S. D.

Proceedings SOR'05 2.

Proceedings of the 8th International Symposium on OPERATIONAL RESEARCH

SOR '05

Nova Gorica, Slovenia
September 28-30, 2005

Edited by:
L. Zadnik Stirn • S. Drobne

S.D.

SOR '05 Proceedings

The 8th International Symposium on Operational Research in Slovenia
Nova Gorica, SLOVENIA, September 28 - 30, 2005

Edited by:

L. Zadnik Stirn and S. Drobne

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

Proceedings of the 8th International Symposium on Operational Research
SOR'05 in Slovenia, Nova Gorica, September 28 - 30, 2005.

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 2005 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

65.012.122(063)(082)
519.8(063)(082)

INTERNATIONAL Symposium on Operational Research in Slovenia (8 ; 2005 ; Nova Gorica)
SOR '05 proceedings / The 8th International Symposium on Operational Research in Slovenia, Nova Gorica, Slovenia, September 28-30, 2005 ; edited by L. Zadnik Stirn and S. Drobne ; [organiser Slovenian Society Informatika, Section for Operational Research]. - Ljubljana : Slovenian Society Informatika (SDI), Section for Operational Research (SOR), 2005

ISBN 961-6165-20-8

1. Zadnik Stirn, Lidija 2. Slovensko društvo Informatika. Sekcija za operacijske raziskave
222031104

Proceedings of the 8th International Symposium on Operational Research in Slovenia (SOR'05) 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 : MIGRAF, Ljubljana, Slovenia

The 8th International Symposium on Operational Research in Slovenia - SOR '05
Nova Gorica, SLOVENIA, September 28 - 30, 2005

Program Committee:

Organizing Committee:

Chairmen and Chairladies:

Preface

The purpose of this volume is to reflect the scientific activities during The 8th International Symposium on Operations Research, which was held in Nova Gorica, Slovenia, from September 28 through September 30, 2005. The symposium on Operations Research is one of the traditional series of the biannual international conferences organized by Slovenian Society INFORMATIKA, Section of Operational Research, and represents the continuity of seven previous symposia. At this symposium the scientists, researchers and practitioners from different areas, like mathematics, economics, statistics, computer science engineering, environment and system theory, often working together on common projects, came together. The 8th International Symposium on Operations Research SOR'05 stood under the high auspices of the Slovenian Research Agency, and was granted by sponsors cited in these Proceedings. The opening address was given by Mr. N. Schlamberger, the President of Slovenian Society INFORMATIKA, Prof. Dr. L. Zadnik Stirn, the President of the Slovenian Section of Operations Research, and the representatives of ministries, representatives of HIT, Nova Gorica, different professional institutions and Operations Research Societies from other countries.

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

As traditionally, also this symposium was an international forum for scientific exchange at the frontiers of Operations Research in mathematics, statistics, economics, engineering, education, environment and computer science. We believe that the presentations reflected the state of the art in Operations Research as well as the actual challenges. Besides contributions on recent advances in the classical fields, the presentations, on new interactions with related fields as well as an intense dialogue between theory and the numerous applications, were delivered at the symposium. Thus, we hope that the division into Invited lectures and 10 sections, reflects on the one hand the variety of fields engaged, on the other hand separating too many subjects which could belong together. The scientific program was divided into the following sections (the number of papers in each section is given in parentheses): Plenary section (6), Scheduling and Control (4), Stochastic and Combinatorial Optimization (4), Algorithms (7), Environment and Human Resources (7), Location Theory and Transport (10), Finance and Investment (8), Multicriteria Decision Making (5), Networks (4), Production and Inventory (5), Education and Statistics (3).

The first part of the Proceedings includes invited papers, presented by 6 prominent scientists: Bernhard Böhm, Vienna University of Technology, Vienna, Austria; Hubertus Th. Jongen, Aachen University, Aachen, Germany; Robert Manger, University of Zagreb, Zagreb, Croatia; Viljem Rupnik, Interacta, Ltd, Business Information Processing, Ljubljana, Slovenia; Dragomir Sundać, International Business Consulting Center, Ltd, Rijeka, Croatia; and Walter Ukovich, University of Trieste, Trieste, Italy. The second part of the Proceedings includes 57 contributed and refereed papers written by 92 authors and co-authors. Most of the authors of the contributed papers came from Slovenia (41), then from Croatia (30), Italy (4), Slovak Republic (3), Czech Republic (3), Austria (2), Belgium (2), Germany (1), Australia (1), Hungary (1), Israel (1), Latvia (1), UK (1), and Ukraine (1).

Further, the symposium was highlighted by two round tables: Globalization – Capital Domination, and Development and Educational Problems in Primorska Region with Regard to European Trends and Demands. They both attracted a great interest of SOR'05 participants. The discussion led to remarkable conclusions concerning regional cooperation and associations with other EU countries, as well as development and use of econometric models in regional and international economies.

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

We would not have succeeded in attracting so many distinguished speakers from all over the world without the engagement and the advice of active members of Slovenian Section of Operations Research. Many thanks to them. Further, we would like to express our deepest gratitude to the members of the Program and Organizing Committees, to the referees, chairpersons and to all the numerous people - far too many to be listed here individually - who helped in carrying out The 8th International Symposium on Operations Research SOR'05 and in putting together these Proceedings. At last, we appreciate the authors' efforts in preparing and presenting the papers, which made The 8th Symposium on Operational Research SOR'05 successful.

Nova Gorica, September 28, 2005

Lidija Zadnik Stirn
Samo Drobne
(Editors)

Foreword

Mathematics with all its theoretical and applied disciplines, in particular with operations research, is a field where people are active to discover new algorithms and approaches from the pleasure of finding things out. Its purpose is not to explain phenomena or to discover rules; this is a mission of other natural sciences. There, however, exists a close and firm relationship between those and mathematics/operations research: without its tools, methods, and insights all other sciences would be more or less descriptive and narrative and they would be able to explain hardly any of the things that they have discovered. Moreover, today, to ever higher degree they depend on mathematics in general and on its branches where we find also operations research, social and humanistic sciences. Without mathematical methods they would not be what they are now. Let us just remind on the frequency analysis of letters in Slovenian language. While it is true that this is not the latest achievement it is surely an illustrative one.

Mathematics in general has seen an intense development in the nineteenth and twentieth century. The description of electromagnetic field and its properties that were able to be predicted from solutions of partial differential equations are hardly imaginable without vector analysis and quaternions. Had Riemann not founded differential geometry of curved spaces which was a result of developing of purely abstract ideas, it should have been invented anew, as without it an analytical approach to Einstein's theory of gravity is impossible. In the twenties the matrix calculus has been developed, so to speak to wait on the shelf, only to be used soon afterwards as an indispensable tool for the first description of quantum mechanics. An overview of quantum states of elementary particles that has been painstakingly put together in a form of tables by W. Heisenberg that are hard to understand even today and which are the foundation of matrix mechanics is now possible to present as matrices that are readily possible to calculate with. As a consequence of need to solve complex problems in technology, economy, and social sciences, a new field of operations research has been invented. Today it is an apparatus that it is hardly imaginable to do without in the areas that at the first sight seem to be too complex to be accessible by mathematical tools – such as medicine, biology or ecology. Topology that has become of age in the second half of twentieth century is a discipline that broadens our insight on properties of multidimensional spaces. Last but not any less important, nearly at the end of the past century the solution to the Fermat great theorem that has molested the greatest minds for a three hundred years, has been found.

What is the agent of the development of mathematics and in particular of operations research?

A simple answer in a form of one condition that is at the same time necessary and sufficient does probably not exist. There are at least two conditions that must be met. The necessary condition is that there exists curiosity, a pleasure of finding things out, as the Nobel laureate R. Feynman has put it. We need not worry about this one; as long as man exists also the human curiosity will exist. The other condition is that new problems arise and that new questions are asked to which no existing knowledge can provide answers. Disciplines where no such problems come up any more that require search for new solutions have completed their evolution. That can even be measured; the index is the number of new publications. By this standard hardly anything else is going on in mechanics and thermodynamics, they both are solving only practical technical problems.

At the beginning of this foreword a thought has been expressed that mathematics and operations research have seen an intensive development in the last century and a half. Has it already lived past its golden age? This is hardly to believe. Aside from the questions to which there are still no answers there are also new challenges that arise from the development of new technologies. Computers and information science are just one of them. The alliance of mathematical methods combined with an immense processing power offers a wholly new perspective and possibilities for solving of special kinds of problems. Operations research is a typical field where such a combination promises an intensive progress. The present volume bears witness that such a belief is realistic.

Nevertheless there is one reservation and one hope. The reservation is that for the progress a motor and a fuel are needed. The motor of the progress are the before described conditions, the curiosity and new problems that bear new challenges. Both will be available in the future as they were available so far. The fuel is the concern of those that are to use the results of the research. Regarding this we may be worried by the dwindling interest of users for solving their problems using methods of operations research. Somewhat disappointed we see that in practice cases have begun to repeat from the time when results of research and development were able to compensate generously and profitably by primitive economy methods. The surprise is genuine as we now that in the long run only knowledge provides a solid basis for adding new value. Nevertheless there is a hope that the decline of motivation to use the methods of operations research is only a passing episode and that the users will understand the importance of this discipline to their advantage and to the benefit of all. The number of contributions in this volume of both domestic and foreign authors bears witness that such hope is not in vain.

Nova Gorica, September 28, 2005

Niko Schlamberger
(President
Slovenian Society INFORMATIKA)

Sponsors of the 8th International Symposium on Operational Research in Slovenia (SOR'05)

The 8th International Symposium on Operational Research in Slovenia was organized under the auspices of the Researh Agency, Republic of Slovenia.

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

- *Austrian Science and Research Liaison Office, Department of Ljubljana, Ljubljana*
- *HIT, Nova Gorica*

S.D.

Contents

Author index

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Invited lectures

Approaches to the measurement of eco-efficiency

Bernhard Böhm
Institute for Mathematical Methods in Economics
University of Technology, Vienna, Austria
(bernhard.boehm@tuwien.ac.at)

Abstract
The present paper presents a selection of approaches to measure eco-efficiency at different levels considering the multi-objective nature of this concept. It shortly reviews the level of implementation in industries and discusses major elements and processes. Different measurement concepts at the industry level will be reviewed. Aspects of eco-efficiency beyond the individual firm are considered next. Micro- and macroeconomic aspects will be integrated by a discussion of the multiple objective nature of the macroeconomic problem and an attempt to measure macro-eco-efficiency.

Keywords: Eco-efficiency, data envelopment analysis, input-output analysis, distance function

1. Introduction

During recent years the term "eco-efficiency" has surfaced in business economics as new criterion among the objectives of a firm. It is now playing an increasing role in firm management where sustainable development has been included among the set of firm targets. While formal decision analysis is often applied to complex economic problems, the development of the eco-efficiency concept and its application in practical management only indirectly relates to such analysis. While looking for an answer as to how the objective of sustainable development can be reached by the firm, the concept itself is often understood as an indirect objective or even an instrument or tool to achieve the sustainability objective. In the definition of the World Business Council for Sustainable Development we read that "eco-efficiency is reached by the delivery of competitively priced goods and services that satisfy human needs and bring quality of life, while progressively reducing ecological impacts and resource intensity throughout the life cycle, to a level at least in line with the earth's carrying capacity".

Approaches differ among countries, industries, and firms. The paper shortly reviews some measurement concepts at the industry level and compares eco-efficiency across individual firms. Micro- and macroeconomic aspects will be integrated by a discussion of the multiple objective nature of the macroeconomic problem and a macro-eco-efficiency measure.

2. Eco-efficiency at the industry level

"Eco-efficiency" is considered an important tool of business practice and management where innovations in technology, production, processes, product design and business organisation and practices can lead to lower unit costs, improved product quality, lower environment-related liability, less material usage and less adverse impact on the environment. The main message in promoting this "tool" is that economy and ecology do not exclude each other but can be made to cooperate profitably for both the firm and for society: more value can be generated with less use of resources and less damages for the environment in general.

Measurement issues of the degree of eco-efficiency achieved do not stand in the foreground of interest. Incorporating the four required strategic elements, de-materialisation, closure of production loops, service extension, enhancing functionality, the identification of indicators and the construction of a unified measure are hard issues. The approach of the WBCSD concentrates on a measure defined by $\left[\frac{\text{product or service value}}{\text{environmental influence}}\right]$, where both, numerator and denominator, have to be derived from indicators. Because of large differences due to products, branches, techniques, customers etc. firms may need to define their own relevant indicators appropriate for external communication and decision making. Thus indicators may belong either to the group of generally applicable or business specific ones. An eco-efficiency performance profile will incorporate both.

A more demanding concept to assess eco-efficiency is provided by life cycle analysis (LCA) of a product or process. It should help to assess potential effects of a product or process on environment starting from the inception of the idea and the use of the raw material through the user phase until the end in a waste deposit or a recycling process. It focuses specifically on energy materials used and wastes released to the environment. Multi-criteria decision techniques can generate dramatic productivity improvements in the application of LCA (Brunn and Rentz (1998)). Results from LCA will contribute to other measurement concepts. A prominent one in Germany is MIPS (**M**aterial **I**nput **P**er **S**ervice unit), developed at the Wuppertal Institute. It calculates the input of total material resources required for the production of goods and services, measured in service units. Material inputs cover all renewable or exhaustible resources, land, water and air, while a service unit represents the utility derived from the use of a particular good.

While many of the concepts discussed so far provide useful steps for the concerned firm, there are many instances when comparisons between firms are desired to judge either policy effectiveness or to design policy instruments appropriately. For such objectives the tool of data envelopment analysis (DEA) has been recommended. For the evaluation of economic efficiency this technique already enjoys widespread applicability as is witnessed by the large number of available publications (cf. Charnes et al. (1994)). To a large extent it avoids problems of valuations of inputs and outputs, or better, provides even valuation estimates. However, applying DEA to ecological and economic efficiency requires reconsideration of the conventional efficiency measure $\left[\frac{\textit{weighted sum of outputs}}{\textit{weighted sum of inputs}}\right]$ in order to adapt it to environmental magnitudes. The WBCSD definition, the MIPS, and similar concept that have been proposed, all relate environmental magnitudes (value of damages, emissions, resources) to some economic magnitude (service unit, utility, product value). Especially the environmental magnitudes cannot be easily classified as inputs or outputs as many of them are undesirable from the point of view of production. If one characterises ecological efficiency as a way of '*producing with the least energy, resources, waste and emissions*' then the lack of prices for emissions and waste or other undesirable outputs makes it difficult to construct measures of ecologically efficient production. But some of these problems can be overcome by using DEA.

3. The DEA approach to eco-efficiency

The use of a nonparametric method for productivity comparisons with non-desirable outputs seems to have started with a paper by Färe et al. (1989). This approach has later been applied to evaluate environmental performance of firms (cf. the review by Tyteca (1996)). A recent paper by Korhonen and Luptáčik (2004) contrasts two types of approaches. The first

decomposes the problem in two parts and measures first technical efficiency along conventional lines and then measures ecological efficiency by using a ratio of a weighted sum of desirable outputs to the weighted sum of undesirable outputs. The efficiency indicators of both models are used as output variables in a new DEA model with inputs equal 1, yielding the eco-efficiency indicator.
The second approach specifies ratios which simultaneously take into account desirable and undesirable outputs. In this case at least four different ratios can be constructed:
Model A maximises the ratio of the difference between a weighted sum of k desirable (y_r) and (p-k) undesirable outputs (y_s) to the weighted sum of m inputs (x_i) for each unit j=1,…,n.

$$\max h_A = \frac{\sum_{r=1}^{k}\mu_r y_{r0} - \sum_{s=k+1}^{p}\mu_s y_{s0}}{\sum_{i=1}^{m} v_i x_{i0}} \quad s.t. \quad \frac{\sum_{r=1}^{k}\mu_r y_{rj} - \sum_{s=k+1}^{p}\mu_s y_{sj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \quad \mu_r, v_i \ge \varepsilon \qquad (1)$$

$j = 1,\dots,n, \quad r = 1,\dots,p, \quad i = 1,\dots,m, \quad \varepsilon > 0 \quad (\text{"}non-Archimedean\text{"})$

Model B treats undesirable outputs as inputs and maximises the ratio of the weighted sum of desirable outputs to the weighted sum of inputs and undesirable outputs

$$\max h_B = \frac{\sum_{r=1}^{k}\mu_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0} + \sum_{s=k+1}^{p}\mu_s y_{s0}} \quad s.t. \quad \frac{\sum_{r=1}^{k}\mu_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij} + \sum_{s=k+1}^{p}\mu_s y_{sj}} \le 1, \quad \mu_r, v_i \ge \varepsilon \qquad (2)$$

$j = 1,\dots,n, \quad r = 1,\dots,p, \quad i = 1,\dots,m, \quad \varepsilon > 0 \quad (\text{"}non-Archimedean\text{"})$

Model C maximises the ratio of the difference between a weighted sum of k desirable outputs and the weighted sum of m inputs to the weighted sum of (p-k) undesirable outputs

$$\max h_C = \frac{\sum_{r=1}^{k}\mu_r y_{r0} - \sum_{i=1}^{m} v_i x_{i0}}{\sum_{s=k+1}^{p}\mu_s y_{s0}} \quad s.t. \quad \frac{\sum_{r=1}^{k}\mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij}}{\sum_{s=k+1}^{p}\mu_s y_{sj}} \le 1, \quad \mu_r, v_i \ge \varepsilon \qquad (3)$$

$j = 1,\dots,n, \quad r = 1,\dots,p, \quad i = 1,\dots,m, \quad \varepsilon > 0 \quad (\text{"}non-Archimedean\text{"})$

Model D is the reciprocal of model B and minimises the ratio of the sum of weighted undesirable outputs and inputs to weighted desirable outputs.
The eco-efficiency indicator of model B cannot be smaller than the one of model A because both inputs and undesirable outputs have to be decreased to increase eco-efficiency. In model C only undesirable outputs have to be decreased, so the indicator of C cannot be greater than that of B.

Applications of such models to firms of the same industry will yield deeper insights into the causes of eco-inefficiency and can indicate potential improvements in inputs and outputs. They are also very useful to identify achievements of emission reduction programs when one compares eco-efficiency indicators before and after the implementation of such programs. A particularly interesting application of this technique can also help to solve allocation problems in the context of preparations for the trading of emission certificates. An application to the French cement industry can be found in Patnaik (2004). He presents alternative calculations of proposed initial allocations of emission rights (allowances), all based on the solution of a variable returns to scale DEA problem. It turns out that it is not so easy to determine how eco-inefficiency should be reflected in the allocation of allowances. The reason why an efficiency approach is desirable at all is given by the fact that a traditional allocation procedure based on historical emission values (so called "grandfathering") would bias in favour of those firms that have contributed to the highest levels of emissions instead of those that have contributed to emission reductions. The DEA results of eco-efficiency can therefore be used to distribute allowances according to the eco-efficiency of a plant. As the various models discussed above all show slight differences in the inefficiency indicators – the efficient units are always the same – it may be left to the policy maker to select the appropriate one. Information requirements are higher when additionally to outputs and emissions also input factors are considered. A DEA based eco-efficiency calculation will have a motivating effect on competing companies because information about the direction of efficiency improvement is provided as well.

