of the disaster (sudden versus gradual); (4) Determining the amount of forewarning associated with the disaster; (5) Estimating the duration of the disaster.

4.1 Qualitative vs. quantitative risk assessment methods

A risk assessment methods should be chosen, not just based on the hazard, but also after consideration of the capabilities of each technique as each may provide different outputs (or parameters) that may be particularly useful towards the solution of the problem. Such outputs could be simple lists of individual failures, or numerical estimates of system failure probabilities, listings of event scenarios and their likelihoods or numerical system failure probabilities and sensitivities to input variables.

Qualitative risk assessment techniques use descriptive terms to define the likelihoods and consequences of risk events. The methods are quick and relatively easy to use and provide a general understanding of comparative risk between risk events. Outputs from qualitative risk analyses are usually evaluated using a risk matrix format (see Fig. 2a). The risk matrix incorporates the pre-determined risk acceptance threshold and is used to determine which risks require treatment and the priorities that should be applied. Using the matrix, a risk rating for a given risk event can be selected by reading across and down the matrix using the assigned likelihood and consequence descriptors. This type of matrix is typically used to compare risk levels for different events and to set priorities for risk treatment actions.

Qualitative approaches are best used as a quick, first-pass exercise where there are many, complex risk issues and low-risk issues need to be screened out for practical purposes. However, they have some shortcomings compared with more quantitative approaches. Key criticisms are that they are imprecise, it is difficult to compare events on a common basis, there is rarely clear justification of weightings placed on severity of consequences and the use of emotive labels makes it difficult for risk communicators to openly present risk assessment findings to stakeholders. Furthermore, the outputs from qualitative approaches are difficult to incorporate into financial business considerations.

		Consequence level				
		1	2	3	4	5
Likelihood level	Descriptor	Insignificant	Minor	Moderate	Major	Catastrophic
A	Almost certain	A1	A2	A3	A4	A5
B	Likely	B1	B2	B3	B4	B5
C	Possible	C1	C2	C3	C4	C5
D	Unlikely	D1	D2	D3	D4	D5
E	Rare	E1	E2	E3	E4	E5

Risk rating
Extreme
High
Moderate
Low

		Consequence level				
		1	2	3	4	5
Likelihood level	Descriptor	Insignificant	Minor	Moderate	Major	Catastrophic
5	Almost certain	5	10	15	20	25
4	Likely	4	8	12	16	20
3	Possible	3	6	9	12	15
2	Unlikely	2	4	6	8	10
1	Rare	1	2	3	4	5

Risk rating
Extreme
High
Moderate
Low

Figure 2: Example of: a) qualitative risk matrix, b) basic semi-quantitative risk rating matrix [3]

Semi-quantitative approaches to risk assessment are currently widely used to overcome some of the shortcomings associated with qualitative approaches. Semi-quantitative risk assessments provide a more detailed, prioritized ranking of risks than the outcome of qualitative risk assessments. It takes the qualitative approach a step further by attributing values or multipliers to the likelihood and consequence groupings and may involve multiplication of frequency levels with a numerical ranking of consequence (Fig. 2b).

An advantage of this approach is that it allows risk ratings to be set based on the derived numeric risk values. A major drawback is that the numeric risk values may not reasonably reflect the relative risk of events, due to possible orders of magnitude differences within the likelihoods and consequences classes.

Quantitative risk assessment is being increasingly applied in the mining and minerals industry due to increasing business requirements to support financial decisions; make comparisons across financial, environmental and social risk profiles; and to demonstrate transparency, consistency and logic of approach. However, the application of quantitative risk outputs, which are often not immediately intuitive, requires some up-front learning investment by decision makers.

Quantitative risk assessment is used across the full range of risk applications – from deriving preliminary, first-pass separation of risk events to much more comprehensive assessments. The comprehensive assessments can derive detailed risk profiles for priority ranking, estimates of the costs that may be incurred due to risk events, input to financial models, and a basis for cost-benefit analysis.

Quantitative risk assessment follows the basic risk assessment approach to its full extent by attributing absolute values to likelihood and consequences. Estimates of *likelihood* are made in terms of event frequency (annual or frequency over the period of a specified project) and/or probability of occurrence of the risk event. Estimates of *consequence* can be made using any consistent measure, selected according to the nature of the application. For example, dams engineering risk assessments often measure risk in terms of frequency and possible number of lives lost, while financial controllers may measure risk in terms of frequency and cost. Where consequences are expressed in financial terms, the risk quotient is equivalent to the commonly used term ‘expected cost’ or ‘expected value’.

Quantitative methods overcome some of the shortcomings associated with qualitative approaches and are ideally applied where system safety and criticality is to be assessed. Catastrophic risks can be assessed using *Fault Tree and Event Tree methods*, ideally as part of a *probabilistic risk analysis (PRA)* provided reliable input data exists and numerical results can be used in estimating the likely range of risks to both employees, plant, society and the environment. Results can also be used in cost benefit studies and demonstration that risks are "as low as reasonably practicable" (*ALARP*) can be supported defensively only by quantitative analysis. Truly probabilistic methods such as "first order reliability methods" (*FORM*) are the most complex type of risk analysis, and its advantage over any other method is the ability not only to successfully cope with the statistical uncertainty in the data but also use it to its advantage. Results from a FORM evaluation also provide further information on system vulnerability as a function of input variables. Analysis utilizing *Monte Carlo* (MC) techniques is more commonplace than FORM and has found widespread acceptance in many professions such as engineering and finance. However, MC methods lack some of the direct leverage that FORM provides for engineering solutions.

5 CONCLUSION

The layout of many mining operations is far from static and changes continuously. All mines and operations are exposed to the danger of fire and explosions, with underground mines, particularly coal mines being extremely vulnerable and endangered by the effects offices and explosions. Therefore, it is necessary to assess the risk from different mining operations and take cost effective suitable measures to prevent, eliminate and minimize risk.

Hazard identification usually establishes what risk assessment techniques should bemused and care needs to be taken in the selection of a technique, as similar techniques may not necessarily yield the same results. Risk assessment techniques may be either qualitative or quantitative and both streams, ideally in combination, can be very effective in the process of hazards management.

The mining industry is now familiar with the use of qualitative techniques, but there is a growing recognition of the value and effectiveness of quantitative studies particularly when assessing system hazards. However, use of quantitative techniques will require a more disciplined approach to recording and interpreting incident, accident and maintenance information to provide accurate and auditable inputs to those studies. Successful application of quantitative risk assessment is however dependent on the availability of necessary data, and the capacity and commitment of the organisation to manage the process and to source.

References

[1] Ericson, C. A. II, Hazard analysis techniques for system safety, Wiley, New Jersey 2005.

[2] Jafari, J., Khorasani, N., Danehkar, A., 2010. Ecological Risk Assessment of Lead (Pb) after Waste Disposal from Metallurgical Industries, Research Journal of Environmental and Earth Sciences, 2(3), pp. 139-145.

[3] Risk Assessment and Management. Leading Practice Sustainable Development Program for the Mining Industry, Australian Government Department of Resources, Energy and Tourism, 2008.

[4] Vujošević, M., 2008. Risk Management as a Part of the Integrated Management System, International Journal "Total Quality Management & Excellence", Vol. 36, No. 4.

TEAMWORK IN SIMULTANEOUS PRODUCT REALIZATION PROCESS

Lidija Rihar[1]**, Stanislav Gorenc**[2]**, Janez Kušar**[1]**,** and **Marko. Starbek**[1]
[1]University of Ljubljana Faculty of Mechanical Engineering, Aškerčeva 6, Ljubljana, Slovenia
(lidija.rihar@fs.uni.lj.si, janez.kusar@fs.uni-lj.si, marko.starbek@fs.uni-lj.si)
[2]PLASTA d.o.o., Šentrupert 35, 8232 Šentrupert
(stane.gorenc@gmail.com)

Abstract: The paper presents transition from sequential to simultaneous product realization. Such a transition is not possible without prior well-organised teamwork or virtual teamwork.
The article demonstrates the team structure in simultaneous product realization. A two-level team structure, with a core team on the first level and several project teams in simultaneous product realization loops on the second level.
The results of organising teamwork and virtual teamwork are shown on a case study of simultaneous realization of a pedal assembly.

Keywords: Virtual team, product realization, loops, communication matrix, pedal assembly

1 INTRODUCTION

The essence of modern production is to make a product that a customer needs, as quickly and as cheaply as possible.

Under these conditions, only a company that can provide customers with the right products in terms of functionality and quality, produced at the right time, in the right location, of required quality and at an acceptable price, can expect global market success. A product that is not produced in accordance with the wishes and requirements of customers, which hits the market too late and/or is too expensive, will not survive competitive pressure [1]. The customers should therefore participate in the process of simultaneous realization of a product as early as possible. They can participate by expressing their wishes and requirements regarding project definition. The customer should be a temporary member of project teams in simultaneous product realization loops.

The main feature of sequential product realization is the sequential execution of stages in the product realization process [2]. The observed stage of the product realization process can only begin after the preceding stage has been completed. Data on the observed process stage are built gradually and are completed at the end of the stage. The data are then forwarded to the next stage (Figure 1).

In contrast with sequential product realization, the main feature of simultaneous product realization is the concurrent execution of stages in the product realization process [2]. In this case, the observed stage can begin before the preceding stage has been completed. Data on the observed process stage are collected gradually and are forwarded continuously to the next stage (Figure 1).

A transition from sequential to simultaneous product realization considerably reduces the time and costs of product realization [3].

The cost of product definition is much higher in simultaneous product realization, due to the parallel execution of activities (more work is done during this stage), while production costs are much lower than in sequential realization, due to short iteration loops for carrying out changes and eliminating errors.

In simultaneous product realization, there are interactions between individual stages of the product realization process. Track-and-loop technology has been developed for executing these interactions [1]. The type of loop defines the type of co-operation between overlapping stages of the simultaneous product realization process. Winner [4] suggests that 3-T loops should be used where interactions exist between three levels of a simultaneous product realization process [3,4].

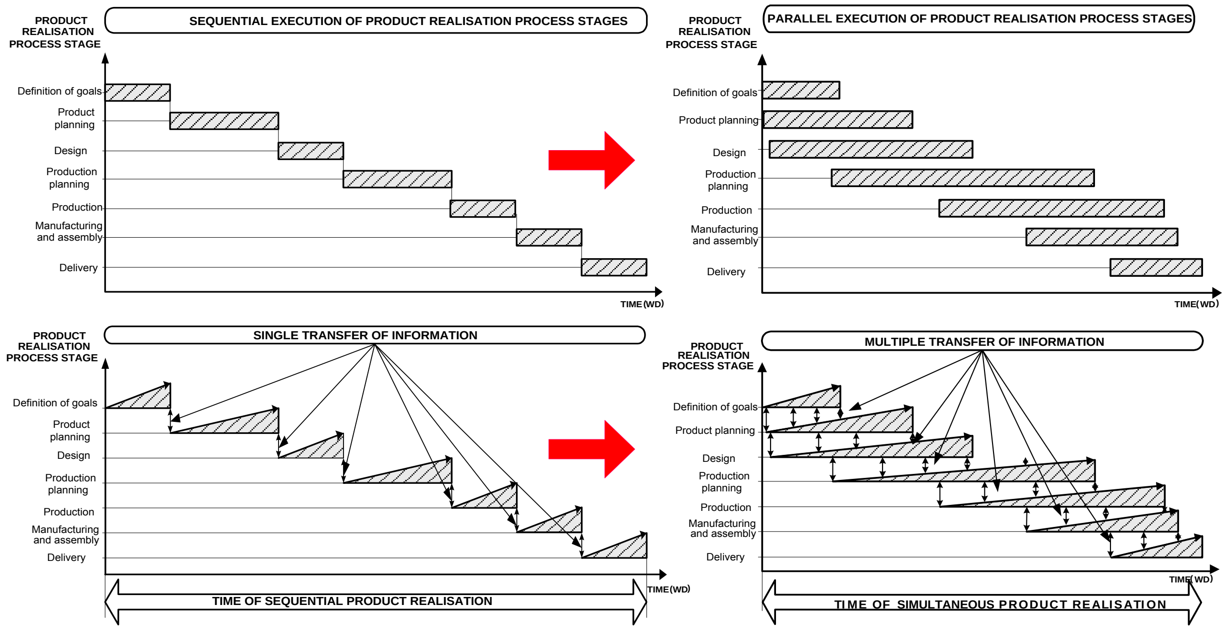


Figure 1: Sequential product realization and Simultaneous product realization

A transformation of input into output is made in every loop, on the basis of requirements and restrictions [2,3].

A two-level team structure is planned for execution of 3-T loops of a simultaneous product realization process [3,5], with a variable structure of core and project teams, for all simultaneous loops.

The task of the core team is process support and control, while the task of (virtual) project teams is execution of the tasks defined within the simultaneous product realization process.

It is obvious that simultaneous product realization is not possible without well-organized teamwork or virtual teamwork, which is the means for organization integration.

It incorporates:

- the formation of a core team, project teams or virtual project teams in product realization loops,
- the selection of communication tools for the core team, project teams or virtual project teams,
- the definition of a communication matrix.

Teamwork is a precondition for transition to simultaneous product realization.

2 SIMULTANEOUS REALIZATION OF A PEDAL ASSEMBLY

The goal of the project was to make a competitive pedal assembly, suitable in terms of quality, reliability, mass, price and realization time.

Simultaneous realization of the pedal assembly was divided into six stages and five loops of simultaneous realization of the pedal assembly within the six stages of pedal assembly realization and five loops as shown on figure 2.

Figure 2 shows how the loops are formed, and the type of cooperation within realization stages.

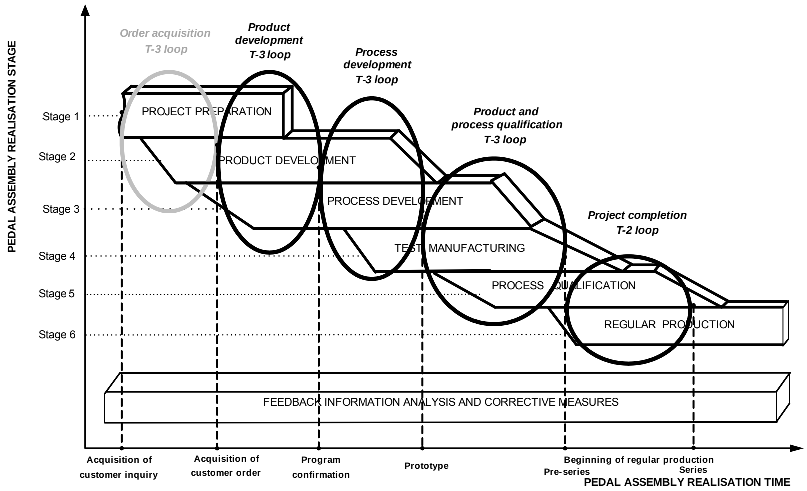


Figure 2: Stages and loops of simultaneous realization of pedal assembly

2.1 Forming Teams / Virtual Teams for Realization of Pedal Assembly

After seeing the presentation of two- and three-level structures of (virtual) teams in product realization loops [3,5] the company management selected a two-level team structure, whereby the core team is on the first level and five virtual project teams are on the second level (Figure 3).

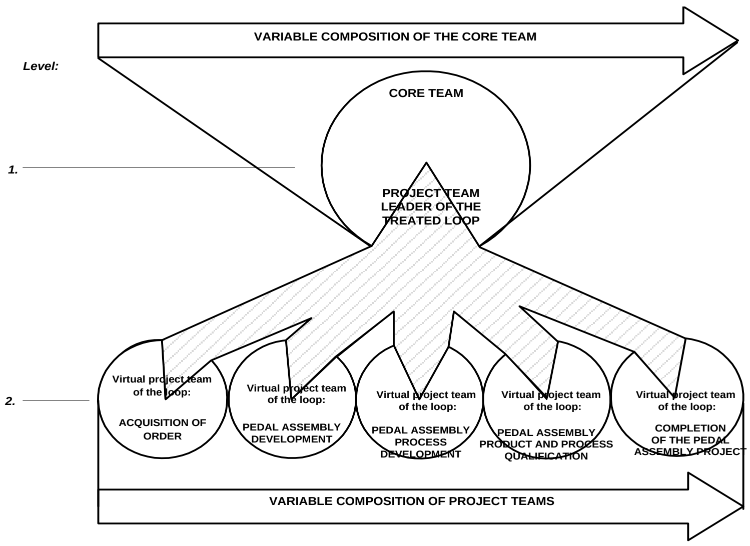


Figure 3: Structure of teams for simultaneous realization of pedal assembly

2.2 Forming the Core Team

The core team for simultaneous realization of the pedal assembly will monitor the whole project, solve organisational issues and coordinate the strategy of performing tasks. The company management decided that the following people would be members of the core team:

- project manager (PM) - permanent member,
- project team leader of a particular loop (VPL) - non-permanent member,
- heads of all company departments - permanent members

2.3 Forming Virtual Project Teams for the Loops of Simultaneous Realization of the Pedal Assembly

As shown in Figure 3, there will be five virtual project teams in loops of simultaneous realization of the pedal assembly.

Members of virtual teams will be experts from all company departments and two representatives from strategic suppliers and customers, depending on the level of assigned responsibility for execution of activities within a particular loop.

When the company obtains an offer, loop 1 activities (Order acquisition loop) are started; its three stages are: project preparation, development of the pedal assembly and development of the pedal assembly process. This loop is executed when the sales department considers that it is sensible to make an offer for the realization of the pedal assembly.

Loop 1 is followed by loops 2, 3, 4 and 5.

The project manager decided (in agreement with the company management) that the intensity of responsibility of each virtual team member during the execution of activities would be marked by a 1(member is informed) -3(member participates) – 9(member has responsibility) method.

A creativity workshop was organised with 14 representatives from company departments, as well as representatives from suppliers and customers. The goal of the workshop was to score the intensity of responsibility of virtual team members when executing the activities of the five loops in simultaneous realization of the pedal assembly.

The results of scoring the intensity of responsibility of virtual team members during execution of the first loop of simultaneous realization of the pedal assembly are presented in Table 1.

Table 1: Scoring the intensity of responsibility of virtual team members in the "Order acquisition loop"

ID	Name	1. MNG	2.S	3.PM	4.DEV	5.IND	6.Q	7.MTD	8.SIF	9.PUR	10.MC	11.MP	12.MNT	13.AD	14.LD	15.SUP	16.CUS
0	**SIMULTANEOUS REALISATION OF PEDAL ASSEMBLY**																
1	**Offer preparation project activities**																
2	**Preparation of the project**																
3	**Inquiry review**																
4	Receipt of inquiries and determination of their status		9														3
5	Review of requirements for completeness		9		3												
6	Opening of the project		9	1	1	1	1		1	1					1		
7	Preparation of inquiry		9	1	1	1	1		1	1					1		
8	Definition of project team		9	1	1	1	1		1	1					1		
9	Repartition draft																
10	Initial meeting 1 (inquiry)		9	3	3	3	3		3	3					3		
	…																
282	**SUM**	3	165	19	163	104	61	9	84	87	9	10	0	0	11	10	30

…

It can be seen from the table what the responsibilities of each virtual team member for the execution of activities in the first loop of pedal assembly realization are.

The procedure of scoring the intensity of responsibility of virtual team members was also carried out for the other loops.

From the sum of points assigned to the i-th team member during execution of activity in the j-th loop, a factor of total intensity of responsibility of the i-th member in the j-th loop can be calculated as (1):

$$FTI_{i,j} = \frac{SMP_{i,j}}{SAP_j} \tag{1}$$

$FTI_{i,j}$ factor of total intensity of responsibility of the i-th team member in the j-th loop

$SMP_{i,j}$ sum of the points assigned to the i-th member in the j-th loop

SAP_j sum of all points assigned in the j-th loop

The results of the calculation of the total intensity of responsibility factor of virtual project team members during execution of activities in all five loops of simultaneous realization of pedal assembly are shown in Table 2.

Table 2: Factors of total intensity of responsibility of virtual project team members during execution of loops of "Order acquisition loop"

REALIZATION OF PEDAL ASSEMBLY LOOPS	*VIRTUAL TEAM MEMBERS*	*MNG*	*S*	*PM*	*DEV*	*IND*	*Q*	*MTD*	*SIF*	*PUR*	*MC*	*MP*	*MNT*	*AD*	*LD*	*SUP*	*CUS*	*SUM*
Loop 1: PREPARATION OF ORDER	*SCORING OF INDIVIDUAL TEAM MEMBERS IN LOOP 1*	3	165	19	163	104	61	9	84	87	9	10	0	0	11	10	30	*765*
	Intensity factor of individual team member	0.39	21.5	2.48	21.3	13.6	7.97	1.19	11.2	11.5	1.19	1.33	0	0	1.46	1.33	3.98	*100*
	SELECTED TEAM MEMBERS IN LOOP 1		165		163	104	61		84	87							30	*694*
	Intensity factor of the selected team member	0	23.8		23.4	14.9	8.79		12.1	12.5							4.32	*100*

After they had made an overview of the total intensity of responsibility factors of virtual team members during execution of activities in the loops of pedal assembly realization, the creativity workshop participants reached the following conclusions:

- the i-th member of the virtual project team (VPT) of the j-th loop of realization of the pedal assembly, with the maximum factor of total intensity of responsibility, would be appointed as a team leader of the j-th loop of PTL,
- representatives from departments with a total intensity of responsibility factor above 5% would also be included in the j-th loop of pedal assembly realization,
- representatives of suppliers and customers would also be included in the j-th loop of pedal assembly realization, regardless of their total intensity of responsibility factor in order to avoid misunderstanding suppliers' and customers' requirements.

Figure 4 presents the structure of virtual project teams of five loops and core team members in simultaneous realization of the pedal assembly.

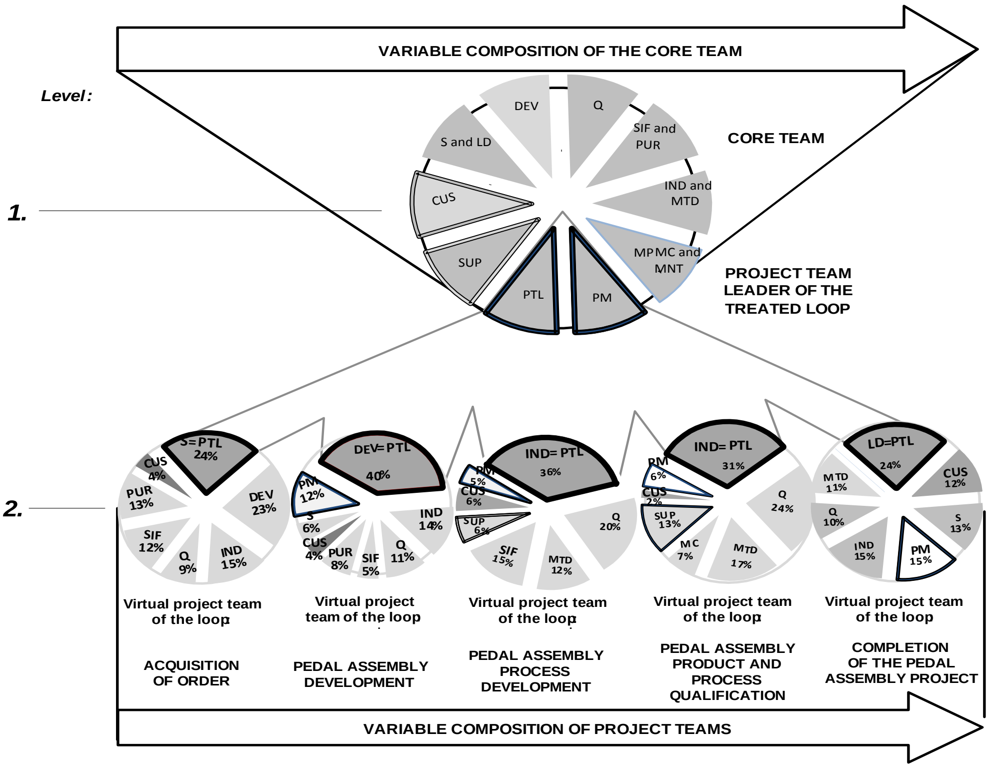


Figure 4: Virtual project teams in the loops of simultaneous realization of the pedal assembly

2.4 Communication Matrix in Product Realization Loops

The communication matrix defines the method of exchanging information and documents in the execution of simultaneous product realization activity loops.

A list must be made for every activity:

- input information with required documents for beginning execution of the activity,
- output information with required documents that arise from execution of the activity,
- tools for creating and storing information,
- sender of the information or document,
- receiver of the information or document,
- communication tool used for information exchange.

The communication matrix defines in advance the mode of information exchange and communication tools required.

3 CONCLUSION

The paper emphasises that simultaneous product realization is not possible without well-organised teamwork or virtual teamwork.

A two-level team structure of a track-and-loop process of simultaneous product realization, suitable for small companies, is presented. An overview is given of available communication tools for teamwork/virtual teamwork, with the advantages and drawbacks of individual tools. The content of the communication matrix of simultaneous product realization is formed, defining the exchange of information/documents in the execution of simultaneous product realization activity loops.

The suggested methodology of forming teams or virtual teams and communication matrix of simultaneous product realization was tested on a study case of a pedal assembly.

Further work on solving simultaneous product realization problems will be focused on making a catalogue of the entire simultaneous product realization process using ARIS—a tool for process modelling and reengineering [6].

References

[1] Dickman, P. (2009): Schlanker Materialfluss, Springer-Verlag, Berlin - Heidelberg.

[2] Prasad, B. (1996) Concurrent Engineering Fundamentals, Volume I, Integrated Product and Proces Organization, New Jersey, Printice Hall PTR, pp 216-276.

[3] Kušar, J., Duhovnik J., Grum J., Starbek, M. (2004): How to reduce new product development time. Robot. comput.-integr. manuf. Vol. 20, No. 1, pp. 1-15.

[4] Winner, R. I. (1988): The Role of Concurrent Engineering in Weapons System Acquisition, IDA Report R-338, Alexandrija, VA: Institut for Defence Analysis.

[5] Rihar Lidija, Kušar Janez, Duhovnik Jožef, Starbek Marko (2010): Teamwork as Precondition for Simultaneous Product realisation, Concurrent engineering: Research and Applications, Vol.18, No.4, pp 261-273.

[6] Scheer, A.W. (1999): ARIS – Business Process Modeling, Springer-Verlag, Berlin – Heidelberg.

AN AXIOMATIC APPROACH TO THE GENERALISED EVALUATION PROCEDURE

Viljem Rupnik
INTERACTA Ltd, Information Processing Engineering
SI-1000 Ljubljana, Parmova 53, Slovenia
Viljem.Rupnik@siol.net

Abstract: Existing valuing business practice shows a series of handicaps, open problems and ambiguities. Therefore the corresponding theory is not consistent as well as powerful enough to meet the demand of broader class of objects. Proposing an evaluation procedure capable to deal with both a set of arbitrary type of objects as well as arbitrary characteristics of objects helps us not only to put the existing business appraisal practice to upper level but also to extend decision theory towards non-price economy. A bridge between appraisal and evaluation is given.

Keywords: appraisal versus evaluation, non-price economy, extended decision theory.

1 PROBLEM STATEMENT

A notion of value usually stems from economics, having been treated in terms of monetary unit. There, narrow limits of value, mostly experienced by financial theory and practice, have been surpassed quite a time ago. Valuing business, for example, very often encounters

- different determinants not being able to be expressed by monetary unit; there are spheres (like production, purchasing, distribution, marketing, ect.) and strata (like technology, organization, information, ect.) which bring non-monetary determinants influencing e.g. the present value of a firm;
- when a firm is treated within the macro-economic framework, nation-wide economic policy brings a variety of methodological embarrassments into the valuing procedure;
- the last, but not the least important group of determinants is affiliated with nondeterministic spaces of valuing.

The three groups from above of non- monetary value-determinants make us to conceive a vectorial concept of value as will be shown below.

2 ASSUMPTIONS

Assumption 1: each determinant is quantifiable; it permits all kinds of quantification.

Assumption 2: a set of determinants of generalised value is arbitrary but finite; this assumption opens the door to any concept of value.

Assumption 3: monetary expressed determinant is a partial component of a set of determinants; consequently, it makes traditional meaure of value competitive to any of the determinants added.

Assumption 4: a procedure of general evaluation is independent of each determinant space; it is crucial requirement towards the approach being generalised.

Assumption 5: An object being valued is arbitrary; this assumption opens the door to valuing of any kind of objects.

Assumption 6: a procedure of generalised evaluation should enable the procedures of comparison, aggregation and optimization of generalised value.

3 OVERALL DESCRIPTION OF THE APPROACH

From the assumptions above we, a first, derive the possibilty of suppression the limits to the number of components (determinants) of the generalised value.These limits are usually a snapshot of subjectivism usually induced by appraisers as well the nature of object appraised. Consequently,an endless series of traditional concepts is generated and thus no consistent theory of evaluation is possible.

Each object we want to evalute is described by the finite but arbitrary number of components being chosen at the beginning of evaluation procedure. Next, we add any arbitrary but finite number of other objects of the same kind and describe by the corresponding values of the same components as chosen at the beginning of evaluation procedure. Thus, the input to the whole procedure is finite dimensional matrix ready for implementation of non-formal modelling (see [1]).

To summarize the properties of evaluation procedure rendered by non-formal modelling we are to review its fundamental features, whereas the details are given in [2]. A particular object is evaluated in general sense of meaning by computing its relative rank with respect to the rest of »competitive« objects from the same family (which is one dimension of its valuing procedure) and this rank being conditioned by all components chosen in advance (which is another dimension of its valuing procedure). Thus, if we have *n* components and *m* competitive objects, we might dare to speak about (n x m) - dimensional ranking, although n-dimensional ranking should be understood as a kind of pseudo-ranking.

As it can be seen from [2], the procedure proposed fully satisfies the requirements on comparison, aggregation and optimization of generalised value.

4 INTERPRETATION OF TRADITIONAL VALUE THROUGH ITS GENERALISED EVALUATION PROCEDURE

According to [3] the traditional value of a given object is associated with its global measure of quality (through its (n x m) - dimensional rank). Thus its traditional value is shadowed by its rank, which shows its position among the traditional values of other objects. This shadow is being produced by all remaining components simultaneously; let us call them *shadowing componen*ts. If there are no additional components but the traditional one, than there is no shadow of traditional value.

Since, in practice, we usually badly need an answer on what component or more of them have the strongest impact on the shadowed value of the particular object. Using *filtering procedure* applied on the set of components we can explore such impacts of different subgroups of components. Most interesting question is what is the impact of individual component on shadowed value of particular object valued. Simulated filtering procedure, starting from the initial set of components for example, computes the change of shadowed value caused by omitting some particular component from the initial set of components. If the change is positive, we eliminate a component in question; and vice versa. We can also act in the opposite order by adding components to the initial single one, e.i. a traditional value (so called *clustering procedure*).

Both procedures give us an insight into the information gain or loss accruing from generalized evaluation procedure in comparison with traditional (scalar) evaluation one. Either filtering or clustering procedures help us to detect an/the optimal set of components for a given type of objects provided the maximum set of components are being decided upon. A shadowed value, being a global measure of quality, at its maximum level possible justifies the content of »overlooked«, but useful collateral components in evaluation procedure.

Next benefit from the generalized approach to evaluation comes from reference objects which enter the whole procedure. Their scope, content and corresponding collateral components influence the global quality measure (shadowed value) of the object being evaluated as a central point of interest. They can therefore act as an environemental condition for the object being evaluated. According to [2] all reference values must be of the same category, like market value, fair value, liquidation value, ect. If we decide upon some representative reference value, like for instance a market value for some subset of reference objects, then the magnitude of shadowed value of a given object being evaluated may be higher/lower then the correponding magnitude of shadowed value belonging to representative reference object. Thus we reveal whether the value of object evaluated is puffed up or withered, even in the case when nominal levels of the two values are the same (here components play their role). The difference between the two levels of shadowed value is a measure of competitiveness of any object with respect to the representative object.

5 HYPOTHETICAL EXAMPLE

Let us have 5 objects from the same family: O_1, …O_5, where traditional value (component) we let to be market price C_0 and additional components be C_1, …C_4 representing in turn: durability in years, cost of maintainance, weight and operating cost. The input data therefore are

objects	C_0	C_1	C_2	C_3	C_4
O_1	20	6	129	35	54
O_2	26	8	153	49	76
O_3	18	5	186	70	56
O_4	22	9	114	27	96
O_5	25	7	112	45	47

Applying [3] over all 4+1 components brings

objects	C_0	shadowed C_0
O_1	20	0,72
O_2	26	0,83
O_3	18	0,59
O_4	22	0,92
O_5	25	0,67

which stands for O_4 as the best choice despite its market value is higher than the correponding value of O_3 which would appear as the best choice when supressing all shadowing components.

What is the effect of C_4 eliminated? Applying the same method as before we get

Objects	C_0	shadowed C_0
O_1	20	0,72
O_2	26	0,83
O_3	18	0,88
O_4	22	0,79
O_5	25	0,67

which means that elimination of C_4 decreases the shadowed market price and shifts to the O_3 as the best choice. Consequently, we have to keep C_4 within the »team« of shadowing components. From this case we see that traditional way of evaluation (decision on O_3) coincides with that belonging to keeping C_4 as shadowing component.

What is the effect of C_3 eliminated? Following the same procedure we get

objects	C_0	shadowed C_0
O_1	20	0,72
O_2	26	0,96
O_3	18	0,57
O_4	22	0,79
O_5	25	0,67

which recommends us C_3 should be filtered out from generalised evaluation precedure, since O_2 has the highest global measure of quality, e. i. the highest shadow of C_0. It happens despite of its C_0 being quite high above that C_0 which might be our decision without shadowing components. Continuing this procedure of filtering we come to the knowledge of C_1 and C_2 being superfluous or not.

6 APPLICATION AREAS

- When reviewing all existing approaches to business appraisal of firms, the metod proposed offers the insight into benefit or loss of the value appraised following traditional concepts. Thus, for example, choosing some representative firm and »similar« firm actually being purchased, we are able to estimate how far the »fair value« from the representative firm really is.
- The method is useful to govern bargaining procedure between the two opponent partners.
- Due to the six assumptions above we can extend the method over a set of objects which do not belong to the same family, provided that there is the same set of shadowing components. Thus, for example, bikes can be evaluated together with cars; or furs and jewelry, ect.
- An outstanding feature of the method proposed refers to the fact that objects may be subject to the evaluation procedure even if the traditional value has no monetary dimension. Consequently, any of the shadowing components might be declared as traditional criterion (like C_0). Thus we came closer to decision theory as it stands today and, furthermore, thus we laid down a ground of non-price economy.
- The pocedure above looks promising in the field of land rent pricing.