4. Macro Eco-efficiency

Let us now turn to the whole economy and the problem to characterise its eco-efficiency. Due to "rebound effects" firm's eco-efficiency not necessarily implies macro-eco-efficiency. Growth effects may more than compensate efficiency gains. The ecological balance may deteriorate due to a larger increase in sales of a product that itself is resource efficient. One of the most useful tools for analysing production relations within an economy considering the interrelatedness of products and sectors is the so-called input-output approach. An input-output table relates all goods and services produced and used in a country to all economic branches of the economy. Because of its detailed structure of the inter-industry relations such a model reveals the role of key sectors or industries, the supplier links and the delivery chains throughout the economy. With the beginning concern about the limits to growth around the seventies of last century and the public awareness of existing environmental problems the input-output approach has been amended to take into account also environmental dimensions. We shall use this framework to obtain measures of eco-efficiency, degrees of achievements of both, economic and environmental, objectives for the whole economy.

Let the production possibility frontier of the economy be determined by the input-output model, primary inputs, pollution generation and abatement, and final demand. The degree by which a net-output vector, for given primary inputs and environmental standards, could be extended, can be considered a measure of eco-inefficiency. Equivalently this could also be achieved by a reduction of primary inputs for given environmental standards and given final demand. We have to take into account that desirable outputs are strongly disposable while undesirable ones are only weakly disposable meaning that their reduction can only be achieved by a reduction of desirable outputs or an increase of primary inputs.

The augmented Leontief model known from the literature (e.g. Luptacik and Böhm(1994)) can be written in partitioned form

$$\begin{aligned} &(I - A_{11})x_1 - A_{12}x_2 \geq y_1 \\ &- A_{21}x_1 + (I - A_{22})x_2 \geq -y_2 \end{aligned} \qquad (4)$$

with A_{11} the k x k input-coefficient matrix of the economic subsystem, A_{12} the k x (p-k) input-coefficient matrix of the abatement system, A_{21} the (p-k) x k emission-coefficient matrix of the economic system, and A_{22} the (p-k x p-k) emission-coefficient matrix of the abatement system. x_1 and x_2 are the (k x 1) gross production vector and the (p-k x 1) abatement vector respectively, y_1 is the k x 1) final demand vector and y_2 the p-k x 1 vector of tolerated pollutants (or pollution standards) vector.

Applying the concept of the distance function to this partitioned system by treating the undesirable outputs similar to inputs, a proportional measure of eco-efficiency can be derived from the following model formulations where the m x 1 vector of primary inputs is given by z.

1. Minimise the use of primary factors for a given level of final demand and tolerated pollution:

$$\begin{aligned} &\min_x \gamma \quad s.t. \\ &(I - A_{11})x_1 - A_{12}x_2 \geq y_1 \\ &A_{21}x_1 - (I - A_{22})x_2 \leq y_2 \\ &B_1x_1 + B_2x_2 - \gamma z \leq 0 \\ &x_1, x_2, \gamma \geq 0 \end{aligned} \qquad (5)$$

It is easily seen that this model corresponds to model A of Korhonen and Luptáčik (2004). This input distance function considers the minimal proportional contraction of the input vector, given the output vector and tolerated pollution.

2. Maximise proportional expansion of final demand for given levels of tolerated pollution and primary factors:

$$\begin{aligned} &\max_x \alpha \quad s.t. \\ &(I - A_{11})x_1 - A_{12}x_2 - \alpha y_1 \geq 0 \\ &A_{21}x_1 - (I - A_{22})x_2 \leq y_2 \\ &B_1x_1 + B_2x_2 \leq z \\ &x_1, x_2, \alpha \geq 0 \end{aligned} \qquad (6)$$

This is the output oriented model D. We note that due to the presence of the pollution subsystem representing undesirable outputs, the optimal values of α and γ are not the reciprocal of each other. However, by treating these undesirable outputs like inputs in the model, i.e. by changing the problem formulation into a proportional reduction of primary inputs *and* undesirable outputs for given final demand, the reciprocal property of the distance function can be established. This happens in the following version which corresponds to model B:

3. Minimise the use of primary factors *and* tolerated pollution for a given level of final demand:

$$\begin{aligned} &\min_x \gamma \quad s.t. \\ &(I - A_{11})x_1 - A_{12}x_2 \geq y_1 \\ &A_{21}x_1 - (I - A_{22})x_2 - \gamma y_2 \leq 0 \\ &B_1x_1 + B_2x_2 - \gamma z \leq 0 \\ &x_1, x_2, \gamma \geq 0 \end{aligned} \qquad (7)$$

Still another version corresponding to model C can be formulated by minimising tolerated pollution for a given level of final demand and primary factors

We note that efficiency indicators derived from the distance function approach are based on optimisation without the possibility to alter the proportions among net outputs or primary inputs. If its slack variable is positive it is still possible to reduce a specific input without reducing any of the net-outputs. By defining a new slack based measure as in Luptáčik and Böhm (2005) we can go beyond the proportional approaches and take into account of changes in the structure of final demand, pollution sources and the composition of primary inputs. The problem is formulated as a goal programming model treating the (p-k) undesirable outputs like inputs:

$$\min_{x} \left\{ \rho = \frac{1 - \dfrac{1}{(p-k)+m}\left(\sum_{j=1}^{m}\dfrac{s_{3j}}{z_j} + \sum_{j=k+1}^{p}\dfrac{s_{2j}}{y_{2j}}\right)}{1 + \dfrac{1}{k}\sum_{i=1}^{k}\dfrac{s_{1i}}{y_{1i}}} \right\} \quad \textit{subject to}$$

$$\begin{aligned} &(I - A_{11})x_1 - A_{12}x_2 - s_1 = y_1 \\ &A_{21}x_1 - (I - A_{22})x_2 + s_2 = y_2 \\ &B_1x_1 + B_2x_2 + s_3 = z \\ &x_1, x_2, s_1, s_2, s_3 \geq 0 \end{aligned} \qquad (8)$$

where we use k desirable outputs, (p-k) undesirable ones, and m inputs. Values on the right hand side are given. This fractional program can be linearised by a simple transformation.

The macro-eco-efficiency models (5), (6), and (7) can be related to the micro-eco-efficiency concept by reformulating them as DEA models under a particular assumption: Construct a fully eco-efficient economy that can serve as a standard against which the actual economy is evaluated. While the microeconomic application of DEA uses inputs and outputs of different *independent* decision making units, the macroeconomic I-O model uses data of usually only one economy composed of *interrelated* sectors. To have these sectors performing the role of independent firms one needs to explore the potential output of each sector given all other outputs and inputs. This generates as many efficient artificial economies as there are outputs and inputs. These will establish the production possibility set (or the input requirement set). Therefore, a multi-objective optimisation problem is formulated in which each output is separately maximised subject to constraints on the production of other outputs and required inputs, and each input is minimised under the same constraints. Denoting by s_1 the vector of k slack variables of the k sectors, by s_2 the slacks in the (p-k) undesirable output relations and by s_3 the slacks in the m input relations, the following model is solved p+m times for given values of outputs and inputs to obtain the maximal values of each slack variable:

$$\begin{aligned} &\max_{x} s_j \quad s.t. \\ &(I - A_{11})x_1 - A_{12}x_2 - s_1 = y_1 \\ &A_{21}x_1 - (I - A_{22})x_2 + s_2 = y_2 \\ &B_1x_1 + B_2x_2 + s_3 = z \\ &x_1, x_2, s_1, s_2, s_3 \geq 0 \\ &j = 1,\ldots,k,k+1,\ldots,p,p+1,\ldots,p+m \end{aligned} \qquad (9)$$

Individually optimal output and input values are calculated from $y^*_1 = y_1 + s_1$, $y^*_2 = y_2 - s_2$, and $z^*_3 = z_3 - s_3$, and are arranged to form a payoff matrix P.

$$P = \begin{bmatrix} y_1 + s_1^1 & y_1 + s_1^2 & \cdots & y_1 + s_1^{p+m} \\ y_2 - s_2^1 & y_2 - s_2^2 & \cdots & y_2 - s_2^{p+m} \\ z - s_3^1 & z - s_3^2 & \cdots & z - s_3^{p+m} \end{bmatrix}$$

This matrix is used to establish the frontier of the production possibility set (or the input requirement set) and thus, the efficient envelope.

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} y_1^{*(1)} & y_1^{*(2)} & \cdots & y_1^{*(p)} \\ y_2^{*(1)} & y_2^{*(2)} & \cdots & y_2^{*(p)} \\ z^{*(1)} & z^{*(2)} & \cdots & z^{*(p)} \end{bmatrix}$$

This efficient envelope is used to evaluate the relative inefficiency of the economy given by the actual output and input data (y_1^0, y_2^0, z^0) in the following input oriented DEA problem:

$$\begin{aligned} &\min_{\mu} \theta \quad s.t. \\ &P_1\mu \geq y_1^0 \\ &P_2\mu - \theta y_2^0 \leq 0 \\ &P_3\mu - \theta z^0 \leq 0 \\ &\mu \geq 0, \theta \geq 0 \end{aligned} \qquad (10)$$

The efficiency score, θ, gives the proportions of all input of the economy which must be sufficient – compared to the production frontier – to achieve the given output levels. (1-θ) describes the necessary reduction of all inputs to achieve the efficiency frontier. Vector μ provides the weighting pattern in the projection point of the efficient surface derived from a radial input contraction.

It has been shown in Luptáčik and Böhm (2005) that the minimum value of θ is the same as the minimal value of γ in the LP model version B of problem (7).

References

Brunn H. and Rentz O. (1998), LCA as a decision-support tool for product optimization, ch.17 in J.E.M. Klostermann and A. Tukker (eds.), *Product Innovation and Eco-efficiency*, Kluwer, 1998

Charnes A., Cooper W.W., Levin A., Seiford L. (eds.), (1994), *Data Envelopment Analysis: Theory, Methodology, and Applications*, Kluwer Academic Publishers, Boston.

Färe R., Grosskopf S., Lovell K., Pasurka C. (1989), Multilateral productivity comparisons when some outputs are undesirable: A nonparametric approach, *The Review of Economics and Statistics* 71 (1)

Korhonen P., M. Luptáčik (2004), "Eco-Efficiency Analysis of Power Plants: An extension of data envelopment analysis", *European Journal of Operational Research* 154 (2004), 437-446

Luptáčik M., Böhm B. (1994), "Reconsideration of Non-negative solutions for the Augmented Leontief Model", *Economic Systems Research* 6, No.2, 167-170

Luptáčik M., Böhm B. (2005), The Analysis of Eco-Efficiency in an Input-Output Frame-work, 9th European Workshop on Productivity and Efficiency Analysis, Brussels, 29. June – 2. July 2005, Belgium

Patnaik S. (2004), *Emissions Allowance Trading; with a special investigation of initial allocation mechanisms*. Master Thesis, University of Technology, Vienna

Tyteca D. (1996), On the measurement of the environmental performance of firms – a literature review and a productive efficiency perspective. *Journal of Environmental Management* 46, 281-308

WBCSD (2000), Measuring eco-efficiency – a guide to reporting company performance, June

WBCSD (1995), Achieving Eco-Efficiency in Business. Report prepared for the Second Antwerp Efficiency Workshop, March 14-15,

ON GRADIENT FLOWS IN GLOBAL OPTIMIZATION

Hubertus Th. Jongen

RWTH Aachen University (D) and Maastricht University (NL)
Department of Mathematics, D-52056 Aachen, Germany
jongen@rwth-aachen.de

Keywords: nonconvex optimization, global optimization, transition points of first order, gradient flows, min-max digraph, adapted metric, global interior point approach

Extended Abstact. We consider smooth finite dimensional optimization problems with a compact, connected feasible set *M* and objective function *f*. The basic problem, on which we focus, is: how to get from one local minimum to all others. Transition points of first order (i.e., Karush-Kuhn-Tucker points of index 1) play a crucial role in this context.

Firstly, we consider the unconstrained case with moderate asymptotic behaviour of the objective function at infinity. It is shown that (generically) any two local minima can be connected via an alternating sequence of local minima and transition points of first order. In particular, the graph with local minima as its nodes and first order transition points representing the edges turns out to be connected. On the other hand, *any* connected (finite) graph can be realized in the above sense by means of a smooth function of three variables having a minimal number of stationary points.

Secondly, we introduce a bipartite graph Γ as follows. Its nodes are formed by the set of local minima and maxima of the objecive function f, respectively. Given a smooth Riemannian (i.e. variable) metric, there is an arc from a local minimum x to a local maximum y if the *ascent* (semi-) flow induced by the projected gradients of f connects points from a neighborhood of x with points from a neighborhood of y. The existence of an arc from y to x is defined with the aid of the *descent* (semi-) flow. Strong connectedness of the digraph Γ ensures that, starting from one local minimum, we may reach any other one using ascent and descent trajectories in an alternating way. In case that *no inequality* constraints are present or active, it is well known that for a generic Riemannian metric the resulting min-max digraph Γ is indeed strongly connected.

However, if *inequality* constraints are active, then there might appear obstructions. The latter phenomenon is due to *active set stategy* In particular, we show that Γ may contain *absorbing* two-cycles. If one enters such a cycle, one cannot leave it anymore via ascent and descent trajectories. Moreover, the appearance of such cycles is stable with respect to small perturbations of the Riemannian metric.

By means of a global adaptation of the metric involved, the appearance of absorbing cycles may be prevented. This adaptation makes the metric *singular* at the boundary of the feasible set and it is automatically performed by means of the constraints. In particular, the *interior* of the feasible set is now *invariant* under the corresponding ascent (descent) flow. The resulting bipartite graph Γ becomes (generically) connected. Finally, the underlying ideas can be interpreted as a global *interior point* approach for *nonconvex* constrained optimization.

References

(1) H.Th.Jongen and A.Ruiz Jhones: *Nonlinear Optimization: On the min-max digraph and global smoothing*. In: Calculus of Variations and Differential Equations (Eds.: A.Ioffe, S.Reich, I.Shafrir), Chapman and Hall/ CRC Research Notes in Mathematics Series, Vol. 410, CRC Press, Boca Raton, Fl, pp. 119-135 (1999).

(2) H.Guenzel and H.Th.Jongen : *Global Optimization: On Pathlengths in Min-Max Graphs*. Journal of Global Optimization, Vol.17, pp.161-165 (2000).

(3) H.Th.Jongen, P.Jonker and F.Twilt: *Nonlinear Optimization in Finite Dimensions*. Kluwer Academic Publishers (2000).

(4) H.Guenzel and H.Th.Jongen: *On absorbing cycles in min-max digraphs*. Journal of Global Optimization, Vol. 31, pp. 85–92 (2005).

(5) H.Th.Jongen and O.Stein: *On the complexity of equalizing inequalities*.Journal of Global Optimization, Vol.27, pp.367-374 (2003) .

(6) H.Th.Jongen and O.Stein: *Constrained Global Optimization: Adaptive Gradient Flows*. In: Frontiers in Global Optimization (Eds.: C.A.Floudas, P.Pardalos), Kluwer Academic Publishers, Boston, pp. 223-236 (2004).

(7) C.A.Floudas, H.Th.Jongen: *Global Optimization: Local Minima and Transition Points*. Journal of Global Optimization (to appear).

SOME RESULTS DEALING WITH THE ALGEBRAIC APPROACH TO PATH PROBLEMS IN GRAPHS

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

Abstract. A wide variety of path problems in graphs can generally be formulated and solved by algebraic means. For this purpose, an abstract algebraic structure is introduced whose instances are called semirings. Each particular type of path problem is characterized by a different instance of the structure. This paper presents two recent results dealing with the algebraic approach to path problems based on semirings. The first part of the paper introduces a method for combining already known semirings into new ones. The obtained composite semirings correspond to relatively complex path problems involving explicit identification of optimal paths or multi-criteria optimization. The second part of the paper describes a distributed algorithm for solving path problems, which has been implemented on a network of computers. Thanks to its specification in terms of an arbitrary semiring, the algorithm can be applied to different types of simple or complex path problems.

Keywords: directed graphs, path problems, algebraic approach, semirings, optimization, distributed algorithms.

1. Introduction

Path problems are a family of optimization and enumeration problems, which reduce to generation or comparison of paths in directed or undirected graphs. Some examples are: checking path existence, finding shortest or most reliable paths, finding paths of maximum capacity, listing all paths, etc.

Each particular type of path problem can be treated separately, and solved by dedicated algorithms [3]. However, a more efficient approach is to establish a general framework for the whole family of problems, and to use general algorithms. The latter can be achieved by introducing a suitable abstract algebraic structure.

Few variants of the algebraic approach for solving path problems have been proposed [1,4,8,9]. Our favorite variant from [1] uses a structure whose instances are called semirings. The approach from [1] relies heavily on matrices and on analogies with ordinary linear algebra. Each type of path problem is formulated by using a different semiring. Solving a concrete problem reduces to computing with matrices over the corresponding semiring.

The aim of this paper is to present two recent results dealing with the algebraic approach to path problems. The first result is a method for building more complex semirings from simpler semirings. The obtained composite semirings can be applied to solve relatively complex but still meaningful path problems. The second result presented in the paper is a specification of a distributed algorithm for solving path problems. Since it works over an arbitrary semiring, the algorithm can be applied to solve different types of simple or complex path problems.

The paper is organized as follows. Section 2 gives preliminaries about semirings and graphs. Section 3 shows by examples how our adopted algebraic approach based on semirings can be used to solve simple path problems. Section 4 introduces composite semirings. Section 5 illustrates again by examples how composite semirings can be applied to solve more complicated tasks, such as optimization with explicit identification of optimal paths or multi-criteria optimization. Section 6 describes our distributed algorithm for solving path problems. Section 7 reports on an actual network implementation of the algorithm. The final Section 8 gives a conclusion.

2. Semirings and graphs

We start with the definition of our algebraic structure. A *semiring* is a set P equipped with two binary operations, $\vee$ (join) and $\circ$ (multiplication), which have the following properties: $\vee$ is idempotent, commutative and associative; $\circ$ is associative, left-distributive and right-distributive over $\vee$; there exist a zero element ϕ and a unit element ϵ, such that (for any x) $\phi \vee x = x$, $\phi \circ x = \phi = x \circ \phi$, $\epsilon \circ x = x = x \circ \epsilon$. When evaluating algebraic expressions over P, we will always assume that $\circ$ takes precedence over $\vee$, unless otherwise regulated by parentheses. Concrete instances of our algebraic structure can be found in [1,6,8,9], and some of them are also repeated in Table 1.

notation	P	$x \vee y$	$x \circ y$	ϕ	ϵ
P_1	$\{0, 1\}$	$\max\{x, y\}$	$\min\{x, y\}$	0	1
P_2	$\mathbf{R} \cup \{\infty\}$	$\min\{x, y\}$	$x + y$	∞	0
P_3	$\mathbf{R} \cup \{-\infty\}$	$\max\{x, y\}$	$x + y$	$-\infty$	0
P_4	$\{r \in \mathbf{R} \mid 0 \leq r \leq 1\}$	$\max\{x, y\}$	$x \cdot y$	0	1
P_5	$\{r \in \mathbf{R} \mid r \geq 0\} \cup \{\infty\}$	$\max\{x, y\}$	$\min\{x, y\}$	0	∞
P_6	$\mathcal{P}(\Sigma^*)$	$x \cup y$	$\{w_x * w_y \mid w_x \in x, w_y \in y\}$	$\emptyset$	$\{\lambda\}$
P_8	$\mathcal{B}(\Sigma^*)$	$\mathrm{bas}(x \cup y)$	$\{w_x * w_y \mid w_x \in x, w_y \in y\}$	$\emptyset$	$\{\lambda\}$

Table 1: Well known semirings.

The first five examples in Table 1 are "extremal" semirings dealing with real numbers, infinity symbols and conventional arithmetic operations. The semirings P_6 and P_8 use "linguistic" concepts, namely Σ denotes a finite alphabet, and Σ^* is the set of all words (finite sequences of letters) over Σ. Consequently, $\mathcal{P}(\Sigma^*)$ is the set of all languages (sets of words) over Σ. The operation $\vee$ is based on the set union $\cup$, and $\circ$ on word concatenation $*$. The symbol λ stands for the empty word, and $\emptyset$ is the empty language. The semiring P_8 deals with basic languages. We define an abbreviation of a word w as any word that can be obtained from w by removing at least one of its letters. For any language $L \subset \Sigma^*$, bas(L) is the basis of L, i.e. the language consisting of all words from L that do not have abbreviations in L. If bas(L) $= L$, then L is a basic language. $\mathcal{B}(\Sigma^*)$ denotes the set of all basic languages over Σ.

Let X be a square matrix whose entries belong to a semiring P. Then we consider the following two expressions:

$$X^* = E \vee X \vee X^2 \vee X^3 \vee \ldots, \qquad (1)$$

$$\widehat{X} = X \vee X^2 \vee X^3 \vee \ldots. \qquad (2)$$

Here, E denotes the unit matrix (ϵ on the diagonal, ϕ elsewhere). X^2 stands for $X \circ X$, X^3 for $X \circ X \circ X$, etc. The matrix operations $\vee$ and $\circ$ are derived from the corresponding scalar operations similarly as in ordinary linear algebra, provided that $\vee$ is analogous to the conventional addition and $\circ$ to the conventional multiplication.

Te expressions (1) and (2) are called the *strong* and *weak closure* of X, respectively. In most situations the involved matrix X is *stable*, meaning that (1) and (2) become saturated after adding enough powers of X. Thus the closures X^* and $\widehat{X}$ can usually be computed in a finite number of algebraic operations. Note that X^* and $\widehat{X}$ are equivalent in the sense that any of them can easily be obtained from the other one. Indeed, $X^* = E \vee \widehat{X}$, while $\widehat{X} = X \circ X^*$. Thus any procedure for computing the strong closure can be modified to compute the weak closure, and vice versa.

In this paper we consider directed graphs whose arcs are labeled with non-zero elements from a semiring P. An n-node labeled graph G is fully described by its $n \times n$ *adjacency matrix* A over P. The (i, j)-th entry of A is the label of the arc from node i to node j if such arc exists, or ϕ otherwise.

In path problems, we explore paths in graphs, i.e. sequences of consecutive arcs. A circular path is called a *cycle*. A path is *elementary* if it does not traverse any node more than once. For convenience, it is supposed that each node in a graph is connected to itself by a *null path* containing no arcs. A null path is considered to be elementary.

3. Solving simple path problems

Now we are ready to explain our algebraic approach to path problems. For a certain problem posed in a graph G, we choose a suitable semiring P and assign appropriate arc labels so that the adjacency matrix A of G is stable over P. We compute the strong or weak closure A^* or $\widehat{A}$, respectively, and read from it the solution to the original problem. The difference between the two types of closures is that the first one takes into account null paths, while the second one restricts to non-null paths. Note that each type of problem requires a different semiring, although the overall problem structure remains the same. Intuitively, P contains values that are used to describe arc properties, $\circ$ defines how the values assigned to arcs along a path are combined to describe that path, while $\vee$ specifies how the values of paths connecting the same pair of nodes are finally combined together.

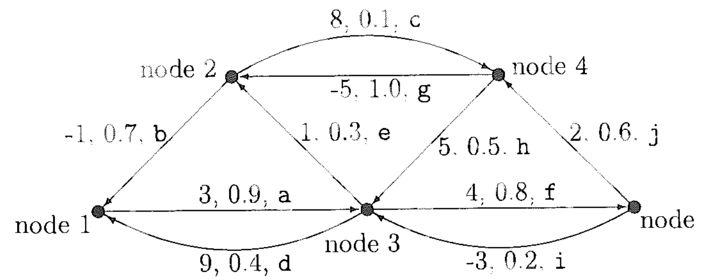


Figure 1: A graph G with given arc lengths, arc reliabilities and arc identifiers.

To illustrate our algebraic approach, let us consider the graph G in Figure 1, whose arcs are assigned triplets of values. For each arc, the first assigned value (an integer) is interpreted as its length, the second value (a real number) as its reliability, and the third value (a letter from an alphabet $\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \ldots\}$) as an identifier.

Suppose that we want to solve the maximum reliability problem in G, i.e. we want to determine the reliability of a path with maximum reliability between any pair of nodes. Then we should use arc reliabilities as arc labels and treat G as labeled with the semiring P_4. The corresponding adjacency matrix A and its strong closure A^* are:

$$A = \begin{bmatrix} 0 & 0 & 0.9 & 0 & 0 \\ 0.7 & 0 & 0 & 0.1 & 0 \\ 0.4 & 0.3 & 0 & 0 & 0.8 \\ 0 & 1.0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.6 & 0 \end{bmatrix}, \quad A^* = \begin{bmatrix} 1.0 & 0.432 & 0.9 & 0.432 & 0.72 \\ 0.7 & 1.0 & 0.63 & 0.3024 & 0.504 \\ 0.4 & 0.48 & 1.0 & 0.48 & 0.8 \\ 0.7 & 1.0 & 0.63 & 1.0 & 0.504 \\ 0.42 & 0.6 & 0.378 & 0.6 & 1.0 \end{bmatrix}.$$

The (i,j)-th entry of A^* is equal to the maximum reliability of a path from node i to node j. Note that analogously we could also solve the shortest distance problem in G. Then instead of arc reliabilities and the semiring P_4 we should use arc lengths and the semiring P_2.

Next, let us suppose that we want to identify all elementary paths in G. Then we can interpret arc identifiers as one-letter words over the alphabet Σ, and one-letter words as one-word languages. In this way, we can consider G as labeled with the semiring P_8. The adjacency matrix A and its weak closure $\hat{A}$ are then the following:

$$A = \begin{bmatrix} \emptyset & \emptyset & \{\mathtt{a}\} & \emptyset & \emptyset \\ \{\mathtt{b}\} & \emptyset & \emptyset & \{\mathtt{c}\} & \emptyset \\ \{\mathtt{d}\} & \{\mathtt{e}\} & \emptyset & \emptyset & \{\mathtt{f}\} \\ \emptyset & \{\mathtt{g}\} & \{\mathtt{h}\} & \emptyset & \emptyset \\ \emptyset & \emptyset & \{\mathtt{i}\} & \{\mathtt{j}\} & \emptyset \end{bmatrix}.$$

$$\hat{A} = \begin{bmatrix} \left\{\begin{matrix}\mathtt{ad},\\ \mathtt{aeb},\\ \mathtt{afjgb}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{ae},\\ \mathtt{afjg}\end{matrix}\right\} & \{\mathtt{a}\} & \left\{\begin{matrix}\mathtt{aec},\\ \mathtt{afj}\end{matrix}\right\} & \{\mathtt{af}\} \\ \left\{\begin{matrix}\mathtt{b},\\ \mathtt{chd}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{bae},\\ \mathtt{bafjg},\\ \mathtt{cg},\\ \mathtt{che}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{ba},\\ \mathtt{ch}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{bafj},\\ \mathtt{c}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{baf},\\ \mathtt{chf}\end{matrix}\right\} \\ \left\{\begin{matrix}\mathtt{d},\\ \mathtt{eb},\\ \mathtt{fjgb}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{e},\\ \mathtt{fjg}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{da},\\ \mathtt{eba},\\ \mathtt{ech},\\ \mathtt{fjgba},\\ \mathtt{fi},\\ \mathtt{fjh}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{ec},\\ \mathtt{fj}\end{matrix}\right\} & \{\mathtt{f}\} \\ \left\{\begin{matrix}\mathtt{gb},\\ \mathtt{hd},\\ \mathtt{heb}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{g},\\ \mathtt{he}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{gba},\\ \mathtt{h}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{gc},\\ \mathtt{gbafj},\\ \mathtt{hec},\\ \mathtt{hfj}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{gbaf},\\ \mathtt{hf}\end{matrix}\right\} \\ \left\{\begin{matrix}\mathtt{id},\\ \mathtt{ieb},\\ \mathtt{jgb},\\ \mathtt{jhd},\\ \mathtt{jheb}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{ie},\\ \mathtt{jg},\\ \mathtt{jhe}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{i},\\ \mathtt{jgba},\\ \mathtt{jh}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{iec},\\ \mathtt{j}\end{matrix}\right\} & \left\{\begin{matrix}\mathtt{if},\\ \mathtt{jgbaf},\\ \mathtt{jhf}\end{matrix}\right\} \end{bmatrix}.$$

As we can see, the (i,j)-th entry of the weak closure $\hat{A}$ over P_8 lists all non-null elementary paths from node i to node j.

Correctness of the above examples is guaranteed by some theorems from [1]. More precisely, those theorems assure that the involved adjacency matrices are stable, so that their closures can be computed, thus giving the desired solutions. Stability in the case of shortest distances relies on the fact that each elementary cycle in G has nonnegative length.

4. Composite semirings

The examples in Section 3 have shown strengths and weaknesses of the classical semirings from Table 1. By computing in extremal semirings such as P_2, P_3, P_4 or P_5 it is possible to find optimal values in a graph, but one cannot directly identify paths where those optimal values are achieved. On the other hand, by computing in linguistic semirings such as P_6 or P_8 one can identify all paths, but those paths are not necessarily optimal. In order to get more specific values, the semirings from Table 1 should be combined. Now we will show how this can be done.

Let $\bar{P}$ be a semiring whose binary operations are $\bar{\vee}$ and $\bar{\circ}$. Let the zero element of $\bar{P}$ be $\bar{\phi}$ and the unit element $\bar{\epsilon}$. Suppose that

- the operation $\bar{\vee}$ is a "choice operation", i.e. for all $\bar{x}, \bar{y} \in \bar{P}$:

$$\bar{x} \mathbin{\bar{\vee}} \bar{y} = \bar{x} \quad \text{or} \quad \bar{x} \mathbin{\bar{\vee}} \bar{y} = \bar{y};$$

- the operation $\bar{\circ}$ has the "cancellation property", i.e. for all $\bar{x}, \bar{y}, \bar{z} \in \bar{P}$:

$$(\bar{x} \mathbin{\bar{\circ}} \bar{z} = \bar{y} \mathbin{\bar{\circ}} \bar{z} \ \text{ or } \ \bar{z} \mathbin{\bar{\circ}} \bar{x} = \bar{z} \mathbin{\bar{\circ}} \bar{y}) \Rightarrow (\bar{x} = \bar{y} \ \text{ or } \ \bar{z} = \bar{\phi}).$$

Let $\tilde{P}$ be any other semiring whose binary operations, zero and unit element are denoted with $\tilde{\vee}$, $\tilde{\circ}$, $\tilde{\phi}$, $\tilde{\epsilon}$, respectively. Then we can construct a set $P = \bar{P} \otimes \tilde{P}$ and two binary operations, $\vee$ and $\circ$, in the following way.

- $\bar{P} \otimes \tilde{P} = \left\{(\bar{x}, \tilde{x}) \mid \bar{x} \in \bar{P} \setminus \{\bar{\phi}\}, \tilde{x} \in \tilde{P}\right\} \cup \left\{(\bar{\phi}, \tilde{\varphi})\right\}$.
- For all $(\bar{x}, \tilde{x}), (\bar{y}, \tilde{y}) \in \bar{P} \otimes \tilde{P}$:

$$(\bar{x}, \tilde{x}) \vee (\bar{y}, \tilde{y}) = \begin{cases} (\bar{x}, \tilde{x}) & \text{if } \bar{x} \mathbin{\bar{\vee}} \bar{y} = \bar{x} \neq \bar{y} \\ (\bar{y}, \tilde{y}) & \text{if } \bar{x} \mathbin{\bar{\vee}} \bar{y} = \bar{y} \neq \bar{x} \\ (\bar{x} \mathbin{\bar{\vee}} \bar{y}, \tilde{x} \mathbin{\tilde{\vee}} \tilde{y}) & \text{if } \bar{x} = \bar{y} \end{cases}.$$

$$(\bar{x}, \tilde{x}) \circ (\bar{y}, \tilde{y}) = (\bar{x} \mathbin{\bar{\circ}} \bar{y}, \tilde{x} \mathbin{\tilde{\circ}} \tilde{y}).$$

It has been proved in [7] that the proposed $\vee$ and $\circ$ are correctly defined binary operations on the set $\bar{P} \otimes \tilde{P}$, and that $\bar{P} \otimes \tilde{P}$ with these operations really constitutes a semiring whose zero element is $\phi = (\bar{\phi}, \tilde{\phi})$ and unit element is $\epsilon = (\bar{\epsilon}, \tilde{\epsilon})$. The newly constructed $P = \bar{P} \otimes \tilde{P}$ is called a *composite* semiring. The paper [7] also gives some sufficient conditions, which can easily be checked in practice, and which can assure that an adjacency matrix over $\bar{P} \otimes \tilde{P}$ is stable. The examples in the next section will illustrate how composite semirings can eliminate the previously mentioned drawbacks of simple semirings.

5. Solving complex path problems

Composite semirings can be applied for explicit identification of optimal paths in graphs. Another application is multi-criteria optimization. Composition of semirings can be iterated - in this way one can accomplish multi-criteria optimization with explicit identification of paths that are optimal according to several criteria.

Some concrete applications are listed in Table 2. All proposed compositions are based on simple semirings from Table 1. There are of course many other possibilities for combining simple semirings which are not listed due to space limitations. Table 2 describes how a graph should be labeled with a particular semiring, and what results can be obtained by computing the strong or weak closure of the corresponding adjacency matrix in that semiring. If the given assumptions about the graph are satisfied, then feasibility and correctness of the involved computations can be guaranteed by [7].

Composite semiring P	Assumptions about the graph G	Components of the (i,j)-th entry of the adjacency matrix A	Components of the (i,j)-th entry of the closure matrix A^* or $\hat{A}$
$P_2 \otimes P_6$	any elementary cycle has positive length	- length of arc (i,j) - identifier of arc (i,j)	- shortest distance from node i to node j - identifiers of all shortest paths from node i to node j
$P_3 \otimes P_6$	there are no cycles	- length of arc (i,j) - identifier of arc (i,j)	- longest distance from node i to node j - identifiers of all longest (critical) paths from node i to node j
$P_4 \otimes P_6$	any elementary cycle contains an arc with reliability < 1	- reliability of arc (i,j) - identifier of arc (i,j)	- maximum reliability of a path from node i to node j - identifiers of all most reliable paths from node i to node j
$P_2 \otimes (P_4 \otimes P_8)$	any elementary cycle has non-negative length	- length of arc (i,j) - reliability of arc (i,j) - identifier of arc (i,j)	- shortest distance from node i to node j - maximum reliability of a shortest path from node i to node j - identifiers of all shortest paths from node i to node j achieving maximum reliability

Table 2: Some applications of composite semirings.

To illustrate the applications from Table 2 in more detail, let us consider again the graph G in Figure 1 whose arcs are assigned lengths, reliabilities and identifiers. Take first into account only arc lengths and arc identifiers, and treat G as labeled with $P_2 \otimes P_6$. Then the adjacency matrix A and its strong closure A^* are the following:

$$A = \begin{bmatrix} (\infty,\emptyset) & (\infty,\emptyset) & (3,\{\mathtt{a}\}) & (\infty,\emptyset) & (\infty,\emptyset) \\ (-1,\{\mathtt{b}\}) & (\infty,\emptyset) & (\infty,\emptyset) & (8,\{\mathtt{c}\}) & (\infty,\emptyset) \\ (9,\{\mathtt{d}\}) & (1,\{\mathtt{e}\}) & (\infty,\emptyset) & (\infty,\emptyset) & (4,\{\mathtt{f}\}) \\ (\infty,\emptyset) & (-5,\{\mathtt{g}\}) & (5,\{\mathtt{h}\}) & (\infty,\emptyset) & (\infty,\emptyset) \\ (\infty,\emptyset) & (\infty,\emptyset) & (-3,\{\mathtt{i}\}) & (2,\{\mathtt{j}\}) & (\infty,\emptyset) \end{bmatrix},$$

$$A^* = \begin{bmatrix} (0,\{\lambda\}) & \left(4,\begin{Bmatrix}\mathtt{ae}, \\ \mathtt{afjg}\end{Bmatrix}\right) & \left(3,\{\mathtt{a}\}\right) & \left(9,\{\mathtt{afj}\}\right) & \left(7,\{\mathtt{af}\}\right) \\ \left(-1,\{\mathtt{b}\}\right) & (0,\{\lambda\}) & \left(2,\{\mathtt{ba}\}\right) & \left(8,\begin{Bmatrix}\mathtt{bafj}, \\ \mathtt{c}\end{Bmatrix}\right) & \left(6,\{\mathtt{baf}\}\right) \\ \left(0,\begin{Bmatrix}\mathtt{eb}, \\ \mathtt{fjgb}\end{Bmatrix}\right) & \left(1,\begin{Bmatrix}\mathtt{e}, \\ \mathtt{fjg}\end{Bmatrix}\right) & (0,\{\lambda\}) & \left(6,\{\mathtt{fj}\}\right) & \left(4,\{\mathtt{f}\}\right) \\ \left(-6,\{\mathtt{gb}\}\right) & \left(-5,\{\mathtt{g}\}\right) & \left(-3,\{\mathtt{gba}\}\right) & (0,\{\lambda\}) & \left(1,\{\mathtt{gbaf}\}\right) \\ \left(-4,\{\mathtt{jgb}\}\right) & \left(-3,\{\mathtt{jg}\}\right) & \left(-3,\{\mathtt{i}\}\right) & \left(2,\{\mathtt{j}\}\right) & (0,\{\lambda\}) \end{bmatrix}$$

In accordance with Table 2, the (i,j)-th entry of A^* gives the shortest distance from node i to node j, together with the identifiers of all corresponding shortest paths.