References:

[1] Rupnik V., An Attemt to Non-formal Modelling, The IVth International Symposium on Operations Research, Preddvor, Slovenia, Oct 1-3, 1997, pp. 207-214.

[2] Rupnik V., Multidimensional Business Evaluation, 11the International Conference on Operational Research, KOI 2006, Pula, Croatia, Sept 27-29, 2006, pp.73-79.

[3] Rupnik V., The shadowed decisions; IV. International Symposium on Operations Research, Preddvor, Oct 1-3, 1997, pp. 139-144.

COMPUTER MANIPULATION METHODS, THEIRS DETECTING, REPORTING AND SANCTIONING

Slavko Šimundić [1], **Siniša Franjić** [2] and **Danijel Barbarić** [3]
[1] Pravni fakultet Split, Domovinskog rata 8, 21000 Split, Republic of Croatia
[2] Elektrotehnički fakultet Osijek, Kneza Trpimira 2 b, 31000 Osijek, Republic of Croatia
[3] Pravni fakultet Split, Domovinskog rata 8, 21000 Split, Republic of Croatia
E-mail: sinisa.franjic@etfos.hr

Abstract: Modern information and computer technology has made a series of socially unacceptable behavior that is needed to adequately criminalized. This was done in Croatia on several occasions through the Law on Amendments to the Penal Code in which they built the Convention on Cybercrime. The Convention on Cybercrime, describes a large number of crimes in the area of information and computer technologies, however, this article deals with the elaboration of the most widespread manifestations of them, and that they are crimes which are under the common denominator- the methods of manipulating computers, though, regarding the prevalence, or other crimes are not negligible.

Keywords: Law on Amendments to the Criminal Law, Convention on Cybercrime, Computer System, Computer Crime, Unauthorised hackers intrusion, Criminal act

1 INTRODUCTION

The eighties and nineties of the twentieth century brought a wealth of useful information in all spheres of life around the world because it is precisely in that year there was a sudden expansion of modern computer technology. Then there was the internet phenomenon that people from one end of the world associated with the other for a few seconds. Like any new technological developments, and the internet has brought a number of new forms of socially unacceptable behavior which is in the national legislation should criminalize adequately. National legislation were initially found unprepared for the fight against socially unacceptable behavior that has brought the internet. However, the situation is changing and today more and more countries have incorporated in their legislation offenses arising from the use of modern information and computer technologies. Croatia has on several occasions through the Law on Amendments to the Criminal Code [1] prescribes a series of offenses described in the Convention on Cybercrime [2]. This article analyzes the illicit social behavior from Article 223 Croatian Criminal Code of the Republic, which is the same as "Violation of the confidentiality, integrity and availability of computer data, programs and systems." This means that in Article 223 Criminal law is directly incorporated provisions of the Convention on Cybercrime.

2 METHODS OF MANIPULATING COMPUTERS

Computers can appear as an object of attack or as a means of attack [3]. The most common types of attacks are:

2.1 Unauthorized access to computer systems (hacking)

Intrusion into the computer system does not necessarily include elements criminal act. For example, the intrusion into it can happen randomly and out of pure curiosity. It is difficult to accept that someone would believe that the intruder the purpose of his entrance to the dwelling only sightseeing.

2.2 Viruses infection of systems

Virus infection is a major problem today. Fear of virus infection is present by many peoples who have open systems. The virus often called programs that have less homework to make certain manipulation on the computer system. Initially they were created idle developers. However, they were not created just bored programmers. Viruses are often spread when using unauthorized (pirated) software but in other situations in whose using mediums that had been in contact with other compatible systems or via telecommunications came into contact with other systems.

2.3 Manipulation of data

Manipulating data in a computer system is possible when collecting and entering data into the system, and possibly later. The objective of such manipulation is usually the acquisition of the illegal gain. For example, increasing the (correction) the state bank account, unauthorized update debts, obtaining false documents, which achieves a right, etc.

2.4 Trojan horse - insert a set of instruction

Insert a set of instruction by other programs using the so-called Trojan horse, a form of manipulation of computer code that is illegal or hidden part of the instruction is inserted in some of the programs that are running on the system. Some viruses are spread by the same method.

2.5 Unauthorized use of software tools

Occasion of the creation of security system needs to be careful because it is possible that with the help of the so-called extra tools and universal software tool made modifications to otherwise protected data or files. If, for example, the IBM personal computer holds a complete version of DOS or OS / s operating system, uninvited people can do anything on that system.

2.6 Entering a logical (software) bombs

Like a Trojan horse, the system can enter a program that is activated only when certain conditions are met. This program destroys or deletes all files in the system when the name appears on the programmer.

2.7 Hacking or salami technique

A program that performs processing can be developed so that, for example, all amounts rounded to the lower, a difference that is placed in a separate account or accounts with each strike out a small amount, which is supposed to be to customers of the bank or institution will not complain, and entered into a special account, etc.

2.8 Scavengening

Scavengening form of unauthorized collection of data remaining on the computer or the computer processing. Scavengening as a concept was created when the system information gathered from the discarded material in trash cans.

2.9 Wiretapping System

One of the ways you can get to the data contained on the system is eavesdropping. Just record the sound of printer paper so that later processing comes to text that is printed on your printer. Eavesdropping can also be used on all communication lines are connected to the system. Therefore, it is necessary to protect that data is encrypted and demodulates the help of special devices and algorithms.

2.10 Disguise that enter the system using a stolen password

The system can be entered using the appropriate cards, false identity cards or similar documents that have been stolen or copied due to negligence or recklessness of employees who work on the system. Such raids are dangerous because they sometimes do not even know when they were given or the way he made breaking into the system.

2.11 Worm [4]

A computer program that when the start of multiplying on computer or network to take using of system resources to the extent that this can not function normally.

2.12 Other ways to incursions into the system

The system can go other ways and using different methods. One method is called asynchronous attack in which the system stops multiprogramming program (in a split second) that can attack and that is entered into the system of its data.

3 DETECTING COMPUTER CRIME

Notwithstanding the enormous efforts to combat computer crime at the global level, the forms of computer crime is very frequently happened. With crime in the area of computer crime often happens that there is no observable physical consequences, and it may take some time until it is determined that any unauthorized hacker intrusion into an information system company if it is legal or intrusions into the personal computer in the case of physical person. If an unauthorized hacker intrusion quickly been established, there is a possibility that the passage of time can not determine the perpetrator. Šimundić [3] in the book "Pravna informatika" / (Legal Informatics) says it is necessary to determine the time of planning for crisis management situations like this. Unforeseen event planning and crisis management considered various names for the same process, the planning process, identify and resolve various types of disasters. A computer crime is just that - a disaster and suggests that special care of the following activities:

- define different measures to identify the shortest possible time recognized that the computer crime occurred,
- creation of a team of people capable to respond to any incident from the domain of computer crime,
- defining the team's reaction to different situations

Anyone who is in the advance team must know how to react to some of the levels of risk. Regardless of the quality of the plan, it is necessary to change over time that the perpetrators are not able to predict whether or not to be controlled during the unauthorized hacker attacks. Since this article talking about various methods of manipulating computers, we must highlight the need to document everything that happens. Information collected in this way will greatly assist in the prosecution of perpetrators, but also in setting up and analyzing security measures. For the detection of crimes in the area of computer crime can be set different measures at the level of a computer system or the level at which it is the same. Detection measures are divided into two groups: proactive and reactive. Proactive measures detect illicit or criminal activity before or during the execution of a reactive detect illicit or criminal activity during or after the execution of computer crimes. Modern and now quite developed computer technologies now offer special programs (expensive detection model) and even the systems (expensive detection model) for the detection of unauthorized hacker attacks and internet service providers have developed a system to report crimes in the area of computer crime. In other words, perpetrators of unauthorized hacker intrusions are exposed to great risk of being discovered during the commission of crimes as a method for detecting unauthorized hacker attacks now developed in almost the same extent as the developed modern computer technology in general, and continues to diligently work on its improvement which is to say that the measures of detection every day are becoming better and more successful in detecting crime in the area of computer crime.

4 REPORTING CRIME IN THE AREA OF COMPUTER CRIME

Article 204 [5]

(1) All state authorities and all persons are obliged to report criminal acts that are prosecuted ex officio, who are notified to them or which they find themselves.
(2) Filing of criminal charges by the police is regulated by special law.
(3) Submitting an application, government bodies and legal persons should indicate the evidence that they are familiar with and to take in order to preserve traces of the crime, objects on which or whom the offense committed and other evidence.
(4) Citizens are obliged to report criminal acts that are prosecuted ex officio.
(5) The law shall determine the cases in which the failure to report the crime, a crime.
(6) Data on the identity of the person against whom criminal charges and the data on which to conclude on the identity of that person's official secret.

Criminal charge has been written or otherwise communicated information on crimes that are prosecuted ex officio, directly or indirectly addressed to Attorney General [6]. Criminal charges is the main cause for initiating criminal proceedings. A special type of criminal charges is a criminal complaint police authorities, which may be different with regard to content, especially bearing in mind whether the application was the reason for undertaking surveys of criminal acts or the result of the survey. Detecting crime includes a large number of different actions with each other. Besides the most important activities of the police authorities, the detection of crime to protect their own interests established companies and businesses to protect assets and people. The law did not explicitly resolve the method of registration in case the offense is committed by a child. In this case, the police authorities would still have to file criminal charges.

The duty to report commits government bodies, regardless of whether they had information on crimes or are notified to the criminal offense to find yourself another way. The body of state power under paragraph 1 has an obligation to give evidence, which are known to take measures to preserve traces of the crime, objects on which or whom the act was committed and all other evidence. In the event of an indication of organized crime, the responsibility belongs to the prosecution office to combat corruption and organized crime [7].

Article 205 [5]

(1) The application must be submitted to the competent state attorney in writing, orally or by other means.

(2) If the application is submitted orally, the applicant will be warned of the consequences of a false report. The oral report will be minutes, and if the application was communicated by telephone or other telecommunications device, is provided, where possible, its electronic records and compiles the official record.

(3) If the application is filed with the court, police or not competent public prosecutor, they will accept the report and immediately to the competent public prosecutor.

(4) The state attorney charges recorded in the register of charges as filed, except in the case of Article 206 Paragraph 7 and 8 this Act.

(5) The Minister of Justice shall prescribe the manner of keeping the register of charges.

Criminal charges is an oral or written [6]. Falsely reporting a criminal offense under the terms of Article 302 Criminal Code offense. When presenting an oral report, a body that receives such application must indicate this to the applicant. The oral report, depending on how you filed, shall be made or an official note. Receipt of application Article 94 governs Regulations on the conduct of police officers [8]. Criminal complaint which was filed with another authority, they are submitted to the competent public prosecutor.

5 SANCTIONING OF CRIMINAL ACTS IN THE FIELD OF COMPUTER CRIME

Criminal law is very accurately describes the unauthorized hacker intrusions through the description of offenses in Article 223 "Violation of the confidentiality, integrity and availability of computer data, programs and systems," a characterization method of manipulating computers are found in this article.

"(1) Who, despite the protective measures for unauthorized access to computer data or programs,

shall be punished by fine or imprisonment not exceeding three years.

(2) Who is disabled or hinders the work or the use of computer systems, computer data or programs or computer communications,

shall be punished by fine or imprisonment not exceeding three years.

(3) Whoever damages, alters, deletes, destroys or otherwise renders unusable or inaccessible someone else's computer data or computer programs,

shall be punished by fine or imprisonment not exceeding three years.

(4) Who intercept or record a non-public transmission of computer data it is not intended to, from or within a computer system, including electromagnetic emissions of a computer system that transmits the data, or who allow unauthorized person to become familiar with such data,

shall be punished by fine or imprisonment not exceeding three years.

(5) If the criminal offense referred to in paragraphs 1, 2, 3 or 4 this Article is committed in relation to the computer system, data or program of government bodies, public institution or a company of special public interest, or has caused substantial damage,

shall be punished by imprisonment from three months to five years.

(6) Whoever without authorization manufactures, procures, sells, possesses or makes available to another special devices, equipment, computer programs or computer data created or adapted to commit the offense referred to in paragraphs 1, 2, 3 or 4 this article

shall be punished by fine or imprisonment not exceeding three years.

(7) Special devices, equipment, computer programs or data created, used or adapted to commit crimes, and used to commit the criminal offense referred to in paragraphs 1, 2, 3 or 4 this Article shall be forfeited.

(8) An attempted criminal offense referred to in paragraphs 1, 2, 3 or 4 this Article shall be punished. "

Article 223, paragraph 1, behavior is described by referring to illegal access. It is determined that the law protects only the system that has safeguards. If the system is left completely open and unprotected, or if the computer is fully available, will be achieved through elements of the offense.

In paragraph 2 the same article described a behavior that is related to disruption of the system. In practice there are frequent instances in which such incriminating. An important element here is the intended meaning to be relieved of responsibility who, for example, inadvertently turned off the key computer internet service providers and lead to interruption of communication. It is important to emphasize that it is here one can find the specifics related to computer crime. An attacker may be released into the distribution of worms that will extend to thousands of other computers, which will then, without knowledge of their users, participate in the attack. This would mean that there is the possibility of using other people's infrastructure in an offense without any knowledge of the owner or persons using this infrastructure. They can not be held liable for criminal misuse of their equipment, but will it be affected by the "infected" computer to work more slowly and their network connection to work under heavy load.

Paragraph 3 disruption of data which describes this form of property protected so as to protect the material things. Computer data in digital format nowadays are extremely important assets. Damages arising from damage to, rotation, removing or destroying other people's data can be huge.

Paragraph 4 describes the illegal seizure that requires sanctioning of unauthorized interception of computer data transmission to an information system from or within it (including electromagnetic emissions from the information system, which carries the same computer data) that was done by technical means. In other words, it would mean that this paragraph specifically prohibits the interception of wireless data transmission and interception of wire transfer that can be exported without a direct connection to a telecommunication line.

Paragraph 5 introducing a qualified form of this offense when the subject acts computer data, programs or system of government bodies, public institution or a company of special public interest, or has caused substantial damage, and as reasonably anticipated to severe a penalty.

The goal of legal protection under paragraph 6 is to prevent the creation and market expansion devices in a very frequent practice of specialized programs for committing criminal acts described in paragraphs 1, 2, 3 and 4 Criminalization of these offenses can often be implemented with the assistance of legal devices and legal software and there are additional problems may arise.

Paragraph 7 is quite clear.

Paragraph 8 says that the Convention on Cybercrime in its Article 11 Paragraph 2 requires that States Parties to prescribe punishment for attempted crimes in this article. This would mean that it is necessary to prescribe punishment for the attempt.

6 CONCLUSION

Modern information and computer technology every day progresses and develops. Accordingly, unfortunately, every day more and more advanced methods of manipulating computers that can cause chaos in the system someone, for example, the economic entity and the personal computer of any ordinary man. As an unauthorized hacker intrusions appear in practice, described the entries in section 2nd of this paper. Because of this, as much as possible, unauthorized hacker attacks must be reporting to the competent authorities. The competent authorities should as soon as possible to take everything you can to be as soon as possible adequately punish the perpetrators of unauthorized hacker attacks. If the penalty is appropriate, for to hope is that the convicted offender in the unauthorized hacker attacks and to reappraise the conscience of this and similar crimes will no longer do. Verdicts in the field of computer crime should be publicized so the public would be so warned the public that crime in the area of computer crime belong to the socially unacceptable behavior and thus prevent potential new hackers to engage in these illegal activities. Methods of manipulating computers uninitiated may seem harmless, but they are by no means because it can cause huge problems in the computer system or personal computer if you see any of them. Article 223 of Croatian Criminal law is quite clearly elaborates the methods of manipulating computers, describing what is not working with computers and modern computer technology (see section 5th of this paper). When you collect all the material evidence, criminal proceedings for the purpose of prosecuting perpetrators of unauthorized hacker attacks can begin.

References

[1] Kazneni zakon *(Criminal Law)* - NN 110/97.; 27/98.; 50/2000.; 129/2000.; 51/2001.; 111/2003.; 190/2003.; 105/2004.; 84/2005 .; 71/2006.; 110/2007.; 152/2008.

[2] Convention on Cybercrime - NN - MU 9 / 2002.

[3] Šimundić, Slavko: "Pravna informatika" *(Legal Informatics)*, Pravni fakultet Sveučilišta u Splitu, Split, 2007.

[4] www.cert.hr

[5] Zakon o kaznenom postupku *(Criminal Procedure Act)* - NN 152/2008.; 76/2009.

[6] Pavišić, Berislav: "Komentar Zakona o kaznenom postupku" *(Commentary on the Law of Criminal Procedure)*, Peto izdanje, Tiskara i grafika Žagar, Rijeka, 2005.

[7] Zakon o Uredu za suzbijanje korupcije i organiziranog kriminaliteta *(Law on the Office for the Prevention of Corruption of organized crime)* - NN - 76/2009.; 116/2010.; 145/2010.

[8] Pravilnik o načinu postupanja policijskih službenika *(Regulations on the conduct of police officers)* - NN 89/2010.

ALLOCATIVE EFFICIENCY AND OPTIMAL ADJUSTMENTS OF A PRODUCER

Ilko Vrankić, Mira Oraić and **Zrinka Lukač**
University of Zagreb, Faculty of Economics and Business
Trg J. F. Kennedy 6, 10000 Zagreb, Croatia
{ivrankic, moraic, zlukac}@efzg.hr

Abstract: Inefficiency of the producer is the result of the inefficient use of technology and production planning which does not take into account market constraints. This article clearly describes the efficiency measures in different market circumstances. Special attention is given to the dual relationships that allow a description of the efficiency of the producer in different ways. Decomposition of the overall efficiency into the technical and allocative inefficiency enables distinguishing causes of inefficiency. We give the original economic interpretation of allocative efficiency which is the basis for determining the direction in which technology allows maximum savings and describing measures that a company should take.

Key words: technological and market constraints, technical and allocative efficiency, distance functions, economic savings, optimal adjustments of the production plan

1 INTRODUCTION

Technology is directly described by a set of technologically available production plans which is called the production possibility set. The description of technology is based on the primary variables, the quantity of inputs and outputs. Technologically efficient production plans are set on the boundary of the production possibility set. Interior of this set contains a technologically inefficient production plans. We can define the distance functions and the measures of technical efficiency based on projections of production plans to the efficient technology frontier.

The rational producer chooses the economically most efficient plan from all the technologically available production plans. The optimum choices depend on the economic context in which the producer meets market constraints. These constraints can be determined by the economic costs or revenue. From this context we derive alternative descriptions of the technology, by which we think of the cost function, the revenue function and the profit function. Alternative descriptions of the technology include dual variables, prices that are determined in markets.

Sources of inefficiency are ones of the technological and economic nature. Their distinction is allowed by the decomposition of the overall efficiency into the technical and allocative components. This article clearly describes the measures of efficiency and appropriate distance functions. It emphasizes the dual relationships that allow a description of the efficiency of the producer in different ways. In this context, special attention is paid on the allocative efficiency which is the basis for determining the optimal adjustment of the production plans of the producer. The producer is interested in changing the quantities of inputs and outputs in a way that enables him or her to achieve economic savings. The direction for which technology allows maximum savings is optimal and it is in this article determined on the basis of the allocative efficiency. In this way we can detect the reasons for the inefficient production of the producer and describe measures that a company should take.

2 INPUT AND OUTPUT DISTANCE FUNCTIONS

The cost function is derived from the model of minimizing the costs for a given level of production,

$$c(\mathbf{w},\mathbf{y}) = \min_{\mathbf{x}} \mathbf{wx} \quad \text{s.t. } (\mathbf{x},\mathbf{y}) \in Y, \tag{1}$$

where Y is the set of technologically feasible production plans, $Y = \{(\mathbf{x},\mathbf{y}) : \mathbf{x} \text{ can produce } \mathbf{y}\}$.

Note that the upper constraint includes technologically feasible production plans which has fixed levels of outputs. It is assumed that a set of technologically feasible production plans meets the following characteristics: it is nonempty, closed and convex set and free disposal of inputs and outputs is satisfied. These properties make it possible to establish dual relationships in production. The cost function is the distance function for the indirect production function in a way that it measures the distance of a normalized price vector from the indirect isoquant and can be interpreted as a measure of a overall cost efficiency. The proportional change of input where the factor of proportionality is reciprocal to the minimal costs of any combination of input prices returns the firm to the isoquant in the space of normalized prices and the level of production does not change. This is a consequence of a linear homogeneity of a cost function in prices.

We move now into the space of inputs and consider the isoquant, which includes all combinations of inputs that produce a given output level. In the same way as before we can define the distance of any bundle in the input space from a given isoquant. The proportional change of inputs where the factor of proportionality is reciprocal to the distance of any combination of factors of production from a given isoquant returns the firm to the isoquant and the level of production does not change. If we analyze a producer that produces more products, the input distance function can be generally defined in the following way:

$$D_I(\mathbf{x},\mathbf{y}) = \underset{\delta}{\text{maks}} \ \{\delta : (\mathbf{x}/\delta, \mathbf{y}) \in Y\}. \tag{2}$$

Knowing the distance function enables us to describe the cost function in the following way:

$$c(\mathbf{w},\mathbf{y}) = \min_{\mathbf{x}} \frac{\mathbf{wx}}{D_I(\mathbf{y},\mathbf{x})}. \tag{3}$$

The input distance function is derived from the cost function by solving the following optimization problem:

$$D_I(\mathbf{y},\mathbf{x}) = \min_{\mathbf{w}} \frac{\mathbf{wx}}{c(\mathbf{w},\mathbf{y})}. \tag{4}$$

This duality relationships are first defined by Shepard [9,10] who also introduced the radial distance functions as an alternative representation of a technology.

Since the reciprocal value of the input distance function describes how many times the efficient production can decrease the amount of factors of production, it can be seen as a measure of technical efficiency. In this way technical efficiency was defined by Farrell [3] who demonstrated how total cost or economic efficiency could be decomposed into technical and allocative efficiency. The allocative efficiency is defined as the ability to choose optimum input levels for given factor prices. The basis for this decomposition is the duality between the cost and input distance function [1,3,9,10].

The overall Farrell input based measure of cost efficiency can be written as

$$\frac{c(\mathbf{p},\mathbf{y})}{\mathbf{px}} = \frac{1}{D_I(\mathbf{y},\mathbf{x})} \cdot AE, \tag{5}$$

where $D_I(\mathbf{y},\mathbf{x})$ is the input based measure of technical efficiency and *AE* is the allocative efficiency.

Similar duality relationship can be established between the revenue function and the output distance function and the revenue efficiency can be defined that can be decomposed into technical and allocative components [4]. The revenue function is derived from the model of maximizing revenue for a given combination of production factors,

$$\begin{aligned} R(\mathbf{x},\mathbf{p}) &= \max_{\mathbf{y}} \mathbf{py} \\ \text{s.t. } &(\mathbf{x},\mathbf{y}) \in Y, \end{aligned} \tag{6}$$

Note that the upper constraint includes technologically feasible production plans which has fixed levels of inputs.If a firm with a given combination of production factors produces less than the maximum possible production level, we can talk about the output distance function whose reciprocal value describes how many times the efficient production can increase the amount of output. If we analyze a producer that produces more products, the output distance function can be generally defined in the following way:

$$D_O(\mathbf{x},\mathbf{y}) = \min_{\delta} \left\{\delta : (\mathbf{x},\mathbf{y}/\delta) \in Y\right\}. \tag{7}$$

Knowing the output distance function the revenue function can be derived by solving the following model:

$$R(\mathbf{x},\mathbf{p}) = \max_{\mathbf{y}} \frac{\mathbf{py}}{D_O(\mathbf{x},\mathbf{y})}, \tag{8}$$

and the output distance function is derived from the revenue function solving the following optimization model:

$$D_O(\mathbf{x},\mathbf{y}) = \max_{\mathbf{p}} \frac{\mathbf{py}}{R(\mathbf{x},\mathbf{p})}. \tag{9}$$

The overall Farrell output based measure of revenue efficiency can be written

$$\frac{R(\mathbf{x},\mathbf{p})}{\mathbf{py}} = \frac{1}{D_O(\mathbf{x},\mathbf{y})} \cdot AE \,, \tag{10}$$

where $(1/D_o(\mathbf{x},\mathbf{y}))$ is the output based measure of technical efficiency and *AE* is the allocative efficiency.

3 GENERALIZED DISTANCE FUNCTION AND ALLOCATIVE EFFICIENCY

The derived results can be extended by defining the directional distance function [2],

$$\vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y}) = \max\left\{\boldsymbol{\beta} : (\mathbf{x}-\boldsymbol{\beta}\mathbf{g_x}, \mathbf{y}+\boldsymbol{\beta}\mathbf{g_y}) \in Y\right\}, \tag{11}$$

determined by the directional vector $\mathbf{g} = (-\mathbf{g_x},\mathbf{g_y})$ which projects the initial combination of production factors and products on the boundary of a set of technologically feasible production plans. The directional distance function was introduced by Chambers, Chung and Färe [2] based on a benefit and shortage functions provided by Luenberger [6,7,8]. It can be shown that the directional technology distance function encompasses all known distance functions as special cases [2].

Dual relationship between the profit function and the directional distance functions can be established which allows us to define the profit efficiency, which can be decomposed into technical and allocative efficiency. Profit function is the result of the following optimization model:

$$\pi(\mathbf{p},\mathbf{w}) = \sup_{\mathbf{x},\mathbf{y}} \{\mathbf{py} - \mathbf{wx} : (\mathbf{x},\mathbf{y}) \in Y\}, \tag{12}$$

where $\mathbf{w} = (w_1,...,w_N)$ is the vector of prices of production factors, and $\mathbf{p} = (p_1,...,p_M)$ is the vector of product prices. The maximum profit and the difference between revenue and cost of any technologically available combination of factors of production and products links the following inequality:

$$\pi(\mathbf{p},\mathbf{w}) \geq \mathbf{py} - \mathbf{wx} + \vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y})(\mathbf{wg_x} + \mathbf{pg_y}). \tag{13}$$

We can interpret $\mathbf{wg_x} + \mathbf{pg_y}$ as an increase in profit from moving in the direction $\mathbf{g} = (-\mathbf{g_x},\mathbf{g_y})$ in one step. Similarly $\vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y})(\mathbf{wg_x} + \mathbf{pg_y})$ is the increase in profit from moving in the direction $\mathbf{g} = (-\mathbf{g_x},\mathbf{g_y})$ in $\vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y})$ steps. The directional distance function is derived from the profit function by solving the following optimization problem:

$$\vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y}) = \inf_{\mathbf{p},\mathbf{w}} \left\{ \frac{\pi(\mathbf{p},\mathbf{w}) - (\mathbf{py} - \mathbf{wx})}{\mathbf{pg_y} + \mathbf{wg_x}} \right\}. \tag{14}$$

From Chambers, Chung and Färe [2] the profit efficiency is defined by

$$\frac{\pi(\mathbf{p},\mathbf{w}) - (\mathbf{py} - \mathbf{wx})}{\mathbf{pg_y} + \mathbf{wg_x}} = \vec{D}_T(\mathbf{x},\mathbf{y};-\mathbf{g_x},\mathbf{g_y}) + AE. \tag{15}$$

Thus the overall profit efficiency is equal to the sum of technical efficiency which is measured by the directional distance function and the allocative efficiency. The cost and revenue measures of efficiency are special cases of the profit efficiency measure [4]. When the directional vector is $\mathbf{g} = (\mathbf{0},\mathbf{y})$, the relationship (15) collapses to the (10) and when on the other hand the directional vector $\mathbf{g} = (-\mathbf{x},\mathbf{0})$ is chosen, the relationship (15) implies (5).

4 OPTIMAL ADJUSTMENTS OF THE PRODUCER'S TECHNOLOGICAL PROCESS

In determining the optimal adjustment of the production plans of the producer it is sufficient to know the allocative efficiency. The allocative efficiency depends on the initial combination of inputs and outputs, prices of inputs and outputs and the directional vector. We are interested in the direction which will lead us to the optimum. When we choose that direction, the allocative efficiency is equal to zero. Thus, the optimal adjustments are determined by solving the equation $AE = 0$.

Let the technological process of a producer be described by the production function $y = \sqrt{x}$. The combination of input and output that maximize profit is represented by the following input demand function $x^* = \frac{p^2}{4w^2}$ and output supply function $y^* = \frac{p}{2w}$. The maximal profit function is $\pi = \frac{p^2}{4w}$.

If we start from combination of input and output inside the production possibility set (x_0, y_0), and if we represent the directional distance function by β, the technologically efficient combination of input and output is obtained by projecting the initial combination of input and output by the directional distance function on the boundary of a set of technologically feasible production plans in the direction $\mathbf{g} = (-\mathbf{g_x}, \mathbf{g_y})$,

$$(x_0, y_0) + \beta(-g_x, g_y) = (x_0 - \beta g_x, y_0 + \beta g_y). \tag{16}$$

Taking the production function from our example leads us to the following

$$y_0 + \beta g_y = \sqrt{x_0 - \beta g_x}. \tag{17}$$

To get the directional distance function, we solve the following equation

$$\beta^2 g_y^2 + \beta(2y_0 g_y + g_x) + (y_0^2 - x_0) = 0. \tag{18}$$

Thus the directional distance function for the chosen production function is represented by the following

$$\beta = \frac{\pm\sqrt{D} - (g_x + 2y_0 g_y)}{2g_y^2} \geq 0. \tag{19}$$

Taking into account the prices of input and output, we can in the same way define the projection of the initial combination of input and output on the tangent to the production possibility set in the direction $\mathbf{g} = (-\mathbf{g_x}, \mathbf{g_y})$. The tangent which can be called the isoprofit line represents all the combinations of input and output that give the profit equal to the maximal profit for the chosen technology,

$$y = \frac{w}{p}x + \frac{p^2}{4w}. \tag{20}$$

The projection of the initial combination (x_0, y_0) to the above isoprofit line gives

$$y_0 + \alpha g_y = \frac{w}{p}(x_0 - \alpha g_x) + \frac{p}{4w}, \tag{21}$$

where α is the number of steps of the initial combination to the isoprofit line in the direction $\mathbf{g} = (-\mathbf{g_x}, \mathbf{g_y})$. Solving for α gives $\alpha = \dfrac{\frac{p^2}{4w} - (py_0 - wx_0)}{pg_y + wg_x}$.

Thus, α represents the measure of the overall or total profit efficiency since it is equal to the ratio of the difference between the maximal profit and profit from the initial combination of input and output, $\frac{p^2}{4w} - (py_0 - wx_0)$, which can be interpreted as the overall possible savings, and savings that correspond moving in the direction $\mathbf{g} = (-\mathbf{g_x}, \mathbf{g_y})$ in one step, $pg_y + wg_x$.

The allocative efficiency is equal to the difference between the overall efficiency and the technical efficiency which is measured by the directional distance function,

$$AE(x_0, y_0, g_x, g_y, p, w) = \alpha - \beta = \frac{\frac{p^2}{4w} - (py_0 - wx_0)}{pg_y + wg_x} - \frac{\pm\sqrt{D} - (g_x + 2y_0 g_y)}{2g_y^2}. \tag{22}$$

Setting the equation $AE(g_x, g_y) = 0$ gives the direction of the projection of the initial combination of input and output to the efficient profit maximizing combination of input and output. In other words, solving the equation $AE(g_x, g_y) = 0$ gives us the optimal adjustments of the producer.

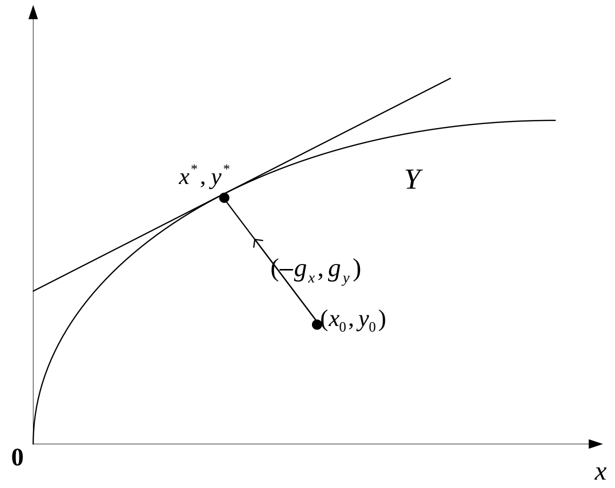


Figure 1. Allocative efficiency and adjustments of the technological process

We can set $pg_y + wg_x = \frac{p^2}{4w} - (py_0 - wx_0)$. In words, let the number of steps in direction $\mathbf{g} = (-\mathbf{g_x}, \mathbf{g_y})$ to the isoprofit line be one. In that case the directional distance function is equal to one, so we get:

$$4g_y^2(g_y^2 + 2y_0 g_y + g_x - x_0 + y_0^2) = 0. \tag{23}$$

If the term in brackets is equal to zero, $g_y^2 + 2y_0 g_y + g_x - x_0 + y_0^2 = 0$,the components of the optimal direction vector are

$$g_x = x_0 - \frac{p^2}{4w^2},\ g_y = \frac{p}{2w} - y_0. \tag{24}$$

Note that in the previous analysis $g_y \neq 0$. If the initial production plan already contains the optimal quantities of outputs, the input quantities would only be adjusted. In that case the previous results would be confirmed and the solution would be $y_0 = \frac{p}{2w}$, $g_y = 0$. Thus, inefficient producer should in general simultaneously affect the cost and the revenue. Changes in costs and revenues are described by a vector of optimal adjustment of the technological process.

5 CONCLUSION

Inefficiency of the producer is the result of the inefficient use of technology and production planning which does not take into account market constraints. Distinguishing sources of inefficiency is enabled by the decomposition of the overall efficiency into the technical and allocative components. Technological and market constraints narrow the possibilities of adjustment of production plans of the producer. This article discusses various market constraints and clearly describes the measures of efficiency in various ways on the basis of dual relationships. Special attention is given to the allocative efficiency which is the basis for determining the optimal adjustments of the production plans of the producer. The starting point of the view is that the choices of the producer doesn't have to be the consequence of the rational decisions. The allocative efficiency depends on technology, current production plan, the direction of change of inputs and outputs and market prices. On the basis of savings of each adjustment of the technological process the number of adjustments which eliminates causes of inefficiency is determined. The direction for which the technology allows such a number of adjustments is optimal. Analysis that starts from the allocative efficiency and describes optimal adjustments of the producer is accompanied with an illustrative example of measures that an inefficient producer should take.

References

[1] Chambers, R. G., Chung, Y., Färe, R.,1996. Benefit and Distance Functions, *Journal of Economic Theory*, 70, 407–419.

[2] Chambers, R.G., Chung, Y., Färe, R., 1998a. Profit, Directional Distance Functions, and Nerlovian Efficiency, *Journal of Optimization Theory and Applications*, 98, 351–364.

[3] Farrell, M.J., 1957. The Measurement of Productive Efficiency, Journal of the Royal Statistical Society, 129A, 253–281.

[4] Färe, R., Grosskopf, S., 2000. Theory and Applications of Directional Distance Functions, *Journal of Productivity Analysis*, 13, 93–103.

[5] Färe, R., Primont, D., 2006. Directional Duality Theory, *Economic Theory*, 29, 239–247.

[6] Luenberger, D. G., 1992a. Benefit Functions and Duality, *Journal of Mathematical Economics,* 21, 461–481.

[7] Luenberger, D. G., 1992b. New Optimality Principles for Economic Efficiency and Equilibrium, *Journal of Optimization Theory and Applications*, 75, 221–264.

[8] Luenberger, D. G., 1995. *Microecomic Theory*. New York: Mc Graw-Hill, Inc.

[9] Shepard, R.W., 1953. *Cost and Production Functions*. Princeton: Princeton University Press.

[10] Shepard, R.W., 1970. *Theory of Cost and Production Functions*. Princeton: Princeton University Press.

Section VIII:

Finance and Investments

PRICE AND VOLUME MEASURES IN NATIONAL ACCOUNTS

Draženka Čizmić
Faculty of Economics, University of Zagreb
Kennedyev trg 6, 10 000 Zagreb, Croatia
dcizmic@efzg.hr

Abstract: A major concern in economic analysis is to measure economic growth in volume terms between different periods. It is then necessary to distinguish, in the value changes for certain economic aggregates, the changes arising solely from price changes from the remainder which is called the change in "volume". The Commission Decision 98/715 divides methods, that can be used for the estimation of prices and volumes, into three groups: A methods (most appropriate methods), B methods (those methods which can be used in case an A method cannot be applied), and C methods (those methods which shall not be used).

Keywords: price measure, volume measure, A/B/C classification

1 INTRODUCTION

Price and volume measurement relates to the decomposition of transaction values in current prices into their price and volume components. The price components should include changes arising solely from price changes, while all other changes (relating to quantity, quality and compositional changes) should be included in the volume components. The volume index can therefore in principle be broken down into the following three components: 1) changes due to changes in the quantity of the products, 2) changes due to changes in the characteristics of the products, and 3) changes due to compositional changes in an aggregate.

The primary objective is not simply to provide comprehensive measures of changes in prices and volumes for the main aggregates of the System[1] but to assemble a set of interdependent measures which make it possible to carry out systematic and detailed analyses of inflation, economic growth and fluctuations.

2 SCOPE OF PRICE AND VOLUME INDICES IN THE ACCOUNTS SYSTEM

SNA 1993[2] and ESA 1995 each contain one relatively short chapter on price and volume measures. Renewed demand for more harmonised national accounts' price and volume data came when the European Council in July 1997, agreed on the so-called "Stability and Growth Pact". It was concluded that the existing guidance given by the SNA 1993 and ESA 1995 was not sufficient to guarantee harmonised price and volume data. Commission Decision 98/715 specified three main principles that price and volume measurement should follow: 1) In the measurement of prices and volumes a detailed level of aggregation of products shall be used. This level of aggregation, which is referred to as the elementary level of aggregation, shall be at least as detailed as the P60 level of ESA 1995, for output as well as all categories of use. 2) Volume measures available at the elementary level of aggregation

[1] System of National Accounts 1993 (SNA 1993)

[2] In 2003, the United Nations Statistical Commission called for an update of the SNA 1993. The update was to bring the accounts into line with the new economic environment, advances in methodological research and the needs of users. European Union countries and a great majority of OECD countries have indicated their intention to implement the SNA 2008 by 2014.

shall be aggregated using the Laspeyres formula[3] to obtain the volume measures of all national accounts aggregates. Price measures available at the elementary level of aggregation shall be aggregated using the Paasche formula[4] to obtain the price measures of all national account aggregates. 3) Volume measures derived at the elementary level of aggregation shall be aggregated using weights derived from the previous year.

The Commission Decision divides methods, that can be used for the estimation of prices and volumes, into three groups: A methods (most appropriate methods), B methods (those methods which can be used in case an A method cannot be applied), and C methods (those methods which shall not be used). The classification of methods can differ from product to product. What is considered a good method for one product can be a less good, or even unacceptable, method for another.

There is duality in the measurement of prices and volumes: one can either deflate a current year value with a price index, or alternatively extrapolate a base year value with a volume index. Deflation with a price index is generally preferred. There are two reasons for this: 1) A sample of price observations is normally more representative than an equally sized sample of quantity observations. 2) It is more difficult to control for quality changes when volume indicators are used.

3 A/B/C METHODS BY TRANSACTION CATEGORY

In the output approach (market output and output for own final use) using appropriate Producer Price Indices (PPIs) will be an A method. Each product should be deflated separately by an appropriate PPI. If an appropriate PPI is not available, several alternative types of indicators exist (B methods): a less appropriate PPI, a Consumer Price Index (CPI), and output volume indicators that are not fully representative. The use of some other possible indicators will usually be C methods: input methods; secondary indicators, i.e. indicators not directly related to the output; and PPIs, CPIs or volume indicators that do not correspond at all with the products in question.

Non-market output can be subdivided into two types of output: individual goods and services, and collective services. The only methods that will be able to meet the criteria for A or B methods are methods that measure output. Input methods are C methods.

An A method for deflating intermediate consumption satisfies the following criteria: deflation takes place product-by-product; and domestically produced products and imported products are deflated separately. B methods may fail to distinguish between domestic produce and imports. If intermediate consumption is deflated at the aggregate level, without product detail, this is a C method.

There is a great number of different methods for the estimation of value added at constant prices. A broad classification into two groups might be considered: 1) methods of double indicators, 2) methods of single indicators. Theoretically speaking, methods of double indicators are considered as more reliable because they take into consideration the changes in output and intermediate consumption for calculating value added at constant prices. There are three separate methods of double indicators: 1) Double deflation, in which the value added at constant prices is determined as the difference between output and intermediate consumption, both at constant prices, with each series obtained by deflating current values with appropriate price indices. 2) Double extrapolation, in which separate

[3] The Laspeyres volume index is defined as a weighted arithmetic average of the quantity ralatives using the values of the earlier period 0 as weights.

[4] Paasche index can be described as current-weighted in the sense that the weights come, not from the year from which the comparison starts, but from the other year. This other (or non-base) year can be called the current year.

estimates are made for output and intermediate consumption by projecting base year values by appropriate volume indicators, again with value added derived as difference. 3) Extrapolation/deflation is a mixed method in which output is deflated and intermediate consumption calculated by extrapolation or vice versa.

Single indicator methods may be classified into two variants: 1) Direct deflation of current price value added by an output price index or other relevant or estimated price series. 2) Direct extrapolation of base year value added using an indicator series which aims to approximate to the movement in constant price value added.

The A method for value added is the use of a double indicator method. A single indicator method can be used as a B method if intermediate consumption is a large share of output and the data required to apply a double indicator method are not reliable enough.

Those methods based on deflation of household expenditure using appropriate CPIs, with adjustments where necessary, are A methods. An index is appropriate if it: a) is representative of all the products covered in national accounts; b) takes proper account of changes in the quality of products; c) is valued at purchasers' prices[5] including VAT; and d) follows the same concepts as the national accounts. Use of a CPI which does not meet all of the standards is a B method. Any method using an index that does not correspond at all to the product in question is a C method.

Whilst the value of government and non-profit institutions serving households (NPISHs) final consumption is measured using an "input" approach, this does not mean that the only method for deriving constant price data is to deflate the value of inputs by suitable deflators. The only methods that will be able to meet the criteria for A or B methods are methods that measure output. Input methods are C methods. Purchases by general government of goods and services produced by market producers that are supplied to households as social transfers in kind, which also form part of government final consumption, are to be deflated by appropriate price indices corresponding to these (market) goods and services.

The use of genuine investment price indices (IPIs) for the deflation of gross fixed capital formation (GFCF) will be an A method, providing they satisfy the following criteria: a) it is an index with a coverage of exactly that group of products; b) it takes proper account of changes in quality of the products; c) it is valued in purchasers' prices including non-deductible VAT; and d) the concepts underlying the index correspond to those of the national accounts. The use of PPIs adjusted to purchasers' prices would also be an A method provided they satisfy the criteria for investment price indices. The very different nature of the products covered within GFCF makes it unlikely that criteria for an A method will be satisfied for all products. Therefore when considering the criteria for B methods it seems sensible to specify a minimum level of product detail to be used in the deflation of GFCF. Methods that do not reflect adequately the level of product detail are C methods.

The calculations of changes in inventories (CI) at current and constant prices are often closely interlinked. A true A method can only be achieved if good information is available from enterprises. Furthermore, appropriate price information is required. If assumptions have to be made concerning the bookkeeping system, or if deflation is undertaken with less

---

[5] Basic price is the amount receivable by the producer from the purchaser for a unit of a good or service minus all taxes on products plus all subsidies on products. It excludes transport charged invoiced separately by the producer. Producer's price is the amount receivable by the producer from the purchaser for a unit of a good or service produced as output minus any VAT, or similar deductible tax, invoiced to the purchaser. It excludes any transport charges invoiced separately by the producer. Purchaser's price is equal to basic price together with all transport charges paid separately by the purchaser to take delivery at the required time and place, and including all trade margins, and taxes less subsidies on products, but excluding any deductible tax.

appropriate price indices, the calculation method becomes a B method. If deflation of CI is carried out with proxy or inappropriate indicators then the method becomes a C method.

The A method applicable to production of valuables is if there is a suitable PPI for an industry producing valuables, which is adjusted properly for changes in quality. B methods for measuring constant price flows for production of valuables are comparison with closely related products. Any measure using deflation of inputs with no adjustment for quality is considered a C method.

A methods for exports and imports of goods should be based on the use of quality adjusted prices for all exports and imports. The price indices should be consistent with the product classification used in the value data being deflated. Price indices that do not adequately reflect changes in quality should be considered B methods. For product groups that are sufficiently homogeneous over time, unit value indices (UVIs)[6] can also be considered B methods. The use of PPIs is also a B method providing the coverage of the indices used is appropriate. The use of foreign country export prices for the deflation of imports is a B method where the coverage of the product is exact and adjustment for exchange rates has been made. All other methods should be considered C methods.

The expenditure of non-residents[7] on the domestic territory should be deflated using CPIs consistent with the range of products that are purchased by non-residents (A method). If CPIs that have a narrower or wider coverage of products than those purchased by non-residents are used, this would be a B method. For the expenditure of domestic residents abroad, the use of detailed and appropriate CPIs for the country visited, adjusted for exchange rates, would be an A method. The use of foreign country CPIs that have a wider or narrower coverage would be a B method.

The amount of taxes on products is measured in terms of volume by applying to the quantities of products produced or imported the taxation prices of the base year or by applying to the value of output or imports, revalued at the prices of the base year, the tax rates of the base year. Similarly, the amount of subsidies on products is measured in terms of volume by applying to the quantities of products produced or imported the subsidy prices of the base year or by applying to the value of output or imports, revalued at the prices of the base year, the rates of subsidy of the base year, taking into account different subsidy prices for different uses. Non-deductible VAT at constant prices can be calculated by applying the VAT rates in force in the base year to the flows expressed in the prices of the base year. If different taxes or subsidies are not fully separable then the methods described become B methods.

4 INDEX NUMBERS OF PRICES AND VOLUMES IN CROATIA

The Central Bureau of Statistics of the Republic of Croatia publishes composite price indices (producers’ price indices of industrial products, by main industrial groupings and NKD 2007. divisions; producers’ price indices of agricultural, forestry and fishing products, by groups of products; and consumer price indices) and composite volume indices (indices of

---

[6] It may sometimes happen, especially in the field of foreign trade statistics, that as a result of lack of information the data on which price and volume indices have to be calculated are not adequate for the purpose. The “price” indices associated with such indices are usually described as average or “unit value” indices as they measure the change in the average value of units that are not homogeneous and may therefore be affected by changes in the mix of items as well as by changes in their prices.

[7] An institutional unit is resident in a country when it has a centre of economic interest in the economic territory of that country. It is said to have a centre of economic interest when there exists some location within the economic territory on, or from, which it engages, and intends to continue to engage, in economic activities and transactions on significant scale either indefinitely or over a finite but long period of time.

agricultural production, the industrial production index, and volume indices of construction works).

Producers' prices of industrial products are monitored in about 577 legal entities classified into sections B, C, D and E according to the NKD 2007. for about 340 industrial products. Producers' prices of agricultural, forestry and fishing products are monitored in agricultural, forestry and fishing legal entities for 74 products. The Consumer Price Index is calculated and published on the basis of about 740 representative goods and services. Price indices are calculated on the basis of index lists of representative products and services as well as related weights according to the modified Laspeyres formula.

Volume indices of agricultural production are calculated on the basis of production data for 65 agricultural products. The moving average of producers' prices in the last three years has been taken as a weighting factor in the index calculation. Industrial production volume indices are chain indices of total industrial production, which are calculated using the Laspeyres formula. The basis for the calculation of industrial production indices for the NKD 2007. industry divisions are data on total quantities of industrial finished products defined by the Nomenclature of Industrial Products (NIPUM 2009.) for the Monthly Survey on Industrial Production, presented in determined measurement units and weighting coefficients for individual NIPUM 2009. products. Data on total volume indices of construction works are based on monthly and quarterly reports on construction. They are filled in by legal entities and parts thereof employing 20 or more persons, classified in the Register of Business Entities into section F Construction, according to the NKD 2007.

5 CROATIAN ANNUAL GDP AT CONSTANT PRICES

Movements in the volume of GDP are always calculated by recalculating the values of the various components of GDP at the constant prices either of the previous year or of some fixed base year. Thus, volume measure of GDP is frequently referred to as "GDP at constant prices". It is preferable to avoid the term "real GDP" as this may suggest the deflation of GDP by some general price not necessarily that of GDP itself.

For the calculation of GDP according to the production approach at constant prices the concept of double deflation is used. Current values are deflated by price indices of the previous year (that is, GDP for a current year at constant prices has been expressed in previous year prices). The volume indicators are used for the transportation and telecommunication activities, while the input indicators (the number of employees) are used for most services.

Individual components of GDP by expenditure approach at constant prices were calculated by deflating the current price data, using chain price indices. Household consumption data at constant prices were calculated by using disaggregated consumer price indices. Concerning government expenditures, an assumption of constant productivity was applied, so that the wages and salaries in constant prices were calculated by dividing the data on current expenditures on wages and salaries by indices of the number of persons employed. Expenditures on other goods and services were deflated by the corresponding components of consumer price indices and producers' price indices of industrial products. Imports and exports of goods were deflated by using Fisher-type[8] unit value indices. Exports of services were deflated by corresponding price indices in the domestic market. Imports of services were deflated by relevant prices indices in the markets of the most significant

[8] Fisher's Ideal Index is defined as the geometric mean of the Laspeyres and Paasche indices. Fisher described this index as "ideal" because it satisfies various tests that be considered important, such as the "time reversal" and "factor reversal" tests.

partner countries of the Republic of Croatia. Gross fixed capital formation data were deflated by producer's price indices of domestic and import capital goods, and implicit deflator for the construction from the GDP production approach. Changes in inventories of finished goods and work in progress were deflated by producer's price indices. Changes in inventories of raw materials were deflated by producer price indices and import price indices of raw materials. Corresponding consumer price indices were used for deflation of changes in inventories of goods purchased for resale.

6 CONCLUSION

Economic and monetary policy of the European Union is becoming increasingly integrated, especially since the introduction of the Euro. This requires higher and higher standards of national accounts data as a solid foundation for the formulation and monitoring of economic policy. Within the EU, national accounts data are also increasingly important for more administrative purposes such as the determination of countries' contributions to the EU budget, assessment of economic convergence, regional funds, etc.

In a system of economic accounts, all the flows and stocks are expressed in monetary units. The monetary unit is the only common denominator which can be used to value the extremely diverse transactions recorded in the accounts and to derive meaningful balancing items. The problem when using the monetary unit as a measuring unit is that this unit is neither a stable nor an international standard. A major concern in economic analysis is to measure economic growth in volume terms between different periods. It is then necessary to distinguish, in the value changes for certain economic aggregates, the changes arising solely from price changes from the remainder which is called the change in "volume".

The A/B/C classification is aimed at improvement of current practice. It sets out in what direction improvements can be made. It is therefore important that the criteria for distinguishing A, B and C methods are absolute criteria, i.e. that they do not depend on the present availability of data.

References

[1] Allen R.G.D., Index Numbers in Theory and Practice, The Macmillan Press Ltd., London and Basingstoke, 1975

[2] Commission Decision clarifying Annex A to Council Regulation (EC) No 2223/96 on the European system of national and regional accounts in the Community as concerns the principles for measuring prices and volumes, OJ L 340/33, 16. 12. 1998.

[3] European system of accounts – ESA 1995, Office for Official Publications of the European Communities, Luxembourg, 1996

[4] Guidelines on principles of a system of price and quantity statistics, United Nations, New York, 1977

[5] Handbook on Price and Volume Measures in National Accounts, Eurostat, 2004

[6] Manual on national accounts at constant prices, United Nations, New York, 1979

[7] Statistical Yearbook of the Republic of Croatia 2010, Central Bureau of Statistics of the Republic of Croatia, Zagreb, 2010

[8] Stone R., Quantity and price indexes in national accounts, OECD, Paris, 1956

[9] System of National Accounts 1993, Inter-Secretariat Working Group on National Accounts: Commission of the European Communities – Eurostat, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, World Bank, Brussels/Luxembourg, New York, Paris, Washington, D.C., 1993

THE OPEN ECONOMY NEW KEYNESIAN PHILLIPS CURVE WITH ADJUSTED MEASURES OF REAL MARGINAL COST: ESTIMATES FOR CROATIA

Alan Domić
Faculty of Economics and Business, University of Zagreb
Trg J. F. Kennedyja 6, 10000 Zagreb, Croatia
alandomic@gmail.com

Abstract: In this paper we examine the inflation dynamics of Croatia and test different marginal cost proxies for the hybrid New Keynesian Phillips Curve (NKPC) by estimating three specifications of the hybrid NKPC for Croatia: a conventional specification, an energy-augmented specification, and an open economy specification. We find that the marginal cost proxy for the energy-augmented specification is statistically significant and quantitatively the largest, whereas those for the other two are statistically significant, but quantitatively negligible.

Keywords: New Keynesian Phillips curve, inflation, open economy, Croatia, real marginal cost, GMM.

1 INTRODUCTION

The modelling of inflation in the short run is an area of macroeconomics where answers are both urgently needed yet difficult to obtain. They are urgently needed because accurate forecasts of short-run inflation are vital for effective monetary policy; however, there are as yet no forecast methods that perform equally reliably in all situations. Phillips curve forecasts (in their current form, i.e. those forecasts that rely on a real activity variable such as unemployment, the output gap, or real marginal costs) appear to achieve better results than any other currently used method (Stock & Watson, 2008).

The New Keynesian Phillips Curve (NKPC) has been the focus of a great deal of recent research in inflation forecasting, and (consequently, to some degree) it is now a part of the monetary policy toolset for many countries. Having evolved from A.W. Phillips' relation of unemployment to inflation, the NKPC, in the form it takes today, relates the output gap or real marginal cost to inflation. That real activity variable is arguably the most disputed and most actively researched aspect of the NKPC, and the definitive form of the relation is far from being a settled issue.

This study's aim is to contribute to the literature by estimating the NKPC for a small open economy, with a corrected measure of marginal cost; most empirical NKPC studies have been carried out for large and highly developed countries, using the labor share of income as a proxy for marginal cost, which has shown to be problematic (Wolman, 1999, Rudd and Whelan 2005, 2007). In section 2, we describe our specifications of the NKPC and discuss the issues surrounding proxies of real marginal cost. In section 3, we describe our data set and methodology, and in section 4 we present our results using the generalized method of moments (GMM) estimator. We conclude in section 5 with a review of the study and a discussion of the possible directions of future research.

2 THEORETICAL FRAMEWORK

The most widely used specification of the NKPC is based on the model of price adjustment introduced by Calvo (1983). This model assumes an economy with monopolistically competitive firms that are perfectly identical, except for the differentiated products they produce and for their pricing history; and being subject to the same constant elasticity demand function. Only some firms, whose proportion is given by $(1-\theta)$, adjust prices in period t, and that probability is independent of the firm's pricing history up to period t. The pricing decisions are formulated as a monopolistic competitor's profit maximization problem, given a stream of expected future marginal costs and given the ability to set a frictionless optimal markup over marginal costs.[1] After including a discount factor β, the NKPC is given by

$$\pi_t = \beta E_t \pi_{t+1} + \lambda mc_t^r \tag{1}$$

where

$$\lambda = \frac{(1-\theta)(1-\theta\beta)}{\theta} \tag{2}$$

The NKPC can also be expressed with the output gap (the difference between real and potential output) substituted for marginal costs. This choice is problematic, primarily because in empirical data the output gap leads inflation for a given period, rather than the other way around as the NKPC implies; because of this, real marginal cost and not the output gap is typically used in the literature.

However, accurate measurement of real marginal cost is one of the most problematic areas of NKPC research. We shall give a fuller overview of this, including the measures we use to attempt to obtain a better fit for the NKPC, in section 2.3. We now turn to the NKPC specification that is the most widely used in the literature, and that we shall also use in this study.

2.1 Hybrid new Keynesian Phillips curve

The so-called hybrid NKPC, which incorporates a lagged inflation term, was introduced by Gali & Gertler (1999), and its specification is as follows:

$$\pi_t = \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1} + \lambda mc_t^r \tag{3}$$

where

$$\begin{aligned}
&\lambda \equiv (1-\omega)(1-\theta)(1-\beta\theta)\phi^{-1} \\
&\gamma_f \equiv \beta\theta\phi^{-1} \\
&\gamma_b \equiv \omega\phi^{-1} \\
&\phi \equiv \theta + \omega[1-\theta(1-\beta)]
\end{aligned} \tag{4}$$

In this model, there is a fraction of firms, given by 1- ω, that behave exactly like all firms in Calvo's model. The remaining firms in the economy, given by ω, set their prices using a backward-looking rule of thumb. This rule is defined by two features: there are no persistent deviations between that rule and optimal behavior; and the price decision for period t depends only on information from t-1 and earlier. Firms described by ω set their price according to $p_t^b = p_{t-1}^* + \pi_{t-1}$, where p_{t-1}^* is the average price level set in the round of price adjustments at t-1 (note that, in this way, the backward-looking firm indirectly takes into account information about the future, since the average price level includes prices of forward-looking firms). When $\omega = 0$, that is, when there are no backward-looking firms in the economy, the hybrid NKPC converges to the conventional NKPC form.

2.2 Real marginal cost and our approach

Since real marginal cost cannot be observed, a proxy must be found. Following Gertner and Gali (1999), who assumed a Cobb-Douglas production function, the choice of proxy has in recent literature overwhelmingly fallen on labor's share of income (or, equivalently, real unit labor costs). The empirical results obtained since then, from diverse data sets, have led to this choice of proxy being strongly criticized (Wolman, 1999, Rudd and Whelan 2005, 2007) on the grounds that only using labor's share of income fails to create a good fit for the NKPC. The studies cited do not go so far as to assert that the theoretical basis must be scrapped, but they strongly affirm that labor's share alone is inadequate and that, for the given specification of the NKPC, a better proxy for real marginal cost must be found.

This has provided the stimulus for a spate of research. A number of alternative measures have been proposed, the most promising of which may be those of Muto (2009)[2] and Bratsiotis and Robinson (2009), who developed successful frameworks of labor adjustment costs, and capital and labor marginal costs, respectively; other authors, such as Leith and Malley (2003) and Dabušinskas and Kulikov (2007) used open economy models that account for international trade. In this paper, we shall empirically test models from the latter two papers; regrettably, the models of Muto (2009) and Bratsiotis and Robinson (2009) are non-trivial to adapt to a Croatian framework, and remain beyond the scope of this study.

2.3 Energy-augmented hybrid new Keynesian Phillips curve

Dabušinskas and Kulikov (2007), in their empirical analysis of inflation in Estonia, Latvia and Lithuania, use two different NKPC specifications that are of interest to us. The first is what they refer to as the energy-augmented Phillips curve, and which explicitly includes the price of energy in its measure of marginal cost, given by:

$$\hat{mc}_t^r = \hat{s}_t - (1-\mu s)(\rho - 1)(\hat{p}_t^E - \hat{w}_t) \tag{5}$$

where μ is the markup, ρ is a measure of the elasticity of substitution between labor and imported intermediate goods, s is labor's share of income, and $(\hat{p}_t^E - \hat{w}_t)$ is the price of energy relative to wages. All hatted variables represent log-deviations from their steady state (μ and ρ, of course are constants).

2.4 Open economy hybrid new Keynesian Phillips Curve

The second model considered is an expansion of the hybrid NKPC into an open economy hybrid NKPC in which intermediate goods are taken into account for marginal cost. In this model, λ is replaced with $\tilde{\lambda}$, and the additions to the NKPC are given by:

$$\begin{aligned} \tilde{\lambda} &\equiv \frac{(1-\theta)(1-\theta\beta)\zeta_t}{\theta} \\ \zeta_t &\equiv \frac{1}{\psi_t \mu - 1} \\ \psi &= \frac{1}{1-\alpha_t} \end{aligned} \tag{6}$$

where $1-\alpha_t$ is the output share of labor and of imported intermediate goods, given by $1-\alpha=\mu\frac{s+i_t^s}{1+i_t^s}$, and real marginal cost[3] is given by

$$mc_t^r = f(\hat{s}_t, \psi_t, i_t^s, \hat{y}_t, \rho, \hat{w}_t - \hat{p}_t^f, \hat{p}_t^d - \hat{p}_t^f) \qquad (7)$$

where i_t^s is the share of imported intermediate goods in output, $\hat{y}_t$ is the output gap, $(w_t - p_t^f)$ is the difference between the wage level and the price of foreign goods, and $(p_t^d - p_t^f)$ is the difference between the price of domestic goods and foreign goods. In all respects other than these, this specification is identical to the hybrid NKPC.

Having defined our theoretical framework, we now turn to the description of our data and methodology, followed by the presentation of our results.

3 DATA DESCRIPTION AND METHODOLOGY

Our sample consists of 11 years of quarterly data for Croatia, beginning with 2000Q1 and ending with 2010Q4. Data are sourced from Eurostat, the Croatian National Bank and the Croatian Bureau of Statistics, and (for the price of Brent Crude oil) the U.S. Energy Information Administration. All time series data are seasonally adjusted, and all variables that are hatted in a given equation are represented in the model as log-deviations from their steady state (for which we take the sample mean).

Inflation, π_t, is the GDP deflator. The difference between the price of energy and wages in equation (5), $(\hat{p}_t^E - \hat{w}_t)$, is the logarithmic difference of the price of Brent Crude oil on European markets (adjusted by the USD/HRK exchange rate for the period) minus nominal wages. The labour income share, $\hat{s}_t$, is the ratio of employee compensation to GDP, and the output gap, $\hat{y}_t$, is obtained by detrending real GDP with the Hodrick-Prescott filter. The difference between wages and domestic prices, ($\hat{w}_t - \hat{p}_t^f$), is the logarithmic difference between nominal wages and the GDP deflator, and that between domestic and foreign prices $(p_t^d - p_t^f)$ is the logarithmic difference of the GDP deflator and the import deflator. The steady-state share of imported intermediate goods in GDP, i_t^s, consistently rises in our sample and is thus allowed to be time-varying, and since data on imports of intermediate goods are unavailable for Croatia, we follow the rule-of-thumb adjustment of Dabušinskas and Kulikov (2007) and assume that 50% of imported goods are intermediates. The rising steady-state value is obtained by applying the Hodrick-Prescott filter. Following relatively common choices in the literature, we set the discount factor β to 0.98, the markup μ to 1.2, and the elasticity of substitution ρ to 0.5.

4 EMPIRICAL RESULTS

Table 1 shows the resuls obtained for our three models: the hybrid NKPC described by (3) and (4) with real marginal cost equal to the labor share of income, the energy-augmented hybrid NKPC described by (3) and (4) with its real marginal cost given by (5), and the open economy hybrid NKPC described by (3), (4) and (6) with its real marginal cost given by (7).

The hybrid NKPC open economy model appears to have less explanatory power for inflation than either the conventional hybrid NKPC or the energy-augmented NKPC. It is possible to speculate extensively on the reasons for this, and two possible explanations we offer are that Croatia's relatively low consumption and export of domestically manufactured goods, and the large size of its service sector (the most notable component of which is tourism), suggest that the cost of labor makes up much of marginal cost; and that the rule of thumb we adopted due to insufficient data (the assumption that 50% of imported goods are intermediate inputs) may give inaccurate results.

The λ value of the energy-augmented NKPC (6) is, as increasingly predicted in the literature, higher than the conventional NKPC which takes into account for marginal cost only the labor share of income. Note also that γ_f is higher in this model: this is in line with the theoretical prediction that, as marginal cost becomes a more significant driver of inflation, forward expectations will play a larger role.

The proportion of backward-looking firms, ω, at close to 40% is near the range found in Krznar's (2011) comprehensive Croatian NKPC study. It is also, very roughly, in the range of values observed for the Baltic countries (Dabušinskas and Kulikov, 2007), and in line with Italy's but not that of other European G7 countries (Leith and Malley, 2003). The percentage of firms that do not adjust prices in period t, given by θ, is fairly high, but very close to the values in Krznar (2011); with such high θ, the higher γ_b compared to other cases in the literature does not come as a surprise (note that θ is lowest for the energy-augmented model, where forward expectations are strongest).

Finally, the J-stat values show that the models do not suffer from overidentification of instruments, and the Cragg-Donald F-stat test separately confirms that the models do not appear to be susceptible to the problem of weak instruments.

Table 1: Empirical hybrid NKPC models for Croatia, 2000Q1 to 2010Q4.

	Structural Parameters		***Reduced-form Parameters***			***Statistical Tests***	
	θ	ω	γ_b	γ_f	λ	*J-stat*	*Cragg-Donald F-stat*
Model (3)	0.9152 (0.0077)	0.3858 (0.0006)	0.4992 (0.0050)	0.5058 (0.0047)	0.0162 (0.0042)	1.5108	8.8010
Model (5)	0.9090 (0.0092)	0.3878 (0.0000)	0.4911 (0.0004)	0.5487 (0.0004)	0.0434 (0.0003)	1.0009	4.8044
Model (6)	0.9565 (0.0059)	0.3768 (0.0005)	0.4997 (0.0048)	0.5067 (0.0046)	0.0098 (0.0030)	1.3364	7.9744

Notes: Standard errors are in parentheses. The instrument set includes both the current value and two lags of the marginal cost series; and four lags of the GDP deflator, of wage growth, of the output gap, and of the import deflator. The GMM estimation uses Newey-West weighting matrices with a Bartlett kernel.

5 CONCLUSION

In this paper, we study the movement of inflation in Croatia with a conventional hybrid NKPC, an energy-augmented hybrid NKPC, and an open economy hybrid NKPC, with the aim of examining the processes that drive inflation in Croatia and the aim of testing methods of more accurate estimation of real marginal cost. The results of our conventional hybrid NKPC model are in line with expectations based on the literature, the one exception being somewhat higher influence from backward-looking inflation than is commonly found, and the results of the energy-augmented NKPC seem promising for energy augmentation as an adjusted measure of real marginal cost. The performance of the open economy NKPC, meanwhile, is contrary to expectations and makes real marginal cost a less significant driver of inflation than the conventional model does. We offer suggestions for the reasons behind this in section 4.

This exploration is intended as a survey of interesting territory, and not (yet) that territory's meticulous charting. More work on the dynamics of inflation in Croatia, and on refinements to the NKPC, is highly needed. To name some pertinent examples, it would be invaluable to adapt Muto's (2009) model of labor market frictions and wage rigidities to countries other than Japan, and to adapt Bratsiotis and Robinson's (2009) model of labor and capital marginal costs to the open economy. On the subject of methodology, flaws in GMM are becoming increasingly discussed in the literature, and estimating the NKPC with Full Information Maximum Likelihood (FIML) appears to be a promising alternative that overcomes some of GMM's limitations. All these avenues of investigation are eminently worhy - and for Croatia, in particular, the nascent study of inflation in a New Keynesian framework (so far largely defined by Krznar's (2011) paper) needs the blooming of a thousand flowers.

---

[1] For the sake of simplicity and brevity, we only provide the closed form of the NKPC with Calvo model pricing in this paper, and not the full derivation. In addition to the original description of the pricing structure in Calvo (1983), the full derivation is available in, among others, Gali and Gertner (1999). and in Leith and Malley (2003) for an open economy model.

[2] But note that Muto does not use GMM, only OLS and NLS.

[3] The full equation for real marginal cost is given by (5) in Dabušinskas and Kulikov (2007). It is too lengthy to be reproduced fully in a paper of this scope.

References

[1] Bratsiotis, G.J., and Robinson, W.A., 2009. Non-Labor Unit Costs, Marginal Costs and the New Keynesian Phillips Curve, *Centre for Growth and Business Cycle Research Discussion Paper Series, University of Manchester,* No. 123.

[2] Calvo, Guillermo A. 1983. Staggered Prices in a Utility Maximizing Framework, *Journal of Monetary Economics*, 12, pp.383-398.

[3] Dabušinskas, A. and Kulikov, D., 2007. New Keynesian Phillips curve for Estonia, Latvia and Lithuania. *Bank of Estonia Working Paper Series,* 7(2007)

[4] Gali, J. and Gertler, M., 1999. Inflation Dynamics: A Structural Econometric Analysis, *Journal of Monetary Economics*, 44(2), pp.195-222.

[5] Ichiro, M., 2009. Estimating a New Keynesian Phillips Curve with a Corrected Measure of Real Marginal Cost: Evidence in Japan, *Economic Inquiry*, 47(4), pp.667-684

[6] Krznar, I., 2011. Analiza kretanja domaće stope inflacije i Phillipsova krivulja, *Croatian National Bank Working Paper Series*, i-031.

[7] Leith, C. & Malley, J., 2003. Estimated Open Economy New Keynesian Phillips Curves for the G7, *CESifo Working Paper Series,* 834.

[8] Rudd, J., and Whelan K., 2005. Does Labor's Share Drive Inflation? *Journal of Money, Credit, and Banking,* 37(2), pp.297-312.

[9] Rudd, J., and Whelan, K., 2007. Modelling Inflation Dynamics: A Critical Review gof Recent Research, *Journal of Money, Credit, and Banking,* 39(s1), pp.155-170.

[10] Wolman, A., 1999. Sticky Prices, Marginal Cost, and the Behavior of Inflation, *Federal Reserve Bank of Richmond Economic Quarterly,* 85(4), pp.29-48.