As another illustration, let us consider once more the graph G in Figure 1, but now take into account all values assigned to its arcs, i.e. arc lengths, arc reliabilities and arc identifiers. Each triplet of values can be regarded as a label from $P_2 \otimes (P_4 \otimes P_8)$. The matrices A and A^* now look as follows:

$$A = \begin{bmatrix} (\infty,0,\emptyset) & (\infty,0,\emptyset) & (3,0.9,\{\mathtt{a}\}) & (\infty,0,\emptyset) & (\infty,0,\emptyset) \\ (-1,0.7,\{\mathtt{b}\}) & (\infty,0,\emptyset) & (\infty,0,\emptyset) & (8,0.1,\{\mathtt{c}\}) & (\infty,0,\emptyset) \\ (9,0.4,\{\mathtt{d}\}) & (1,0.3,\{\mathtt{e}\}) & (\infty,0,\emptyset) & (\infty,0,\emptyset) & (4,0.8,\{\mathtt{f}\}) \\ (\infty,0,\emptyset) & (-5,1.0,\{\mathtt{g}\}) & (5,0.5,\{\mathtt{h}\}) & (\infty,0,\emptyset) & (\infty,0,\emptyset) \\ (\infty,0,\emptyset) & (\infty,0,\emptyset) & (-3,0.2,\{\mathtt{i}\}) & (2,0.6,\{\mathtt{j}\}) & (\infty,0,\emptyset) \end{bmatrix},$$

$$A^* = \begin{bmatrix} \begin{pmatrix}0, \\ 1.0, \\ \{\lambda\}\end{pmatrix} & \begin{pmatrix}4, \\ 0.432, \\ \{\mathtt{afjg}\}\end{pmatrix} & \begin{pmatrix}3, \\ 0.9, \\ \{\mathtt{a}\}\end{pmatrix} & \begin{pmatrix}9, \\ 0.432, \\ \{\mathtt{afj}\}\end{pmatrix} & \begin{pmatrix}7, \\ 0.72, \\ \{\mathtt{af}\}\end{pmatrix} \\ \begin{pmatrix}-1, \\ 0.7, \\ \{\mathtt{b}\}\end{pmatrix} & \begin{pmatrix}0, \\ 1.0, \\ \{\lambda\}\end{pmatrix} & \begin{pmatrix}2, \\ 0.63, \\ \{\mathtt{ba}\}\end{pmatrix} & \begin{pmatrix}8, \\ 0.3024, \\ \{\mathtt{bafj}\}\end{pmatrix} & \begin{pmatrix}6, \\ 0.504, \\ \{\mathtt{baf}\}\end{pmatrix} \\ \begin{pmatrix}0, \\ 0.336, \\ \{\mathtt{fjgb}\}\end{pmatrix} & \begin{pmatrix}1, \\ 0.48, \\ \{\mathtt{fjg}\}\end{pmatrix} & \begin{pmatrix}0, \\ 1.0, \\ \{\lambda\}\end{pmatrix} & \begin{pmatrix}6, \\ 0.48, \\ \{\mathtt{fj}\}\end{pmatrix} & \begin{pmatrix}4, \\ 0.8, \\ \{\mathtt{f}\}\end{pmatrix} \\ \begin{pmatrix}-6, \\ 0.7, \\ \{\mathtt{gb}\}\end{pmatrix} & \begin{pmatrix}-5, \\ 1.0, \\ \{\mathtt{g}\}\end{pmatrix} & \begin{pmatrix}-3, \\ 0.63, \\ \{\mathtt{gba}\}\end{pmatrix} & \begin{pmatrix}0, \\ 1.0, \\ \{\lambda\}\end{pmatrix} & \begin{pmatrix}1, \\ 0.504, \\ \{\mathtt{gbaf}\}\end{pmatrix} \\ \begin{pmatrix}-4, \\ 0.42, \\ \{\mathtt{jgb}\}\end{pmatrix} & \begin{pmatrix}-3, \\ 0.6, \\ \{\mathtt{jg}\}\end{pmatrix} & \begin{pmatrix}-3, \\ 0.2, \\ \{\mathtt{i}\}\end{pmatrix} & \begin{pmatrix}2, \\ 0.6, \\ \{\mathtt{j}\}\end{pmatrix} & \begin{pmatrix}0, \\ 1.0, \\ \{\lambda\}\end{pmatrix} \end{bmatrix}.$$

In accordance with Table 2, the (i, j)-th entry of A^* now contains three values: the shortest distance from node i to node j, the maximum reliability that can be achieved by considering only shortest paths from node i to node j, and the set of identifiers of only those shortest paths from node i to node j that achieve the maximum reliability.

6. A distributed algorithm

According to the adopted algebraic approach, an algorithm for solving path problems is simply a procedure that computes a matrix closure. Such algorithm should be general in the sense that it works over an arbitrary semiring and can therefore be applied to different types of simple or complex path problems. Our particular algorithm computes the strong closure A^* of a given $n \times n$ matrix A over a semiring P. The expression of the form (1) is evaluated by *iterative squaring and updating* of a suitable matrix B. The algorithm starts with $B = E \vee A$ and stops when B stabilizes.

Our algorithm is *distributed* in the sense that it consists of a ring of m concurrent processes, where $1 \leq m \leq n/2$. Each process communicates with its predecessor and its successor along the ring. There is no shared memory, but each process has its own private memory. Figure 2 refers to a ring of $m = 4$ processes.

To enable distributed computing, the algorithm maintains two copies of the matrix B, denoted by $R = [r_{ij}]$ and $C = [c_{ij}]$. The matrix R is divided into $2m$ blocks, so that every block consists of a roughly equal number of adjacent rows. Similarly, C is divided into $2m$ blocks of columns. The range of column indices assigned to one particular block of C is the same as the range of row indices assigned to the *corresponding* block of R. The blocks of R and C are distributed among processes, so that one process keeps exactly two blocks of R and the corresponding two blocks of C.

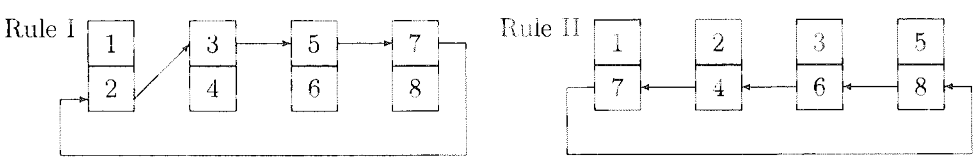


Figure 2: Block exchange rules for $m = 4$.

As previously explained, the algorithm consists of iterations. However, due to distributed computing, each iteration is further divided into $2m - 1$ smaller parts called *phases*. In one phase a process generates row-column pairs from its locally available blocks. In each phase, all possible pairs from non-corresponding blocks are generated. In the first phase within an iteration, additionally, all possible pairs from the corresponding blocks are also generated. For each generated pair, consisting of, say, the i-th row of R and the j-th column of C, the process computes the "inner product". The obtained value is used to update both the j-th element of the i-th row of R and the i-th element of the j-th column of C; thus:

$$r_{ij} := c_{ij} := \bigvee_{k=1}^{n} r_{ik} \circ c_{kj}. \qquad (3)$$

After any phase the processes exchange blocks, in order to form new combinations of blocks for the next phase. Any block of C moves together with the corresponding block of R. The exact block exchange procedure is shown in Figure 2, for the case where $m = 4$. Rules I and II are alternately applied. Each index in Figure 2 denotes one block of R together with the corresponding block of C. Arrows in Figure 2 indicate block moves. It has been proved in [5] that for a stable matrix A the algorithm terminates after a finite number of iterations, with the final matrices R and C equal to A^*.

7. Network implementation

The presented algorithm has been implemented as a distributed C program by using the *PVM library* [2]. The program code has been designed so that it can easily be adjusted to solve various types of path problems. In order to monitor performance, the program has been extended so that it counts its own iterations and records its own execution time.

Our distributed program runs as a set of concurrent processes, which can be allocated to different computers, and which exchange data through the network in a ring-like fashion. The program has been tested on a "virtual parallel machine" assembled of four UNIX computers. We have been able to run the program with up to eight truly concurrent processes.

Performance of the program has been measured on few hundreds of path problems of different type and size. The obtained performance measurements are given in [5]. A small part of the measurements is also presented here.

graph density	number of processes m	total execution time in seconds	speedup compared to the sequential case
1%	1	292.7	1.00
	2	153.3	1.91
	4	82.1	3.56
	8	46.3	6.32
3%	1	316.7	1.00
	2	164.7	1.92
	4	88.3	3.59
	8	49.9	6.35
10%	1	324.8	1.00
	2	167.2	1.94
	4	89.9	3.61
	8	51.0	6.37
30%	1	374.3	1.00
	2	191.6	1.95
	4	101.9	3.67
	8	57.7	6.49

Table 3: Performance for shortest distance problems with size $n = 256$.

Table 3 summarizes the performance data for 40 randomly generated shortest distance problems of the same size $n = 256$. The problems have been further divided into groups of 10 according to their graph density (percentage of non-zeros in A), which can be 1%, 3%, 10% or 30%. Each particular problem has been solved four times, i.e. with the number of concurrent processes equal to 1, 2, 4 or 8. Entries in Table 3 are mean values computed over a whole group of 10 similar problems. A table row shows the total execution time, expressed as an absolute value (seconds) and as well as a relative value (speedup compared to the sequential case).

Table 3 clearly indicates that the algorithm is quite efficient when applied to larger shortest distance problems. Satisfactory speedups are achieved even with 8 processes. Performance is better for denser graphs.

8. Conclusion

In this paper we have presented two results dealing with the algebraic approach to path problems. The first result is a mechanism for combining simpler semirings into composite semirings. The second result is a distributed algorithm for solving path problems.

Both results extend the applicability of the adopted algebraic approach. Namely, with composite semirings it is possible to use the same algebraic formulation not only for simple problems but also for more complex tasks, such as explicit identification of optimal paths or multi-criteria optimization. On the other hand, the described distributed algorithm allows solving the whole range of simple and complex problems on a variety of computer networks.

The idea of extending the applicability of the algebraic approach has an obvious aesthetic appeal. Moreover, the described extensions bring some practical benefits. For instance, it becomes possible to solve a wider class of problems by already known and tested algorithms. Or the same class of problems can be solved more efficiently by employing more computers.

References

1. Carré, B., *Graphs and Networks*, Oxford University Press, Oxford, 1979.
2. Geist, A. et al., *PVM: Parallel Virtual Machine - A Users' Guide and Tutorial for Networked Parallel Computing*, The MIT Press, Cambridge MA, 1994.
3. Hu, T.C., Shing, M.T., *Combinatorial Algorithms*, Dover, London, 2002.
4. Huckenbeck, U., *Extremal Paths in Graphs*, Akademie Verlag, Berlin, 1997.
5. Manger, R., "Solving path problems on a network of computers", *Informatica*, Vol 26 (2002), 91–100.
6. Manger, R., "A new path algebra for finding paths in graphs", in: Lužar-Stiffler, V., Hljuz Dobrić, V. (editors), *Proceedings of the 26th International Conference on Information Technology Interfaces - ITI 2004, Cavtat, Croatia, June 7-10, 2004*, University Computing Centre, Zagreb, 2004, 657–662.
7. Manger, R., "Composite semirings and their applications dealing with path problems in graphs", *Grazer Mathematische Berichte* (to appear in 2005).
8. Mohri, M., "Semiring frameworks and algorithms for shortest-distance problems", *Journal of Automata, Languages and Combinatorics*, Vol 7 (2002), 321–350.
9. Rote, G., "Path problems in graphs", *Computing Supplement*, Vol 7 (1990), 155–189.

ON SOME TRAPPING STATES TO OR

by Viljem Rupnik
INTERACTA, LTD, Business Information Processing
Parmova 53, Ljubljana
e-mail: Viljem.Rupnik@siol.net

Abstract:
There can always be some handicaps to OR to be successful, but to remain honest, we are not persistently entitled to blame OR users for not being efficient. We are first obliged to recognize the problem in its intrinsic scope and essence; consequently, we also have to find an efficient solution for any OR trapping state.. This very last task is in our focus in this paper.

Keywords: decision theory, stochastic and chaotic processes, parameter activated optimisation, flexible economic modelling, entropy measurement.

0. An Introduction

There are two significant and crucial scientific disciplines which influence pragmatical reputation of OR: gnoseology and hermeneutics. Both of them should be taken into account by OR problem solver: the first one determines his problem solving procedure and the latter one shapes the scope and depth of its implementation in real business lives. They interact strongly; if we are not paying enough attention to such an interdependence, it may well happen that our OR produce will not be accepted by practitioners/managers/users. If such an event persists again and again, we are due to check it as a trapping state to OR. We assume that resposibility is imposed on OR designer; we picked up some most chalenging reasons for such trapping states.

1. Embedience problem

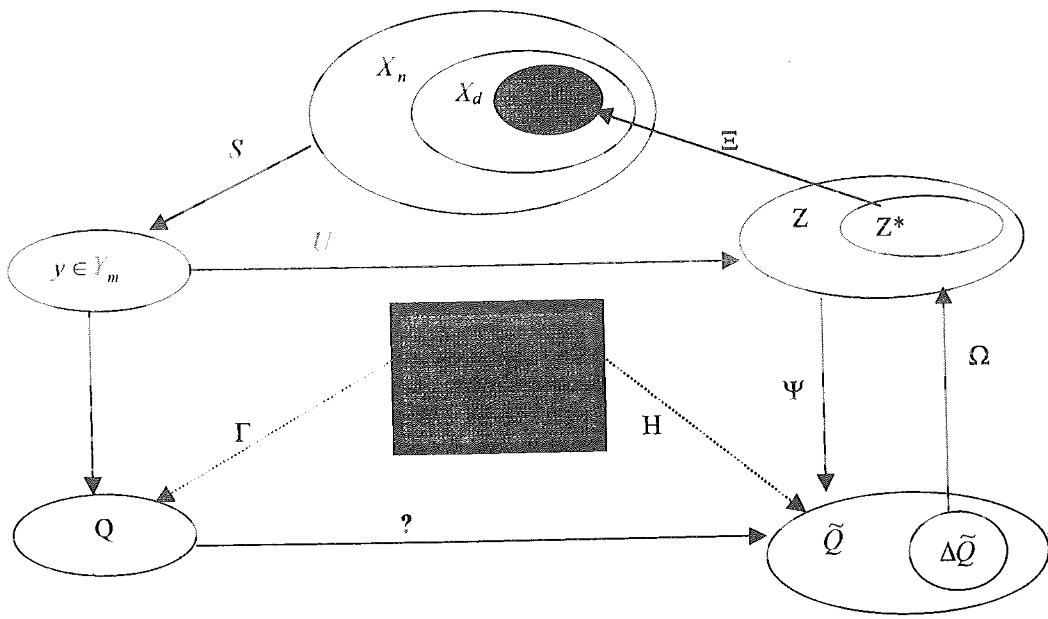
When solving some well defined problem, it may appear that the solution obtained just not satisfy the management, although it is perfectly defined from the mathematical point of view. As we shall consider various potential reasons of it later on, it is safe to start with most common issue on so called embedience phenomenon.

To illuminate it, let us start with general decision theory and its fundamental equation, whose constituents are as follows:

- let X_n be an arbitrary finite dimensional vector space, representing all conceivable decision variables $x \in X_n$;
- let $\mathfrak{R}$ be an object of decision making process (which, in general, is not a problem itself);
- there is always at least one consequence $y \in Y_m$, again from finite dimensional vector space, corresponding to input $x \in X_n$;
- we also introduce an operator $Y = SX$ producing output from input; let us call it as decision-generator;

- X_a should serves as a space of admissible decision variables, $X_a \subset X_n$;
- let $z \in Z$ be an estimate of a consequence of decision variable $y \in Y_m$ (of an arbitrary dimension);
- there is a mapping U of $y \in Y_m$ into $z \in Z$;
- let $q = (y_1, ..., y_m) \in Q$ be a primary (backwards) construction of a problem, based on $\Re$, under the conditions of certainty;
- as an analogy, let $\tilde{q} = (z_1, ..., z_m) \in \tilde{Q}$ be a secondary (backwards) construction of the same problem, based on $\Re$, under the conditions of uncertainty;
- let Φ is a mappping of Y into Q (a primary generator of problem constructions); in real situations, this mapping reduces Q into some part ΔQ;
- for the case of uncertainty, let Ψ maps Z into $\tilde{Q}$; here again, this mapping may produce some shrinkage Δ of $\tilde{Q}$;
- the shrinkage of $\tilde{Q}$ is then projected onto $Z^* \subset Z$ via operator Ω;
- Ξ is an operator of induced subspace X^*_d of alternative admisible decisions, being mapped from $Z^* \subset Z$;

Based on these minimum categories of decision making process the following graph might be useful (see/1/):



For the case of uncertainty we derive the fundamental equation of decision theory (FED):

$$X^*_d = (\Xi\Omega\Delta\Psi US)X$$

A series of potential standard failures is as follows:

- we usually simplify »our« space of consequences to be Z=(US)X and thus X=$(US)^{-1}Z$ (if possible?!), and consequently $\Xi(US)^{-1}$: it means that a) we neglect any alternative solutions, b) we neglect admissible solutions, c) we are too bold to assume that all inputs could be solutions;
- is S really known to us?
- is U really known to us?
- the two operators U and S in FED are tacitly assumed to be uncertain: do we use it as deterministic operators?
- are we sure that $\Delta\tilde{Q}$ is sufficient for our decision on X^*_d?
- a similar doubt as to the operators U and S to hold true for all other operators in FED ,
- a transition from Q to $\tilde{Q}$ has not been examined whatsoever; it is a reflection of our dangerous oversimplification of decision proces;
- a bridge between hard and soft sciences is often demolished by using U= identity operator which is, in general, very far from being realistic and adequate approach, it may well hapen that an/the solution $y \in Y_m$ is acceptable from technical aspects only, but from the others;
- are we not worried about making some adequate snapshots of $\Re$ to get problem space Q, or, even worse, $\tilde{Q}$?