OPTION LEVERAGE

Margareta Gardijan
Faculty of Economics – Zagreb,
Trg J.F. Kennedy-ja 6, 10 000 Zagreb, Croatia
mgardijan@efzg.hr

Abstract: Options are the most complex financial derivatives which open numerous possibilities to investors. One of those possibilities is achieving option leverage effect. By option leverage we mean the multiplication of profit from an option strategy with respect to changes in market price. In this paper we are analysing option leverage in terms of an elasticity coefficient of profit functions for option strategies long call and short put and comparing the values of those coefficients to elasticity of profit from long position in asset. It should be helpful for investors in decision making.

Keywords: options, leverage, elasticity, derivatives.

1 THE LEVERAGE EFFECT

Options as financial derivatives can be used for hedging, arbitrage and speculation. The benefits of options are high earning potential, flexibility, enabling investors to stay on the market without the need of owning certain asset and the leverage effect.[1] The term "leverage" is taken from natural sciences and it means achieving greater action by using less power.[2] In economics, the term leverage stands for more intensive change in one variable with regards to the change in other variable that has caused the change.[3] In case of options, the leverage could be observed as the ability of profit from certain option strategy to change in greater degree in with regards to change in price of the underlying asset. The paper considers the option leverage of chosen basic option strategies using the function of elasticity of profit defined for these strategies and compares it to the elasticity of profit from long positions in underlying asset.

2 ELASTICITY

The degree of leverage in economics is actually the elasticity of one variable with respect to changes in another variable at one point, i.e. "the ability of one economic variable to react on changes of a related economic variable"[4]. The value of elasticity coefficient is interpreted as a rate of change of one variable if another variable changes. Point elasticity $E_{y,x}$ is a ratio between relative changes in a dependent variable $y=f(x)$ and in an independent variable x :[5]

$$E_{y,x} = \frac{\text{relative change in } y}{\text{relativne change in } x}, \tag{2.1}$$

assuming that $\Delta x \to 0$ and $\Delta y \to 0$, therefore the point elasticity is calculated using Marshall's formula:

$$E_{y,x} = \frac{x}{y} \cdot \frac{dy}{dx}. \tag{2.2}$$

The interpretation of the coefficient is as follows:

---

[1] Andrijanić, 1997, p. 110, there are some other advantages, but we mentioned only few here.
[2] Orsag, 2011, p. 868
[3] Rosenberg, 1993
[4] Neralić & Šego, 2009, str. 267-269
[5] Ibid.

$$|E_{y,x}| = \begin{cases} 0 & \text{variable } y \text{ is perfectly inelastic} \\ >1 & \text{variable } y \text{ is elastic} \\ <1 & \text{variable } y \text{ is inelastic} \\ 1 & \text{variable } y \text{ is unit elastic} \end{cases} \quad . \tag{2.3}$$

If $\lim_{x \to x_0} |E_{y,x}| = \infty$, variable y is perfectly elastic at a point x_0.

3 BUILDING A MODEL

We shall analyze the option leverage by considering the elasticity of profit from basic positions (option strategies) in European call and put option on asset with no current income and compare the values to their, in terms of market expectations, equivalent position in asset. Therefore, we have chosen the long asset position and its, in terms of market expectations, equivalent option strategies short put and long call. The assumption is that profit π from any market position as a function of a market price is as follows:

$$\pi = f(S_t), \;\; \pi \in \mathbb{R}, S_t \in [0, +\infty\rangle. \tag{3.1}$$

Let S_0 be the asset price at t=0. Then the profit from long position in asset is

$$\pi_{di}(S_t) = -S_0 + S_t \;\; . \tag{3.2}$$

Profits of long call and short put, expressed as functions of market price are given in following forms:

$$\pi_{dc}(S_t) = \begin{cases} -c_0 & , S_t \le X_{dc} \\ -c_0 - X_{dc} + S_t & , S_t > X_{dc} \end{cases}, \tag{3.3}$$

$$\pi_{kp}(S_t) = \begin{cases} p_0 & , S_t \ge X_{kp} \\ p_0 - X_{kp} + S_t & , S_t < X_{kp} \end{cases}, . \tag{3.4}$$

where:

S_t – market price of underlying asset at period t, $t \in \{0,1,2,\ldots,T\}$, $S_t \in [0, +\infty\rangle$,

X – exercise price, $X \in [0, +\infty\rangle$,

c_0 – call option premium at period t=0, $t \in \{0,1,2,\ldots,T\}$,

p_0 – put option premium at period t=0, $t \in \{0,1,2,\ldots,T\}$,

T – time to maturity,
π_{di} – profit from a long asset strategy,
π_{dc} – profit from a long call strategy,
π_{kp} – profit from a short put strategy.

Because of simplicity, we assume that S_0=X_{dc}=X_{kp}. Also, we note:

$$c_0(S_{0,} X, T) > p_0(S_{0,} X, T) \;\; za \;\; \forall S_t, X, T \in [0, +\infty\rangle. \tag{3.5}$$

Moreover, the assumption is that $X=S_0 > c_0$. Hence, $X=S_0 > p_0$.

The elasticity of profits from observed strategies with respect to changes of market price is calculated using the following expression:

$$E_{\pi, S_t} = \frac{S_t}{\pi} \cdot \frac{d\pi}{dS_t}. \tag{3.6}$$

4 THE ELASTICITY OF PROFIT FROM LONG ASSET

According to (3.2) elasticity of profit from long asset is

$$E_{\pi_{di},S_t} = \frac{S_t}{\pi_{di}} \cdot \frac{d\pi_{di}}{dS_t} = \frac{S_t}{S_t - S_0} \quad \text{for } \forall S_t \in \left[0, +\infty\right\rangle = \mathcal{D}_\pi . \tag{4.1}$$

Function π_{di} is elastic when the elasticity coefficient is greater than 1 in its absolute value, and this is when $S_t \geq \frac{S_0}{2}$.

Function E_{π_{di},S_t} is fractioned and has a discontinuity at S_t=S_0. Thus, the elasticity is analysed in intervals $\left[0, S_0\right\rangle$ and $\left\langle S_0, +\infty\right\rangle$.

At a point S_t=S_0 the function has a fraction since

$$\lim_{S_t \to S_0^{-0}} \frac{S_t}{-S_0 + S_t} = -\infty \text{ , and } \lim_{S_t \to S_0^{+0}} \frac{S_t}{-S_0 + S_t} = +\infty , \tag{4.3}$$

and, obviously, there is a vertical asymptote S_t=S_0, i.e. profits are perfectly elastic when S_t=S_0.

Given that

$$\lim_{S_t \to +\infty} \frac{S_t}{-S_0 + S_t} = 1 \text{ and } \lim_{S_t \to 0} \frac{S_t}{-S_0 + S_t} = 0 , \tag{4.4}$$

the line $E = 1$ is horizontal asymptote of the function E_{π_{di},S_t} .

Hence $\frac{\partial E_{\pi,S_t}}{\partial S_t} = \frac{-S_0}{\left(-S_0 + S_t\right)^2} < 0$, $\forall S_t \in \mathcal{D}_E$, function of elasticity of profits from long assets is strictly decreasing. Also, $\frac{\partial^2 E_{\pi,S_t}}{\partial {S_t}^2} = \frac{-2}{\left(-S_0 + S_t\right)^3} \begin{cases} < 0 \text{ , } \left[0, S_0\right\rangle \\ > 0 \text{ , } \left\langle S_0, +\infty\right\rangle \end{cases}$. It follows that in $\left[0, S_0\right\rangle$ function E_{π_{di},S_t} is decreasing and concave, while in $\left\langle S_0, +\infty\right\rangle$ decreasing and convex.

5 ELASTICITY OF PROFIT FROM A LONG CALL

Elasticity of profit from a long call is analysed on $\left[0, X\right]$ and in $\left\langle X, +\infty\right\rangle$.

For $S_t \leq X$, we note that $\pi_{dc}(S_t) = -c_0$ and the elasticity coefficient, according to (2.3), equals 0, meaning. the profit from long call is perfectly inelastic. For $S_t > X$, $\pi_{dc}(S_t) = -c_0 - X + S_t$ and the degree of option leverage, i.e. the elasticity is:

$$E_{\pi_{dc},S_t} = \frac{S_t}{\pi_{dc}} \cdot \frac{d\pi_{dc}}{dS_t} = \frac{S_t}{-c_0 - X + S_t} \cdot 1 = \frac{S_t}{-c_0 - X + S_t} . \tag{5.1}$$

In order to draw some conclusions about elasticity E_{π_{dc},S_t} in $\left\langle X, +\infty\right\rangle$, we need to analyse it as a function of market price in specified interval. The function $\pi_{dc}(S_t)$ is elastic if E_{π_{dc},S_t} in its absolute value is greater than 1, and this applies for $S_t \geq \frac{c_0 + X}{2}$. Since $\frac{c_0 + X}{2} < X$, it is obvious that π_{dc} is elastic in $\langle X, +\infty \rangle$. We note that function of elasticity in $\langle X, +\infty \rangle$ is a discontinuous rational function with a fraction at a point $S_t = c_0 + X$. Therefore, we analyse the function in $\langle X, +\infty \rangle \setminus \{X + c_0\}$.

It follows that:

$$E_{\pi_{dc},S_t} = \begin{cases} 0 & , S_t \le X \\ E_{\pi_{dc},S_t} < -1 & , X < S_t < X + c_0 \\ E_{\pi_{dc},S_t} > 1 & , S_t > X + c_0 \end{cases}. \tag{5.2}$$

Given that $\lim\limits_{S_t \to c_0 + X^{+0}} \dfrac{S_t}{-c_0 - X + S_t} = +\infty$ and $\lim\limits_{S_t \to c_0 + X^{-0}} \dfrac{S_t}{-c_0 - X + S_t} = -\infty$, the parallel to the π-axis containing the point $S_t = c_0 + X$ is a vertical asymptote of function E_{π_{dc},S_t}. The limit value of the function when S_t tends to the lower limit of the interval $\langle X, X + c_0 \rangle$ is: $\lim\limits_{S_t \to X^{+0}} E_{\pi,S_t} = \lim\limits_{S_t \to X^{+0}} \dfrac{S_t}{-c_0 - X + S_t} = -\dfrac{X}{c_0}$. $X > -c_0$, therefore the function is elastic and negative.

Since $\lim\limits_{S_t \to +\infty} \dfrac{S_t}{-c_0 - X + S_t} = 1$, line $E = 1$ is horizontal asymptote of the function E_{π,S_t}. Given that $\dfrac{\partial E_{\pi_{dc},S_t}}{\partial S_t} = \dfrac{-c_0 - X}{(-c_0 - X + S_t)^2} < 0$ for $\forall S_t \in \langle X, +\infty \rangle \setminus (X + c_0)$, function $\pi_{dc}(S_t)$ is strictly decreasing in its domain. Moreover, thus

$$\frac{\partial^2 E_{\pi,S_t}}{\partial S_t^{\,2}} = \frac{2 \cdot (c_0 + X)}{(-c_0 - X + S_t)^3} \begin{cases} < 0 & , S_t \in \langle X, X + c_0 \rangle \\ > 0 & , S_t \in \langle X + c_0, +\infty \rangle \end{cases}, \tag{5.3}$$

we find that for $S_t \in \langle X, X + c_0 \rangle$ the function is decreasing and concave, whereas for $S_t \in \langle X + c_0, +\infty \rangle$ it is decreasing and convex.

6 ELASTICITY OF PROFIT FROM A SHORT PUT

Elasticity of profit from a short put is analysed in $[0, X\rangle$ and in $\langle X, +\infty \rangle$.

For $S_t \in [X, +\infty\rangle$ is follows that $\pi_{kp}(S_t) = p_0$ and elasticity equals zero, i.e. function is perfectly inelastic. For $S_t \in [0, X\rangle$, we have that $\pi(S_t) = p_0 - X + S_t$ and the degree of leverage, i.e. elasticity is:

$$E_{\pi_{kp},S_t} = \frac{S_t}{p_0 - X + S_t}. \tag{6.1}$$

We need to analyse the interval $[0, X\rangle$, in which the elasticity is defined as a function of market price, more thoroughly.

Function $\pi_{kp}(S_t)$ is elastic when the elasticity coefficient is in its absolute value greater than 1 and this applies for $S_t > \dfrac{X - p_0}{2}$. Since $\dfrac{X - p_0}{2} < X$ we note that function

$\pi_{kp}(S_t)$ is elastic in the interval $\left\langle \frac{X-p_0}{2}, X \right\rangle$. Function E_{π_{kp},S_t} is discontinued and has a fraction at $S_t=X-p_0$. Elasticity of profit from short put is analysed in $\left[0, X-p_0\right\rangle \cup \left\langle X-p_0, X\right\rangle$.

Obviously,

$$E_{\pi_{kp},S_t} = \begin{cases} 0 & , S_t \geq X \\ E_{\pi_{kp},S_t} > 1 & , X-p_0 < S_t < X \\ E_{\pi_{kp},S_t} < -1 & , \frac{X-p_0}{2} < S_t < X-p_0 \\ -1 < E_{\pi_{kp},S_t} < 0 & , S_t < \frac{X-p_0}{2} \end{cases} . \tag{6.2}$$

Since $\lim\limits_{S_t \to -p_0+X^{+0}} \frac{S_t}{p_0-X+S_t} = +\infty$, and $\lim\limits_{S_t \to -p_0+X^{-0}} \frac{S_t}{p_0-X+S_t} = -\infty$, it follows that vertical asymptote contains the fraction point of function E_{π_{kp},S_t}.

Given that $\lim\limits_{S_t \to 0} E_{\pi_{kp},S_t} = \lim\limits_{S_t \to 0} \frac{S_t}{p_0-X+S_t} = 0$ and $\lim\limits_{S_t \to X} E_{\pi_{kp},S_t} = \lim\limits_{S_t \to X} \frac{S_t}{p_0-X+S_t} = \frac{X}{p_0}$, where $X > p_0$, function $\pi_{kp}(S_t)$ is elastic and elasticity coefficient is positive.

In addition, we get that $\frac{\partial E_{\pi_{kp},S_t}}{\partial S_t} = \frac{p_0-X}{\left(p_0-X+S_t\right)^2} < 0$ for $\forall S_t \in \left[0, X\right\rangle \setminus \left\{X-p_0\right\}$, so the function E_{π_{kp},S_t} is strictly decreasing.

Furthermore, we have that $\frac{\partial^2 E_{\pi_{kp},S_t}}{\partial {S_t}^2} = \frac{2\left(X-p_0\right)}{\left(p_0-X+S_t\right)^3} = \begin{cases} < 0 \text{ , } \forall S_t \in \left[0, X-p_0\right\rangle \\ > 0 \text{ , } \forall S_t \in \left\langle X-p_0, X\right\rangle \end{cases}$.

Hence, the function of elasticity E_{π_{kp},S_t} is in $\left[0, X-p_0\right\rangle$ decreasing and concave whereas in $\left\langle X-p_0, X\right\rangle$ decreasing and convex.

7 THE COMPARISON OF ELASTICITY OF OBSERVED STRATEGIES

In order for elasticity of option strategies to be greater, in its absolute value, than elasticity of long asset position, the following conditions must be met:

$$\begin{aligned} \left|E_{\pi_{di},S_t}\right| < \left|E_{\pi_{dc},S_t}\right| \Rightarrow \left|\frac{S_t}{S_t-S_0}\right| < \left|\frac{S_t}{-c_0-X+S_t}\right| \Rightarrow \left|S_t-S_0\right| > \left|S_t-c_0-X\right| \\ \left|E_{\pi_{di},S_t}\right| < \left|E_{\pi_{kp},S_t}\right| \Rightarrow \left|\frac{S_t}{S_t-S_0}\right| < \left|\frac{S_t}{p_0-X+S_t}\right| \Rightarrow \left|S_t-S_0\right| > \left|S_t+p_0-X\right| \end{aligned} . \tag{7.1}$$

First condition is fulfilled for all $S_t \in \left[\frac{c_0+X+S_0}{2}, +\infty\right\rangle$, and the second one for all $S_t \in \left[0, \frac{-p_0+X+S_0}{2}\right\rangle$. However, the previous does not mean that long position in asset outside $\left[\frac{c_0+X+S_0}{2}, +\infty\right\rangle$ has a priority over long call which is on $\left[0, X\right]$ perfectly inelastic.

Long asset is on $\left[0, \frac{S_0}{2}\right]$ inelastic, but the elasticity coefficient is positive, thus being less favourable for an investor in long asset. Concerning short put, it is perfectly inelastic in interval in which long asset is elastic, thus being unfavourable for a short put investor.

8 CONCLUSION

Profit functions of long call and short put are elastic, therefore these strategies enable the multiplication of their profit with regards to unit change of market price of the underlying assets because their elasticity coefficients, in their absolute values, are greater than one, but only in certain intervals.

Long call strategy is elastic in $\left\langle X, +\infty \right\rangle$, i.e. when the market price is greater than executive price, while short put strategy is elastic in $\left\langle \frac{X - p_0}{2}, X \right\rangle$, while. inelastic in interval $\left[0, \frac{X - p_0}{2}\right\rangle$. Elasticity coefficients of both option strategies are negative and their absolute value increases up to the point of fraction in which the profits are perfectly elastic, and after which coefficients become positive but start decreasing in their absolute value tending to the horizontal asymptote $E = 1$. In comparison with elasticity of the profit from long position in asset, we conclude that for $S_t > \frac{c_0 + X + S_0}{2}$, profit from long call strategy is more elastic which is favourable for the investor. Short put strategy, on the other hand, is more elastic if $S_t < \frac{-p_0 + X + S_0}{2}$, which is unfavourable and risky for the investor. The option leverage, in terms of elasticity, must be considered with caution, because, as we have shown, it can multiply gains as well as losses. Let us note that option leverage should be analysed separately for each option strategy.

References

[1] Andrijanić, I. (1997). *Trgovanje na robnim burzama.* Zagreb: Mikrorad.

[2] Martić, L. (1992). *Matematičke metode za ekonomske analize* (9.izdanje, I. svezak). Zagreb: Narodne novine.

[3] Neralić, L., Šego, B. (2009). *Matematika.* Zagreb: Element.

[4] Orsag, S. (2011). *Vrijednosni papiri.* Sarajevo: Revicon.

[5] Rosenberg, J. M. (1993). *Dictionary of Investing.* New York: John Wiley & Sons.

MODELLING STOCK EXCHANGE INDEX RETURNS IN 12 NEW MEMBER STATES WITH A TREE-BASED APPROACH

Alenka Kavkler* and **Mejra Festic****
* Faculty of Economics and Business, University of Maribor,
Razlagova 14, 2000 Maribor, Slovenia
** Bank of Slovenia[1]
Slovenska 35, 1505 Ljubljana, Slovenia
e-mails: alenka.kavkler@uni-mb.si, mejra.festic@bsi.si

Abstract: In this paper, we apply model-based recursive partitioning (MOB) to examine the stock exchange index returns for the panel of 12 New Member States in the last ten years. The MOB algorithm uses the generalized M-fluctuation test to check for parameter instability. As explanatory variables we employ GDP, interest rates and other financial and macroeconomic variables. The algorithm yields three terminal nodes pointing to rate of inflation as the splitting variable. The results are discussed and interpreted in light of the current economic situation.

Keywords: model-based recursive partitioning, regression trees, financial markets, stock returns.

1 INTRODUCTION

There is a great deal of empirical literature on the macroeconomic factors influencing stock market indices. Several empirical analyses of stock market integrations revealed that the main economic variables such as real GDP, trade balances, exchange rates, interest rates, consumer price index, public debt and unemployment are significant in their relation to the indices of the stock market. Drawing upon the methods used by the authors who have dealt with the relation of stock market indices, we analyzed the impact of several financial and macroeconomic variables on the returns of stock market indices in the 12 New Member States (NMSs).

Our method of choice is a special kind of tree-models, namely model-based recursive partitioning (MOB) developed by Zeileis, Hothorn and Hornik (2008). The authors carry further the integration of parametric models into trees by providing a framework that embeds recursive partitioning into statistical model selection. After fitting a segmented parametric model, a tree is obtained for which every terminal node is associated with its own model, often a linear regression.

The paper is organized as follows. Chapter 2 describes the technical details about the model-based recursive partitioning algorithm. Data specifications and empirical results can be found in chapter 3 and discussion in chapter 4. The implications of the empirical analysis are revisited in the conclusion.

2 MODEL-BASED RECURSIVE PARTITIONING

We start with a parametric model $M(Y,\theta)$ and *n* observations (vectors) $Y_1, Y_2, \ldots, Y_n$. The estimate of the *k*-dimensional parameter vector θ can be obtained by minimizing some objective function $G(Y,\theta)$, i.e.

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} G(Y_i, \theta). \tag{1}$$

[1] The statements of the author do not present the statements of the Bank of Slovenia nor its councils.

M-type estimators can be implemented in this way, including OLS and maximum likelihood. For example, in case of OLS, the function G is given by the sum of squared residuals. The idea behind the MOB approach is that while it is often not reasonable to assume the existence of a model that fits well on all observations, one may partition the data with respect to one or more variables to obtain locally well-fitting models[2]. For l partitioning variables Z_j, $j = 1,2,\ldots,l$, we assume the existence of the partition

$$\{\mathcal{B}_b\}_{b=1,\ldots,B} \tag{2}$$

consisting of B cells (segments) with the property that in each cell $\mathcal{B}_b$ a local model $M(Y,\theta_b)$ holds. The segmented model is denoted by $M_{\mathcal{B}}(Y,\{\theta_b\})$. Classification and regression trees are special cases of segmented models.

In case of several ($l > 1$) splitting variables, determining the optimal partition is complicated as the number of all possible partitions becomes too large for an exhaustive search. The authors therefore suggest a greedy forward search (where the objective function can be optimized locally in each step) to find a partition that is close to the optimal one. A single parametric model is assigned to each node.

The MOB algorithm uses the generalized M-fluctuation test to examine the parameter stability for a given node. The suggested algorithm is carried out in four steps: (i) the model is fitted to all observations in the current node by minimizing the objective function, (ii) the test for parameter instability is performed, (iii) in case of an existent parameter instability, the variable associated with the highest instability is chosen, (iv) the current node is split into daughter nodes and the procedure is repeated.

When mild regularity conditions hold (for details see Zeileis, Hothorn and Hornik (2008)), the estimate of the parameter vector θ can be obtained by solving first order conditions

$$\sum_{i=1}^{n} \psi\left(Y_i,\hat{\theta}\right) = 0\,, \tag{3}$$

where

$$\psi(Y,\theta) = \frac{\partial G(Y,\theta)}{\partial\theta} \tag{4}$$

stands for the score function. For many interesting models, one can use well-known algorithms for computing the estimate of θ, i.e. OLS estimation with QR decomposition in case of linear regression. We denote the score function evaluated at the estimated parameters as $\hat{\psi}_i = \psi\left(Y_i,\hat{\theta}\right)$.

The next question to answer is whether the parameters of the fitted model are stable over the ordering implied by the partitioning variables or whether one could improve the fit by splitting the sample according to one of the variables Z_j. Parameter instability can be tested by checking if the scores $\hat{\psi}_i$ fluctuate randomly around their mean (which is 0) or if there are any systematic deviations from 0 over Z_j. For this purpose, we evaluate the empirical fluctuation process

$$W_j(t) = \hat{J}^{-\frac{1}{2}} n^{-\frac{1}{2}} \sum_{i=1}^{\lfloor nt \rfloor} \hat{\psi}_{\sigma(Z_{ij})}\,, \qquad 0 \le t \le 1, \tag{5}$$

[2] Our notation and description of this methodological approach is based on the paper by Zeileis, Hothorn and Hornik (2008).

where $\hat{J}$ is na estimate of the covariance matrix $\mathrm{COV}\left(\psi\left(Y,\hat{\theta}\right)\right)$, for example $\hat{J}=\frac{1}{n}\sum_{i=1}^{n}\psi\left(Y_i,\hat{\theta}\right)\psi\left(Y_i,\hat{\theta}\right)^T$. $\sigma\left(Z_{ij}\right)$ stands for the permuation that orders the elements of the vector $Z_j=\left(Z_{1j},\ldots,Z_{nj}\right)^T$ in increasing order. $W_j(t)$ is therefore the partial sum process of the scores ordered by the variable Z_j and scaled by *n* (number of observations) and $\hat{J}$. According to the functional central limit theorem (Zeileis and Hornik 2007), this empirical fluctuation process converges to a Brownian bridge W^0 under the null hypothesis of parameter stability. Brownian bridge is given by equation $W^0(t)=W(t)-tW(1)$, where W is a Brownian motion, characterized by 3 properties: (1) *W(0) = 0*; (2) *W* is almost surely continuous; (3) $W(t)$ has independent increments and $W(t)-W(s)\sim N(0,t-s)$ for $0\le s<t$. One can derive a test statistic by computing $\lambda\left(W_j(.)\right)$, where λ is a scalar functional capturing the fluctuation of the empirical process. The corresponding limiting process is $\lambda\left(W^0(.)\right)$. This general framework for testing parameter instability was developed by Zeileis and Hornik (2007) and is called generalized M-fluctuation test.

For assessing a numerical variable Z_j, the instabilities can be captured with the help of the functional

$$\lambda_{\sup LM}(W_j)=\max_{i=\underline{i},\ldots\bar{i}}\left(\frac{i}{n}\cdot\frac{n-i}{n}\right)^{-1}\left\|W_j\left(\frac{i}{n}\right)\right\|_2^2, \tag{6}$$

defined as the maximum of the squared L_2 norm of the empirical fluctuation process scaled by its variance function. In this way we obtain the well-known sup*LM* statistics of Andrews (1993) that represents the supremum of *LM* statistics against the alternative of a single change point which is shifted over the interval $\left[\underline{i},\bar{i}\right]$. This interval is usually defined with some minimal segment size $\underline{i}$ where $\bar{i}=n-\underline{i}$. The idea behind this test approach could be explained as: to test for a single abrupt change of unknown location, we compute the test statistic for several potential change points and reject if their maximum is too large.

In the next step step, the fitted model is split with respect to the variable Z_j that exhibits the highest instability into a segmented model with *B* segments. *B* can be fixed or determined adaptively. For a fixed number of splits, two rival segmentations can be compared by comparing the segmented objective function. The package party in software R that we use in our empirical analysis implements only the simplest case *B* = 2.

3 EMPIRICAL ANALYSIS

To arrive at appropriate specifications in the spirit of the theoretical suggestions, we had to investigate the time series properties of the data. The following variables were employed in our empirical analysis as possible predictors for the returns of the NMS-12 stock exchange indices (R): rate of inflation (INFL), gross domestic product (GDP), money market interest rate (IR), public debt (DEBT), unemployment rate (U) and real effective exchange rate (REER). Variable names are given in brackets. Quarterly data for the period from the fourth quarter of 2000 till the second quarter of 2010 were obtained from Eurostat, Yahoo Finance database, national central banks and SourceOECD database. Unfortunately, we were not able to obtain the data on monetary aggregates for all of the countries.

In a preliminary analysis, unit root tests were applied to all variables and three variables (GDP, public debt and real effective exchange rate) turned out to be I(1).

Therefore, annual growth rate variables expressed as % changes were employed in our study and the suffix _GRW was added to variable names. Next, we performed the model-based recursive partitioning algorithm. MOB yielded three terminal nodes (leaves) pointing to rate of inflation as the splitting variable. During different inflation regimes, the observed explanatory variables thus impact the stock exchange index returns with varying intensity. The models in the terminal nodes are as follows:

- model 1 is determined by low inflation, i.e. annual rate of inflation less or equal to 2.58 %;
- model 2 is a higher inflation model with annual rate of inflation over 4.65 %;
- model 3 is determined by a medium annual rate of inflation (between 2.58 % and 4.65 %).

Results are shown in Table 1 below.

Table 1: Model-based recursive partitioning results

Model 1: INFL <= 2.58 % 150 observations				
Variable	**Estimate**	**Std. Error**	**t-value**	**p-value**
INFL	6.4665	3.0165	2.1437	0.0339
IR	-8.4026	2.1578	-3.8941	0.0002
GDP_GRW	3.2261	0.8398	3.8418	0.0002
DEBT_GRW	0.1220	0.2089	0.5840	0.5602
U	6.2460	1.4742	4.2370	0.0000
REER_GRW	1.3133	0.6768	1.9403	0.0545
R^2	0.352	F-statistic (p-value)		0.0000
Model 2: INFL > 4.65 % 148 observations				
Variable	**Estimate**	**Std. Error**	**t-value**	**p-value**
INFL	2.1839	1.3762	1.5869	0.1149
IR	-2.0234	1.2629	-1.6022	0.1115
GDP_GRW	5.6386	0.9798	5.7549	0.0000
DEBT_GRW	0.1629	0.2450	0.6648	0.5073
U	7.3393	1.3036	5.6300	0.0000
REER_GRW	0.7750	0.5616	1.3799	0.1700
R^2	0.413	F-statistic (p-value)		0.0000
Model 3: 2.58 % < INFL <= 4.65 % 122 observations				
Variable	**Estimate**	**Std. Error**	**t-value**	**p-value**
INFL	-14.6773	5.5202	-2.6588	0.0091
IR	-9.8598	2.3945	-4.1178	0.0001
GDP_GRW	5.9636	1.1237	5.3071	0.0000
DEBT_GRW	0.6613	0.2615	2.5290	0.0129
U	5.6896	1.0185	5.5865	0.0000
REER_GRW	0.8176	0.5612	1.4570	0.1481
R^2	0.519	F-statistic (p-value)		0.0000

4 RESULTS AND DISCUSSION

Our analysis revealed a significant influence of the chosen explanatory variables on the stock exchange indices of the 12 NMS countries. We can confirm the positive influence of GDP and unemployment. The positive impact of real exchange rate is significant (at the 10 % level) only in the low inflation regime. Interest rate, on the other hand, has a negative impact on stock exchange index returns. Similar results were obtained by Maysami et al. (2004) and Lin (2009).

Low inflation has a positive impact on the stock exchange index returns, while the influence of high inflation is not significant at the 10 % level. Medium level inflation negatively influences stock index returns. There is no consensus in theories and empirical evidence about the influence of inflation on stock exchange. The impact can be either negative or positive, depending on the current economic situation (Fisher 1930, Fama 1981, Baele et al. 2004).

The strong GDP growth in the 12 NMS countries, especially in the period from 2004 to 2008 (together with a growth in capital inflow, trade balances and industrial production), significantly influenced the stock exchange dynamics (Adam and Tweneboah 2009, Babetskii et al. 2007). According to our empirical analysis, the impact of GDP growth on stock returns is always positive. The strongest impact of GDP is seen in the regime of medium inflation rate. The majority of studies have confirmed that the dynamics of stock exchange index returns are procyclical with respect to economic growth.

Unemployment has a positive impact on stock exchange index returns. Our results are partially in line with Boyd et al. (2001) who argue that on average an announcement of rising unemployment is good news for stocks during economic expansions and bad news during economic contractions. Therefore, most of the time stock prices increase on news of rising unemployment. Boyd et al. (2001) also provide an explanation for this phenomenon involving two types of information relevant for valuing stock returns that is incorporated into unemployment rate, namely information about future interest rates and about future corporate earnings and dividends. An increase in unemployment typically signals a decline in interest rates, which is positive for stock returns, as well as a decline in future corporate earnings and dividends, which has a negative connotation.

Interest rate has always a negative impact on stock exchange index returns, with a marginal significance in the regime of high inflation and the strongest influence in the regime of medium inflation rate. Negative interest rates impact is in line with the theory that stock market returns are usually negatively correlated to interest rates. A rather high interest rate is typical for some some NMSs due to insufficient national accumulation and credit supply potential (Baltzer et al. 2008).

Public debt growth has a significant negative impact only in the regime of medium inflation rate. Similar results were obtained by Presbitero (2010) and Muradoglu (2009). The accession of the candidate countries to the EU required the implementation of reforms that lead to further economic expansion. Implementing reforms that includes cutting government spending is a condition in the Lisbon Treaty. Probably the most important factors driving the acceleration of financial integration are related to the policy measures undertaken by the “new” and "marginal" member states in order to meet European financial standards, including the liberalization of capital accounts, as well as legal and institutional reforms (Christiansen and Ranaldo 2008, Poghossian 2008).

5 CONCLUSION

Our analysis confirmed that the financial system of the 12 NMS countries is cyclically related to some macro economic variables. The strength of the relationship depends on different inflation rate regimes. Low inflation has a positive impact on the stock exchange index returns, while the impact of medium inflation is negative. Interest rate has always a negative influence on the stock exchange index returns, with marginal significance in the regime of high inflation and the strongest influence in the regime of medium inflation. Public debt growth has a significant negative impact only in the regime of medium inflation rate.

References

[1] Adam, A.M. and Tweneboah, G. (2009), „Foreign Direct Investment and Stock Market Development: Ghana's Evidence." *International Research Journal of Finance and Economics* Vol 26.

[2] Andrews, D.W.K. (1993), „ Tests for Parameter Instability and Structural Change With Unknown Change Point." *Econometrica*, 61: 821-856.

[3] Babetskii, I., Komârek L. and Komârkovâ Z. (2007), „Financial Integration of Stock Markets among New EU Member States and the Euro Area." *Finance a úver - Czech Journal of Economics and Finance*, 57: 7-8.

[4] Baele, L., Ferrando, A., Hördahl, P., Krylova, E. and Monnet, C. (2004), „Measuring Financial Integration in the Euro Area." *European Central Bank*, Occasional Paper Series, No. 14, pp. 1–93.

[5] Baltzer, M., Cappiello, L., De Santis, R. and Manganelli, S. (2008), „Measureing financial integration in the New EU Member States." *European Central Bank*, Occasional Paper Series, No. 81.

[6] Boyd, J.H., Jagannathan, R. and Hu, J. (2001), "The Stock Market's Reaction to Unemployment News: Why Bad News is Usually Good for Stocks", NBER Working Paper, No. 8092.

[7] Christiansen, C. and Ranaldo, A. (2008)," Extreme coexceedances in New EU Member States' stock markets", Swiss National Bank, Working Paper 2008/10.

[8] Fama, E. (1981), "Stock returns, real activity, inflation, and money". *American Economic Review*, 71: 545-565.

[9] Fisher, I. (1930), *The theory of interest*, Macmillan, New York.

[10] Lin, S.C. (2009), "Inflation and Real Stock Return Revisited. Economic Inquiry", 47: 783–795.

[11] Maysami, R. C., Howe, L. C., and Hamzah, M. A. (2004), „Relationship between Macroeconomic Variables and Stock Market Indices: Cointegration Evidence from Stock Exchange of Singapore‟s All-S Sector Indices.‟ *Jurnal Pengurusan* 24: 47-77.