From this we see that there many reasons to fall into trapping state when solving real life problems to meet the demands of managers. What can we do anyway?

1. when fighting against Y=SX as input-output mapping, we have to consider three possible distintions: is there our problem deterministic, stohastic or chaotic (ultrastohastic). The corresponding tool may bring less disappointments to both OR-specialists and managers.

2.looking at evaluation mapping Z=UY the main target of ours is that the definition of utility which should be paid to our OR clients: what are their measures of satisfaction, what are *their* criteria of *our* solution quality, ect. should be told by them to us in advance.

3. problem mapping $\tilde{Q}$= ΨZ detects our ability of perception: have we passed through a sufficiently deep information analysis of our problem? As a whole, we are to relay on information sciences as a confident and reliable assistant.

4. since the mapping $y \in Y_m$ practice is largely used (regardless of its quality), the explicit use of mapping $Q \to \tilde{Q}$ is almost totally neglected, although promising much richer space of problems under uncertainty. Here, we are in doubt, whether the reversed procedure is more fruitful; such a dilemma reduces to our decision which principle of causality should be used first.

5. the space $\Delta\tilde{Q}$ of reduced problems is very often an origin of misunderstanding between managers and OR specialists: it lies deeply in a dualism of perception (its treasure being $\tilde{Q}$= ΨZ) and aperception (its treasure being $\Delta\tilde{Q}$) of uncertainty-conditioned problems. Consequently, operator Δ is most obscure factor in FED ; if we choose it as identity operator, we exposed ourselves to a lot of blames from the side of users.

6. in the above highly simplified formulation of FED an operator Ω represents an evaluation of solution to a problem under uncertainty; the reversed operator would describe our conceptualisation of problem(s) in an indirect way: this might fill in the space $\Delta\tilde{Q}$ in the sense of its gradual refinement (like auto-adaptive control processes do).

7. a straightforward by-pass of FED is trying to be the expert modelling whose task is to find the mapping $\Delta\tilde{Q} \to X^{*}{}_{a}$; we have to cherish it as a potential way to reduce OR failures.

8. furthermore, an artificial intelligence looks for a mapping $(\Psi, H) \to \tilde{Q}$ and thus helps us to decrease the frequency of OR falling into trapping state.

9. finally, of course, we do not dare to call for help a mapping $\Delta\tilde{Q} \to (\Omega, \Xi)$, where for $\Delta\tilde{Q}$ as a subspace of $\tilde{Q}$ a pair (Ψ, H) is applied (meta-mathematics).

An example: let $\Re$=firm's gross revenue per employee; then Q=$\{Q_1, .. Q_6\}$, where Q_1= economic problem, Q_2= administrative problem, Q_3=psychological problem, Q_4=technological problem, Q_5=organisational problem and Q_6= problem of social policy. Our object thus embeded six (induced) problems; it is likely that each of these six problems is going to be treated as an object again: embedience phenomenon streches over problems, either. Simplifying the case by taking $Q = \tilde{Q}$, our promise to manager to run a project on object $\Re$ has then at least six dimensional criteria of consequences Z and S is six-component operator and Y is 6-tuple »criteriased« output $y \in Y_m$; the same nature is featured by 6-dimensional vector $z \in Z$:

- how to find an/the optimal vector $z \in Z$ in case of its component having different dimensions (it calls for multidimensional ranking procedure);
- how to find an/the optimal vectorial solution to $\Re$ problem, if *dim* $\Re$ is different from all six dimensions of $z \in Z$ components (it calls for non-formal optimisation modelling of $\Re$);
- if we avoid the above two approaches, what is the mapping of 6-dimensional »dimensionally non-congruent« components of $z \in Z$ into *dim* $\Re$.

The example above (see for the details /2/) has shown that the embedience phenomena are different and numerous: so are the potential reasons of OR being trapped.

2. Stochastic/chaotic modelling problem

A transition from stochastic to chaotic modelling is not a trivial task of modelling in general. To start with, we propose a rather non-conventional definition of chaos: the domain of a variable is said to be a chaos, if and anly if there does not exist any probability distribution (we shall sharpen this definition later on). Basic problem, then, when building up a model, is to recognize a/the chaotic nature of variables in a model. From pragmatic point of view, we look for its approximation via some finite series composed by convex composition of finite

number of any arbitrary probability distributions. Consequently, a chaotic model can not be processed in its full extent, but in its probabilistic representative (see /3/).

Let I = [0,T] be the time domain of stohastic process F(t), streched over its empirical realisations f(t); let us furthermore introduce corresponding differential probability function (DPF) p(t) = p[F(t),t], J = (T, ∞) being the future of the same process (where F(t) appears as a concatenation from the past), and

$$\Pi[F(t)] = \sum_{i=N} \xi_i \pi_i(x) \quad , \; x(t) = F(t), \; \sum_{i \in N} \xi_i = 1$$

being the corresponding DPF and $\pi_i(x)$ being any finite set of arbitrary differential probability functions; in case, when this series is finite or infinite, but convergent, the process F(t) is said to be stochastic. In the opposite case such a process is said to be *ultrastochastic* or *chaotic*. Consequently, in practice we have to deal with some ε-approximation of chaos.

$$\Pi[F(t)] = \sum_{i=N} \xi_i \pi_i(x) \quad , \; x(t) = F(t), \; \sum_{i \in N} \xi_i = 1$$

There are the two exstrems of Mandelbroot dilemma:

- p(t) and $\Pi(t)$ are totally independent;
- they are identical, $\Pi(t) \equiv$ p(t), which is found as traditional approach of naive business modelling.

Thus, our task is to discover the existence of mapping $\Pi(t) = \Psi(p(t))$. Thus, the point T serves as switching point from p(t) to $\Pi(t)$. In order to »palpitate« the future behaviour we insert or implant some points of future process and, according to balance between the two extrema, we concatenate all the past points and the future points »chosen«. The production of future points might be leaned upon Monte Carlo simulation procedure and then we mix the two sets of points. In short, the whole procedure may be called as *implantation*, what a set of points should be accepted as a operandum of implantation process? The simpliest way is to use *infimal* and *supremal of processes* both in the past and the future; in most of the practical cases, an additional information is useful, namely that of *modus process* as a turnpike curve. In such a case we deal with 3-parameter implantation process. To simplyfy our exposition, we confine ourselves to the above presented »3-bone skeleton« throughout the implantation procedure. Such a choice says we assume that »a majority« of process lies inside the cone build up by its skeleton, inside of which we have no further skeleton-wise information (see the details: /4/).

To generalize our approach we allow to assure the prediction [$F(t)$, $\Pi(t)$] for any t>T based on implantation described above and obeying a constraint [$F(t) - m \leq F(t) \leq F(t) + M$], m and M being arbitrary finite limits. The corresponding cone is expected at probability $\Pi_{\Sigma}\,(m \leq x(t > \mathrm{T}) \leq M)$ or risk being $\rho = \int_{-\infty}^{m} \Pi(t)dt + \int_{M}^{\infty} \Pi / t)dt$. A series of prediction impulses $t_1, t_2, \ldots$ is matched to a series of $F_0(t_1)$, $F_0(t_2)$, ……, $m(t_1)$, $m(t_2)$, ……and $M(t_1)$, $M(t_2)$, ……, thus creating a *prediction skeleton*. In practice, it is thoughtful to assume it to be finite, having a length l; we shall call it *a prediction window* and denote as $\Theta_\varepsilon(\tau \geq T, l)$.

Our task has been to discover whether prediction window tends to qualify the process observed as stohastic or chaotic one. In addition, if we simplyfy our discussion by assuming $\mu = |m(t_i) - F_0(t_i)| = |F_0(t_i) - M(t_i)|$, then $\Theta_\varepsilon(\tau \geq T, l)$ is a *chaotic window*, if $\Pi_\varepsilon(t)$ for t > T is divergent: in such a case a sequence of localised probability distributions represent $\varepsilon(t)$ - approximation of chaos. Both parameters μ an l are heavily dependent on ρ. Thus, we have to

- construct apriori probability space, out of which we slice-out prediction probability subspace embracing the modus turn-pike trajectory of width 2μ;
- find the corresponding space of apriori random variable, streched over l pairs of supremal values $F(\tau_i, \Pi(\tau_i) + \mu)$ an infimal values $F(\tau_i, \Pi(\tau_i) - \mu)$ respectively for each τ inside chaotic window.

To resolve Mandelbroot dilemma we have to observe the two pairs $[F_i(\tau), \Pi_\varepsilon(\tau)]$ and $[f_i(t), p_i(t)]$. It is convenient to define $\Psi = \| \, |\Pi| - |p| \, \|$ as a measure of prediction quality in terms of probability. It is computed inside the chaotic window provided we have shifted the corresponding »past slice« onto the chaotic window domain: thus, the recording of Ψ supports information on the relationship between *the past* and *future stochasticity* (a measure of Mandelbroot's dilemma). The left Mandelbroot has value zero, the right one refers to infinity.

The issue of prediction needs one more operator, say Φ, which measures a dependence between $f_i(\tau)$ and $F_i(\tau)$. After assuring the same domain we have $\Phi = \| |F| - |f| \|$, which could be considered as a *measure of prediction of realizations*. Consequently, a pair Q = (Φ, Ψ) is a measure of *prediction quality*, operator Φ being a criterion of risk ignorant and operator Ψ being a criterion of risk relevant decision maker. Finally, a complete information needed by decision maker is given by $\Xi = [[F_i(\tau), \Pi_\varepsilon(\tau)], Q]$.

From this formulation, it is obvious what reasons for OR to be stuck in practice. In short, in general the past stochasticity does not hold true in the future; moreover, if the future is chaotic (being ε-approximated), the error produced by stochastic modelling might be disastrous. For example, traditional statistics rests on aposteriori stohastics, since it deals with regression analysis as a tool for extrapolation/forecasting/prediction of the random process.

3. Flexible versus stiff category input application in economic modelling

The traditional economic models, mostly econometrics, deal with a *some* input, which determines *some* output, both defined by some pre-chosen economic categories. Apart from the issue on how and why these inputs and outputs were selected (by the way, we may seriously argue about it), there is a variation of output as a consequence of variation of input, where the latter one reflects its data-history. Such an inaccuracy may be decreased by using wavelets, which brings the quality of regression procedure at arbitrary level. However,

it still may not be satisfactory. The approach which promises some additional increase of output quality (like accuracy, optimality, ect.) could be rendered by the concept of flexible inputs. They are allowed to change its content and not to keep them fixed over time. If we insist on traditional econometric approach, for example, the error acruing from the two possible approaches may be to large to be endured in practice. Following the scheme of solution proposed below we can easily detect all dangers to OR of being stuck in its practical aplications.

To introduce the improvement of today's econometrics we shall refer to simbols in Ch. 2. The inverse of Ξ is in question, but not in a direct way. Following traditional econometric model, the space X of decision variables effects (indirectly) the problem space Q: thus, we look for mapping $X \to Q$ and consequently, it effects Z^* as our output space. For this purpose, let $P(t) \subseteq X(t)$ be a maximum set of economic categories, assumed as stochastic processes of an arbitrary type and having been defined through $x(0) = m_0$ and covariance matrix R_0. If we follow the concept of fluid modelling (FM) (see /5/), we shall, in service of economic modelling, have $P(t) = \bigcup_{i=0}^{\bar{n}} P_i(t), t \in T$, where $P_0(t)$ is being called a »kernel«, if it contains $n_0 \leq N$ linearly independent (or »almost-independent«) processes, where n_0 is maximum number of non-correlated stochastic processes. The rest of processes is then located to a finite number $\bar{n}$ of »satelytes« $P_1, \dots P_{\bar{n}}$ having prescribed their levels of dependancy on the kernel P_0, say $\delta_1, \dots \delta_{\bar{n}}$, which are known to be the values of functional determinants computed for each satelyte set of processes. Here, statistical theory of classification has been applied in order to get clusters as desired. A question on minimum number of satelytes has been left open, but interesting. All satelytes are thus dependent at some arbitrary predetermined levels δ_i and are algebraically expressed through the kernel (as done in linear case). For the service of economic modelling we are »happy« to use the same kernel as a set of indepenedent »variables« to act as »hard« part of the econometric model and a set of satelytes is »swinging« part of it. Apart from the control vector $u(t)$ in case of control oriented econometric model we vary $\delta_1, \dots \delta_{\bar{n}}$ to optimize/improve criteria/goals functional $y(t)$ as you see below.

Thus the state-space operator of control oriented econometric model is a system of differential/difference equations

$$\dot{x} = A(t)x(t) + B(t)u(t), \qquad x(t) \in P_0 \tag{1}$$

where all satelyte variables are algebraically modelled, e.g. in linear case

$$\begin{aligned} x_i(t) &= L_i[x(t)], \qquad i = 1, \dots \bar{n} \\ x(t) &\in P_0 \\ x_i(t) &\in P_i(t) \end{aligned} \tag{2}$$

As partition of $P(t)$ depends on t, (1) and (2) represent a *fluid model* (see /5 /), which is now a substitute of econometric (or any other economic model). Functions chosen within a kernel as well as those within satelytes vary over time: input categories are thus flexible and the kernel is being determined by input for each t. For this reason we may call it as »one-way«

fluid model, in the contrast to »two-way« fluid model, where the kernel is input-iutput determined. In the latter case we need some criterion functional

$$y(t) = Y[x(t), u(t)] \qquad x(t) \in P_0 \qquad (3)$$

which qualifies (1), (2) and (3) as a control oriented econometric model (see /6/). Even in the case of $u(t)$ kept fixed, the two-way fluid model could be improved through adjusting $\delta_0 > 1 - \varepsilon_0, \delta_1, ...\delta_{\bar{n}}$ so that $y(t)$ improves. Evidently, the same procedure is welicome to one-way fluid model.

Any conomic ppolicy should be as good as possible; why not to replace stiff input categories with flexible ones, varying over time, provided the policy goals are improved. We can see from a composition of fluid model, what reasons for OR to get stuck are worth while to be paid our attention.

4. Swapping (interchangeable) optimisation

When formal modelling some problem, there are two distinct groups of model constituents: variables and parameters. In practice, we hit upon the situation when variables become constants and parameters call for variation. If we take such a situation into account, the so called *swapping procedure* takes place.

Let us discuss it in an abstract way first. According to Ch.1 we have $Y=SX$ as a formal presentation of problem. A swapping procedure needs to reshape it in a way $Y= S(\Sigma)X$, where Σ represents a space of all relevant model parameters. If we confine ourselves to finite dimensional vector spaces appearing in (1), we have

$$y=S(\sigma)x \qquad (1)$$
$$\Phi(x) \le 0$$

where x, y and σ are vectors. Let (1') be Hicksian-sense dynamic model of a problem in consideration; its total operating horizon $T=\cup T_i$ is thus being split by the »swapping« behaviour of σ and x vectors. Let us simplyfy our discussion by assuming that S is independent of i; optimizing across T_i we get a functional

$$\text{opt } y(T_i) = y_i^0(\sigma_i^0 \text{ or } x_i^0) \qquad \forall i \qquad (2)$$

subject to some constraints to either x or σ.

If we can not endure either Δ opt $y(T_i)= y_i^0(\sigma_i^0 \text{ or } x_i^0) - y_i^0(\sigma_i^0, x_i = const)$ or Δ opt $y(T_i)= y_i^0(\sigma_i^0 \text{ or } x_i^0) - y_i^0(\sigma_i = const, x_i^0)$, then, refusing to deal with swapping phenomenon, a serious damage may be caused to the problem. This damage may increase in case, when $y_i^0(T_i)=$ opt $y(T_i)$ is a criterion aditive over time. Following FED from Ch.1 this damage may be demonstrated still stronger when introducing space Z (expressing

economic categories, e.g.). However, the swapping procedure extends to $U(\gamma)$ which may play role simultaneously with (2) – a triple swapping procedure.

The trapping state stemming from swapping phenomenon might be today very difficult to overcome. For example, if $S(\sigma = const)$ is linear operator with constant coefficients and z= $U(\gamma)y$ being linear function of y, then the corresponding problem is bi-linear version of Wolfe's program in variables x and γ, provided we added some linear constraints on γ. As a whole, any swapping analogue to (1) as a mathematical programming problem might be a trapping state per se, since the corresponding existence theorems may not exist; a short review of swapping allowed problems is given in /7/. However, the simulation and non-formal modelling (see /8/) may be of use in such cases.

It is worth while to illuminate the simulation of problem (1) and its swapping analogue in their discrete version through the following

EXAMPLE: when running business, manager is obliged to respect balance sheet and income statement which represent very simple model, mostly used as an aposteriori information source. It is very seldom applied as a managerial optimization tool, one of the reasons is a fact that it does not contain many functional relationships; instead, it is more a set of »soft mapping« between economic/financial/technical/technological categories one onto another. To show how this tool could be used for optimal decision making along the strict line of FED structure shown in Ch. 1, we shall combine both sheets in one:

-space X contains the following decision variables x: volume of interets, nominal price of share, EBIT/total assets, sales volume;
-space X contains the following decision parameters σ (X): interest rate, non-risk rate of return, risk rate of return, beta coefficient, expected return of assets, reinvestment coefficient, coefficients of ordinary/priority shares, tax rate;
-space Y consists of: leverage, total assets, number of outstanding shares, fixed costs, variable costs, selling price(s); in adition, it contains also induced (=not primary criteria) vector $\sigma(Y)$: ROA,ROE, dividend paid, market price of share, the ratio between market price of share and its return, the ratio between market price of share and its book price, WACC, debt capital, equity capital;
-space Z of consequences: z = EPS (earning per share).

The policy goal here is a single one, EPS, being able to be expressed analytically through $y \in Y$ (e.i. operator U has formal description), whereas S does not have such a possibility. We face a hybrid of formal and non-formal part of a EPS- optimizing model. To derive $z \in Z$ we apply a finite (but arbitrary) step simulation over the union of spaces X (of both parts) and non-induced part of Y; the simulation procedure is always finite, regardless to what swapping parameters σ (X) and $\sigma(Y)$ were used (even both of them). For certain subperiod of time a firm is not able to vary some component of x; if the corresponding level of EPS is below the manager's expectation, then he activates some component(s) of $\sigma(.)$; for the details you may see /2/. It is also worth to stress that the traditional balance sheet $\cup$ income statement can always be extended towards non-financial spaces, e. i. to the spaces a manager feels as an additional tool to improve EPS. Such enlargements are not vulnerable by category dimension involved neither they are prohibited, when space Z is a vector space; in the latter case we deal with multicriteria swapping procedue, exercised over (in general) a hybrid managerial decision making problem: it still obeys FED philosophy.

6. Statistical and non-statistical notion of uncertainty

The uncertainty is, simply saying, a complement to expectation. To deal with uncertainty of some object in a quantitative way we have to define a) the object, b) the domain of object and c) a measure of uncertainty.Ad a): as an object, each of the variables, mappings and properties of FED could be used. Ad b): depending on the nature of the object we define its domain (universum U). Ad c): a measure of uncertainty/expectation may not only be dependent on the nature of an object, but also/or on the aspect (criterion) of our analysis of the object. If this analysis is a quantitative one, a measure of uncertainty/expectation should be a uniform and real function/functional/operator. There is a significant dichotomy of all possible measures: i) it is either a measure in the sense of the set theory or ii) it is a measure of existence of the object. In the first case we may speak about non-statistical (or non-probability) uncertainty resp. expectation, while in the latter case uncertainty/expectation is of statistical type. However, we shall need some additional condition for modification of the above distinction.

Let $U = \cup U_i$ be a set of n objects U_i and $\mu(U)$ and $\mu(U_i)$ set theory measures; we label it as *absolute measure*. Thus the relative uncertainty or enthropy of U_i is

$$\frac{\mu(U)-\mu(U_i)}{\mu(U)} = \text{non-probability entropy} = H(\mu) \qquad (1)$$

which helps us to compare different objects. In this case $\mu(U_i)$ is one-dimensional information on object U_i (for the sake of simplicity we disregard an issue of more-dimensional non-probability entropies.

If we replace $\mu(U_i)$ by its probability $p(U_i)$, (1) becomes a measure of risk, if a set $\{ p(U_i) \}$ is complete.It besomes a measure of the probability entropy, if a set $\{ p(U_i) \}$ is incomplete.An analogue to (1) as a relative measure of probability entropy can not be used, unless

$$\frac{p(U)-p(U_i)}{p(U)} = \text{probability entropy} = H(p) \qquad \text{provided } i = 1,..n' < n \qquad (2)$$

where n is the number of all objects. The distinction between (1) and (2) is crucial when dealing along FED umbrella. Consequently, in case of (2) we can not expect to have a complete description of probability distribution: we have only an incomplete information on probability space belonging to U. A notion of probability based entropy rests on (2) only. Apart from the ignorance of the difference between (1) and (2) we may commit an additional severe error by not immersing deeply into various types of (1) as well as (2).

To show that it is not a fairy tale, let us first turn our attention to one single type of non-probability entropy, namely that one which refers to optimisation processes, based on formal modelling (see /9/). As a consequence (1) as a definition of non-probability entropy we face a fact that the only measure of entropy is based on the existence of the object. In case of formal optimisation modelling we look for uncertainty of an/the optimal value of criterion function/functional/operator. Let us have scalar field $y = F(x_1,...x_n)$ of optimising variables

which we treat as objects in a way that each of them has its non-probability uncertainty (1). Thus we deal with n-dimensional optimisation where optimisation algorithm assures the optimal direction

$$\textit{optimal direction of } y = \left(\frac{\partial y}{\partial x_1},...\frac{\partial y}{x_n}\right) \qquad (3)$$

The uncertainty $H(\mu)[\, U(x)\,]$ of the universum $U(x)$ is then defined on vector field having its gradient

$$grad\, H = \left(\frac{\partial H}{\partial x_1},...\frac{\partial H}{\partial x_n}\right) \qquad (4)$$

We are interested in the increase of this uncertainty in the direction of y, e.i. its projection on optimal direction (3)

$$H(\mu)[\, U(x)\,] = \left(\frac{\partial y}{\partial x_1},...\frac{\partial y}{\partial x_n}\right) grad\, H = \sum_i \frac{\partial y}{\partial x_i}\frac{\partial H}{\partial x_n} \qquad (5)$$

where $\frac{\partial H}{\partial x_n}$ are non-probability uncertainties of individual optimizing variables x_i, $i = 1,...n$. Thus, the non-probability uncertainty of the optimal solution of any formal problem increases along with the scope of universum and the responsiveness of optimal solution to each variable; when optimisation procedure is carried out through simulation, it increases too, if simulation step is decreasing. From (5), it follows that one and the same decision space X changes its type(1) – entropy, if it serves as an optimisation domain. It is an important finding, if X serves at the same time to some other optimisation problem, bringing a new/different gradient (4). If there is a need for simultaneous optimisation (e.i. in the sense of Bolza control problem) based on finite set of criteria, it might be a problem of detecting »an overall measure of uncertainty« belonging to the given set of optimisation tasks.

On the other hand, when applying definition (2), an entire different handicaps may arise. As known, there exist various probability based entropies. A measure of probability-based entropy havily depends on the type of probability distribution we know only a part of which If there is a severe doubt about some particular probability distribution, the question is on how to find some »better«one. There are only two »boundary« conditions for a $H(p)$ to be found:

1) if the existence of an object is certain, then $H(p)$=0; 2) if the existence of an object is uncertain, $H(p)$=-∞. The remaining behaviour is rather simple: it is monotonically increasing with the increasing number of events. All of its finer properties heaviliy depend on the problem we study: no apriori steps are assured to be the correct ones.

A review of OR possible trapping states has not been finished by now. Let us look at most simple Shanon entropy, $H(X) = -\sum_{i=1}^{i=N} p(X)\log_2 p(X)$; it is very sensible to small probabilities of any type.The way of getting out lies in the redefinition of such an entropy, like adopting Tsallis entropy as a generalisation of Shanon's entropy and being convenient for distributions not from exponential family; another way might be Bayesian priors.

There is a strong justification of our care to be much involved in the classification of probability based entropies. Before we start fulfilling spaces X and Y in Ch.1, it is very important to discover inter-relationships between them, e.g. it may be of great help to use interaction analysis which is essentially based on probability-entropy-measure shown in interaction graphs (see /10/).

References

/1/ Rupnik V., Generalised Economic Modelling and its Applications; Third conference on Operational Research, , Oct. 5-7., 1993, Rovinj, Croatia, pp- 33-49.

/2/http://www.atnet.si/interacta/ssop.html

/3/ Volcjak R., Applicability Limits to the Quality of Economic Modelling; PhD thesis, Department of Economics, University of Ljubljana, 2000.

/4/ Rupnik V., Prediction Chaotic Modelling; unpublished ,Interacta Ltd., 2003.

/5/ Rupnik V., Some Experiences with Fluid Modelling of Economic Systems; Systems Research, The Official Journal of the International Federation for System Research, Vol. 2, No.3, 1985, pp.211-219.

/6/ Mičković S., RupnikV., The Building up of a Functional Optimal Control Oriented Fluid Model; XII Intenational Conference on Dynamic Modelling, Nova Gorica, YU, 1989.

/7/ Rupnik V., The Limits of Structural Control over Production; Proceedings of the Seventh InternationalConference on Operational Research, , Rovinj, Croatia, Sept- 30-Oct.2, 1998, pp- 219-227.

/8/ Rupnik V., An Attempt to Non-Formal Modelling; Proceedings of the Fourth International Symposium on Operational Research, Preddvor Slovenia, Oct.1-3 1997, pp.207-213.

/9/ Rupnik V., Research report on uncertainty of mathematical programming solutions; InterActa Ltd, Parmova 53, Ljubljana, 2004, unpublished

/10/ Jakulin A., Bratko I., Quantifying and Visualising Attribute Interactions; Faculty of Computer and Information Science, University of Ljubljana, August, 2003, unpublished.

DOMINATION OF CAPITAL = TRAPPING STATE OF MANKIND

Viljem Rupnik
Professor emeritus of Mathematics,
Economics and Engineering
Faculty of Economics Ljubljana,
University of Ljubljana
Kardeljeva ploscad 17, 1000 Ljubljana, Slovenia
viljem.rupnik@siol.net

&

Dragomir Sundać
Tenured professor of Economics
Faculty of Economics Rijeka,
University of Rijeka,
I. Filipovića 4, 51000 Rijeka,
Croatia
dragomir.sundac@efri.hr

Keywords: neoliberal globalisation = antiglobalisation, transition countries, "corporation colonialism", globalisation modeling project, intellectual capital, dualism of capital power vs decision power, EU capital allocation strategy.

Extended Abstact.

All the books talking about globalisation, and in connection with that about domination of capital, remain on a description of causes and consequences. This book is currently unique[1] since is simulates the functioning of globalisation for the next 10-15 years. This is done by using metamathematics[2].

The created mathematical model shows that the world's economy is going towards economic downfall unless correctives are implemented, that is if globalisation is spread through liberalisation and if domestic economies are not protected from negative consequences of globalisation.

This particulary applies to a small country and a country in transition such as the Republic of Croatia.

The authors support their claims with mathematical accuracy which presents a great scientific step in shaping the process of globalisation.

The book consists of three parts: 1) a description of the legalities of the process of globalisation, 2) mathematical simulation and 3) proof of economic downfall; as well as a list of needed correctives which should be implemented in order to prevent economic crises of the world economy of unseen proportions.

In the first chapter, titled THE GLOBALIZATION CRITIQUE – WHY ARE PEOPLE FORGOTTEN IN THE MODELS OF ECONOMIC DEVELOPMENT?, authors analyse and describe the consequences of the current process of globalisation, particularly focusing on the destiny of transition countries (among which is the Republic of Croatia) and developing countries, that following a model of "fast integration" into global economic system (forced by world's leading financial institutions: WB, IMF i WTO) experienced a downfall of their own economic, social and political systems.

In the second chapter, GLOBALIZATION MODELLING PROJECT – MATHEMATICAL PROOF OF GLOBALIZATION BREAKDOWN, authors present a mathematical model which they built (based on metamathematics) and by which they project the future and discover scientifically exact possible ways of the development of globalisation.

[1] Last Internet research was completed on 10th of June, 2005.
[2] Metamathematical apparatus which was created for the concrete case.

They are especially retrospective on the transition countries.

According to the words of authors, the goal of the model was "to set formally-consistent doctrine of globalisation, to open potential training ground to various socio-economic aspects and above all, by disclosing negative consequences of globalisation to determine possible defence against negative consequences of globalisation for national economies.

Results which the model gives are daunting and bring to an understanding that the present globalisation works against itself and leads/guides the mankind into an economic self-distruction, as well as any other.

The third part of the book, APPENDICES, contains mathematical evidences of what has been said.

References

/1/ Rupnik, V.: *AN ATTEMPT TO NON-FORMAL MODELLING*; IVth International Symposium on Operations Research, Preddvor, Slovenia, Oct 1997.

/2/ Rupnik, M. and Rupnik, V.: *RKLR method*; InterActa, Ltd,Business Information Proccessing, Ljubljana, 1995

/3/ Rupnik, V.: *THE SHADOWED DECISIONS*; The Fourth International Symposium of Operations Research, October 1997, Preddvor, Slovenia

/4/ Rupnik, V.: *THE RELATIONSHIP BETWEEN CONSUMPTION AND INVESTMENTS AS A METHODOLOGICAL ASPECT IN THE THEORY OF ECONOMIC CONTROL*, Economic Review, Vol. 25, No. 1., 1974.

/5/ Rupnik, V., Strmčnik. S., Karba, R., Bremsak, F.: *A NON – TERMINAL CONTROL APPROACH TO DYNAMIC ECONOMIC MODELING IN NON – COMPACT SPACE*, III. IFAC/IFORS Conference, Warsaw, June 16-19, 1980.

/6/ Rupnik, V.: *AN ATTEMPT TO NON – FORMAL MODELING*, Proceedings to Fourth International Symposium on Operations Research, Preddvor, October 1-3, Slovenia.

/7/ Sundać, D., Pulić, A.: *INTELEKTUALNI KAPITAL – KLJUČNI RESURS 21. STOLJEĆA*, International Business Consulting Center, Rijeka, 2001.

/8/ Sundać, D., Nikolovska, N.: *SCENARIJ ZA EKONOMSKI SLOM ZEMALJA U TRANZICIJI*, International Business Consulting Center, Rijeka, 2001.

/9/ Sundać, D., Nikolovska, N.: *GLOBALIZATION AND ECONOMIC DOWNFALL OF COUNTRIES IN TRANSITION*, MAGOR, Skopje, 2002.

/10/ Sundać, D.: *ESENCIJALNI UZROCI (NE)ZAPOSLENOSTI SA STAJALIŠTA EFIKASNOSTI UPOTREBE KAPITALA*, Zbornik radova Ekonomskog fakulteta Rijeka, br. 14, sv. 2, 1996.

/11/ Sundać, D.: *PRILOZI ZA NOVU EKONOMSKU ZNANOST*, Verzal, Rijeka, 1992.

A REVIEW OF DEA MODELS WHEN THE INTERNAL STRUCTURE OF THE DECISION MAKING UNITS IS CONSIDERED

Lorenzo Castelli [a], Raffaele Pesenti [b], Walter Ukovich [a]

[a] DEEI – Università degli Studi di Trieste, 34127 Via A.Valerio 10, Trieste, Italy

[b] DIAI – Università degli Studi di Palermo, 90128 Viale delle Scienze, Palermo, Italy

castelli@units.it, pesenti@unipa.it, ukovich@units.it

Abstract

Classical Data Envelopment Analysis (DEA) models consider each Decision Making Unit (DMU), whose relative efficiency they evaluate, as a "black box", i.e., its internal structure is disregarded. The paper presents a comprehensive framework of the most advanced theoretical findings in DEA when the internal structure of the DMUs is taken into account, thus giving directions for novel applications of such methodology and introducing it as a powerful tool for complex processes performance evaluation.

Keywords: Efficiency evaluation, Data Envelopment Analysis, Internal structure.

Introduction

Data Envelopment Analysis (DEA) has been a standard tool for evaluating the relative efficiencies of Decision Making Units (DMUs) since the seminal paper by Charnes et al. [1]. Throughout the years, different variations on the original model and interpretation have been proposed (see, e.g., [2] and in particular [3]). However, some underlying assumptions are common to classical DEA models. The *efficiency* of a DMU is defined as the weighted ratio of the outputs (products or outcomes) yielded by the DMU over the inputs (resources used or consumed). All the considered DMUs are *homogeneous*, i.e., they all have the same types of inputs and outputs, and *independent*, i.e., no constraint binds input and output levels of a DMU with the inputs and outputs of the other DMUs. Furthermore, DMUs are seen as *black boxes*, i.e., their internal structure is not considered. As a consequence, generally there is no clear evidence of the transformations the inputs are subject to within the considered units.

In the last two decades, various authors have explored the possibility of measuring efficiency relative to subprocesses or components of the DMUs within the DEA framework. These authors abandon the black box perspective in the assumption that, in some particular contexts, the knowledge of the internal structure of the DMUs can give further insights for the DMU performance evaluation. As an example, such knowledge allows to determine whether better performances can be theoretically obtained by merging the technologies of some substructures of the observed DMUs. In addition, assessing the efficiency of each of the processes or subunits might prevent the inefficiency of some of them being compensated by the efficiency of other ones. In this framework, the aim of this paper is to review the models proposed in the literature that consider internal structures or processes of the DMUs.

1 The basic model

The common building block of the models considering DMU internal processes is represented by a set of N DMUs, each composed of a set of K Decision Making SubUnits (DMSUs). The DMUs are assumed to be homogeneous (i.e., each DMU is composed of the same set of K DMSUs) and independent (i.e., the output of a DMU cannot be the input of another DMU or re-enter the same DMU). In addition, the basic model assumes that:

1. no intermediate flows among DMSUs exist, i.e., the DMSUs are parallel. In other words, the output of a DMSU cannot be the input of another DMSU (and also cannot re-enter the same DMSU), but it must be an output of the whole DMU;
2. any input (output) of the DMU is also an input (output) of one of its subunits;
3. the amount of any input (output) of any DMSU is a-priori fixed.

When a set of N homogeneous and independent DMUs composed of K subunits according with Assumptions $1 \div 3$ is evaluated, "the overall efficient production system can also be improved in technical or scalar efficiency with the aid of information from other DMUs" [5]. In particular, other $N^K - N$ non-observed homogeneous units are added as terms of comparison: they are all the DMUs whose internal structure is composed of a set of K observed DMSUs [6]. As an example, consider $U_1 = (a_1, b_1)$ and $U_2 = (a_2, b_2)$ as two homogeneous and independent DMUs to be evaluated, where (a_i, b_i) are the DMSUs of unit U_i. Since the internal structure is known, the non-existing DMUs (a_1, b_2) and (a_2, b_1) can also be added to the comparison set (Figure 1).

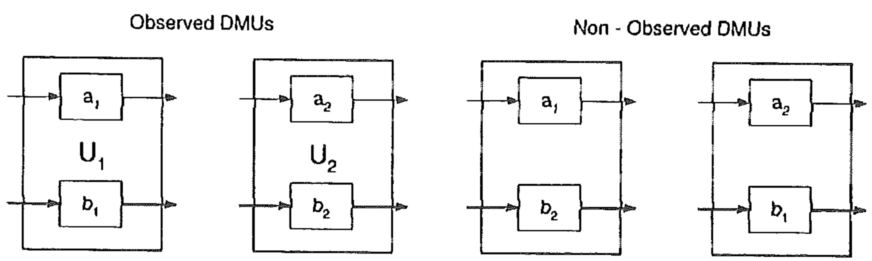


Figure 1: Comparison Set for DMUs U_1 and U_2

It can be easily shown ([5] and [6]) that the maximum relative efficiency of a DMU whose internal structure fulfils Assumptions $1 - 3$ is equal to the maximum of the relative efficiency of its subunits. In particular, such a DMU is [5]:

- weak efficient if and only if there exists at least one of its subunits which is weak efficient relative to the corresponding subunits of other DMUs;
- CCR-efficient if and only if each of its subunits is CCR-efficient relative to the corresponding subunits of other DMUs;

where, according to [4], we define as CCR-efficient a DMU whose optimal objective value of the dual linear problem is equal to 1 and all the slacks are equal to zero. When only the first of these two conditions is achieved, the DMU is referred to as radial, technical or weak efficient.

The introduction of returns to scale of the DMSUs does not affect the above results. By assuming that each of the DMSUs inside a DMU exhibits constant returns to scale independently of the other one, the maximum relative efficiency evaluation of each existing DMU should be assessed by comparing it with all the infinitely many DMUs that can be obtained merging the DMSUs whose inputs and outputs are scaled by a (generally) different positive constant. According to the above arguments, each DMU should be compared with an infinite number of other DMUs. Nevertheless, the problem may be reduced to the comparison of the DMU with all its subunits [6].

When Assumptions $1 \div 3$ are dropped, more complex models are produced. They are going to be discussed in the next section. In particular, three main partially overlapping categories emerge form the literature. The models in the first category deal with DMUs that are still independent and homogeneous but perform several different and clearly identifiable functions, or can be separated into different components [7]. The models in both the second and in the third category consider DMUs that are themselves components of greater structures. The rest of this paper presents a review of results dealing with models of the first type.