[12] Muradoglu, G., Metin, K. and Argac R. (2001), "Is there are long run relationship between stock return and monetary variables: Evidence from emerging market?" *Applied Financial Economics*, 11: 641 – 649.

[13] Poghossian, T. (2008), "Are "New" and "Old" EU Members Becoming More Financially Integrated? A Threshold Cointegration Analysis." *International Economics and Economic Policy*, 6(3): 259 – 281.

[14] Presbitero, F. A. 2010. Total public debt and growth in developing countries. Paper prepared for the conference. Available: http://www.csae.ox.ac.uk/conferences/2011-EDiA/papers/608-Presbitero.pdf.

[15] Zeileis, A. and Hornik, K. (2007), „General M-Fluctuation Tests for Parameter Instability ", *Statistica Neerlandica*, 61(4): 488-508.

[16] Zeileis, A., Hothorn, T. and Hornik, K. (2008), „Model-based Recursive partitioning", *Journal of Computational and Graphical Statistics*, 17(2): 492-514.

ANALYSIS OF CROATIAN TAX REVENUE INDICATORS AND COMPARISON WITH SELECTED COUNTRIES

Snježana Pivac, Blanka Škrabić Perić and **Anita Udiljak**
Faculty of Economics Split, University of Split
Matice hrvatske 31, 21000 Split, Croatia
spivac@efst.hr, bskrabic@efst.hr, anita.udiljak@yahoo.com

Abstract: The goal of this paper is to evaluate Croatian tax revenue indicators through quantitative analysis with selected transition and developed countries in Europe. Taxes are the most important fiscal instruments of revenue collection in each country as well as in Croatia. Tax revenues are of great importance for all countries in the transition and with historical heritage like the Republic of Croatia as a country in development and due to EU accession. The survey includes classification, ranking and comparative analysis of selected countries according to tax revenue indicators.

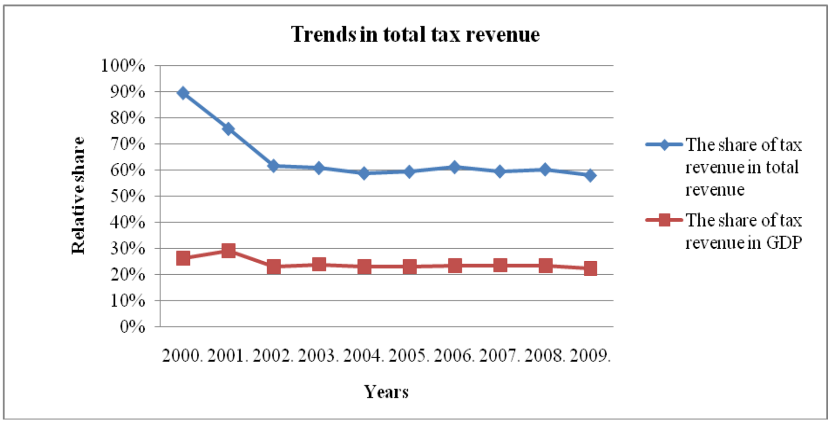
Keywords: tax revenue, quantitative analysis, developed countries, transition countries

1 INTRODUCTION

The role of taxes in financial and hence in the economic life of each country is increasing almost by the day. The financial resources that are collected using many forms of taxation are enormous. Collected taxes are large and significant revenue for the state budget and GDP. Jelčić (1998) concludes that taxes as a financial instrument in some countries in the early 20th century were formed about 10% of GDP, and now there are countries where taxes form for more than 50% of national income [4]. Njavro (2000) pointed out the specifics of the Croatian economy and politics and some features of the eighties: the stagnation of domestic product, the rise in unemployment, the price growth rate over 1000% and a large national debt. Unlike other transition countries, Croatia has gone through the war, in which the expense went up to 40% of GDP; [5] which all had a great influence on fiscal system and the role of tax revenue. Bronic (2009) pointed out that the total tax burden, which is calculated as a proportion of all taxes and social contributions in GDP, in 2007 in Croatia was a little lower than the average in OECD countries [1]. The trend in the world is that developed countries have income-oriented tax systems and the highest taxation is on assets and income with increasing rates of value added tax (VAT). Transition countries have consumption-oriented tax systems, major state tax is VAT and tax revenue is mainly used for short-term government plans [6]. The paper is constructed in five parts. After this introduction part, in first part it is also given the overview of the some former researches about taxes. In second part are presented Croatian tax revenue indicators. The emphasis is on state budget taxes and their contribution to GDP. In the second part of this paper the crisis influences on the tax revenues in Croatia are analyzed using trend models with dummy variables. The tax regulations and the tax burden of selected countries are presented in the third part. In the fourth part of this paper the quantitative analysis of selected countries will be made using the cluster k - means classification method and ranking with the Promethee method for 2008 year with the selected variables. At the end some conclusion remarks are given.

2 ANALYSIS OF CROATIAN TAX REVENUE

Croatian tax system is created on the model of the system of modern states with respect to national specificities. Tax reform ended in 1998 with introduction of VAT. The state taxes as well as all county, cities, municipal and common taxes are prescribed with tax laws and regulations. The backbone of Croatian tax system is VAT and substantial financial income also comes from income tax, excise and income taxes [2]. Figure 1 show that after the collapse in 2001 and 2002, the share of tax revenue in total revenue in Croatia is stable at a level of about 60%.



Source: Yearbooks of the Ministry of Finance, 2000 - 2009

Figure 1: Trends of Croatian total tax revenue

Appropriate trend models are estimated. Regression diagnostics show the validity of the models. The impact of the crisis is examined using dummy variables; coefficient values show a negative impact on total tax revenues and revenues from VAT (Table 1).

Table 1: Trend models of total tax revenue, other tax revenues and revenues from VAT in Croatia since 2000 until 2009 year

Dependent variable	Standardized coefficients		R-square	ANOVA p-values	VIF	Durbin Watson test
	Trend (p-value)	Crisis (p-value)				
Total tax revenues	1.223* (<0.001)**	-0.441 (<0.001)**	0.983	<0.001	1.752	1.778
Other tax revenues	-0.101 (<0.001)**	1.014 (0.410)	0.931	<0.001	1.375	2.346
Revenues from VAT	1.108 (<0.001)**	-0.265 (<0.001)**	0.991	<0.001	1.375	1.547

* -Total tax revenues (t-1)
** p-value<5% - parameters are statistically significant
Note: Testing with Spearman correlation coefficient showed that there is no heteroscedasticity problem in the models.
Note: With Monte Carlo simulations were obtained similar results.

Source: Yearbooks of the Ministry of Finance, 2000 - 2009

3 TAXATION IN SELECTED COUNTRIES

Selected EU developed countries in this analysis are: Denmark, Germany, France, Ireland, Italy and Spain. They all apply a system of taxation in accordance with EU regulations with some specificity. Denmark with the rate of value added tax of 25% and corporate income tax rate of 30% is among the countries with the highest tax burdens. In Germany there are more than 40 types of taxes and tax calculation is extremely complex. In Ireland, corporate income tax is the lowest in Europe and is 12.5%. Italy and France have progressive rates of income taxation with respect to the numerous tax exemptions. In Spain, corporate income tax rate is 35% and there is a special system of taxation in the Canary Islands. The tax burden in selected EU countries is at characteristic levels of each country as a result of historical legacy and political influence. Stationarity of tax burden in observed period indicate the stability of the economy and tax system in selected countries.

Croatian tax system is compared with the tax systems of selected transition countries: Slovenia, Czech Republic, Hungary, Slovakia and Poland. The transition countries are in the process of fiscal adjustment to EU rules. Slovenia has three rates of income tax and two rates of VAT of 20% and 8.5%. In the Czech corporate income tax is paid at a rate of 24%. In Hungary, in corporate taxation there are many tax advantages. Slovakia has the lowest corporate tax rate across the EU of 19%. Poland tax system involves 12 types of taxes and income tax has a complex calculation.

4 COMPARATIVE ANALYSIS ACCORDING TO TAX REVENUE INDICATORS

The selected countries were first classified according to the tax revenue indicators using cluster method. Data taken for this analysis are from the Yearbook Croatian Ministry of Finance [8] and the Organisation for Economic Co-operation and Development [9]. The criteria used for the purpose of this paper were: total tax burden as % of GDP, the share of taxes on income and profits in GDP, the share of taxes on goods and services in GDP and the statutory rate of VAT [10].

Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical variable at a time. In any design and analysis, the technique is used to perform studies across multiple dimensions while taking into account the effects of all variables on the responses of interest [7]. Classification of selected groups of the country is made using k-means cluster method, with assumption of three clusters.

Table 2: Classification of selected EU developed countries (with Croatia) and selected transition countries for 2008 using the k-means cluster method

Developed EU countries (with Croatia)	**Cluster**	**Transition countries**	**Cluster**
Denmark	1	Czech Republic	1
Germany	2	Hungary	1
France	2	Poland	1
Ireland	3	Slovakia	3
Italy	2	Slovenia	1
Spain	3	Croatia	2
Croatia	2		

Source: According to the statistic data of OECD, 2008

The results in Table 2 show that Denmark is classified in a separate group. France, Italy, Germany and Croatia are classified in one group. Ireland and Spain are classified in a separate group. In transition countries classification Slovakia is separated as one group. Czech Republic, Hungary, Poland and Slovenia are classified in the joint group. In another special group was classified Croatia. Empirical p-values for the tax revenue indicators in developed countries are satisfying; except for the share of taxes on goods and services in GDP so we can conclude that other variables contribute significantly to the difference between the clusters (Table 3). For transition countries ANOVA test indicates that the empirical p-value are satisfying, except for statutory VAT rates.

Table 3: K-means cluster ANOVA test for selected tax indicators

ANOVA	Developed countries p-value	Transition countries p-value
Total tax burden as % of GDP	0.052	0.129
The share of taxes on income and profits in GDP	0.019	0.006
The share of taxes on goods and services in GDP	0.285	0.126
Statutory VAT rates	0.145	0.363

Note: The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.
Source: According to the statistic data of OECD, 2008

After classification the countries are ranked according to the same tax revenue indicators using the PROMETHEE method. The PROMETHEE method is appropriate to treat the multicriteria problem of the following type:

$$Max\{f_1(a),..., f_n(a) | a \in K\}, \tag{1}$$

where K is a finite set of possible actions (here countries), and f_j are n criteria to be maximized. For each action $f_j(a)$ is an evaluation of this action. When we compare two actions, $a, b \in K$, we must be able to express the result of this comparison in terms of preference. We, therefore, consider a preference function P:

$$P : K \times K \rightarrow [0,1], \tag{2}$$

representing the intensity of action a with regard to action b. In practice, this preference function will be a function of the difference between the two evaluations $d = f(a) - f(b)$, and it is monotonically increasing. Six possible types (usual, U-shape, V-shape, level, linear and Gaussian) of this preference function are proposed to the decision maker [7]. The effective choice is made interactively by the decision maker and the analyst according to their feeling of the intensities of preference. In each case, zero, one, or two parameters have to be fixed: q is a threshold defining an indifference area; p is a threshold defining a strict preference area; s is a parameter the value of which lies between p and q.

Now, we can define a preference index:

$$\Pi(a,b) = \frac{\sum_{j=1}^{n} w_j P_j(a,b)}{\sum_{j=1}^{n} w_j}, \tag{3}$$

where w_j are weights associated with each criteria.

Finally, for every $a \in K$, let us consider the two following outranking flows: leaving flow: $\phi^+(a) = \sum_{b \in K} \Pi(a,b)$; entering flow: $\phi^-(a) = \sum_{b \in K} \Pi(b,a)$.

The leaving flow ϕ^+ is the measure of the outranking character of a (indicates how a dominates all other actions of K). Symmetrically, the entering flow ϕ^- gives the outranked character of a (indicates how a is dominated by all other actions). When the decision maker requests the complete ranking, the net outranking flow may be considered:

$$\phi(a) = \phi^+(a) - \phi^-(a) \tag{4}$$

And the higher the net flow, the better the action is. All the actions of *K* are now completely ranked (PROMETHEE II). For each criterion, one of the six offered preference function types and its thresholds have been chosen. The choice of the function types, and its thresholds, was carried out according to the economic theory and countries data. In addition to that, the final ranking is obtained by cumulating mutual comparisons of alternative pairs, according to all the criteria, into final leaving and entering flows, i.e. the final rank of alternatives. The group of alternatives consists of 12 countries which are compared according to the 4 previously observed criteria. The types and weight values of all criteria for 2008 are shown in Table 4. The sum of all criteria values equals one hundred, as it should be [3]. Total tax burden as % of GDP dominates the remaining tax revenue criteria. Statutory VAT rates are the lowest weight. The country is attractive to investors if there are fewer taxes, so the function of the minimum is used for all criteria.

Table 4: The weights and types of criteria in PROMETHEE method for selected countries

Variable	**Total tax burden as % of GDP**	**The share of taxes on income and profits in GDP**	**The share of taxes on goods and services in GDP**	**Statutory VAT rates**
Min/Max	Minimize	Minimize	Minimize	Minimize
Weight	0.40	0.25	0.25	0.10
Preference Function	Linear	Gaussian	V-shape	Gaussian

Source: According to the statistic data of OECD, 2008

Table 5: Ranking of selected countries using the PROMETHEE method for 2008

	Country	**Phi**		**Country**	**Phi**
1.	Spain	0.11	**7.**	Italy	0.02
2.	Slovakia	0.08	**8.**	Slovenia	0.01
3.	France	0.06	**9.**	Poland	0.01
4.	Ireland	0.06	**10.**	Hungary	-0.03
5.	Germany	0.05	**11.**	Croatia	-0.16
6.	Czech Republic	0.05	**12.**	Denmark	-0.28

Source: According to the statistic data of OECD, 2008.

Table 6: Mann-Whitney U-test results for all tax indicators according to countries developments in 2008

Variable	**Country**	**Mean Rank**	**P-value**
Total tax burden as % of GDP	transition country	5.17	0.200
	developed country	7.83	
The share of taxes on income and profits in GDP	transition country	4.33	0.037
	developed country	8.67	
The share of taxes on goods and services in GDP	transition country	8.17	0.109
	developed country	4.83	
Statutory VAT rates	transition country	6.83	0.745
	developed country	6.17	

Source: According to the statistic data of OECD, 2008.

After the analysis has been carried out, the final rank of alternatives according to the tax revenue indicators is given in Table 5. Spain holds the leading position. Developed countries are ranked in the first half of the group with the exception of Denmark. Also, the classifying cluster analysis confirms that Denmark constitutes a class for itself. Transition countries are at the bottom of ranking with the exception of Slovakia. Using cluster classification Slovakia is classified as an entropy group. Hungary, Croatia and Denmark are ranked as the last.

Using nonparametric Mann-Whitney U test selected countries were tested for tax indicators rank differences between developed and transition countries in 2008 (Table 6). Based on the performed testing we can conclude that in 2008 there was a statistically significant difference in the size of the share of taxes on income and profits in GDP in developed EU countries compared to countries in transition. There were no significant differences in other variables between these groups of countries.

5 CONCLUSION

This paper statistically analyzes the state budget tax revenues and tax burden in Croatia in the period since 2000 until 2009. The survey also included qualitative and comparative analysis with selected developed EU countries and selected transition countries in Europe. It can be concluded that the transition countries that joined the EU and Croatia have a high proportion of tax revenue to total revenue of the country as well as the high share of GDP (around 35 to 39%). The impact on economic flows was negative in the total tax revenues and revenues from VAT. The introduction of “crisis tax” had a positive impact on growth in other income. The reduction in revenue from VAT, the most important tax for the Croatian economy and taxpayers, we can see that the crisis has negative impact on economic trends. Of course, after publishing data from 2010 year, we will be able to see the full impact of the crisis to the state budget. Croatian tax system, as in most transition countries rely on the VAT as the main tax form and has a consumer orientation, and countries in the EU have the income orientation and main taxes are on profit and income.

References

[1] Bronic, M., 2009. Taxes in Croatia and in OECD Countries, Institute of Public Finance, Press releases No. 16. [Available at http://www.ijf.hr/osvrti/16.pdf].

[2] Budic, M., Koharic, Z., Loncaric-Horvat, O., Mratovic, V. Perkovic, M., 2001. General Tax Law with Commentary, the Institute of Public Finance, Zagreb.

[3] Doumpos, M. and Zopounidis C., 2002. On the Use of a Multi-criteria Hierarchial Discrimination Approach for Country Risk Assessment. Journal of Multi-criteria Decision Analysis, 11(4-5), pp. 279-289.

[4] Jelcic, B., 1998. Financial Law and Financial Science, Zagreb, pp. 72 -75.

[5] Njavro, Đ., 2004. Tax Systems in Competition and Countries in Transition: The International Conference, Zagreb, pp. 721-722.

[6] Švaljek, S., 2005. The 2000 Tax Reform in Croatia: Causes and Consequences, Economic Review, Vol.56, No.12, pp. 1217-1236.

[7] Tomic-Plazibat N., Aljinovic Z., Pivac, S., 2010. Risk Assessment of Transition Economies by Multivariate and Multicriteria Approaches**,** Panoeconomicus, No. 3, pp. 283.-302.

[8] http://www.mfin.hr/hr/godisnjaci-ministarstva; [Croatia State Budget Statistics, Croatia, Ministry of Finance] (accessed April 15, 2011).

[9] http://stats.oecd.org/Index.aspx?DatasetCode=CSP2010; [Country Statistical Profiles] (accessed April 15, 2011).

[10] http://epp.eurostat.ec.europa.eu; [European Commission, Eurostat] (accessed April 15, 2011).

IMPACT OF SOCIAL SECURITY ON FERTILITY – A PANEL DATA ANALYSIS

Josip Arnerić, PhD[1], **Nada Pleli, PhD**[2] and **Tihana Škrinjarić**[3]
University of Split – Faculty of Economics
Matice hrvatske 31, 21000 Split – Croatia
University of Zagreb – Faculty of Economics and Business
Trg J.F. Kennedya 6, 10000 Zagreb – Croatia
josip.arneric@efst.hr, npleli@efzg.hr, tihana.skrinjaric@gmail.com

Abstract Falling fertility rates, ageing population and their consequences on the economy are in center of attention of economists for quite some time. Some of the problems are emerging in social security programs, due to financial vulnerability of PAYG systems of pension insurance, a common system in a great number of countries. Since Croatia has a PAYG system for the first and biggest pillar of pension insurance, we examine theoretical and empirical research of interactions between PAYG benefits and total fertility rate. Results suggest that PAYG systems should have been taken into account when making policies for stimulating fertility.

Keywords social security, pension system, total fertility rate, panel analysis, Croatia, PAYG system

1 INTRODUCTION

Much discussion and controversies have been made in Republic of Croatia concerning pension insurance system and its financial vulnerability, especially since 1999 and its first reform. Croatia has a PAYG (pay as you go) system of pension insurance for the first pillar and mandatory, apropos voluntarily pension insurance based on individual capitalized savings for second, apropos third pillar. Negative trends show that the ratio of taxpayers and beneficiaries of pension insurance system has fallen from 1.66 in 1996 to 1.41 taxpayers on one beneficiary in 2010. Such a low ratio means a higher burden for taxpayers, the government has to incur new debts and in the long run it is not sustainable. Rustowski [21] has been warning about the general problems of PAYG systems of pension insurance, because of the risks of population aging, increasing unemployment, etc., problems Croatia is facing today.

On the other hand, discussions are made about the ageing population. There are numerous consequences of the vanishing population; lack of local labor force and its contributions to pension insurance system is one of them. Total fertility rate in Croatia has been falling since 1950s, and in the 1970s it has fallen below the critical limit of 2.1 children per woman. Croatia is not the only country with these problems. Aging population, falling fertility and concerns about sustainability of the PAYG system have been in the center of attention across European countries. UN projections are predicting that until mid-century about 80% of the countries allover world will have total fertility rate lower than necessary for generation renewal [24]. Regarding the pension insurance system, whilst in 2007 adults (15-64 years) made 67.2% of total EU-27 population; in 2050 the share will fall to 57%. This decreased base will support growing old population (65+), predicted to rise to 29% in total population by 2050 [16].

Bejaković [7] in his paper discussed about main reasons for the lack of perspective of the pension systems, and one of them is the weak relationship between the contributions paid and the pensions earned. The very definition of PAYG system itself explains settlement of

[1] Senior Assistant at Faculty of Economics, Split.
[2] Assistant Professor at Faculty of Economics and Business, Zagreb.
[3] Undergraduate student of Economics at Faculty of Economics and Business, Zagreb.

current pension benefits from current expenditures, meaning there is lack of incentive to save for the future. Moreover, the benefits received by retired people are not related nor with contributions made from their own adult children to the system.

This paper is focused on empirical investigation of models which describe relationships between social security and fertility; two great problems Croatia is dealing with today. Consequently, a summary of papers which elaborate those interactions is given in the next section, with focus on relationships between aforementioned matter.

2 THEORETICAL BACKGROUND AND PREVIOUS EMPIRICAL ANALYSIS

Since the second half of the 20th century, issues such as fertility, family formation, social security and their interaction with economic growth have been in focus of economists. It began with theoretical studies of Becker [5], Becker and Lewis [6], about investing in children; Diamond's [10] study of national debt in growth models. Eventually, theory and models of fertility, economic growth, and influences of social security on main growth determinants became center of attention for many economists.

Initial, i.e. fundamental model regarding economic growth and fertility choice was one developed by Barro and Becker [2], where they showed among other, reductions in fertility caused by increasing social security benefits. Following year, 1989, Barro and Becker [3] extended their analysis on population growth, capital accumulation, and income growth, by showing that expansion in social security system lowers steady-state rates of population, i.e. fertility by raising the net cost of children. As they discussed: "Another example comes from the positive effect of a larger social security program on the cost of children. The increase in the net tax on younger persons lowers the net lifetime incomes of children and thereby raises the net cost of children." [2, p. 493]. An important contribution was made in 1990, when Becker, Murphy and Tamura [4] identified the motivating force - parental altruism – in allocating their resources in human capital. A detailed discussion about these two models is given in Hartmann [19]. Ehrlich and Lui [12] developed an overlapping-generations model, where besides altruism, parents invest in children also for their self-interest when they reach old age and depend of their offspring, but they had not considered in full factors such as social insurance programs. Authors did say these programs could diminish the role of the family insurance they analyzed, and concluding that "social security programs of a "pay as you go" nature can reduce an economy's growth potential by providing an inefficient substitute for voluntary family insurance. " [12, p. 1057]

There has been a lot of research regarding influence of social security on saving and capital accumulation. In 1974 Feldstein [18] presented his results that social security depresses personal saving. Barro and MacDonald [1] got inconclusive results due to different sign of the relationship between social security and consumer spending depending of whether they observed time series movements or cross-sectional. They did not find empirical support in their research. Coates and Humphreys [9] commented on Feldstein's [18] results in their analysis on this matter. In their conclusion, they said that implications of social security on private savings are larger today than before. Zhang and Zhang [27] gave a list of more papers about interactions between the mentioned matters of research. In their investigation, they also found that social security affects fertility very negatively, but it tends to stimulate per capita growth; as they said: "Such empirical evidence differs from the popular view in the literature that regards social security as harmful to growth." [27, p. 494].

Wildasin [25] argued that underfunded social security programs and government borrowing may be necessary to take into account while analyzing determinants of fertility, especially in developed countries since major of them have those characteristics. Many authors have analyzed fertility, altruism, human capital regarding social security in growth models. A good review of a numerous number of papers is given in already mentioned work of Zhang and Zhang [27]. Empirical research in this paper is focused on impacts of social security programs on fertility from these growth models, such as recent ones, made by Ehrlich and Zhong [14] and Ehrlich and Kim [11]. For the analytical framework, one can address papers mentioned above. Ehrlich and Kim [11] developed a model and empirically showed that the decline in total fertility rate has been a partial result of PAYG system of social security in most of the countries they analyzed. They emphasized a weak relationship between contributions paid and benefits (pensions) received in PAYG systems. Pensions do not even rely on the contributions made by one's offspring, or whether a retired person even has any children. Therefore, one can look at this matter as a way of receiving no compensation for raising more children.

Following these previous theories; and papers of Ehrlich and Lui [13] and Ehrlich and Zhong [14], we give a brief overview of the theoretical model. The family is a partnership where altruism and old-age insurance are two motivating forces for parenthood, and overlapping generations are mutual dependent. Parents invest in their children, but old parents can also share a portion of their success. This factor, complementary to altruism, promotes human-capital accumulation and growth. Increases, i. e. changes in social-security tax proportional to earnings negatively affect fertility[4]. Thus, PAYG system can have a negative incentive to form families, due to the tax levied in proportion to earnings and "defined" benefits received by old parents. As mentioned before, their pensions do not depend on whether they have children to contribute to the social fund. Probabilities of survival from childhood to adulthood and from adulthood to old age are two of model's basic parameters, as they affect the demand for having children; explain the transition from stagnant to growth equilibrium and as Ehrlich and Kim [11] said: "they capture the rationale for any old-age insurance scheme." [11, p. 117].

Therefore, the characteristics of the model, explaining fertility and social security in the model are being tested in the next section.

3 EMPIRICAL ANALYSIS

The data for our analysis were obtained from statistical databases of OECD, European Commission, i.e. Eurostat, World Bank and Summers-Heston data set (or Penn World Table) [20, 17, 26, 23], for the period 1980-2007. The sample includes annual data for 15 countries[5] across period of 28 years which produced 386 observations (unbalanced panel data). Variables used in the analysis as exogenous were LOASDI – old age, survivor and disability-insurance portion of social security benefits as a share of GDP, PI1 – survival probability of the population from ages 0-24, LPI2(x0-x4) – survival probability of the population from ages 50-74 and 60-84[6], LGDP_P_C – gross domestic product per capita[7], LN_MARRIAGE_RATE – net marriage rate (crude marriage rate net of crude divorce rate)[8], LS_PARTICIPATION_F – female labor force participation rate[9], LG_GDP – government spending as a share of GDP[10], LGDI_P_C – gross domestic income per

---

[4] Besides fertility, it affects investment in human capital and the savings rate [14]. Here, we are focusing on fertility in those growth models.

[5] Belgium, Czech Republic, Denmark, Germany, Ireland, Greece, Spain, Italy, Luxembourg, Netherlands, Portugal, Finland, Sweden, Norway and Switzerland.

[6] We used two variants of probability to survive from adulthood to old age, due to different specifications used in previous (mentioned) studies.

[7] Real GDP per capita, Constant Prices 2005., International Dollars.

[8] Number of marriages on 1000 people, minus number of divorces on 1000 people.

[9] Proportion of women ages 15 and older that is economically active.

[10] General government final consumption expenditure as a % of GDP.

person[11]. The total fertility rate LTFR is defined as endogenous variable. All variables are transformed using natural logs, and therefore the letter L stands for these transformations. Furthermore, log-linear panel model with fixed effects within cross-section terms has been estimated using OLS to account for country-specific variations in regressors. The results of country-specific effects using dummy variables are suppressed. Table 1 presents estimated parameters with standard errors in parentheses, while * and ** denote statistical significance at 1% and 5% level.

Table 1. Estimated panel models of Total Fertility Rate

Exogenous variable	Explanation	Model 1	Model 2	Model 3
CONSTANT	-	-1,3677 (-1.40)	-1,3235 (-1.39)	-2,0701* (-2.14)
LOASDI	Old age, survivor and disability-insurance portion of social security benefits as a share of GDP	-0,3622** (-6.14)	-0,3589** (-6.10)	-0,3434** (-5.84)
LPI1	Survival probability of the population from ages 0-24	-8,3401** (-2.87)	-8,7169** (-2.99)	-9,1835** (-3.27)
LPI2 (50-74)	Survival probability of the population from ages 50-74	0,6742** (2.97)	-	-
LPI2 (60-84)	Survival probability of the population from ages 60-84	-	0,2852** (3.23)	0,2295* (2.51)
LN_MARRIAGE_ RATE	Net marriage rate	0,0509* (2.57)	0,0512** (2.60)	0,0429** (2.14)
LS_PARTICIPATI ON_F	Female labor force participation rate	0,2466** (2.70)	0,2588** (2.87)	0,2891** (3.16)
LGDI_P_C	Gross domestic income per person	0,1022 (1.44)	0,0945 (1.35)	-
LG_GDP	Government spending as a share of GDP	0,2459** (3.73)	0,2488 (3.79)	0,2383** (2.15)
LGDP_P_C	Gross domestic product per capita	-	-	0,1519* (3.63)
Adjusted R Square Observations	-	0,7175 373	0,7188 373	0,7210 373

Source: Authors' calculations, using software EViews 7
Note: L stands for natural logarithm

In Model 1, LPI2 (60-84) and LGDP_P_C are excluded while the alternative indicators are included such as LPI2 (60-84) in Model 2 and LGDP_P_C in Model 3. In all models variables LOASDI and LPI1 have expected negative influence on total fertility rate and statistically significant at p-value less than 1%. All other exogenous variable have positive and statistically significant influence on total fertility rate except LGDI_P_C. The results of estimated parameters from Model 2 do not differ significantly with respect to Model 1, while from Model 3 it can be seen that the impact of LGDP_P_C is significant at p-value less than 5%. Moreover, in Model 3 all estimated parameters are statistically significant, with the

[11] 2005 International dollar per person, PPP converted. We use GDP per capita and GDI per capita due to different specifications used in previous studies.

highest coefficient of determination. In all estimated models robust standard errors with White correction are used to overcome possible heteroskedasticity and serial correlation of error terms. To overcome possible endogeneity of exogenous variable LOASDI (according with previous studies) we estimated panel model with lagged variable LOASDI(t-1). From estimated model variable LOASDI becomes insignificant, while LOASDI(t-1) becomes significant which provides evidence that LOASDI is correctly specified as an exogenous variable.

As it can be seen, in all models the pension variable (LOASDI) has a negative effect on fertility, indicating that higher benefits paid, the higher tax (for paying more benefits) is levied on young adults, thus discouraging incentive to have children. The probabilities to survive have signs in accordance with previous theories and studies[12]. Other variables, relevant in modeling follow. If the marriage survives, it can have positive effects on fertility; if more women participate in economical activities, they are motivated to have more children[13]. Regarding GDI and GDP per capita, we got inconclusive results, because of the contradiction in statistical significance. The positive sign is probably a result of existence of a common trend among variables. Government spending[14] was included in the model, in order to separate effects of other taxes and benefits from effects of PAYG taxes.

4 CONCLUSION

Concern about falling fertility and growing old population in developed countries has exhorted studies which developed models of economic growth that take into account investing in human capital, fertility and social security costs and benefits. Motivated by problems of our own country and facing same problems as many of other countries in the world, we have investigated previous studies about these problems and tried to empirically examine them, with newer data compared to the previous studies. Due to lack of our countries data, we obtained those available of other European countries. The results appear to be consistent with other studies, and suggest that PAYG systems of social security should have been taken into account when making policies for stimulating fertility. Since the PAYG system itself is facing financial vulnerability, it also needs a reasonable consideration about reforming it. Questions emerge. Is merely 5% of the income paid to pension insurance in second pillar for individual capitalized savings enough? What about future generations of pensioners, who will pay contributions for their benefits, considering low (and falling) fertility? Despite the great number of studies dealing with these topics, Croatia’s pension system and trends in fertility are still unsolved. There is a great need for a better awareness of the bearers of economic policies.

References

[1] Barro R. J., MacDonald, G. M. 1979. Social Security and Consumer Spending in an International Cross Section. Journal of Public Economics 11, pp. 275-289

[2] Barro, R., Becker, G., 1988. A reformulation of the economic theory of fertility. The Quarterly Journal of Economics 103 (1), pp. 1-25

[12] If the probability of surviving to adulthood for a child rises, parents are motivated to invest more in its human capital. This leads to having fewer children, in which parents invest more. If the probability of surviving from adulthood to old age rises, young parents are motivated to have more children due to sharing a part of their children’s success, as theoretical models are predicting.

[13] Although it may seem strange at first, but in recent years studies showed that the sign of the relationship between women’s employment and fertility has changed from negative to positive since 1980s. [23, 8].

[14] Here we follow [12].

[3] Barro, R. Becker, G. S., 1989. Fertility Choice in a Model of Economic Growth: Econometrica, Vol. 57, No. 2, pp. 481-501
[4] Becker, G. S., Murphy, K. M., Tamura, R. 1990. Human Capital, Fertility and Economic Growth, Journal of Political Economy 98, pp. 12-38
[5] Becker, G., 1960. An Economic Analysis of Fertility. In: A.J. Coale (ed.), Demographic and Economic Change in Developed Countries. Princeton University Press, Princeton, NJ.
[6] Becker, G., Lewis, H., 1973. On the interaction between the quantity and quality of children, Journal of Political Economy 81 (2), pp. 279-288
[7] Bejaković, P. 2005. Način prikupljanja mirovinskih doprinosa, Review 'Mirovinsko osiguranje' Number 10/11, 2005., Institute of Public Finance
[8] Brewster, K. L., Rindfuss, R. R. 2000. Fertility and women's employment in industrialized nations, Annual Review of Sociology, Vol. 26
[9] Coates D., Humphreys B.R. 1999. Social Security and Saving: A Comment. National Tax Journal 52, pp. 261-268
[10] Diamond, P., 1965. National Debt in a Neoclassical Growth Model. American Economic Review 55, pp. 1125-1150
[11] Ehrlich, I., Kim, J., 2007. Has social security influenced family formation and fertility in OECD countries? An economic and econometric analysis, Pharmaceuticals Policy and Law 9, pp. 99-120
[12] Ehrlich, I., Lui, F. T. 1991. Intergenerational Trade, Longevity, and Economic Growth, The Journal of Political Economy, Vol. 99, No. 5, pp. 1029-1059
[13] Ehrlich, I., Lui, F. T. 1998. Social Security, The Family and Economic Growth, Economic inquiry, Vol 36, Issue 3, pp. 390-409
[14] Ehrlich, I., Zhong, J-G. 1998. Social Security and the Real Economy: An Inquiry into Some Neglected Issuses, The American Economic Reviews, Vol. 88, No. 2, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association, pp. 151-157
[15] European Commision, Eurostat, Population statistics, http://epp.eurostat.ec.europa.eu/portal/page/portal/population/introduction
[16] European Commission 2009. Europe in Figures - Eurostat yearbook 2009, Population
[17] European Commission, Eurostat, http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/
[18] Feldstein MS 1974. Social Security, Induced Retirement, and Aggregate Capital Formation. Journal of Political Economy 23, pp. 905-926
[19] Hartmann, A., M. 2010. Fertility and Economic Growth, How does the fertility rate influence economic growth in developing countries? Master thesis submitted for the degree of Master of Science in International Economic Consulting Academic, Msc in International Economic Consulting, Aarhus School of Business, University of Aarhus
[20] OECD, Statistical Database, http://stats.oecd.org/Index.aspx
[21] Rutkowski, M. 1998. A New Generation of Pension Reforms Conquers the East—A Taxonomy in Transition Economies, Transition – The newsletter about reforming economies, Volume 9, No 4, World Bank
[22] Sleebos, J. 2003. Low Fertility Rates in OECD Countries: Facts and Policy Responses, OECD Labor Market and Social Policy Occasional Papers, No. 15, OECD Publishing, doi:101787/568477207883
[23] Summers-Heston Data Set (Penn World Table), http://www.bris.ac.uk/Depts/Economics/Growth/summers.htm
[24] UN 2005. The Future of Fertility in Intermediate – Fertility Countries, United Nations Department of Economic and Social Affairs/Population Division
[25] Wildasin, D. E. 1990. Non-Neutrality of Debt with Endogenous Fertility. Oxford Economic Papers 42, pp. 414-42
[26] World Bank, http://data.worldbank.org/
[27] Zhang, J., Zhang, J. 2004. How Does Social Security Affect Economic Growth? Evidence from Cross-Country Data, Journal of Population Economics, Vol. 17, No 3, pp. 473-500

Section IX:
Location and Transport

WHERE TO PLACE CROSS-DOCKING POINTS?

Karlo Bala[a], **Dejan Brcanov**[b] and **Nebojša Gvozdenović**[b1]
[a] University of Novi Sad, Faculty of Philosophy, Dr Zorana Đinđića 2, 21000 Novi Sad, Serbia
kbalayu@gmail.com
[b] Faculty of Economics Subotica, Segedinski put 9-11, 24000 Subotica, Serbia
dejanbrcanov@gmail.com, nebojsa.gvozdenovic@gmail.com

Abstract: Recently, driven by industrial applications, researchers started considering facility location and vehicle routing simultaneously [3] and named it location routing. In this paper we consider an example of location routing problem whose main focus is cross-docking points positioning. We present two approaches for determining promising locations for cross-docking points. The first one is dynamic, i.e. meeting points are chosen during a heuristic, while the second one uses a partitioning approach based on minimum spanning tree with positioning done via calculating p-medians.

Keywords: vehicle routing, facility location, cross-docking, multi-echelon

1 INTRODUCTION

We consider a problem of planing distribution that focuses on location and routing decisions. The problem deals with depots, cross-docking points, customers and a heterogeneous fleet of vehicles. Products stored in a depot are distributed to customers directly or via cross-docking points. A cross-docking point is a place where one vehicle brings products, which are then split and loaded into several vehicles. After being loaded according to capacity constraints, the vehicles either start or continue their routes from a cross-docking, and distribute products to customers. All customers have time windows that cannot be violated. The goal is to find good locations for cross-docking points and to generate cost-effective vehicle routes such that all customers are visited by exactly one vehicle within specified time windows. This research is motivated by real life newspaper distribution problems in Denmark and Sweden. The cross-docking point concept has been invented by dispatchers, since it is often crucial for time window constraint satisfaction and for decreasing the total sum of route lengths.

2 PROBLEM

We consider a problem of planning vehicle routes for distribution of a type of a product from a *depot* to *customers*. Each vehicle has a *capacity*, i.e. the max number of products that it can carry. Each customer ordered certain number of products, and requires that products are delivered within certain time interval called a *time window*. For delivery, one will spend a certain time called a *service time.* We also assume that all customers have a parking lot in which several vehicles can meet and where loading and unloading of vehicles can be performed. In other words, we assume that a vehicle can bring products from the depot to a parking lot for other vehicles. In a standard scenario, some portion of products is first unloaded and loaded to other vehicles at a parking lot and then all vehicles continue with their routes. Hence, some vehicles can start their routes at the depot, while others start at customer's parking lots. For any two destinations in our network we know travel distances and times.

[1] The work presented here was supported by the Serbian Ministry of Education and Science (projects III 44006 and OI 174018)

The main goal is to build a route plan such that the total number of kilometres is minimized, and all vehicle capacities and time window constraints are satisfied.

3 MODEL

We model an underlying network of the problem as a directed graph with nodes *V={0,1,2,...,n}* and arcs $E=\{(i,j)\,|\,i\neq j\}$. Here, *0* is the depot node and *1,2,...,n* are the nodes representing customers. For $i\in V$ we denote by *o(i), s(i), tstart(i), tend(i), t(i)* its number of ordered products, service time, earliest delivery start, latest delivery end, and the start of a delivery, respectively. Obviously, constraints

$tstart(i) \leq t(i) \leq tend(i)-s(i)$,

have to be satisfied for all $i\in V\setminus\{0\}$. Nevertheless, for the node *0* we assume *o(i)=0* and *s(i)=tstart(i)=tend(i)=t(i)=0*. We also denote by *o(V')* the number of ordered products by all nodes from $V'\subseteq V$. With an arc *(i,j)* we associate its travel distance *d(i,j)* and travel time *t(i,j)*. Further, we denote by *C* the set of vehicles, and by *c(v)* the capacity of a vehicle *v* from *C*. Vehicles in our model correspond to route plans, hence w.l.o.g. we assume that a vehicle *v* is a tuple of nodes from *V*. For example if $v=(v_0, v_1,\dots, v_k)$ we write $v_j\in v$, for *j=0,1,2,...,k*, we say that vehicle *starts* from v_0, and we assume that vehicle *v* delivers $o(v):=\sum_{j=1}^{k} o(v_j)$ products to customers. We assume that vehicle is loaded in v_0, and products for its order, if exists, are not loaded into vehicle. We call a node *i* a *cross-docking point* if $i\neq 0$ and there exists a vehicle from *C* which starts from *i*. In order to satisfy orders, each customer node has to be visited exactly ones, i.e. it has to belong as a non starting node to exactly one vehicle. We restrict vehicles to contain at most one cross-docking point, and for a vehicle *v* we denote it by *cd(v)*. Capacity constraints then read:

$o(v)\leq c(v)$ if vehicle *v* starts at a cross-docking point *cd(v)*, and

$\sum_{cd(v)\in u} o(u)\leq c(v)$ if vehicle *v* does not start at *cd(v)*.

In other words, a vehicle which starts from a cross-docking point carries only products that it delivers to customers, while a vehicle which starts from a depot in addition contains products that are unloaded in its cross-docking point (if such exists).

For a vehicle *v* we assume that it starts from $t(v_0)$, and we require

$t(v_{j-1}) + t(v_{j-1}, v_j) \leq t(v_j)$ for $j\geq 0$.

Finally, the optimization problem is to minimize $\sum_{v}\sum_{j\geq 0} d(v_{j-1}, v_j)$ with respect to time and capacity constraints listed above, and we call it *cross-docking placement problem* or shortly CDP. Note, that one can not expect to have efficient exact algorithms for CDP (unless P=NP), since a version with a single vehicle and dropped capacity and time constraints is the famous travelling salesman problem.

4 HEURISTICS

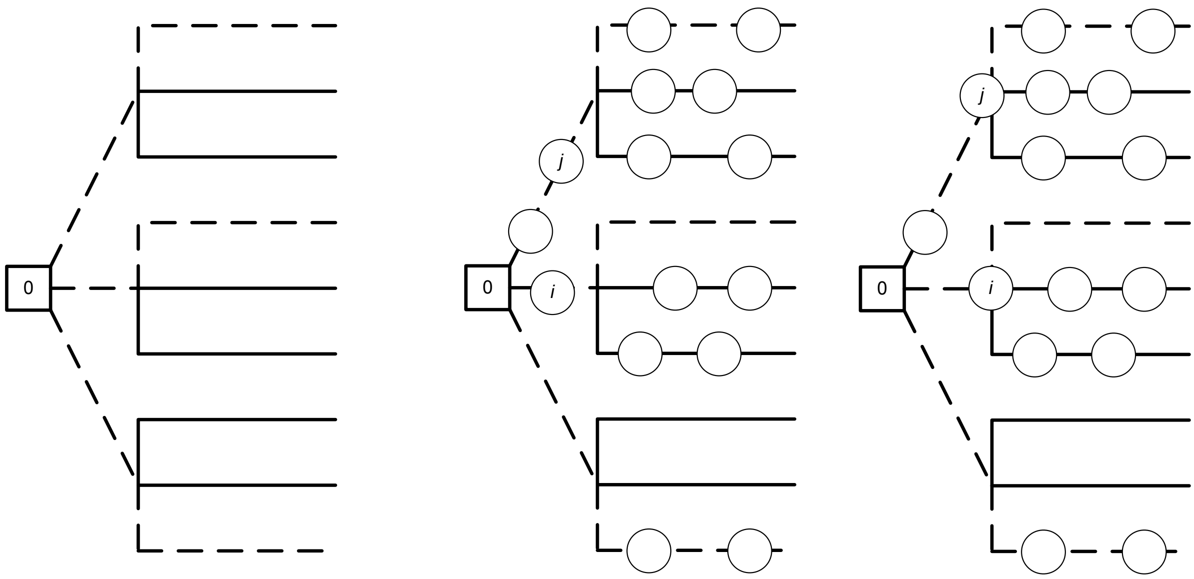
The vehicle routing problem with time windows (see e.g. [1,2]), can be considered as a parent of the problem considered in this paper, and as a core engine we used simulated annealing heuristics (see e.g. [4]) which is known to work well for the family of vehicle routing problems. In this section we present how one can produce different solution for the CDP via different approaches.

4.1 Dynamic approach

A standard and most popular heuristic approaches for vehicle routing problems (simulated annealing, tabu search, neighbourhood search) are based on iteratively applying transformations known as traditional move operators (insertion, swap, 2-opt etc). By these transformations, i.e. by changing places for customers on vehicles, one tries to find a good solution regarding provided evaluation (objective) function. Roughly speaking, vehicles are lists of customers, and by reshuffling customers on vehicles one tries to improve an evaluation function.

The main idea of our dynamic approach to CDP problem is to use traditional move operators on a redesigned data structures. Instead of having simple lists of vehicles, as for standard vehicle routing problems, we introduce *splitters*. We introduce a *tree-list*, a structure beginning with the depot node *0* continued with the *starting sublist* of a tree-list ended with a splitter. A splitter is simply a place where starting sublist splits into several sublists, with the one called the *main sublist* of a tree-list.

Recall that a solution for CDP are vehicles that satisfy all time and capacity constraints, and such that each customer is served by exactly one vehicle. Basically, in a solution for CDP, an initial list contains information about customers served until the cross-docking point is reached. The last node before a splitter is considered to be a cross-docking point, the main sublist contains information about customers served by a vehicle after unloading at a cross-docking point, and the remaining sublists are lists of nodes of vehicles starting at the cross-docking point.



a) *b)* *c)*

Figure 1: Tree-list structures, dashed lines corresponds to vehicles leaving the depot 0; a) at the start, b) during the algorithm, and c) the way it is used for CPP.

The main advantage of this approach is that cross-docking points are determined dynamically during an algorithm. Moreover, all approaches on traditional move operators can be applied.

4.2 P-median approach

In contrast to the previous approach, here we start by clustering customers. We assume that our fleet is homogeneous and that the vehicle capacity equals *c*. The technique is described using the following steps:

a) Remove the arcs connecting customers with no time window intersection.
b) Determine the minimum spanning tree T based on the distances $d(i,j)$ for i,j from V.
c) Assign a weight $w(e)=d(0,i)+d(0,j)$ to each $e=(i,j)$ in T, and set $k := 1$.
d) Let e be the longest arc in T w.r.t. w, and $T–\{e\}=T_1 U\, T_2$. Set $c' =(1–k\lambda)c, (0\leq\lambda<1)$.
If $c' \leq o\,(T_i) \leq c$, group customers from T_i into a cluster and set $T:= T–T_i$.
e) Repeat step d) until T is not empty.
f) Set $k:=k+1$ and repeat steps d) and e) until stopping criteria.

The clustering procedure is based on the underlying spanning tree T in the reduced network. In the third step of the procedure we assign a weight to each arc of the tree as a sum of the distances from its endpoints to depot and set $k = 1$. The arc furthest from a depot is chosen and the graph $T–\{e\}$ is considered. We exploit the fact that deleting an arc from a tree results in a two connected components T_1, T_2. If the aggregated demand of component T_i ($i = 1, 2$) can be fitted in the interval (c', c), where c' equals $(1-k\lambda)c$, with $(0 \leq \lambda < 1)$, we group the customers in component T_i into a single cluster and proceed with $T = T–T_i$. Step e) is repeated until empting the tree T. In step f) we decrease the lower limit c' by increasing parameter k and repeat steps d) and e).

Since demand in each cluster does not exceed the vehicle capacity c, there exists a necessity for only one cross-docking point per cluster. In this paper, we decide to extract a *1*–median for each cluster with metrics

$$c(i, j) = a(i)^{\alpha} + dist(i, j)^{\beta}\,. \tag{1}$$

The metrics is motivated by the fact that customers with an earlier time window opening and closer to a depot, leave larger space for combining nodes in a cluster into a feasible solution for CDP. Consequently, we get a better candidate for a cross-docking point. By tuning the parameters α and β, more importance can be given to either time or distance component.

5 RESULTS

In order to test the proposed approaches, we first run the simulated annealing algorithm for a problem with no cross-dockings and compared it with approaches described in section 4. We used the standard extended Solomon benchmarking instances and two real-life problems generated from a newspaper distribution. The problem sizes are 400 and 600 customers, and 633 and 2249 customers for the real world problem instances.

Table 1: Results of the algorithms

Problem	Problem Size	(H1)	Heuristics CD (H2)		Predefined CD (H3)		Savings		
		Distance	Distance	CD	Distance	CD	((4)-(3))/(3)	((6)-(3))/(3)	((6)-(4))/(4)
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
RC2_4_1	400	6563,44	5587,11	7	5675,76	7	-14,88%	-13,52%	1,56%
RC2_4_2	400	5971,58	5139,53	5	5224,15	7	-13,93%	-12,52%	1,62%
C2_4_1	400	4248,93	4030,33	5	4223,10	8	-5,14%	-0,61%	4,56%
C2_4_2	400	4408,89	4104,97	3	4281,94	10	-6,89%	-2,88%	4,13%
R2_4_1	400	7889,43	6727,59	9	7077,12	7	-14,73%	-10,30%	4,94%
R2_4_2	400	7159,40	6133,21	5	6173,57	10	-14,33%	-13,77%	0,65%
RC2_6_1	600	12910,4	10733,2	12	11166,9	12	-16,86%	-13,50%	3,88%
RC2_6_2	600	11542,9	10720,0	8	10433,5	12	-7,13%	-9,61%	-2,75%
RC2_6_3	600	9767,47	9004,82	3	9121,41	11	-7,81%	-6,61%	1,28%
D_633	633	1771,30	1305,88	5	1311,07	4	-26,28%	-25,98%	0,40%
D_2249	2249	8953,00	8617,00	6	7828,00	19	-3,75%	-12,57%	-10,08%
Average	-	-	-	6,1	-	9,7	-11,98%	-11,08%	0,93%

Table 1 presents the results of the proposed algorithms. First two columns represent the problem name and the corresponding number of customers. Column (3) contains the distance achieved by the algorithm with serving customers only via a depot. Columns (4) and (5) represent distance and number of cross-docking points provided with algorithm described in 4.1. Similarly, columns (6) and (7) contain the distance and number of cross-docking points for the algorithm described in 4.2. Columns (8), (9) are the relative savings where distances from column (3) are taken as a base. It is obvious that presented algorithms can obtain significant savings in total distance travelled, ranging from 0,61% to 26,28% and in average 11,98% and 11,08%.

Column (10) can be used to compare the quality of algorithms. It can be seen that the algorithm H2 gives slightly better results for a theoretical test instances, but in a large scale real life example the clustering approach gives a better result.

6 CONCLUSION

We defined a cross-docking placement problem and proposed two solution approaches. Preliminary calculations were performed for standard extended Solomon benchmarking instances and for two real-life problems generated from newspaper distribution. Results show that significant savings in total distance travelled can be achieved if cross-docking points are used and placed intelligently. The performance of our algorithm is, due to having similar move operators as basic iterations, comparable to the standard vehicle routing algorithms, hence they can be customized and used for real world applications.

References

[1] Bräysy, O., Gendreau, M., 2005. Vehicle Routing Problem with Time Windows, Part I: Route Construction and Local Search Algorithms, Transportation Science 39, 104 - 118.

[2] Bräysy, O., Gendreau, M., 2005. Vehicle Routing Problem with Time Windows, Part II: Metaheuristics, Transportation Science 39, 119 - 139.

[3] Nagy, G., Salhi, S., 2007. Location - Routing: Issues, models and methods, European Journal of Operational Research 177, 649 - 672.

[4] Kirkpatrick, S., Gelatt, C., Vecchi, M., 1983. Optimization by simulated annealing. Science 220 671–680.

DYNAMICS AND LOCAL POLICY IN COMMUTING: ATTRACTIVENESS AND STICKINESS OF SLOVENIAN MUNICIPALITIES

Samo Drobne*, Marija Bogataj, Mateja Zupan***** and **Anka Lisec******
* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: samo.drobne@fgg.uni-lj.si
** University of Ljubljana, Faculty of Maritime Studies and Transport, Portorož,
& MEDIFAS, Nova Gorica, Slovenia, e-mail: marija.bogataj@guest.arnes.si
*** Jesenovo 20, Čemšenik, Slovenia, e-mail: mateja5zupan@gmail.com
**** University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia,
e-mail: anka.lisec@fgg.uni-lj.si

Abstract: In this paper, we analyse the dynamics of parameters in the gravity models of inter-municipal labour commuters in Slovenia in the period 2000 – 2009. For this purpose, extended gravity model was calculated for each year of the analysed period. The change of the impacts of parameters on commuting flows was studied separately for each year and in general for whole analysed period. The approach to the policy of commuting dynamics is presented where some factors are compensated with the change of the others.

Keywords: commuting, attractiveness, stickiness, dynamics, decision-making, municipalities, Slovenia.

1 INTRODUCTION

The separation of residential location and location of working places has led to complex commuting patterns which have extended in geographical scale over the past decades. As a consequence, home-job-home trips have adopted multi-regional network configurations and have thus led to complex interactive networks. Commuting has become an important field of study in geography, transportation science and regional science [4]. Models of commuting flows have become an increasing important topic within regional science [3,4]. One obvious use for such models is in making predictions about how changes in the spatial distribution of jobs and workers or the infrastructure connecting them might affect a region's economy. A variety of models have been developed with the aim of modelling commuting flows. One popular class of models is the class of gravity models. According to Sen and Smith [5], the analogy of gravity forces in social sciences can be dated back to 1858 on the sociological phenomena of group behaviour. In modern times, the gravity model has many applications. It is mostly used in areas such as transportation, physical planning, environmental studies, regional economics and geographical analysis [2]. This paper focuses on their use in policy implementation.

2 THE PROBLEM

In this paper, we analyse dynamics of parameters in the gravity models of inter-municipal labour commuters in Slovenia in the period 2000 – 2009. For this purpose, extended gravity model was calculated for each year of the analysed period. The change of the impacts of parameters on commuting flows were studied separately for each year and in general for the whole analysed period. We analysed the following parameters: population of municipality, travel time by car between municipalities' centres, employment in the municipality, and average gross earnings in the municipality. Data on average population, employment and workforce population, average gross earnings in the municipality and in the state are collected by Statistical Office of the Republic of Slovenia (SORS) and published at SI-STAT

Data Portal [6] and/or in Statistical Yearbook [7]. Data on state road network by year are collected by Slovenian Road Agency [8]. Using data on state roads, we constructed network models, which were the basis for calculation of optimal (shortest) time-spending distances between municipal centres of Slovenia. Ten OD matrixes were calculated considering conditions on state roads for each year separately (construction of new highways per year, toll stations on highways and abolition of the toll system in 2008). In the analysed period of ten years from 2000 – 2009, the number of Slovenian municipalities was changed from 192 in 2000 to 193 in 2002 and to 210 municipalities in 2009. The change was reflected in statistical research next year.

In the analysed period 2000 – 2009, there were 348,339 labour commuters between Slovenian municipalities in average. Tab. 1 shows that the number of labour commuters was increased from 299,188 in 2000 to 390,500 in 2009. The numbers of inter-municipal labour commuters increased for the travel times of 0 – 150 minutes, but it decreased for the travel times higher than 150 minutes. Tab. 2 shows the percentage of labour commuters between Slovenian municipalities by travel times in years of 2000 – 2009. Comparing Tab. 2 and linear trend coefficients of percentage of labour inter-municipal commuters by travel times' intervals in Tab. 3, one can notice that the percentage of labour commuters slightly decreased for travel times of less than 15 and more than 135 minutes. It is also evident that the percentage of labour commuters for the time-spending distances between 30 and 90 minutes increased in general. The attraction and stickiness of municipalities can be measured by labour migration index. Labour migration index (LMI) [6]:

$$LMI = \frac{persons\ in\ employment\ by\ municipality\ of\ workplace}{persons\ in\ employment\ by\ municipality\ of\ residence} \cdot 100 \Rightarrow LMI_i = 100 \cdot \frac{\sum_{j \neq i} DC_{ji}}{\sum_{j \neq i} DC_{ij}}$$

where we denote with i the municipality of origin and with j the municipality of destination, and DC_{ij} is the number of inter-municipal labour commuters from the municipality i to the municipality j. Territorial units, of which LMI is more than 96.0, are labour-oriented, and others are residential-oriented [6].

Table 1: Number of labour commuters between Slovenian municipalities by travel times and by year in the period 2000 – 2009 (source: [6] and own calculation).

Travel time (min)	***Year***									
	2000	***2001***	***2002***	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***
[0-15)	116,116	117,047	117,425	117,959	118,086	118,441	120,803	130,243	135,905	127,374
[15-30)	115,058	119,197	122,953	126,535	131,871	136,433	141,973	150,024	152,615	151,006
[30-45)	31,114	31,097	33,235	34,658	35,805	39,624	41,345	44,380	46,514	46,693
[45-60)	12,159	13,805	16,167	17,598	17,936	19,497	21,547	24,229	26,489	25,972
[60-75)	8362	9213	9447	10,839	13,947	14,002	16,342	18,407	18,029	16,455
[75-90)	6028	5614	7629	8134	7662	8796	9824	10,867	11,685	10,422
[90-105)	3804	4590	3222	3515	4030	3887	4342	5041	5155	4766
[105-120)	2461	2105	1812	2050	2312	2392	2627	3267	5238	4581
[120-135)	1230	1195	1175	1285	1306	1966	2251	2525	1917	1331
[135-150)	905	1216	1345	1682	1841	1511	1685	1901	1066	1189
[150-165)	1109	1020	1081	892	834	754	830	922	325	444
[165-180)	571	454	192	203	210	174	161	194	276	220
[180-195)	125	116	115	118	130	127	129	179	61	43
[195-210)	69	62	59	84	81	85	90	104	12	3
[210-225)	46	48	29	40	30	23	25	25	2	1
[225-240)	23	18	18	12	9	2	3	2	0	0
> 240	8	9	3	1	1	1	1	1	0	0
Slovenia	**299,188**	**306,806**	**315,907**	**325,605**	**336,091**	**347,715**	**363,978**	**392,311**	**405,289**	**390,500**

Table 2: The percentage of labour commuters between Slovenian municipalities by travel times' interval in the period 2000 – 2009 (source: [6] and own calculation).

Travel time (min)	***Year***									
	2000	***2001***	***2002***	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***
[0-15)	38.8%	38.2%	37.2%	36.2%	35.1%	34.1%	33.2%	33.2%	33.5%	32.6%
[15-30)	38.5%	38.9%	38.9%	38.9%	39.2%	39.2%	39.0%	38.2%	37.7%	38.7%
[30-45)	10.4%	10.1%	10.5%	10.6%	10.7%	11.4%	11.4%	11.3%	11.5%	12.0%
[45-60)	4.1%	4.5%	5.1%	5.4%	5.3%	5.6%	5.9%	6.2%	6.5%	6.7%
[60-75)	2.8%	3.0%	3.0%	3.3%	4.1%	4.0%	4.5%	4.7%	4.4%	4.2%
[75-90)	2.0%	1.8%	2.4%	2.5%	2.3%	2.5%	2.7%	2.8%	2.9%	2.7%
[90-105)	1.3%	1.5%	1.0%	1.1%	1.2%	1.1%	1.2%	1.3%	1.3%	1.2%
[105-120)	0.8%	0.7%	0.6%	0.6%	0.7%	0.7%	0.7%	0.8%	1.3%	1.2%
[120-135)	0.4%	0.4%	0.4%	0.4%	0.4%	0.6%	0.6%	0.6%	0.5%	0.3%
[135-150)	0.3%	0.4%	0.4%	0.5%	0.5%	0.4%	0.5%	0.5%	0.3%	0.3%
[150-165)	0.4%	0.3%	0.3%	0.3%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%
[165-180)	0.2%	0.1%	0.1%	0.1%	0.1%	0.1%	<0.0%	<0.0%	0.1%	0.1%
> 180	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%	<0.0%
Slovenia	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**

Table 3: Linear trend coefficients of percentage of labour inter-municipal commuters by travel times' interval in Slovenian in the period 2000 – 2009.

Travel time (min)	***<15***	***[15-30)***	***[30-45)***	***[45-60)***	***[60-75)***	***[75-90)***	***[90-105)***	***[105-120)***	***[120-135)***
Linear trend coefficients	$-7*10^{-03}$	$-6*10^{-04}$	$2*10^{-03}$	$3*10^{-03}$	$2*10^{-03}$	$1*10^{-03}$	$-3*10^{-05}$	$5*10^{-04}$	$1*10^{-04}$
Travel time (min)	***[135-150)***	***[150-165)***	***[165-180)***	***[180-195)***	***[195-210)***	***[210-225)***	***[225-240)***	***>240***	
Linear trend coefficients	$-5*10^{-05}$	$-3*10^{-04}$	$-1*10^{-04}$	$-2*10^{-05}$	$-2*10^{-05}$	$-2*10^{-05}$	$-9*10^{-06}$	$-3*10^{-06}$	

We analysed the dynamics of exponents in extended gravity model of commuting between Slovenian municipalities by year in the period 2000 – 2009 using model (1). The same model was used for calculating the average powers for the same period:

$$DC_{ij} = \alpha \cdot P_i^{\beta_1} \cdot P_j^{\beta_2} \cdot d(t)_{ij}^{\gamma} \cdot K_{EMP,i}^{\delta_1} \cdot K_{EMP,j}^{\delta_2} \cdot K_{GEAR,i}^{\varepsilon_1} \cdot K_{GEAR,j}^{\varepsilon_2} \tag{1}$$

where we denote with i the municipality of origin and with j the municipality of destination, DC_{ij} is the number of inter-municipal labour commuters, P_i and P_j are the populations in the municipality of origin respectively destination, $d(t)_{ij}$ is travel time between the municipality of origin and municipality of destination, $K_{EMP,i}$ and $K_{EMP,j}$ are the coefficients of employment in the municipality of origin respectively destination, $K_{GEAR,i}$ and $K_{GEAR,j}$ are the coefficients of gross salary in the municipality of origin respectively destination. α, which is the constant, and $\beta_1, \beta_2, \gamma, \delta_1, \delta_2, \varepsilon_1, \varepsilon_2$, which are the powers, are estimated in the regression analysis. Coefficients were calculated as follows:

$$K_{EMP} = \frac{EMP_{mun}/WFP_{mun}}{EMP_{SI}/WFP_{SI}} \quad \text{and} \quad K_{GEAR} = \frac{GEAR_{mun}}{GEAR_{SI}},$$

where EMP_{mun} is number of employed persons in municipality, EMP_{SI} is number of employed persons in Slovenia, WFP_{mun} is number of workforce population in municipality, WFP_{SI} is number of workforce population in Slovenia, $GEAR_{mun}$ is average gross earnings in municipality, and $GEAR_{SI}$ is average gross earnings in Slovenia. Tab. 4 shows the results of power trends in the period 2000 – 2009.

Table 4: Powers (exponents) of the extended gravity model (2) between Slovenian municipalities by year and in average in the period 2000 – 2009 (parentheses denote exponents which significance is bad – P-value is greater than 0.1; in the most of other cases the significance is very good – P-value is much lower than 0.0001).

Parameter	**Symbol of regression coefficients in the model (2)**	**Year**										**Period**
		2000	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2000-2009**
N		36,672	36,672	36,672	37,056	37,056	37,056	37,056	43,890	43,890	43,890	389,910
R^2 (%)		63.8	64.2	64.8	65.3	65.7	66.5	66.8	66.6	67.2	67.1	65.8
Adjusted R^2 (%)		40.7	41.2	42.0	42.6	43.1	44.2	44.7	44.3	45.1	45.0	43.3
constant	α	1.79	1.64	1.74	1.69	1.32	1.60	1.37	(1.02)	0.83	(1.07)	1.373
P_i	β_1	0.15	0.15	0.16	0.16	0.17	0.17	0.18	0.19	0.21	0.20	0.175
P_j	β_2	0.26	0.26	0.27	0.28	0.30	0.30	0.31	0.32	0.34	0.32	0.296
$d(t)_{ij}$	γ	-0.83	-0.85	-0.88	-0.90	-0.91	-0.93	-0.95	-0.92	-0.96	-0.96	-0.909
$K_{EMP,i}$	δ_1	-0.05	-0.05	-0.07	-0.06	-0.08	-0.05	-0.06	-0.08	-0.10	-0.08	-0.065
$K_{EMP,j}$	δ_2	-0.08	0.07	0.05	0.09	(0.02)	0.08	0.07	0.08	0.04	0.06	0.065
$K_{GEAR,i}$	ε_1	0.12	(0.04)	0.05	(0.02)	(0.04)	0.16	(0.06)	-0.10	-0.11	(-0.03)	0.033
$K_{GEAR,j}$	ε_2	0.46	0.37	0.45	0.44	0.44	0.82	0.69	0.41	0.36	0.32	0.464

Inter-municipal gravity model of labour commuting in Slovenia in the period 2000 – 2009 in average is:

$$DC_{ij}^{(2000-2009)} = 1.373 \cdot P_i^{0.175} \cdot P_j^{0.296} \cdot d(t)_{ij}^{-0.909} \cdot K_{EMP,i}^{-0.065} \cdot K_{EMP,j}^{0.065} \cdot K_{GEAR,i}^{0.033} \cdot K_{GEAR,j}^{0.464} \qquad (2)$$

Model (2) shows that the power of the population in municipality of destination is 1.7-times powerful than of the population in municipality of origin, the power of the coefficient of employment is equal for the municipalities of origin and destination, and the power of the coefficient of average gross earnings in the municipality of destination is 14-times powerful than equal coefficient in the municipality of origin. The results show that analysed parameters more attract than dispatch commuting flows. From Tab. 4, we can also see that the exponents are changing in time.

Let us introduce technical time $k = y - 2000$, where y is year. The trend analysis of γ gives

$$\gamma(k) = -0.848 - 0.014 \cdot k \qquad (3)$$

and trend analysis of ε_2 gives

$$\varepsilon_2(k) = 0.487 - 0.003 \cdot k \qquad (4)$$

The power of travel distance in equation of labour commuting flows decreases in time. Let us suppose (a) that this is more the result of the reluctance to day migration (and less because of the improved road conditions) and (b) that γ is constant for all distances as it is supposed in the model (1). In this case especially more distanced commuters influenced such results against those, which are very close. If this is the case, the question appears: "How the other parameters of gravity model (1) should change to keep the intensity of flow on the same level?"

For the municipality of destination (j), the decision variable, which can compensate a lower willingness to commute, is gross- or net-earnings or their relative change p. The general formula considering this condition follows from (2):

$$DC_{ij}^{(2000-2009)} \doteq a^{(2000-2009)} \cdot K_{GEAR,j}^{\varepsilon_2^0} \cdot d(t)_{ij}^{-\gamma^0} \tag{5}$$

where ε_2^0 and γ^0 denote the exponents for total time period (2000- 2009).

To keep the intensity of in-flows on the same level, the following has to be obeyed:

$$DC_{ij} = a(k) \cdot \left[K_{GEAR,j}(1+p(k))\right]^{\varepsilon_2(k)} \cdot d(t)_{ij}^{\gamma(k)} = const \tag{6}$$

In case that $a(k)$ is constant, from (4) it follows

$$\left[K_{GEAR,j}(1+p(k))\right]^{\varepsilon_2(k)} \cdot d(t)_{ij}^{\gamma(k)} = K_{GEAR,j}^{\varepsilon_2^0} \cdot d(t)_{ij}^{-\gamma^0}$$

from where we get

$$p(k) = \left(\frac{d(t)_{ij}^{-\gamma(k)-\gamma^0}}{K_{GEAR,j}^{\varepsilon_2(k)-\varepsilon_2^0}} \right)^{\frac{1}{\varepsilon_2(k)}} - 1, \tag{7}$$

where $p(k)$ is the dynamic decision variable (depends on time k).

3 NUMERICAL EXAMPLE

Let us take municipality of Ljubljana, which geometric mean of LMI from 2000 – 2009 is equal to 167.51; it follows that Ljubljana is very labour oriented municipality (the average number of labour commuters who commute to Ljubljana in 2000 – 2009 was 88,358). ***Policy:*** Let us suppose that Ljubljana municipality wish to keep the intensity of flows which originate from the travel distance of d minutes at least on the same level as it was in the average in the analysed period 2000 – 2009. Ljubljana wishes to realize this policy in 2012 ($k=12$). From Tab. 5 it follows that the geometric mean of K_{GEAR} of Ljubljana in the period 2000 – 2009 equals $K_{GEAR,Ljubljana}^{(2000-2009)} = 1.23$ and $K_{GEAR,Ljubljana}$ in 2012 can be estimated using linear trend:

$$K_{GEAR,Ljubljana}(k) = 1.265 - 0.008 \cdot k \tag{8}$$

Table 5: $K_{GEAR,Ljubljana}$ in the period 2000 – 2009.