2 Multicomponent models

The models described in this section deal with DMUs that are independent and homogeneous but perform several different and clearly identifiable functions, or can be separated into different components. The literature refers to such models as multicomponent [7], joint efficiency [8], or multi–activity [9] models. Formally, according to [7], we define as *component* the bundle of outputs and inputs that characterizes a function of a DMU (see Fig. 2).

In [10], Beasley introduced one of the first examples of a multicomponent DEA model. It is applied to university departments concerning the same disciplines. It was not originally referred as a multicomponent model, but it is nowadays acknowledged as part of the literature on the subject. The considered departments are homogeneous and independent DMUs. However, within them, the teaching and the research activities define two different clearly separable functions. DMU outputs are split: the number of undergraduates and of taught postgraduates are outputs of the teaching function; the number of research postgraduates, research income, and research rating are outputs of the research function. One input, again research income, is specifically *dedicated* to the research function. The other inputs, general and equipment expenditure, are *shared*

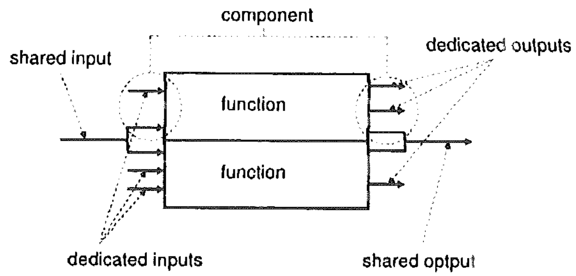


Figure 2: A multicomponent DMU.

(joined) between the two functions.

Note that Roll and Cook in [11] introduced earlier the concept of partial efficiency measures for a DMU. However, in [11], each component is a bundle of independent outputs and inputs. No shared inputs exist. A similar model was later also proposed in [5] by Yang *et al.*

In the following, a notation similar to the one proposed in [7] is adopted. In particular, for each DMU k, define

- i, j, r: the indices of respectively the generic input, output, and component,
- $X_k^r = \{x_{ik}^r\}$: the vector of the dedicated inputs of component r,
- $X_k^S = \{x_{ik}^S\}$: the vector of shared inputs,
- $Y_k^r = \{y_{jk}^r\}$: the vector of the outputs of component r,
- $\nu^r = \{\nu_i^r\}$: the vector of weights of the dedicated inputs of component r,
- $\nu^S = \{\nu_i^S\}$: the vector of weights of shared inputs,
- $\mu^r = \{\mu_j^r\}$: the vector of weights of the outputs of component r,
- $\alpha^r = \{\alpha_i^r\}$: the vector of proportions of the shared inputs allocated to component r.

With a little abuse of notation, $\alpha^r X_k^S$ is also defined as the column vector whose generic entry is $\alpha_i^r X_{ik}^S$. In this context, $\alpha_i^r X_{ik}^S$ is the amount of shared input i allocated to component r by DMU k to maximize its efficiency. However, when a shared input cannot be divided, then α_i^r can be seen as the proportion of the (virtual) value of the input i allotted to component r.

By using the above notation, the aggregated efficiency of the whole DMU k is expressed as $e_k = \frac{\sum_r \mu^r Y_k^r}{\sum_r \nu^r X_k^r + \sum_r \nu^S(\alpha^r X_k^S)}$, and the *partial* efficiency of the single component r as $e_k^r = \frac{\mu^r Y_k^r}{\nu^r X_k^r + \nu^S(\alpha^r X_k^S)}$.

In [10], a department k *aggregate* efficiency $\hat{e}_k = q_k^1 e_k^1 + q_k^2 e_k^2$ is defined as the weighted combination of the partial efficiencies of its two components, where the weight q_k^r of each component r equals to the proportion of the total inputs it receives. In the spirit of DEA, Beasley proposes that each DMU can allocate the shared inputs among its different components so that its aggregate efficiency is maximized. Under such a hypothesis, he proves that, for each DMU, the expression of the weighted combination of the component partial efficiencies is equal the usual efficiency considered in DEA models, i.e., $\hat{e}_k = e_k$. Beasley proves that such a result generalizes to the case in which more than two components are considered. The general Beasley's model is

$$
\begin{aligned}
e_0^* &= \max e_0 && (1a)\\
e_0 &\le 1 && (1b)\\
e_k^r &\le 1 \quad \forall k, r && (1c)\\
\sum_r \alpha_i^r &= 1 \quad \forall i && (1d)\\
(\mu^r, \forall r) &\in \Omega_{out} && (1e)\\
(\nu^S, \nu^r, \forall r) &\in \Omega_{in} && (1f)\\
\nu_i^r, \nu_i^S, \alpha_i^r, \mu_j^r &\ge \varepsilon \quad \forall i, j, r, && (1g)
\end{aligned}
$$

where $\varepsilon > 0$ is the usual non-Archimedean constant, and the sets Ω_{in} and Ω_{out} are *assurance regions* as defined in [12]. Differently form the classical DEA models, Beasley includes conditions (1c) imposing that even the cross efficiency [13] of each DMU components cannot exceed 1. Moreover, Beasley points out that it may turn useful to incorporate the additional constraints (1f) and (1e) involving value judgments concerning the proportions α^r and the weights μ^r and ν^r of the different DMU components. These constraints are not strictly necessary for the definition of a multicomponent DEA model, but prevent the model from yielding unreasonable results. In this contest, the author provides an example where, in absence of constraints (1f)-(1e), one research postgraduate was worth about 880,000 undergraduates for a given department.

As for the classical DEA models, (1) can be rewritten as follows

$$
\begin{aligned}
e_0^* &= \max \sum_r \mu^r Y_0^r && (2a)\\
\sum_r \nu^r X_0^r + \sum_r \nu^S(\alpha^r X_0^S) &= 1 && (2b)\\
\mu^r Y_k^r &\le \nu^r X_k^r + \nu^S(\alpha^r X_k^S) \quad \forall k, r && (2c)\\
\sum_r \alpha_i^r &= 1 \quad \forall i && (2d)\\
(\mu^r, \forall r) &\in \Omega_{out} && (2e)\\
(\nu^S, \nu^r, \forall r) &\in \Omega_{in} && (2f)\\
\nu_i^r, \nu_i^S, \alpha_i^r, \mu_j^r &\ge \varepsilon \quad \forall i, j, r. && (2g)
\end{aligned}
$$

Even provided that the assurance regions are expressed in terms of linear constraints, model (2) is not a linear programming problem. More precisely, inequalities (2b)-(2c) include the rectangular terms involving variables ν_i^S, α_i^r. Below it is shown how different authors have proposed possible ways of overcoming such a difficulty.

Many authors have introduced variants of model (2) either changing some of the assumptions made by Beasley or generalizing them. A common feature of the different variants is that in all of them the aggregate efficiency of the generic DMU cannot exceed the unity and a DMU is efficient if and only if it is efficient in all its components.

Molinero and Tsai, in [8, 14], propose an approach dual to (1). In addition, the authors include *shared outputs*, i.e., outputs yielded synergically by two or more components. Their output oriented model can be written as

$$e_0^* = \max \sum_r q_0^r \theta_0^r \tag{3a}$$
$$\sum_k \lambda_k^r x_{ik}^r \leq x_{i0}^r \quad \forall i, r \tag{3b}$$
$$\sum_k \sum_r \lambda_k^r (\alpha_i^r x_{ik}^S) \leq x_{i0}^S \quad \forall i \tag{3c}$$
$$\sum_k \lambda_k^r y_{jk}^r \geq \theta_0^r y_{j0}^r \quad \forall j, r \tag{3d}$$
$$\sum_k \sum_r \lambda_k^r (\beta_j^r y_{jk}^S) \geq \sum_r \theta_0^r (\beta_j^r y_{j0}^S) \quad \forall i \tag{3e}$$
$$\sum_r \alpha_i^r = 1 \quad \forall i \tag{3f}$$
$$\sum_r \beta_j^r = 1 \quad \forall j \tag{3g}$$
$$\sum_r q_0^r = 1 \tag{3h}$$
$$\lambda_k^r, q_0^r, \alpha_i^r, \beta_j^r \geq \varepsilon \quad \forall i, j, r, k, \tag{3i}$$

where y_{jk}^S are the shared outputs of DMU k, β_j^r are the proportions of the shared outputs attributed to component r by DMU k, q_0^r are positive numbers representing the relative importance of each component r for DMU 0, and θ_0^r are measures of the inefficiencies of the DMU 0 components. Actually, θ_0^r are the reciprocals of the *distance functions* [15] from the frontier of the production set defined by the conic combination of the components of the observed DMUs in the hypothesis that the assumptions on free disposal and minimum extrapolation hold [16].

Notice that the values β_j^r, differently from α_i^r, should not be seen as the proportion of the amount of output j yielded by the component r. Actually, no component can produce a shared output by itself but needs synergy with other components. Instead, β_j^r can be seen as the proportion of the (virtual) value of output j that can be attributed to component r.

If the values α_i^r, β_j^r, and θ_0^r are hold as constants, model (3) is a linear programming problem and its dual can be determined. The model obtained in this way is the output oriented version of (2). However, two main differences occur. The overall weighted sum of the outputs of the component r of the generic DMU k becomes now

$$\mu^r Y_k^r + \mu^S(\beta^r Y_k^S), \tag{4}$$

where $Y^S = \{y_{jk}^S\}$ and $\beta^r = \{\beta_j^r\}$. Also, additional constraints on the outputs of DMU 0 and the numbers representing on the relative importance of each component are

present

$$\mu^r Y_0^r + \mu^S(\beta^r Y_0^S) = q_0^r \quad \forall r \tag{5a}$$
$$\sum_r q_0^r = 1 \tag{5b}$$
$$q_0^r \geq \varepsilon \quad \forall r. \tag{5c}$$

Conditions (5) are particularly interesting. They mathematically state a precise relationships between the relative importance attributed to a component and the optimal amount of outputs allocated to it (respectively, the optimal amount of allocated inputs in an input oriented version of the model as (2)). Then, conditions (5) give a mathematical justification to the choice in [10] of expressing the weight q_k^r of the generic component r. As already mentioned, q_k^r is fixed equal to the proportion of the total inputs component r receives. Without (5), such a choice might appear arbitrary, although reasonable.

Molinero and Tsai, in [8], prove that the feasible solutions of (3) define a convex set and (3a) is a convex function. In [9, 17], considering university departments as a reference example, they introduce and discuss a variable returns to scale version of (3). The efficiency of each component r of DMU k is then defined as

$$e_k^r = \frac{\mu^r Y_k^r + \mu^S(\beta^r Y_k^S)}{\nu^r X_k^r + \nu^S(\alpha^r X_k^S) + \delta_k^r} \tag{6}$$

where the variable δ_k^r is unrestricted and its optimal value defines the components returns to scale status. Depending on δ_k^r being negative, null, or positive, a component homogeneous to r, but with proportionally smaller inputs and outputs, results to be more, equally, or less efficient than component r, hence component r is in an area of decreasing, constant, or increasing returns to scale, respectively. The aggregate efficiency of DMU k is

$$e^r = \frac{\sum_r \mu^r Y_k^r + \sum_r \mu^S(\beta^r Y_k^S)}{\sum_r \nu^r X_k^r + \sum_r \nu^S(\alpha^r X_k^S) + \sum_r \delta_k^r}. \tag{7}$$

Note that the optimal value of $\sum_r \delta_k^r$ may be zero even if some or all elements in the sum are different form zero. In this case, as Tsai and Molinero point out, DMU k may appear to be operating under constant returns to scale and technically efficient when analyzed as a single activity DMU, but when its individual activities are analyzed it may be found that the DMU is scale inefficient in each activity [9]. Tsai and Molinero stress that in general a DMU, that turns efficient when considered as performing a single activity, may result inefficient when its different components are taken into account, independently of its returns to scale status.

Tsai and Molinero, in [9], introduce the performance evaluation of National Health Service (NHS) trusts in England as a case study. They analyze the trusts from two different points of view. The so called system perspective considers the trusts as single activity DMUs. The so called trust perspective considers the trusts as mutlticomponent DMUs. The system perspective is the one of a general authority, such as the Department of Health, supervising the trusts. On the other hand, the second perspective could be the one of an inefficient trust interested in determining which are its more efficient components. Sixteen of the considered trusts are system efficient, but only one of them results efficient from a trust perspective.

In [18], Diez-Ticio and Mancebon report and comment in detail an application of a multicomponent DEA model to evaluate the efficiency of Spanish Police Service. The authors use the variable returns to scale model proposed in [17]. In [19], Arcelus and Coleman introduce a DEA application to review the efficiency of the departments of the University of New Brunswick. The authors use classical DEA models, but cite the models in [10, 14] as possible ways to overcome the dichotomy observed between the scientific-technical departments and the remaining ones. In particular, the authors mention the probable existence of different production functions per DMU.

Cook *et al.*, in [7], allow a same shared input i to be weighted differently by the components of the same DMU. The rationale behind such a choice is that different components may disagree on the importance of a same input. Consequently, the multicomponent model as in [7] includes a set of vectors ν^{Sr}, one for each component r, instead of a single one in (2b)-(2c). The rectangular terms are no more $\nu_i^S \alpha_i^r$ but $\nu_i^{Sr} \alpha_i^r$. Also, a change of variables is proposed. In particular, let $i = 1, \ldots, s$ be the index of the shared inputs, then $\bar{\nu}_i^{Sr} = \nu_i^{Sr} \alpha_i^r$ for $i = 1, \ldots, s-1$ and $\bar{\nu}_s^{Sr} = \nu_s^{Sr}(1 - \sum_{i=1}^{s-1} \alpha_i^r)$. Thanks to the new variables, Cook *et al.* obtain a linear model, when constraints (2e)-(2f) are disregarded. The rectangular terms $\nu_r^S(\alpha^r X_k^S)$ in the conditions (2b)-(2c) become $\bar{\nu}^{Sr} X_k^S$ and $\nu_r^S \geq \varepsilon$ in (2g) turns $\bar{\nu}_i^{Sr} \geq \varepsilon \alpha_i^r$. Unfortunately, non linearity may arise again when additional constraints concerning value judgements as (2e)-(2f) are necessary. If such judgements are expressed also in terms of ν_r^S, the variable substitution may not lead to a linear model. In fact, in [7], the authors present a case study on 20 Canadian banks branches. They take into account two components for each DMU, the first one related to service specific activities, the second one related to product specific sales activities. They obtain (and solve) a non linear model, since they impose ratio constraints on the weights of the shared inputs of type $\nu_{i_1}^{Sr} \geq a \nu_{i_2}^{Sr}$.

Also Cook and Hababou deal with Canadian banks branches in [20]. The model proposed presents variables and constraints as in [7] but it is an additive one. The authors discuss how to formulate an additive objective function that represents an aggregare measure the efficiencies of all the DMU components. In the classical additive DEA models, a possible measure of inefficiency for a generic DMU k is given by the difference between the weighted sum of the inputs minus the weighted outputs of DMU k. Then the authors suggest a multiobjective approach where the partial inefficiencies of all the components are considered. For each component, the weighted sum of its inputs minus the weighted sum of its outputs is considered. In particular, Cook and Hababou minimize the maximum partial inefficiency in order to give equal importance to each component, i.e., their objective function is

$$\min \max\{\nu^r X_k^r + \nu^{Sr}(\alpha^r X_k^S) - \mu^r Y_k^r : \forall r \text{ component of } \mathrm{DMU}_0\}. \quad (8)$$

Finally, the authors linearize their model with the same variable changes proposed in [7]. The results concerning the Canadian banking industry in [7, 20] are cited in [21, 22, 23].

Cook and Green in [24] deal with a manufacturing multi-plant company. They point out that in such a context some outputs of different components of a same DMU can partially overlap, i.e., some outputs may be common to different components. Note that the overlapping outputs are different form the shared outputs considered in [8, 14, 25]. In [24] any component can yield a given amount of each overlapping output j, with no need of synergy with the other components and with no possibility of attributing the

considered amount to other components. Then, Cook and Green cannot approach what they call the *overlap problem* by introducing variables β_r as in (3) to determine which proportions of shared outputs are attributed to each component. In [24], the efficiency of a single component r remains $e_k^r = \frac{\mu^r Y_k^r}{\nu^r X_k^r + \nu^{Sr}(\alpha^r X_k^S)}$ and, consequently, the aggregate efficiency of a whole DMU k remains $e_k = \frac{\sum_r \mu^r Y_k^r}{\sum_r \nu^r X_k^r + \sum_r \nu^{Sr}(\alpha^r X_k^S)}$. However, in [24], shared inputs are no more allocated to the components, as such task could be hardly performed without introducing some ambiguities due to the component overlapping. Shared inputs are allocated directly to the outputs. In particular, consider model (1) and the extension proposed in [7]. Cook and Green introduce a new set of variables α_i^j as the proportions of the shared inputs i allocated for outputs j. In addition, they replace condition (1d) with $\sum_j \alpha_i^j = 1$, for all i. Finally, they define α_i^r as $\alpha_i^r = \sum_{i \in O^r} \alpha_i^j$, where O^r is the set of shared outputs of component r. Note that now in general $\sum_j \alpha_i^r \geq 1$.

In [24], the authors also address the problem of determining in which areas a generic DMU k would perform better. Such areas form the DMU k *core business*. In a perspective of a multi-plant company general manager, the core business areas of the different DMUs should be privileged even at the cost of possibly forcing some DMUs to abandon the components with less satisfactory performances. With this aim, Cook and Green further modify the original model (1). In this case, they also introduce as objective function $\hat{e}_k = \frac{\sum_r d_k^r \mu^r Y_k^r}{\sum_r d_k^r \nu^r X_k^r + \sum_r d_k^r \nu^{Sr}(\alpha^r X_k^S)}$, where d_k^r are binary variables which assume value 1 if component r is attributed to DMU k, 0 otherwise. When such an objective function is considered, a DMU is assigned only its most efficient components. Some constraints are also added to the multicomponent model. In fact, each DMU must have assigned at least a component and each component must be assigned to at least a DMU.