Year	***2000***	***2001***	***2002***	***2003***	***2004***	***2005***	***2006***	***2007***	***2008***	***2009***	***GeoMean 2000-2009***
$K_{GEAR,Ljubljana}$	1.26	1.25	1.25	1.26	1.22	1.21	1.21	1.21	1.20	1.19	1.23

From Tab. 1 and equations (3) and (4) we calculate $\gamma(12) = -1.011$ and $\varepsilon_2(12) = 0.455$. Tab. 3 shows that the intensity of commuting flows in the interval of 0 – 30 min decrease. To keep the intensity of flows on the distance of $d=15$ minutes of travel time on the same level as it was in the average in 2000 – 2009, equation (7) requires

$$p(12) = \left(\frac{d(t)_{ij}^{-\gamma(12)-0.909}}{K_{GEAR,j}^{\varepsilon_2(12)-0.464}} \right)^{\frac{1}{\varepsilon_2(12)}} - 1 = \left(\frac{15^{1.011-0.909}}{1.23^{0.455-0.464}} \right)^{\frac{1}{0.455}} - 1 = 0.85 \tag{9}$$

From (8), it follows that expected $K_{GEAR,Ljubljana}(12) = 1.17$, if the existing policy would not change. It means that expected percentage of change in Ljubljana for 2012

$\hat{p}_{Ljubljana}(12) = \frac{K_{GEAR,Ljubljana}(12)}{K_{GEAR,Ljubljana}^{(2000-2009)}} - 1 = \frac{1.17}{1.23} - 1 = -0.05$. From here we can conclude that we cannot stop the decrease of labour commuting flows from the nearest municipalities $(d \dot{<} 30\,\text{minutes})$ to Ljubljana $(-0.05 < 0.85)$ (a) if there won't be introduced any new policy which increase the gross earnings in Ljubljana relatively to other municipalities in Slovenia, or (b) if the investments to improve the accessibility to Ljubljana won't be realized. We have to note here that γ is constant and does not depend on distance, which should be additionally consider; see e.g. [1].

4 CONCLUSIONS

In the paper, the parameters in the extended gravity models of inter-municipal labour commuters in Slovenia in the period 2000 – 2009 were studied. We found out the following: **(a)** The stickiness of population size in municipality of origin and attractiveness of the population in the municipality of destination were increased in the analysed period, having power of stickiness $\beta_1 = 0.20$ and power of attractiveness $\beta_2 = 0.32$ in 2009. **(b)** The impact of distance on commuting increased during the analysed period from $\gamma = -0.83$ in 2000 to $\gamma = -0.96$ in 2009. It follows that the accessibility between Slovenian municipalities should increase, or some other (economic) parameters should be changed, to keep the flows at the same level, as presented in numerical example. In our study, the logit model with constant exponentials was analysed. According to our studies [1] the distance power could be stepwise constant. In this case probably more accurate results would be achieved.

References

[1] Bogataj M, D. Tuljak Suban and S. Drobne (2011). Samo. Regression-fuzzy approach to land valuation. Cent. Eur. j. oper. res., 13 pp., http://www.springerlink.com/content/r08r73t2611831j2/fulltext.pdf, doi: 10.1007/s10100-010-0188-x; accessed: 25.3.2011)

[2] Gitlesen J. P. and K. Jörnsten (2000). A disaggregated gravity model. Norwegian School of Economics and Business Administration. Department of Finance and Management Science. Discussion paper 2000:28. (http://hdl.handle.net/2330/787; accessed: 25.3.2011)

[3] Gorman S. P., R. Patuelli, A. Reggiani, P. Nijkamp, R. Kulkarni and G. Haag (2007). An Application of Complex Network Theory to German Commuting Patterns. In: Friesz T. L. (Ed.), Network Science, Nonlinear Science and Infrastructure Systems, Springer, New York, pp. 167-185.

[4] Rouwendal J. and P. Nijkamp (2004). Living in Two Worlds: A Review of Home-to-Work Decisions. Growth and Change, 35(3):287-303.

[5] Sen, A. and T. E. Smith (1998). Gravity Models of Spatial Interaction Behaviour. Springer-Verlag, Heidelberg.

[6] SORS (2011a). SI-STAT Data Portal. Statistical Office of the Republic of Slovenia. Ljubljana. (http://pxweb.stat.si/pxweb/Dialog/statfile1.asp; accessed: 1.10.2010)

[7] SORS (2011b). Statistical yearbook of the Republic of Slovenia, Statistical Office of the Republic of Slovenia. Ljubljana. (http://www.stat.si/eng/pub_letopis_prva.asp; accessed: 1.10.2010)

[8] SRA (2010). Data on State Roads of the Republic of Slovenia, Slovenian Road Agency, Slovenia, Ljubljana.

MONTE CARLO SIMULATION APPLIED TO THE LOGISTICS OF CERAMICS

Eloy Hontoria *, M.Victoria de-la-Fuente Aragon *,Lorenzo Ros-McDonnell * and **Marija Bogataj****
* GIO Universidad Politécnica de Cartagena, 30202 Cartagena, Spain.
e-mail: eloy.hontoria@upct.es, marivi.fuente@upct.es, lorenzo.ros@upct.es.
** University of Ljubljana & MEDIFAS, Sempeter pri Gorici, Slovenia,
e-mail: marija.bogataj@guest.arnes.si

Abstract: The Capacitated Vehicle Routing Problem CVRP is presented giving a new solution to the problem, with small variations to the basic constraints added to the mathematical model developed by Ros-McDonnell et al. in 2010 describing realistic variations of constraints in a working day of logistic operators in the ceramic industry in Southern Spain. Here the orders must be grouped by route with the objective of improving the efficiency of the transported load and the reduction of associated costs. The fleet variations are modelled on the different demands and constraints which arose where the Monte Carlo simulation approach has been used. The aim is to find a variety of operative solutions which are given quickly when validated using the Google Maps web application.

Keywords: Logistics, Capacitated Vehicle Routing Problem, Cargo assignments, Transportation costs, Monte Carlo Simulation

1 INTRODUCTION

Tough competition in today's market is forcing companies to study and rethink all their processes in order to make them more efficient. Improved transport-related logistics leads to greater efficiency, serving clients by improving satisfaction in terms of time and cost. In the case of ceramic companies, this study is oriented toward the processes connected to the *grouping* of merchandise and customer orders. The lack of appropriate tools to support optimal logistics of such items at highly volatile demand and supply levels has led to the development of different techniques to facilitate both the calculation and implementation of these tasks. Our plan is to improve "an approximate algorithm for optimal logistics of heavy and variable size items" described in [7]. This is why, with the objective of picking up all orders from all clients on a certain date, we ran a Monte Carlo simulation aiming to establish a new, more well-organized vehicle assignment solution, with lower costs than those predicted by the mathematical model initially proposed by the cited authors, when constraints could be slightly relaxed.

2 THE CAPACITATED VEHICLE ROUTING PROBLEM

The Vehicle Routing Problem (VRP) needs to determine the optimal routes to be used by a fleet of vehicles stationed at one or several depots. These vehicles serve a group of clients. The problem has been thoroughly studied in several papers like in [3] [4] [5] or [1]. Many of the requirements and operative constraints, such as route specifications and length, number of vehicles in the fleet and their loading capacities, composition of orders, type of demands, number of warehouses, etc., may be imposed on the practical applications of the VRP problem [6].

In everyday operations, certain variables often rapidly change throughout the day and new solutions must be hastily given to solve the new context. Examples of such variables include travel time (it is very important to consider the rush hour in urban logistics), customer demand and many other characteristics. Important constraints are given by the Capacitated Vehicle Routing Problem (CVRP). Especially in the ceramic industry CVRP can be used to test the efficiency of optimization procedures. In the basic CVRP the demands are predetermined, the service only involves deliveries or collections but not both, the vehicles are based at a single depot, the capacity restrictions for the vehicles are imposed and normally, the objective is to minimize the total cost needed to serve all the clients. Generally, the travel cost between each pair of customer locations is the same in both directions, i.e., the resulting cost matrix is symmetric, whereas in some applications, such as the distribution in urban areas with one-way directions imposed on the roads, the cost matrix is asymmetric [5], such as those which could appear in our “out of city” applications.

When investigating acceptable solutions in this context, researchers favour the optimization techniques based on mathematical models, which provide simplifications of all possible constraints. These methods have important difficulties in their practical application, particularly when changes in the definition of the problem appear, such as additional constraints to the problem or their relaxation. At this stage, the problem could cease to be a robust one, and not have a solution via the previous optimization techniques. In this situation, the application of simulation models permits descriptions of complex systems without multiple assumptions, obtaining near-optimal solutions and guaranteeing their feasibility [1].

This paper describes a new procedure, which improves the results of the mathematical model described in [7] Starting from the solution of the approximate algorithm, which is based on the previous results of the Branch & Bound technique [2], some relaxations are permitted and rules are applied for the assignment of loads in a fleet of vehicles. Orders are grouped by route with the objective of improving the efficiency of the transported load, with the consequent reduction of associated costs. This new procedure changes the previous model setting by introducing additional constraints and relaxations, which must be checked after the assignment of vehicles is carried out by means of a simulation process. Solutions will be obtained using the Monte Carlo method. A subsequent process used to validate the solutions found is determined using the Google Maps web application.

3 PROBLEM DESCRIPTION

The client is a logistics operator receiving daily orders from distributors. These orders must be collected from manufacturers and grouped according to companies and routes in order to eventually assign a means of transport for their collection. This strategy grants to the logistic operator a better use of logistical resources, and minimizes their associated costs.

However, there are a number of constraints that the company must bear in mind when it comes to providing their services: (a) Capacities of vehicles for road transport; (b) Pick-up routes acceptable and constrained by the logistics operator; (c) Number of suppliers that can be visited per vehicle per route; (d) Cost of each vehicle and (e) Time windows (with relaxations). The variables and constrains are given in Tab. 1.

Table 1: Capacities of vehicles

VEHICLES	T1	T2	T3	T4	T5	T6	T7
CAPACITY (Kg)	25000	16000	8000	5500	4500	4000	1100

There are 6 possible routes $R_k = [A, B, C, D, E, F]$ (Tab. 2):

Table 2: The maximal number of suppliers that can be visited per vehicle according to route.

	A	B	C	D	E	F
T1	12	10	8	6	6	9
T2	12	10	8	6	6	9
T3	15	14	12	6	6	14
T4	15	14	12	6	6	14
T5	17	18	13	6	6	15
T6	17	18	13	6	6	15
T7	18	18	14	6	6	15

The daily cost per vehicle is given in Tab. 3:

Table 3: Daily cost per vehicle Daily cost per vehicle.

VEHICLE	T1	T2	T3	T4	T5	T6	T7
EUROS/DAY	300	220	170	160	160	160	125

There are multiple combinations of these variables and constraints even for attainable solutions, which have to be known before the previous working day is finished and the assignment of vehicles is conducted. For these reasons, it is necessary to use a quick procedure to find the best solutions.

4 MONTE CARLO SIMULATIONS

The first solution to this problem was proposed by [7] by using an approximate algorithm formalization. These authors developed a set of decision-making rules which carried out an efficient load assignment in a vehicle fleet based on the routes determined in advance by the logistics operator. The previous algorithm was robust enough to carry out the assignment, but another method was needed to find solutions as quickly as possible and available to make changes in any additional restriction or changing variables throughout the same day. Therefore a new, additional method which combines Monte Carlo Simulation using Excel and validates results with Google Maps has been developed to solve this logistic problem at the second stage. This improved version, in which all the objectives are met in terms of order collection, aims to find as good or better solutions (based on cost minimization) than those given by the approximate algorithm of [7]. To this end, fleet variations are modelled for the different cases studied, by using the Monte Carlo simulation technique. With this purpose, the aim is to find a variety of solutions that can be used under the additional constraints or relaxations. The procedure for searching for solutions is as follows:

1. Calculation of the load capacity to be hired on a daily basis. This capacity of the fleet must be greater or equal than the total load to be picked up on any particular day.
2. Calculation of the possible solutions, by means of a Monte Carlo simulation, for the fleet to be used to collect the daily orders.
3. Cost study of each solution found in the previous step. Those solutions granting lower costs than those of the approximate algorithm must be selected.
4. The final choice of the cheapest solution and route definition for each vehicle, along several alternative routes, on condition that the constraints are observed and the collections of all received orders are carried out satisfactorily.

5 MONTE CARLO SIMULATION SOLUTION

The vehicle leaves from one single depot and must travel one or several routes, with a limitation of routes. As an example, Fig. 1 shows the itinerary for a vehicle that is making two routes and collects suppliers' orders on those routes in a working day. The terms included in Fig. 1 are given as follows:

TDep O_i	time elapsed from the depot to the start of route i
$Troute_i$	travelling time needed to complete route i
Tpr_i	time needed to collect all orders from all suppliers on route i (calculated service time at each supplier's facilities is 30 minutes).
$TDiO_j$	Time consumed from the completion of route i to the start of route j.
$Troute_j$	travelling time of route j
Tpr_j	time needed to collect all orders from all suppliers on route j (30 minutes).
TD_jDep	return time from the end of the last route to the depot.

Service time is defined as the time needed for the vehicle to leave the route, enter the supplier's facilities, pick up the order and return to the route. This has been calculated to be 30 minutes for each supplier. This is additional to the time needed for the vehicle to travel the whole route, which has been calculated by using the Google Maps web application. The time that a vehicle would take to collect all the orders on the routes that have been assigned to it is calculated by the formula: *Ttotal= TDep Oi+ Trutai+ Tpri+ TDiOj+ Trutaj+ Tprj+ TDjDep.*

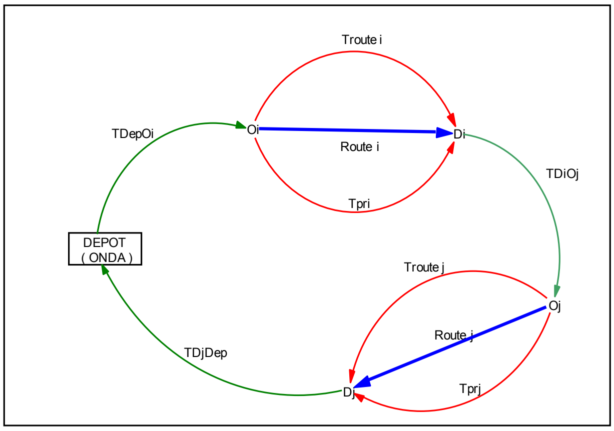


Figure 1: Diagram for the calculation of vehicle times

The optimal logistics on January 19th 2006 with the quantity collected (in kg) on each road is presented in Tab.4.

Table 4: Loads per route (in k.g).

ROUTE A	ROUTE B	ROUTE C	ROUTE D	ROUTE E	ROUTE F	TOTAL
40906	24402	4577	12517	1885	23898	108185

Table 5: Number of suppliers per route.

route	**ROUTE A**	**ROUTE B**	**ROUTE C**	**ROUTE D**	**ROUTE E**	**ROUT E F**	**TOTAL**
number	**14**	**22**	**4**	**2**	**1**	**18**	**61**

Starting from the approximate algorithm [7] the assignments had been determined with the following use of fleet: 3 x T1, 2 x T2, 1 x T4, 1 x T5 and 1 x T6 and 1820 € of costs. Using Monte Carlo Simulation by Excel under given constraints the solution has been obtained, which is given in Fig. 2.

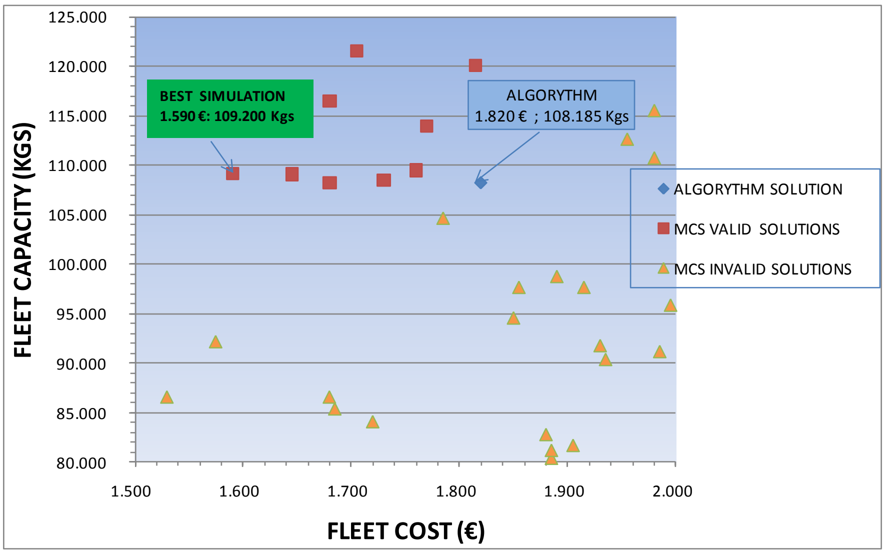


Figure 2: Solutions obtained with Monte Carlo Simulation method (MCS).

Once the assignment of routes to vehicles is completed, the solution must be validated to verify that the fleet can in fact carry out the collection of all the merchandise in the prescribed time window of 8 hours, which could be exceeded by no more than 3% (Tab. 6). Using Google Maps the admissible solutions can be given as presented at Fig. 3.

Table 6: The optimal solution.

VEHICLE	ROUTE	SUPPLIERS VISITED	TIME	
1 T1	A	10	6,07	OK
2 T1	A-B	9	5,98	OK
3 T1	F-B	10	6,85	OK
1 T2	F-C-B	10	8,02	OK
2 T2	E-D-F	6	6,32	OK
1 T7	F-C	7	5,8	OK
2 T7	A	8	6,1	OK

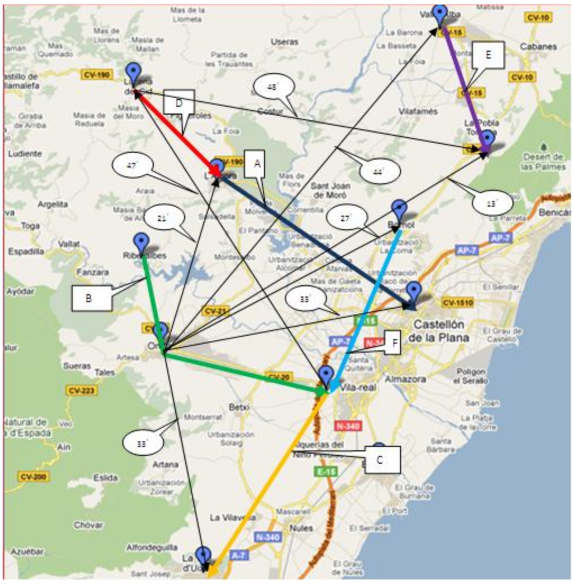


Figure 3: Route Map and travelling times on Google Maps web application.

6 CONCLUSIONS

Capacitated vehicle routing, having a time window and other constraints, is an important job that must be done quickly, with relevant economic implications for a company. Orders being collected on one particular day will determine the size of the fleet for the following day. This situation, together with the lack of an appropriate method, may result in solutions that are quick and effective, but not efficient. The Monte Carlo-based calculation method improves the results obtained by the approximate algorithm of [7], allowing a higher volatility of parameters. The improvements obtained are reflected in a reduction of costs close to 20%, and in the increase of efficiency of the load of vehicles in the fleet. The advantages given by the method (as opposed to other existing methods) is an easy and faster way to obtain the results. All the obtained solutions are operative and they have been verified and validated by the Google Maps time & distance tool.

References

[1] Bachelet, B., Yon, L., 2007. Model enhancement: Improving theoretical optimization with simulation, Simul Model Pract Th 15:703-715.

[2] Fischetti, M., Toth, P., Vigo, D., 1994. A branch-and-bound algorithm for the capacitated vehicle routing problem on directed graphs, Oper. Res., 42:846-859.

[3] Laporte, G., 1992. The vehicle Routing Problem: An overview of exact and approximate algorithms. Eur J Oper Res 59:345-358.

[4] Laporte, G., Gendreau, M., Potvin, J. Y., Semet, F., 2000. Classical and Modern Heuristics for the Vehicle Routing Problem, Int Tra Oper Res 7:285-300.

[5] Toth, P., Vigo, D., 2002. Models, relaxations and exact approaches for the capacitated vehicle routing problem. Discrete Appl Math 123:487-512.

[6] Ralphs, T. K., Kopman, L., Pulleyblank, W. R., Trotter, L. E., 2003. On the Capacitated Vehicle Routing Problem, Math Program 94:343-359.

[7] Ros-McDonnell, L., de-la-Fuente Aragon, M. V., Bogataj, M., 2010. An approximate algorithm for optimal logistics of heavy and variable size items, Cent Eur J Oper Res doi: 10.1007/s10100-010-0155-6.

PROPOSING MIXED NON LINEAR PROGRAMMING MODEL FOR NETWORK DESIGN PROBLEM UNDER PERSPECTIVE OF THIRD PARTY LOGISTICS

Arash Motaghedi-Larijani and **Mohamad-Saied Jabalameli**
Iran University of science and technology, Department of industrial engineering
{amotaghedi, jabal }@ind.iust.ac.ir

Abstract: Third party logistics (3PLs) services providers' importance is increased in supply chain management continuously. Particularly in warehousing and transportation services, a number of clients expect for 3PLs to improve lead times, fill rates, inventory levels, etc. The main objective of this paper is to study the logistics network design problem from 3PLs perspectives with considering inventory cost under the dynamic environment. This paper proposes a nonlinear mixed integer programming model for an integrated logistics network design that capacitated facility location-allocation and inventory location problem are taken into account. Finally this model was solved by Lingo.

Keywords: Location and allocation models, Third party logistics, Network design

1 INTRODUCTION

As the complexity of a supply chain continuously grows and the importance of prompt delivery to end customers increases, an independent company, so called a third party logistics (3PLs) provider, has emerged that performs all or part of a manufacturer's product distribution function. The 3PLs is an outer provider who manages, controls, and delivers logistics activities on behalf of a manufacturer [1]. According to [3], 3PLs involves ''the use of external companies to perform logistics functions that have traditionally been performed within an organization. The functions performed by the third party can surround the entire logistics process or selected activities within that process''. Their functions can involve any aspect of logistics and is generally more than outsourcing warehousing or transportation alone. 3PLs integrate more than one function within the overall supply chain [2]. Bask [5] declares 3PLs as a set of three dyadic relationships linking seller, buyer and logistics service provider in a supply chain.

3PLS often have to design an integrated reverse logistics facility system to satisfy multi-clients' forward and reverse logistics demands at the same time. The product return process involves the determination of the number and location-allocation of repair facilities for returned products in such a way that total reverse logistics costs such as warehousing and transportation costs are minimized, capacity of repair facilities are fully utilized, and the convenience of customers who return products is maximized.

With considering previous explains about 3PLs, can be found that logistic network design have a key role in reducing cost for 3PLs. A logistics network is a series of nodes and transportation links. Logistics networks can range from direct shipments to complex, multi-site operations. The design of a logistics network depends on factors such as the type of products, range and volume of products, the level of service required, and the number and type of customers. A logistics network must meet a specific set of requirements over a given planning horizon. A good design creates the best network for providing the customer with the right goods, in the right quantity, and at the right place, in the right condition, at the right time, and at the right cost.

Despite of wide literature in various domains of 3Pls, there are a few literatures for network design under respect of 3PLs. In literature have been studied a lot of research about reverse logistic problem. Min et al. [6] propose nonlinear mixed integer programming model and a genetic algorithm that can solve the reverse logistics problem involving product returns. But, there are few researches to study network design for 3PLs. Papers that are studding logistic network design under perspective of 3PLs are divided into two categories. The first category, involving dynamic parameter in it problems and the second one considering uncertainty.

Ko et al. [7] presented a mixed integer nonlinear programming model for the design of a dynamic integrated distribution network to account for the integrated aspect of optimizing the forward and return network simultaneously. After that ko et al. [8] proposed a hybrid optimization/simulation approach to design a distribution network for 3PLs in consideration of the performance of the warehouses. Min et al [9] presented a mixed-integer programming model and a genetic algorithm that can solve the reverse logistics problem involving the location and allocation of repair facilities for 3PLs. Their model can optimally create the reverse logistics network linking repair facilities, warehouses, and manufacturing facilities.

Zhang et al. [11] proposed one fuzzy chance constrained programming model to study the remanufacturing logistics network design problem from third party logistics supplier(3PLS)' perspectives under the uncertain environment. Ke et al. [13] focused on the multiple of object and uncertainty of environment for 3PLS to design the customizing logistics facility network, a fuzzy multi-object programming model. They studied, the integration of the forward and reverse logistics demand for 3PLS' clients, and multi-commodity, multi-customer and capacitated facility location-allocation problem are taken into account. Literature of network design under respect of 3PLs is summarized in table1.

2 PROBLEM FORMULATION

In this section, is gave the details of the overall mathematical model. This model involve shipping, inventory handling, lost sailing, and ordering cost. Index sets are as follows:

P : Index of clients' product types
I : The index of production sites
J : The index of existing warehouses and new potential sites
L : The index of existing repair facilities and new potential sites
K : The index of fixed customer locations
T : The index of time periods

Several assumptions are considered for the proposed model: Capacity constraints for vehicles do not exist. Carry all products from manufacturer i to warehouse j one type of vehicle is used. Customer demands are deterministic. If the capacities of warehouse or repair center are increased, they are no to be closed in the ext periods. Planning horizon is the same for all customers. All products of manufacturer b is given order by warehouse j together All of the returned products at the end of period will be delivered to manufacturer. Inventory has linearly reduced and has suddenly entered. Inventory control costs for reverse logistics is to be negligible. For each warehouse in each period a number cycle ordering are there. Term contract is the same for all customers.

The notation of parameters is as follow:

mc_{pit} : The maximum production capacity of client's plant i for product p; $i \in I$, $t \in T$, $p \in P$

mc^{w}_{jt} : The maximum capacity of warehouse j ; $j \in J$, $t \in T$,

mc^{r}_{lt} : The maximum capacity of repair facility l; $l \in L$, $t \in T$,

ub_j^w : The upper bound for modular expansion size of warehouse j ; $j \in J$,

ub_l^r : The upper bound of modular expansion size repair facility l; $l \in L$,

β_p : Per unit storage capacity by product p, $p \in P$,

d_{pkt} : Demand of product p at customer k in period t ; $p \in P, k \in K, t \in T$,

p_{pkt} : The percent of returns of product p from customer k in period t ; $p \in P, k \in K, t \in T$,

B_t : Assigned budget in period t; $t \in T$

Cost parameter

A_{ijt} fixed cost of ordering from warehouse j to manufacturer i at period t; $i \in I$, $j \in J$, $t \in T$

h_{pjt} : unit inventory holding cost of product p in warehouse j at period t; $p \in P$, $j \in J$, $t \in \mathbf{T}$

s_{pkt} : unit lost sales penalty cost of product p in customer k at period t**;** $p \in P$, $k \in K$, $t \in T$

fs_{jt}^w : the setup cost for installing warehouse j in period t; $j \in J, t \in T$,

fe_{jt}^w : the fixed cost of expanding modular size on warehouse j in period t; $j \in J, t \in T$,

ve_{jt}^w : the operating cost for modular expansion of warehouse j in period t; $j \in J, t \in T$,

CC_{jt}^w : cost of closing warehouse j in period t; $j \in J, t \in T$

fs_{lt}^r : the setup cost for installing repair center l in period t; $l \in L, t \in T$,

fe_{lt}^r : the fixed cost of expanding repair center l in period t; $l \in L, t \in T$,

ve_{lt}^r : the variable cost associated with the expansion of repair center l in period t; $l \in L$, $t \in T$,

CC_{lt}^r : cost of closing repair center l in period t; $l \in L, t \in T$,

wr_{jt} : the costs of savings associated with opening an integrated warehouse-repair facility in period t; ; $j \in J, t \in T$,

c_{pijt}^f : the unit variable cost of transportation of product p from manufacturer i to warehouse j in period; $p \in P, i \in I, j \in J, t \in T$,

c_{pklt}^r : the unit variable cost of transportation of return product p from costumer k to repair centre j in period t; $p \in P, k \in K, l \in L, t \in T$,

c_{pjkt}^f : the unit variable cost of transportation of product p from warehouse j to costumer k in period t; $p \in P, j \in J, k \in K, t \in T$,

c_{plit}^r : the unit variable cost of transportation of return product p from repair center l to manufacturer i in period t; $p \in P, l \in L, i \in I, t \in T$,

The decision variables of the problem are as follow:

X_{pijt}^f : Amount of transported product p from manufacturer i to warehouse j in period t ; $p \in P, i \in I, j \in J, t \in T$,

X^{r}_{pklt} : Amount of returned product p from customer k to repair center l in period t ; $p \in P$, $k \in K$, $l \in L$,$t \in T$,

X^{f}_{pjkt} : Amount of transported product p from warehouse j to manufacturer i in period t ; $p \in P$, $j \in J$, $k \in K$, $t \in T$,

X^{r}_{plit} : Amount of returned product p from repair center l to repair center l in period t ; $p \in P$, $l \in L$, $i \in I$,$t \in T$,

$$Z_{jt} = \begin{cases} 1 & \text{if warehouse } j \text{ is open in period } t,\ j \in J, \\ 0 & \text{otherwise,} \end{cases}$$

$$WE_{jt} = \begin{cases} 1 & \text{if warehouse } j \text{ is expanded in period } t,\ j \in J, \\ 0 & \text{otherwise,} \end{cases}$$

$$G_{lt} = \begin{cases} 1 & \text{if repair center } l \text{ is open in period } t,\ l \in L, \\ 0 & \textit{otherwise}, \end{cases}$$

$$RE_{lt} = \begin{cases} 1 & \text{if repair center } l \text{ is expeaded in period } t,\ l \in L, \\ 0 & \textit{otherwise}, \end{cases}$$

E^{w}_{jt} : The value of expansion for warehouse j in period t if warehouse j is installed; $j \in J$, $t \in T$

E^{r}_{lt} : The value of modularized expansion for repair center l in period t if repair center l is installed; $l \in L$, $t \in T$,

N_{ijt} : The integer number of orders from warehouse j to producer i in period t

Now mathematical models are as follow:

$$\text{Min}=\sum_{t=1}\left[\sum_{j}\sum_{p}\sum_{i} X^{f}_{pijt}.c^{f}_{pijt} + \sum_{j}\sum_{p}\sum_{k} X^{f}_{pjkt}.c^{f}_{pjkt} + \sum_{l}\sum_{p}\sum_{k} X^{r}_{pklt}.c^{r}_{pklt} + \sum_{l}\sum_{p}\sum_{i} X^{r}_{plit}.c^{r}_{plit}\right]$$
$$+\sum_{t}\sum_{j}\left(\sum_{i}\left(\frac{\sum_{p} X^{f}_{pijt}}{2N_{ijt}}\right)h_{pjt}\right) + \sum_{t}\sum_{i}\sum_{j} N_{ijt}.A_{ijt} + \sum_{k}\sum_{t}\sum_{p} s_{pkt}\left(d_{pkt} - \sum_{j} X^{f}_{pjkt}\right) \tag{1}$$

$$\sum_{j}\left[fs^{w}_{jt}.Z_{jt}.(1-Z_{jt-1}) + CC^{w}_{jt}.Z_{jt}.(1-Z_{jt+1})\right] + \sum_{l}\left[fs^{r}_{lt}.G_{lt}.(1-G_{jt-1}) + CC^{r}_{lt}.G_{lt}.(1-G_{lt+1})\right]$$
$$+\left[\sum_{j} fe^{w}_{jt}.WE_{jt} + \sum_{j} ve^{w}_{jt}.E^{w}_{jt}\right] + \left[\sum_{l} fe^{r}_{lt}.RE_{lt} + \sum_{l} ve^{r}_{lt}.E^{r}_{lt}\right] - \left[\sum_{j=l=1} wr_{jt}.Z_{jt}.G_{lt}\right] \le B_t \quad \forall t \in \mathrm{T} \tag{2}$$

$$\sum_{j} X^{f}_{pijt} \le mc_{pit} \qquad \forall p \in P, i \in I, t \tag{3}$$

$$\sum_{p}\sum_{i} \beta_p.\left(\frac{X^{f}_{pijt}}{N_{ijt}}\right) \le mc^{w}_{jt}.Z_{jt} + \sum_{x=1}^{t} E^{w}_{jx} \qquad \forall j \in J, t \tag{4}$$

$$\sum_{k}\sum_{p} \beta_p.X^{r}_{plkt} \le mc^{r}_{lt}.G_{lt} + \sum_{x=1}^{t} E^{r}_{lx} \qquad \forall l \in L, t \in T \tag{5}$$

$$\sum_{j} X^{f}_{pjkt} \le d_{pkt} \qquad \forall p \in P, k \in K, t \tag{6}$$

$$\sum_{l} X^{r}_{pklt} \geq p_{pkt} \sum_{j} X^{f}_{pjkt} \qquad \forall p \in P, k \in K, t \tag{7}$$

$$\sum_{i} X^{f}_{pijt} = \sum_{k} X^{f}_{pjkt} \qquad \forall p \in P, j \in J, t \in T \tag{8}$$

$$\sum_{k} X^{r}_{pklt} = \sum_{i} X^{r}_{plit} \qquad \forall p \in P, l \in L, t \in T \tag{9}$$

$$E^{w}_{jt} \leq ub^{w}_{j}.WE_{jt} \qquad \forall j \in J, t \in T \tag{10}$$

$$E^{r}_{lt} \leq ub^{r}_{l}.RE_{lt} \qquad \forall l \in L, t \in T \tag{11}$$

$$(NT - t + 1)WE_{jt} \leq \sum_{x=t}^{NT} Z_{jx} \qquad \forall j \in J, t \in T \tag{12}$$

$$(NT - t + 1)RE_{lt} \leq \sum_{x=t}^{NT} G_{lx} \qquad \forall l \in L, t \in T \tag{13}$$

$$X^{f}_{pijt} \leq M \times Z_{jt} \qquad \forall p, \text{i,j,t} \tag{14}$$

$$X^{f}_{pjkt} \leq M \times Z_{jt} \qquad \forall p, \text{ j, k, t} \tag{15}$$

$$X^{r}_{pklt} \leq M \times G_{lt} \qquad \forall p, k, l, t \tag{16}$$

$$X^{r}_{plit} \leq M \times G_{lt} \qquad \forall p, l, i, t \tag{17}$$

$$X_{pijt} \leq M \times A_{ijt} \qquad \forall p, \text{i,j,t} \tag{18}$$

The objective function (1) minimizes total transportation, warehousing, ordering and lost sales costs. Constraint (2) assures that the total investment costs that consist of the costs of opening, closing and expansion of facilities costs, the savings from integrated facilities, and expansion costs in forward and reverse flows, does not exceed the budget. In this model supposed that there is no facility at the start and end of planning horizon hence Zj,0 and Zj,T+1 is zero . Constraint (3) assures that the manufacturing plant have limited capacities during contract terms. Constraint (4) is the capacity limitations on warehouses including expansion size across the time period. Also constraint (5) is the capacity limitations on repair centres including expansion size across the time period. Constraint (6) guarantees that shipping product to customer doesn't exceed from customer demands. Constraint (7) ensures that the returned products should be sent to customers. Constraint (8) insures that the total products into each warehouse equal to the total product out the same warehouse at each period. Really this constraint clarifies that inventory should not be existed in warehouse at the end of the period. Similarly Constraint (9) insures that the total products into each repair centre equal to the total product out the same repair centre at each period. Really this constraint clarifies that inventory should not be exist in repair centre at the end of the period. Constraint (10) limits the added capacity of an existing warehouse during expansion. Similarly constraint (11) limits the added capacity of an existing repair centre during expansion. Constraint (12) also assures that if there is an expansion decision at any facility, this facility would not be closed. Constraint (13) also assures that if there is an expansion decision at any repair centre, this facility would not be closed. From constraints (14) to (17) are relate to flows products. If warehouse or repair centre aren't generated in candidate location, there isn't any flows in or out of candidate locations. Finally constraint (18) explains that if the number of orders from warehouse to manufacturer equals to zero, total of shipped products from manufacturer to warehouse should be zero.

3 EXPERIMENTAL RESULT

To illustrate the validity of the proposed model, a small example was solved by lingo 8 on a vostro 1500 with 2.2 GHz CPU and 2 GB Ram. One instance is generated randomly in small size. This example is involved, two production site, five warehouse site, five repair centre, three costumer, two product and two period. Totally it has 90 integer variables that 30 of them are binary variables and also 262 constraints that 19 of them is nonlinear. Input parameters are shown in table 2. Lingo solved this problem by local solver in one second. The best objective is 11660.7.

4 CONCLUSION

Because of the increasing importance of outsourcing companies, 3PLs have a key role in logistic services. In 3Pls one thing that should be marked is logistic network design. Pervious researches are studied it. But a few of them consider all of the logistic cost. In this article proposed a new model that optimizes logistic cost together. Our contribution is that involving inventory cost. Really it has been supposed that in each period can have some cycle time. Proposed model gives optimum cycle time and lost sales for each warehouse in each period. The future researches of this paper are as follows:

- Considering nondeterministic parameter for this model such demand of costumer.
- Solving problem with meta-heuristic algorithms and hybrid them with exact method.
- Incorporating this problem with vehicle routing problem.

References

[1] Bask, A.H., 2001. "Relationships among TPL providers and members of supply chains: A strategic perspective", Journal of Business & Industrial Marketing, Vol. 16, No. 6, pp. 470–486

[2] Hertz, S., & Alfredsson, M., 2003. "Strategic development of third party logistics providers". Industrial Marketing Management, Vol. 32, pp. 139–149

[3] Ke, W., Xiao-jiao, C., and Yong, Z., 2008. "Fuzzy Multi-Object Model and Solution to Design Logistics Facility Network for 3PLS", Proceedings of the IEEE International Conference on Automation and Logistics

[4] Ko, H.J., Evans, G.W., 2005. "A genetic algorithm-based heuristic for the dynamic integrated forward/reverse logistics network for 3PLs", Computers & Operations Research, Vol. 34, pp. 346–366

[5] Ko, H.J., Ko, C.S., and Kim, T., 2006. "A hybrid optimization/simulation approach for a distribution network design of 3PLS", Computers & Industrial Engineering, Vol. 50, pp. 440–449

[6] Lieb, R.C., 1992. "The use of third-party logistics services by large American manufacturers". Journal of Business Logistics, Vol. 13, No.2, pp. 29–42

[7] Marasco, A., 2008. "Third-party logistics: A literature review", International Journal of Production Economics, Vol. 113, pp. 127–147

[8] Min, H., Ko, H.J., Ko, C.S, 2004. "A genetic algorithm approach to developing the multi-echelon reverse logistics network for product returns", Omega, Vol. 34, pp. 56 – 69

[9] Min, H., Ko, H.-J., 2008. "The dynamic design of a reverse logistics network from the perspective of third-party logistics service providers", international journal of Production Economics, Vol. 113, pp. 176–192

[10] Zhang, Y., Xie, L., Hang, W., and Cui, X., 2007. "A Robust Model for 3PLS to Design a Remanufacturing Logistics Network under the Uncertain Environment", Proceedings of the IEEE International Conference on Automation and Logistics

A FLEXIBLE OPTIMIZATION FRAMEWORK FOR DRIVER SCHEDULING[1]

Attila Tóth and **Miklós Krész**
Institute of Applied Sciences
Juhász Gyula Faculty of Education
University of Szeged
Boldogasszony sgt. 6, 6725 Szeged, Hungary
{attila,kresz}@jgypk.u-szeged.hu

Abstract: In this paper we work out an application-oriented flexible optimization framework for driver scheduling. The new framework is general enough for the known optimization methods to be integrated and also take into consideration most of the practical requirements occurred in real-life public transportation systems. The potential usability of our approach is presented through the case of a Hungarian local transportation company.

Keywords: optimization in transportation, driver scheduling, real-life applications

1 INTRODUCTION

Driver scheduling is a central logistic problem in public transportation, since driver related expenses make a large part of total transportation costs. The basic driver scheduling problem is defined in the following way. There are vehicle trips with their fixed departure and arrival time and place, and the task is to assign drivers to each trip without overlapping, fulfilling the rules and regulations with minimal cost. The classical mathematical formulation of the above situation results in the set covering problem known to be NP-complete [1]. In literature, several solution methods and applications can be found; for an overview see [2].

In real life situations, a solution approach for driver scheduling is expected to be integrated into a decision support and transportation planning system, thus flexibility and modularity are particularly important for these methods. The above mentioned features are rarely captured in published results: While academic research concentrates on the optimization methods and mathematical results, the real life conditions are missing from the models. Concerning the case of driver scheduling, several frameworks are worked out (see a nice example in [3]), but these approaches are usually based on narrow circumstances and built inside a compact application. Nevertheless, in commercial application, the main demand is the flexibility of adaptation and usage: the entirely practical feasibility of the solution is more important than its high quality (at least from mathematical optimization point of view). Therefore, the solution methods usually act for the two opposite cases: the academic way with high quality but low complexity, and the commercial way with complex regulations but poor optimization. The aim of this paper is to bridge this duality by working out a flexible, application-oriented optimization framework for driver scheduling, which provides a model for integrating general optimization methods and problem-specific constraints (defined by the particular application or company). Our framework is independent from both the commercial requirements and the optimization method.

2 A FLEXIBLE MODEL FOR DRIVER SCHEDULING

Since the staff scheduling in different application areas are really dissimilar in regulations, requirements and preferences, it is inevitable to divide the scheduling into two main phases.

---

[1] This work was partially supported by Szeged City Bus Company (Tisza Volán, Urban Transport Division).

The first phase is the *generation* of the shifts producing so-called rough shifts containing only the basic tasks; while the *completion* phase results in the final shifts fulfilling all the rules with being acceptable in industrial usage. The first step is general, responsible for the quality of the shifts by applying some advanced optimization methodology, while the second step is usually a deterministic method being application-specific (or company-specific). Although this duality is an elementary attribute of any industrial staff scheduling problem, this approach is usually not considered in the published solution methodologies. The indisputable advantage of our framework is that the first general phase can be adapted to almost any application area by defining shift-events and their parameters, while the second area-dependent phase is separated from the optimization process. This flexible model gives a useful framework for any transportation operators' hand. Since the set of shift-events are varied in different application areas, the main questions with respect to such a model are the following: How can we define the events of the shifts such that their relations are also described? How can we set the rules for the events?

2.1 Events

Most of the solution methods in the literature handle only basic tasks in driver scheduling, such as passenger trips or simple breaks, while the situation of industrial staff scheduling is significantly different. Usually employees have several further duties in a shift, some of them are obligatory and some of them are potential tasks. Furthermore, a shift must dispose about the idle time between the activities (e.g. how it is paid). One of the main objectives of such a framework is to produce a serried system for defining these particular events determined by local circumstances. In our approach, operators can define the local proper events listed and categorized in 0-5 levels.

0. *Shift*
   This dummy event stands for the shift itself. The reason for introducing this level is to define rules concerning the global shift like maximum working time etc.
1. *Basic tasks*
   This type of activity is generally defined for vehicle duties (e.g. trips). These tasks are different from the later ones, as coming from the problem instances with fixed parameters. Their specification is not about to define them in the shift, but to select which ones are assigned to the currently built shift.
2. *Obligatory activities*
   Assigning of these activities is already the part of the completion phase. Each shift must contain these tasks, but the exact place can be fixed or freely selected. An example for obligatory activity is the daily cleaning duty.
3. *Potential activities*
   Presence of an activity from this type depends on some kind of other conditions of the shift, e.g. its structure. These activities must be inserted if a defined situation is occurred. A typical example for a potential activity is the break prescribed if the working time exceeds a limit.
4. *Optional activities*
   Activities which can be optionally used with some restrictions. For example, a driver can have a longer break around the middle of the shift.
5. *Idle event*
   This event type defines artificial events filling up the idle periods inside the shift. A finalized shift should not contain "gaps", each time period must be tagged with some characteristics (e.g. payment). These attributes can depend on either the length or the place of the idle periods. As an example, an idle time with a length of a few minutes

needs to be spent on the vehicle with full payment, while in a longer period the driver can move to a resting room with partial payment.

The members of 2-5 types are called together as *collateral activities*.

The allocation of an activity usually depends on some other type of event. As an example, if a certain task appears in the shift, another duty must follow it. If an activity has no such a dependence, we can say that the activity is connected to the dummy event (i.e. the shift itself). Therefore, in addition to categorization, the relationships of the events also need to be defined. The structure of the above categorization is a tree with its root being the shift event such that further activities are located in the tree connected to the event according to the dependency. We distinguish two basic types of connections: *pre-activity* and *post-activity*. In the case of pre-activity connections, the children must be inserted before the parent one, while for the post-activity case we are in the inverse situation. We also introduce a third connection type, the *inner-activity*, which can be used only between the shift and its child nodes with the meaning that the place of the activity is not prescribed inside the shift. The structure of the tree also represents the precedence of the events: for each node first check the current node, then its child nodes from left to right. It starts with the root (the shift) and then continues recursively. We can say that the general sequence of the activities in each level follows the order of the event types. See a sample event tree in Figure 1 of Szeged city public bus transportation in Hungary. (The events are from Table 1 in Section 4)

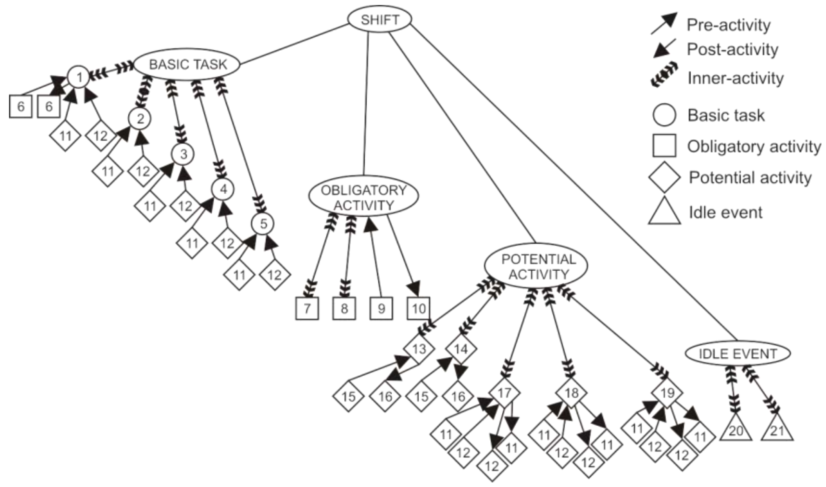


Figure 1: Event tree of Szeged public bus transportation.

2.2 Settings

The other important question of a general framework is the definition of the rules and preferences. The rules are about an event (including the shift event) and can be characterized by parameters. In many areas, different shift types are distinguished (e.g. early and late shift etc.), the parameters below usually should be defined to each type independently. If a parameter does not get a user value, it uses its default value found in the brackets.

Shift event

(S1) MinLength — Minimal length of the shift in minutes (default is 0).
(S2) MaxLength — Maximal length of the shift in minutes (1440).
(S3) WorkingTime — Maximal working time in minutes (1440).
(S4) DrivingTime — Maximal driving time in minutes (1440).
(S5) ShiftType[2] — Identifier of the shift type

[2] The list of the shift types should be defined earlier.

Basic tasks

(B1)	Driving	Signs if the task is driving or not (TRUE)
(B2)	Cost	Cost multiplier (1)
(B3)	Qualification[3]	Identifier of the required qualification (0)

Collateral activities

(C1)	MinLenght	Minimal length of the activity in minutes (0)
(C2)	MaxLength	Maximal length of the activity in minutes (1440)
(C3)	Multi	Signs if the activity can appear more than once (FALSE)
(C4)	MaxPiece	Maximum number of pieces into which it can be divided (1)
(C5)	PieceMinLenght	Minimal length of one piece of the activity in minutes (0)
(C6)	PieceMaxLength	Maximal length of one piece of the activity in minutes (1440)
(C7)	PlaceList[4]	Identifiers of the possible geographical places (ALL)
(C8)	MeasureRef[5]	Defines the reference time for the allocation (WORKING)
(C9)	FromTime	Earliest time when the activity can be started in the shift (0)
(C10)	UntilTime	Latest time when the activity can be started in the shift (1440)
(C11)	Working	Signs whether the activity is working time or not (TRUE)
(C12)	Cost	Cost multiplier (1)

Idle event

(I1)	MinLenght	Minimal length of the idle period in minutes (0)
(I2)	MaxLength	Maximal length of the idle period in minutes (1440)
(I3)	Multi	Signs if the event can appear more than once (FALSE)
(I4)	Working	Signs whether the event is working time or not (TRUE)
(I5)	Cost	Cost multiplier (1)

In an application, classifying all required events appearing in a shift, locating them in the event tree and using the above parameters, most of the rules and constraints can be defined and considered during the optimization process.

3 METHODOLOGY

As mentioned before, our driver scheduling approach is divided into two phases. The generation is the first phase being responsible for producing effective rough shifts containing basic tasks only. Completion as a second step produces the final shift for the real life usage. The main advantage of our framework is that it makes connection between the above phases. In industrial situation, the completion phase is generally too difficult to handle during an optimization process, since the solution is needed in reasonable time. However, executing an optimization method without taking into consideration of all the rules can lead to poor or even infeasible solutions. With the above detailed system, the generation phase can handle the rules, estimate the final cost and prepare the rough shifts for the completion method. With our categorization, the optimization process can be independent from the application area and only the completion algorithm must be revised.

In literature, several methods are known for crew scheduling containing a shift building phase. In Generate and Select approach, the generate phase produces a high number of shifts [2]. The constructive approach produces the shifts along a guided optimization way

[3] The list of the qualification should be defined earlier. 0 means no qualification is needed.
[4] The list of the geographical places should be defined earlier.
[5] It has three values: working, driving and shift length.

[4]. The iteratively improving approach starts with initial shifts and performs modifications for the iterative improvement [5]. In each approach, changes of a shift means removing or adding basic tasks, during which feasibility needs to be permanently checked. Therefore, this methodology for the generation phase in our model is useful for each optimization algorithm.

Modifications such as described above have a local and a global effect. For example, if a task is added to the shift, it locally modifies the idle gap before and after, the sequence of the basic tasks, and some collateral activities might also need to be “attached”. On the other hand, the new task globally changes the working time and the driving time in the shift. In order to keeping the feasibility both effects must be handled.

The methodology of keeping the feasibility during the shift modification action works in three steps:

1. *Initialization*

   In this step, the shift rules are checked with respect to parameters (S1)-(S5). The length of those obligatory activities which are connected to the shift with inner type and have no exact place, i.e. (C9) and (C10) are empty, can be added to the shift values, since they must be contained in the shift.

2. *Activities*

   Each collateral activity is handled sequentially by its location in the event tree. Obligatory pre- and post-activities are obviously joined before or after its parent activity, taking into consideration the corresponding parameters (the inner connection can appear only with the shift event and discussed above). This action is deterministic, quite easy and fast. Nevertheless, the case of the potential and optional activities can be more difficult. It needs to check other conditions concerning either local or global circumstances. For example, the current activity must be inserted before the parent activity, if the considered time period is long enough (local condition), but it can appear once in the shift, i.e. C3=FALSE (global condition). Obviously, after each activity action the global variables are updated.

3. *Idle events*

   In this phase, the appropriate event is assigned to each idle period. This action could be easy and fast, but possibly also leads to infeasibility. If an idle event is working time (I4=TRUE), it can change the conditions of some activities assigned before, which, for example, should appear in the shift at a given time (C9 and C10 are set, while C8=WORKING). In that case, these two events depend on each other, and the best option is to consider the latter idle event during the activity assignment by an estimation or statistical measuring.

A general measurement of the crew scheduling is the number of employees is needed, i.e. the number of shifts. However, the smaller number of employees does not necessarily decrease the cost. Since there is a duality in the model, the optimization works on rough shifts, but the real operational cost can be measured on the final shifts after the completion phase. The appropriate cost function is crucial for the quality. The generation phase is general and application independent for evaluation; it does not use any special information from the completion phase. However, the completion is always specific with respect to the application area, and the evaluation needs adjusting. The evaluation of both phases should be consonant, thus the “best” solution in the generation should be approximately the “best” after the completion.

The main difficulty is to solve the above conflict. It is mainly caused by handling of those activities in generation phase which have no fixed place. This means that they are not connected to any other event before or after, i.e. they are connected to the shift with inner connection. All other activities can be calculated and stored quite easily when a modification is performed. Therefore, the cost of the “gaps” in the shift can not be measured exactly during the generation, since the mentioned activities can get into them later in the completion phase. Accordingly, the cost of these periods is determined using a statistical ranking. This ranking is always depends on the local application area and needs professional consultations.

4 APPLICATION

The correctness of our framework is confirmed in a real life application of Szeged city local bus transportation with approximately 2700 trips and 166 drivers working in a general workday. Table 1 shows the list of the operative events, while the corresponding event tree is represented by Figure 1.

Table 1: Event list of Szeged city local bus transportation

Basic task		Obligatory activity		Potential activity		Idle event	
0	SHIFT	6	GETOFFON	11	DRCHANGE	20	TECHTIME
1	TRIP	7	VEHMAINT	12	TRAVEL	21	AVAIL
2	VEHSTART	8	ADMIN	13	DAYREST		
3	VEHEND	9	DRSTART	14	SPLITTIME		
4	PATROL	10	DREND	15	MIDDRSTART		
5	DEADHEAD			16	MIDDREND		
				17	BREAK 1		
				18	BREAK 2		
				19	BREAK 3		

The model fits to the requirements of the company as well as that of the national rules, and the evaluation system was proved to be robust. The applied optimization is based on a greedy method which produced feasible solutions with good quality in 2-3 seconds. Though concerning the number of drivers, the algorithm does not provide an improvement, it results in 2-3% cost reduction in average compared to the present scheduling of the local bus company. For further details of the application see [4].

5 CONCLUSION

We have presented an application-oriented flexible optimization framework for driver scheduling. The approach is strictly implementation oriented, but integrates the results of academic research as well. The methodology was tested at Szeged city bus company, Hungary, which showed the good applicability of the framework.

References

[1] Garey, M.R. Johnson, D.S., Computers and Interactability: *A Guide to the Theory of NP-Completness*, Freeman, San Fransisco, 1979.

[2] A.T. Ernst, H. Jiang, M. Krishnamoorthy, D. Sier, *Staff scheduling and rostering: A review of applications, methods and models*, European Journal of Operational Research 153 (2004), pp. 3–27.

[3] A. Wren, S. Fores, A. Kwan, R. Kwan, M. Parker, L. Proll, A flexible system for scheduling *drivers*, Journal of Scheduling, Vol. 6, No. 5. (2003), pp. 437-455.

[4] V. Argilan, Cs. Kemeny, G. Pongracz, A. Toth, B. David, *Greedy heuristics for driver scheduling and rostering*, in Proceedings of Information Society 2010, Proceedings A (2010), pp. 397-400.

[5] L. Cavique1, C. Rego, I. Themido, *Subgraph Ejection Chains and Tabu Search for the Crew Scheduling Problem*, Journal of the Operational Research Society 50 (1999), pp. 608-616.

The 11th International Symposium on
Operational Research in Slovenia

SOR ’11

Dolenjske Toplice, SLOVENIA
September 28 - 30, 2011

Appendix

Authors' addresses

Addresses of SOR'11 Authors

(The 11th International Symposium on OR in Slovenia, Dolenjske Toplice, SLOVENIA, September 28 – 30, 2011)

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
1.	Erling D.	Andersen	MOSEK ApS	Fruebjergvej 3, Box 16	2100	Copenhagen	Denmark	e.d.andersen@ mosek.com
2.	Josip	Arnerić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	jarneric@efst.hr
3.	Zoran	Babić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	babic@efst.hr
4.	Karlo	Bala	University of Novi Sad, Faculty of Philosophy	Dr. Zorana Đinđića 2	21000	Novi Sad	Serbia	kbalayu@gmail.com
5.	Danijel	Barbarić	Pravni fakultet Split	Domovinskog rata 8	21000	Split	Croatia	danijel.barbaric@pravst. hr
6.	Martina	Basarac	Economic Research Division, Croatian Academy of Sciences and Arts	Strossmayerov trg 2	10000	Zagreb	Croatia	mbasarac@hazu.hr
7.	Tomaž	Berlec	University of Ljubljana, Faculty of Mechanical Engineering, Laboratory for Manufacturing Systems and Production Process Planning	Aškerčeva ulica 6	1000	Ljubljana	Slovenia	tomaz.berlec@fs.uni-lj.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
8.	Marija	Bogataj	University of Ljubljana, Faculty of Maritime Studies and Transport; MEDIFAS			Portorož; Nova Gorica	Slovenia	marija.bogataj@ guest.arnes.si
9.	Dejan	Brcanov	Faculty of Economics Subotica	Segedinski put 9-11	24000	Subotica	Serbia	dejanbrcanov@ gmail.com
10.	Andrej	Bregar	Informatika d.d.	Vetrinjska ulica 2	2000	Maribor	Slovenia	andrej.bregar@ informatika.si
11.	Alenka	Brezavšček	University of Maribor, Faculty of Organizational Sciences	Kidričeva cesta 55a	4000	Kranj	Slovenia	alenka.brezavscek@ fov.uni-mb.si
12.	Kristijan	Cafuta	University in Ljubljana, Faculty of Electrical Engineering	Tržaška 25	1000	Ljubljana	Slovenia	kristijan.cafuta@ fe.uni-lj.si
13.	Mehmet	Can	International University of Sarajevo, Faculty of Engineering and Natural Sciences	Hrasnićka Cesta 15	71000	Sarajevo	Bosnia and Herzegovina	mcan@ius.edu.ba
14.	Maja	Copak	University of Zagreb, Faculty Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	maja.copak@ hotmail.com
15.	Vesna	Čančer	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	vesna.cancer@uni-mb.si
16.	Anita	Čeh Časni	University of Zagreb, Faculty of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	aceh@efzg.hr
17.	Marjan	Čeh	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	mceh@fgg.uni-lj.si
18.	Draženka	Čizmić	University of Zagreb, Faculty of Economics	Kennedyev trg 6	10000	Zagreb	Croatia	dcizmic@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
19.	Mirjana	Čižmešija	University of Zagreb, Faculty Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	mcizmesija@efzg.hr
20.	Alan	Domić	University of Zagreb, Faculty Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	alandomic@gmail.com
21.	Samo	Drobne	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	samo.drobne@ fgg.uni-lj.si
22.	Ksenija	Dumičić	University of Zagreb, Faculty of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	kdumicic@efzg.hr
23.	Igor	Đukanović	Univerza v Mariboru, Ekonomska-poslovna fakulteta	Razlagova 14	2000	Maribor	Slovenia	igor.mat@uni-mb.si
24.	Milena	Đurović	University of Belgrade, Faculty of Organizational Sciences	Jove Ilica 154		Belgrade	Serbia	djurovicmilena86@ gmail.com
25.	Mejra	Festic	Bank of Slovenia	Slovenska 35	1505	Maribor	Slovenia	mejra.festic@bsi.si
26.	Siniša	Franjić	Elektrotehnički fakultet Osijek	Kneza Trpimira 2b	21000	Split	Croatia	sinisa.franjic@etfos.hr
27.	M. Victoria	de la Fuente Aragon	Universidad Politécnica de Cartagena	Campus Muralla del Mar	30202	Cartagena	Spain	marivi.fuente@upct.es
28.	Margareta	Gardijan	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	mgardijan@efzg.hr
29.	Katarina	Gerhardt	Department of Mathematics, University of Kaiserslautern				Germany	gerhardt@ mathematik.uni-kl.de
30.	Frank	Gielsdorf	Technet GmbH			Berlin	Germany	frank.gielsdorf@technet-gmbh.com

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
31.	Stanislav	Gorenc	PLASTA proizvodnja in trgovina d.o.o.			Šentrupert	Slovenia	stane.gorenc@gmail.com
32.	Edvard	Govekar	University of Ljubljana, Faculty of Mechanical Engineering, Laboratory for Synergetics	Aškrčeva 6	1000	Ljubljana	Slovenia	edvard.govekar@fs.uni-lj.si
33.	Jelena	Govorčin	Fakulteta za informacijske študije v Novem mestu	Novi trg 5	8000	Novo mesto	Slovenia	jelena.govorcin@fis.unm.si
34.	Petra	Grošelj	University of Ljubljana, Biotechnical Faculty	Jamnikarjeva 101	1000	Ljubljana	Slovenia	petra.groselj@bf.uni-lj.si
35.	Walter J.	Gutjahr	University of Vienna, Department of Statistics and Operations Research				Austria	walter.gutjahr@univie.ac.at
36.	Nebojša	Gvozdenović	Univerzitet u Novom Sadu, Ekonomski fakultet Subotica	Segedinski put 9-11	24000	Subotica	Serbia	nebojsa.gvozdenovic@gmail.com
37.	Horst W.	Hamacher	Department of Mathematics, University of Kaiserslautern				Germany	hamacher@mathematik.uni-kl.de
38.	Eloy	Hontoria	GIO Universidad Politécnica de Cartagena	Campus Muralla del Mar	30202	Cartagena	Spain	eloy.hontoria@upct.es
39.	Mitja	Hrast	Kovinski izdelki Hrast Milan s.p.			Breginj	Slovenia	mitja.hrast@gmail.com
40.	Dušan	Hvalica	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	dusan.hvalica@ef.uni-lj.si
41.	Mohamad-Saied	Jabalameli	Iran University of science and technology, Department of industrial engineering				Iran	jabal@ind.iust.ac.ir

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
42.	Amir	Jamak	International University of Sarajevo, Faculty of Engineering and Natural Sciences	Hrasnićka Cesta 15	71000	Sarajevo	Bosnia and Herzegovina	amir.jamak@ bhtelecom.ba
43.	Alenka	Kavkler	University of Maribor, Faculty of Economics and Business	Razlagova 14	2000	Maribor	Slovenia	alenka.kavkler@ uni-mb.si
44.	Igor	Klep	University of Maribor, Faculty of Natural Science and Mathematics; Universtity of Ljubljana, Faculty of Mathematics and Physics	Koroška 160 Jadranska 19	2000 1111	Maribor, Ljubljana	Slovenia	igor.klep@ fmf.uni-lj.si
45.	Jelena	Knezović	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	jelenaknezovic@ yahoo.com
46.	Arie M. C. A.	Koster	RWTH Aachen University, Lehrstuhl II für Mathematik	Wüllnerstr. 5b	52062	Aachen	Germany	koster@ math2.rwth-aachen.de
47.	Danijel	Kovačić	University of Ljubljana, Faculty of Economics, Department of Mathematics and Operations Researchs	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	danijel.kovacic@ ef.uni-lj.si
48.	Miklós	Krész	Institute of Applied Sciences, Juhász Gyula Faculty of Education, University of Szeged	Boldogasszony sgt. 6	6725	Szeged	Hungary	kresz@ jgypk.u-szeged.hu
49.	Nataša	Kurnoga Živadinović	University of Zagreb, Faculty Economics and Business,	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	nkurnoga@efzg.hr
50.	Janez	Kušar	University of Ljubljana, Faculty of Mechanical Engineering	Aškrčeva 6	1000	Ljubljana	Slovenia	janez.kusar@fs.uni-lj.si
51.	Bojan	Kuzma	FAMNIT, University of Primorska	Glagoljaška 8	6000	Koper	Slovenia	bojan.kuzma@upr.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
52.	Marija	Kuzmanović	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	mari@fon.rs
53.	Anka	Lisec	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	anka.lisec@fgg.uni-lj.si
54.	Zrinka	Lukać	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	zlukac@efzg.hr
55.	Dragana	Makajić-Nikolić	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	gis@fon.rs
56.	Minja	Marinović	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	marinovic75709@ fon.bg.ac.rs
57.	Milan	Martić	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	milan@fon.rs
58.	Arash	Motaghedi-Larijani	Iran University of science and technology, Department of industrial engineering				Iran	amotaghedi@ ind.iust.ac.ir
59.	Mira	Oraić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	moraic@efzg.hr
60.	Irena	Palić	University of Zagreb, Faculty of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	ipalic@efzg.hr
61.	Biljana	Panić	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	biljana.panic@fon.bg.ac.rs
62.	Polona	Pavlovčič Prešeren	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	polona.pavlovcic@ fgg.uni-lj.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
63.	Tunjo	Perić	Bakeries Sunce	Rakitska cesta 98	10437	Bestovje	Croatia	tunjo.peric1@ zg.t-com.hr
64.	Ulrich	Pferschy	Department of Statistics and Operational Research, University of Graz	Universitaets-strasse 15	8010	Graz	Austria	pferschy@ uni-graz.at
65.	Snježana	Pivac	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	spivac@efst.hr
66.	Nada	Pleli	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	npleli@efzg.hr
67.	Primož	Potočnik	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva ulica 6	1000	Ljubljana	Slovenia	primoz.potocnik@fs.uni-lj.si
68.	Janez	Povh	Faculty of Information Studies in Novo mesto	Novi trg 5	8000	Novo mesto	Slovenia	janez.povh@ fis.unm.si
69.	Željko V.	Račić	University of Banja Luka, Faculty of Economics	Majke Jugovića 4	78000	Banja Luka	Bosnia and Herzegovina	zeljko.racic@efbl.org
70.	Maja	Remic	University of Ljubljana, Faculty of Computer and Information Science	Tržaška 25	1000	Ljubljan	Slovenia	maja.remic@siol.net
71.	Lidija	Rihar	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva ulica 6	1000	Ljubljana	Slovenia	lidija.rihar@fs.uni.lj.si
72.	Franco	Robledo Amoza	Instituto de Computación, Facultad de Ingeniería, Universidad de la República	Julio Herrera y Reissig 565	11300	Montevideo	Uruguay	frobledo@fing.edu.uy
73.	Lorenzo	Ros-McDonnell	Universidad Politécnica de Cartagena	Campus Muralla del Mar	30202	Cartagena	Spain	lorenzo.ros@upct.es

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
74.	Viljem	Rupnik	INTERACTA Ltd, Information Processing Engineering	Parmova 53	1000	Ljubljana	Slovenia	viljem.rupnik@siol.net
75.	Stefan	Ruzika	Department of Mathematics, University of Kaiserslautern				Germany	hamacher@ mathematik.uni-kl.de
76.	Pablo	Sartor Del Giudice	Instituto de Computación, Facultad de Ingeniería, Universidad de la República	Julio Herrera y Reissig 565	11.300	Montevideo	Uruguay	psartor@un.edu.uy
77.	Alen	Savatić	International University of Sarajevo, Faculty of Engineering and Natural Sciences	Hrasnićka Cesta 15	71000	Sarajevo	Bosnia and Herzegovina	alen@savatic.net
78.	Gordana	Savić	University of Belgrade, Faculty of Organizational Sciences	Jove Ilica 154		Belgrade	Serbia	goca@fon.rs
79.	Marko	Starbek	University of Ljubljana, Faculty of Mechanical Engineering, Laboratory for Synergetics	Aškerčeva 6	1000	Ljubljana	Slovenia	marko.starbek@ fs.uni-lj.si
80.	Milan	Stanojević	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	milans@fon.rs
81.	Bojan	Stopar	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	bojan.stopar@ fgg.uni-lj.si
82.	Slavko	Šimundić	Pravni fakultet Split	Domovinskog rata 8	21000	Split	Croatia	slavko.simundic@pravst. hr
83.	Blanka	Škrabić-Perić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	bskrabic@efst.hr
84.	Tihana	Škrinjarić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	tihana.skrinjaric@ gmail.com

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
85.	Petra	Šparl	University of Maribor, Faculty of Organizational Sciences	Kidričeva cesta 55a	4000	Kranj	Slovenia	petra.sparl@fov.uni-mb.si
86.	Attila	Tóth	Institute of Applied Sciences, Juhász Gyula Faculty of Education, University of Szeged	Boldogasszony sgt. 6	6725	Szeged	Hungary	attila@ jgypk.u-szeged.hu
87.	Danijela	Tuljak Suban	University of Ljubljana, Faculty of Maritime Studies and Transport			Portorož	Slovenia	danijela.tuljak@ fpp.uni-lj.si
88.	Anita	Udiljak	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	anita.udiljak@ yahoo.com
89.	Mario	Varović	Zagrebačka škola ekonomije i managementa	Jordanovac 110	10000	Zagreb	Croatia	mario.varovic@ zsem.hr
90.	Višnja	Vojvodić Rosenzweig	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	vvojvodic@efzg.hr
91.	Hrvoje	Volarević	Zagrebačka škola ekonomije i managementa	Jordanovac 110	10000	Zagreb	Croatia	hrvoje.volarevic@ zsem.hr
92.	Ilko	Vrankić	University of Zagreb, Faculty of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	ivrankic@efzg.hr
93.	Slobodan	Vujić	University of Belgrade, Faculty of Mining and Geology	Djušina 10	11000	Belgrade	Serbia	vujic@rgf.bg.ac.rs
94.	Mirko	Vujošević	University of Belgrade, Faculty of Organizational Sciences	Jove Ilića 154	11000	Belgrade	Serbia	mirko.vujosevic@fon.bg.ac.rs
95.	Lidija	Zadnik Stirn	University of Ljubljana, Biotechnical Faculty	Jamnikarjeva 101	1000	Ljubljana	Slovenia	lidija.zadnik@bf.uni-lj.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
96.	Berislav	Zmuk	University of Zagreb, Faculty of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	bzmuk@efzg.hr
97.	Mateja	Zupan		Jesenovo 20		Čemšenik	Slovenia	mateja5zupan@gmail.com
98.	Urška	Železnik		Dolenja vas 30	1410	Zagorje ob Savi	Slovenia	u.zeleznik@gmail.com
99.	Gašper	Žerovnik	Jozef Stefan Institute, Reactor Physics Division	Jamova 39	1000	Ljubljana	Slovenia	gasper.zerovnik@ijs.si
100.	Janez	Žerovnik	Institute of Mathematics, Physics and Mechanics; University of Ljubljana, FME	Jadranska 19; Aškerčeva 6	1000	Ljubljana	Slovenia	janez.zerovnik@fs.uni-lj.si; janez.zerovnik@imfm.uni-lj.si