Jahanshahloo *et al.*, in [25], extend the model proposed in [7]. They introduce shared outputs and consider panel data. The same way [25] relates to [7] as [8] relates to [10]. Also, Jahanshahloo *et al.* prove that the aggregate efficiency of the whole DMU is a convex combination of the efficiency of the DMU components even in presence of shared output. In [26], the same authors further extend their model introducing non-discretionary inputs as something different from the normal inputs. The former inputs are not under decision making control, then they are considered as negative terms in the numerator of the fraction that describe a DMU efficiency value. The efficiency of each component r of a generic DMU k is then defined as

$$e_k^r = \frac{\mu^r Y_k^r + \mu^{Sr}(\beta^r Y_k^S) - \rho^{Nr}(\gamma^r X_k^N)}{\nu^r X_k^r + \nu^{Sr}(\alpha^r X_k^S)}, \quad (9)$$

where $X^N = \{x_{ik}^N\}$ is the vector of the non-discretionary inputs, $\gamma^r = \{\gamma_i^r\}$ is the vector of the proportions of the non-discretionary inputs allocated to component r by DMU k, and $\rho^{Nr} = \{\rho_i^{Nr}\}$ is the vector of the weights of the non-discretionary inputs. Weights may differ in the different components of a same DMU. Jahanshahloo *et al.* linearize the models proposed in [25, 26] changing variables as in [7]. The case study in [25, 26] deals with Iranian commercial bank branches. Panel data are considered to measure possible progress and regress (see [27]) of the bank sectors.

References

[1] A. Charnes, W. Cooper, E. Rhodes, Measuring the efficiency of decision making units, European Journal of Operational Research 2 (6) (1978) 429–444.

[2] W. Cooper, L. Seiford, J. Zhu, Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston, Ma, 2004.

[3] W. Cooper, L. Seiford, J. Zhu, Data envelopment analysis: History, models and interpretations, in: W. Cooper, L. Seiford, J. Zhu (Eds.), Handbook on Data Envelopment Analysis, Kluwer Academic Publishers, Boston, 2004, Ch. 1, pp. 1–39.

[4] W. Cooper, L. Seiford, K. Tone, Data Envelopment Analysis, Kluwer Academic Publishers, Boston, Ma, 2000.

[5] Y. Yang, B. Ma, M. Koike, Efficiency–measuring DEA model for production system with k independent subsystems, Journal of the Operations Research Society of Japan 43 (3) (2000) 343–354.

[6] L. Castelli, R. Pesenti, W. Ukovich, DEA–like models for the efficiency evaluation of hierarchically structured units, European Journal of Operational Research 154 (2) (2004) 465–476.

[7] W. Cook, M. Hababou, H. Tuenter, Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches, Journal of Productivity Analysis 14 (3) (2000) 209–224.

[8] C. Molinero, P. Tsai, Some mathematical properties of a DEA model for the joint determination of efficiencies, Journal of the Operations Research Society 48 (1) (1997) 51–56.

[9] P. Tsai, C. Molinero, A variable returns to scale data envelopment analysis model for the joint determination of efficiencies with an example of the UK health service, European Journal of Operational Research 141 (1) (2002) 21–38.

[10] J. Beasley, Determining teaching and research efficiencies, Journal of the Operational Reseach Society 46 (4) (1995) 441–452.

[11] W. Cook, Y. Roll, Partial efficiencies in data envelopment analysis, Socio–Economic Planning Sciences 37 (3) (1993) 171–179.

[12] R. Thompson, L. Langemeier, C. Lee, E. Lee, R. Thrall, The role of multiplier bounds in efficiency analysis with application to Kansas farming, Journal of Econometrics 46 (1–2) (1990) 93–108.

[13] J. Doyle, R. Green, Efficiency and cross efficiency in DEA: Derivations, meanings and uses, Journal of the Operations Research Society 45 (5) (1994) 567–578.

[14] C. Molinero, On the joint determination of efficiencies in a data envelopment analysis context, Journal of the Operations Research Society 47 (10) (1996) 1273–1279.

[15] R. Shephard, Theory of Cost and Production Frontier, Princeton University Press, Princeton, NJ, 1970.

[16] R. Banker, A. Charnes, W. Cooper, Some models for estimating technical and scale inefficiencies in DEA, Management Science 30 (9) (1984) 1078–1092.

[17] P. Tsai, C. Molinero, The joint determination of efficiencies in DEA: An application to the UK health service, Department of Management Discussion Paper, University of Southampton .

[18] A. Diez-Ticio, M. Mancebon, The efficiency of the Spanish police service: an application of the multiactivity DEA model, Applied Economics 34 (3) (2002) 351–362.

[19] F. Arcelus, D. Coleman, An efficiency review of university departments, International Journal of System Sciences 28 (7) (1997) 721–729.

[20] W. Cook, M. Hababou, Sales performance measurement in bank branches, Omega–International Journal of Management Science 29 (4) (2001) 299–307.

[21] K. Bala, W. Cook, Performance measurement with classification information: an enhanced additive DEA model, Omega–International Journal of Management Science 31 (6) (2003) 439–450.

[22] M. Asmild, J. Paradi, V. Aggarwall, C. Schaffnit, Combining dea window analysis with the malmquist index approach in a study of the Canadian banking industry, Journal of Productivity Analysis 21 (1) (2004) 67–89.

[23] J. Paradi, C. Schaffnit, Commercial branch performance evaluation and results communication in a Canadian bank: a DEA application, European Journal of Operational Research 156 (3) (2004) 719–735.

[24] W. Cook, R. Green, Multicomponent efficiency measurement and core business identification in multiplant firms: A DEA model, European Journal of Operational Research 157 (3) (2004) 540–551.

[25] G. Jahanshahloo, A. Amirteimoori, S. Kordrostami, Multi–component performance, progress and regress measurement and shared inputs and outputs in DEA for panel data: an application in commercial bank branches, Applied Mathematics and Computation 151 (1) (2004) 1–16.

[26] G. Jahanshahloo, A. Amirteimoori, S. Kordrostami, Measuring the multi–component efficiency with shared inputs and outputs in data envelopment analysis, Applied Mathematics and Computation 155 (1) (2004) 283–293.

[27] H. Tulkens, P. Eeckaut, Non–parametric efficiency progress and regress measures for panel data: methodological aspects, European Journal of Operational Research 80 (1995) 474–499.

Section 1:
Scheduling and Control

OPTIMISATION OF SCHEDULING: A CASE OF INDUSTRIAL CHAIR MANUFACTURING

Janez Abram
Novoles d.d., Na žago 6, 8351 Straža, Slovenija
Janez.Abram@novoles.si

Lidija Zadnik Stirn, Mirko Tratnik, Leon Oblak
Univerza v Ljubljani, Biotehniška fakulteta
Jamnikarjeva 101, 1000 Ljubljana
lidija.zadnik@bf.uni-lj.si
mirko.tratnik@bf.uni-lj.si
leon.oblak@bf.uni-lj.si

Abstract

Industrial chair manufacturing can be looked at as a complex production system requiring organisation and technical skills. Heller-Logemann algorithm was tested for better surveying and organisation of production. The production times for each operation were time co-ordinated. A genetic algorithm based computer programme to optimise manufacturing cycle times was developed. The cycle times for the selected products were reduced and forecast information for leading production were obtained.

Keywords: *scheduling, Heller-Logemann algorithm, genetic algorithm, chair manufacturing*

1. INTRODUCTION

Because of the market demands, gone are the good old days when companies held high stocks of raw materials and use inflexible production planning techniques based on stock control and long range forecasting. Companies have to offer more options in products with greatly reduced life cycle and product complexity. All this has had a great effect on the production organisation and the manner of scheduling and launching products into the manufacturing process.

Production scheduling is the activity performed in manufacturing companies to manage and control the execution of the production process. The basic task is to perform the production as planned while at the same time trying to satisfy the overall goals of the company. Production scheduling involves master production scheduling (MPS), materials requirements planning and shop floor scheduling. Each of these levels of scheduling impact on the others.

There exists much interest from industry in using software systems to support the scheduling process but the application of such systems has shown to be problematic. This paper reviews area of production scheduling and outlines the Heller-Logemann and genetic algorithm implemented in complex production system.

2. SCHEDULING METHODS

The scheduling approach and methods that are suitable for a production environment depend on the characteristics of the environment, the complexity, uncertainties and randomness of the production system, the scheduling objectives and the organisation around the scheduling function. Scheduling methods can range from simple rules for choosing which job to execute next often called dispatching rules or individual heuristic approaches, to sophisticated optimising methods [6].

For solving the problems of individual product operation schedule on an individual machine, where scheduling is performed on a fixed set of orders, there exist many methods. For the time scale planning, we can use the method of linear programming, dispatching rules, the Johnson algorithm, the method branch and bound... For dynamic scheduling where new orders are continuously added during scheduling, we should use other method such as the Heller-Logemann algorithm or heuristic priority rules [2]. It is up to the production characteristics as to which method to use.

The most basic scheduling method is to use dispatching rules (also called priority sequencing rules) to determine which order to run next at a work centre. These rules are applied when jobs arrive at a work centre or when fixed set of orders are planed at a work centre, to choose the next task to be executed. Since dispatching rules only use information that is available at the moment when the next activity shall be selected, they work equally well in systems with a high degree of uncertainty as in more stable environments. When there are high levels of randomness and uncertainty in the production environment, dispatching rules may be the only viable way to schedule the production. There exist many dispatching rules, some of the most common are [6]:

- *First come, first served (FCFS).* Jobs are processed in the order they arrive at the work centre.
- *Shortest processing time (SPT).* The job with the shortest processing time is processed first.
- *Earliest due date (EDD).* The job with the earliest due date is processed first.
- *Critical ratio (CR).* A priority index is calculated using (time remaining/work remaining). A ratio less than 1 means that the job is late. The job with the lowest ratio is processed first.
- *Least work remaining (LWR).* Priority based on all processing time remaining until job is completed.
- *Fewest operations remaining (FOR).* Priority based on number of remaining operations.
- *Slack time (ST).* Jobs run in the order of the smallest amount of slack.
- *Slack time per operation (ST/O).* Slack time is divided by the number of remaining operations. Jobs are sequenced in order of smallest value.
- *Next queue (NQ).* The queues in front of successive work centres are measured (in hours or number of jobs). The job that is going to the smallest queue is processed first.
- *Least set-up (LS).* The job with the least set-up time is processed first.

The general properties of these rules are different. SPT, and its variations LWR and FOR, reduces work in process inventory, average job completion time and average job lateness but can cause starvation of jobs with long processing times and thus cause missed due date. EDD, and its variations ST and ST/O, reduce job lateness but result in higher average time in the system. NQ and LSU maximise machine utilisation. There exist many other dispatching rules and also variations of the above rules. To combine rules, for instance using different rules for different work centres is also possible. Scheduling using these rules can, depending on the scheduling problem, give good results but there is a risk of sub-optimisation since the information used is local and no consideration is given to the global state of the production system.

Johnson's algorithms, have been developed for single machine, two or three machine and job-shop scheduling problems. These algorithms could be used in the multi-machine set-up in the circumstance that one of the machines is a bottleneck to production for example, or where two machines are very closely associated based on product routings. The algorithms could then be applied to the appropriate machine and the rest of the schedule developed around this core by forward and backward scheduling [5].

In a flowshop scheduling problem there is a set of n jobs to be processed in a set of m machines in the same order. First in machine 1 then on machine 2 and so on until machine m. The objective is to find a sequence for the processing of the jobs in the machines so that a given criterion is optimised [10]. In the chair manufacturing process there is a set of n jobs to be processed in a different set of machines in the different order. Assembling of a product influences the manufacturing and the choice of algorithm (e.g. we cannot assemble the final chair joint if we do not have all joint elements, which is typical of connected working assignments). The scheduling problem is solvable with the Heller-Logemann algorithm in which we should also consider a shifts utilisation.

The Heller-Logemann algorithm is a derivative of network planning, allowing scheduling of parallel network plans where operations are dealt with as knots in a network, the relations among them or the technological sequence of operations is shown with arrows. All other characteristics of a classical network planning are true also for the Heller-Logemann algorithm. The algorithm not only considers the production time but also the intertwining of operations and inter-operational deadlocks and allows scheduling ahead (running to the right) and backwards (running to the left) [2].

In future account ahead (running to the right) schedules are defined from the anticipated starting deadlines of work orders onwards. What is formed are tables with operations needing to be performed in a term unit. The table is then arranged according to codes of working places in the way that all the operations performed in the same work place are joined in groups, irrespective of the work order they appear in. The calculation of the initial and the final term of operation is performed in many transitions through the table of operations. When operations performed in the same work place at the same time, external - costumer and internal - scheduling priority rules should be considered. Internal priority rules are set by assemble technological characteristics of products (figure 1) and basic scheduling methods like first come, first served or shortest processing time. External priority rules influence on completion time of the ordered products. With a job permutation we could search a schedule where a given criterion is optimised.

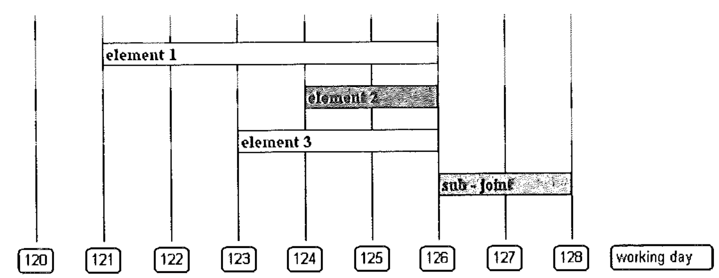


Figure 1. Influence of assembling of sub-joint on scheduling process

Elements must be produced before assembling into sub-joint. Element 1 is processed first and have highest priority because completion time of elements 2 and 3 is shorter. Critical path of assembling product determines the completion time and include element 1 and sub-joint.

2.1 Genetic algorithms for scheduling

Genetic algorithms are a general methodology for searching in a discrete solution space in a way that is similar to process of natural selection procedure in biological systems. The algorithm can be applied to different problems [7].

Genetic algorithms have many variants, many of the basic ideas exist, and individual applications may be highly different. Standard genetic algorithms at random or heuristically generate an initial population of individuals (chromosomes). At every evolutionary step, known as a generation, the individuals in the current population are evaluated according to some predefined quality criterion, referred to as the "fitness" function. To form a new population (the next generation), individuals are selected according to their fitness. Then some or all of the existing members of the current solution pool is replaced with the newly created members. Creation of new members are done by crossover and mutation operations [9]. Genetic algorithm strings encoding the solutions are often binary coded. In our case a chromosome is represented as a list of ordered products [1], whose production should be scheduled according to prescribed criteria and restrictions in the production process.

The genetic algorithm object determines which individuals should survive, which should reproduce, and which should die. It also records statistics and decides how long the evolution should continue. Typically a genetic algorithm has no obvious stopping criterion. One must tell the algorithm when to stop. Often the number of generations is used as a stopping measure, but goodness-of-best-solution, convergence-of-population, or any problem-specific criterion can be used [8].

When creating new population by crossover and mutation, we have a big chance that we will loose the best chromosome. Elitism is a method, which first copies the best chromosome (or a few best chromosomes) to new population. The rest is done in classical way. Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution [9].

The mutation operator defines the procedure for mutating each genome. Mutation means different things for different data types. For example, a typical mutator for a binary string genome flips the bits in the string with a given probability. A typical mutator for a tree, on the other hand, would swap sub trees with a given probability. Mutation should be able to introduce new genetic material as well as modify existing material [8].

The crossover operator defines the procedure for generating a child from two parent genomes. The standard crossover operator called simple crossover has numerous variants such as partially-mapped, position-based, order-based, sub tour chunking, cyclic, acyclic, inversion, and edge-recombination crossovers. All of these involve two parents [9]. Better results have been obtained by rejecting the conventional binary representations and using more direct encoding (a schedule genotype is a list specifying the order). But simple crossover applied to such strings would nearly always result in illegal offspring with some orders missing, while other orders presented twice. Hence more sophisticated crossover operations are needed. In our scheduling problem we used operator called liner order crossover which was suggested by

some authors, such as Bernik [3]. Crossover is not usually applied to all pairs of individuals selected for mating. A random choice is made, where the likelihood of crossover being applied, it is typically between 0.6 and 1.0. If the crossover is not applied, offspring are produced simply by duplicating the parents.

Fitness function is responsible for evaluation of quality of given code string. In the considered cases the cycle time or completion time of the last job is a value, which has to be minimised. This time is calculated for specified number of jobs, consisting of a large number of operations.

3. RESULTS

In the "hand made" schedule different dispatching or priority sequencing rules were used to determine which order to run next at a work machine, where utilisation level is the highest. Because of the complex production scheduling without computer support, we focus on bottlenecks in the production system. The bottleneck resource determines the throughput of the whole system and therefore its utilisation should be maximised. A buffer is put before the bottleneck resource to ensure that it never has to wait for jobs to execute. Resource-based schedule requires that there exist just a few bottleneck resources and that they should be well defined. In reality happens that the bottleneck resource shifts over time and thus, there appear problems.

Comparing the actual schedule of products in the manufacturing process with optimal computer schedule, we can observe that in the latter results are better. The lowest cycle times for individual products are achieved in the schedule found with the genetic algorithm. By looking for the optimal schedule, the average cycle time for manufacturing products compared to a planned schedule determined by a monthly plan is shortened by up to 30 % [4].

The biggest savings in cycle time of operation in manufacturing chairs is achieved by rearranging certain working operations to optimal terms, considering the utilisation of individual, i.e., key working places. Success in searching for good schedule depends also on the type of product being simultaneously manufactured. Similar products produced in the same working places in a similar sequence of operations burden only certain machines, which therefore become overloaded. By planning mutually different products, we can partly reallocate the burdening of production to different working places.

The results of the application are shown in Gantt charts and graphic displays of the capacity burden of the machines and workers.

4. CONCLUSIONS

Due to the fact that the sequence of products influences the cycle time, we tried to find a better schedule with the help of Heller-Logemann algorithm and stochastic techniques such as genetic algorithms. Comparing the "hand made – manual" schedule with optimal computer schedule, we can observe that in the latter results are better.

The next step is modified Heller-Logemann algorithm to consider machine and worker utilisation. With further comparison of various scheduling methods and quality function we could achieve better production results. The planned deadlines of work operations vary due to the breakdown of machines, the variable quality of input raw material and the cancellation of orders. Therefore, the organisation model should be supplemented with the reallocation of activities in the manufacturing process.

REFERENCES

[1] BRESKVAR, U., KLJAJIĆ, M.: Prednosti planiranja proizvodnje s simulacijo in genetskimi algoritmi, Organizacija, 2001, letn. 34, št. 10, str. 670 – 675

[2] LJUBIČ, T.: Planiranje in vodenje proizvodnje, Moderna organizacija, Kranj, 2000, 443 str.

[3] BERNIK, I.: Večkriterijsko razporejanje z genetskimi algoritmi in vizuelno simulacijo s Petri mrežami, Doktorska disertacija, Kranj, FOV, 2001, 75 str.

[4] ABRAM, J.: Optimizacija terminskega in kapacitetnega načrtovanja pri proizvodnji stolov, Magistrsko delo, Ljubljana, BF, 2005, 82 str.

[5] TOAL, D., COFFEY, T., SMITH, P.: Expert systems and simulation in scheduling, Department of Electronics and Computer Engineering, University of Limerick, http://www.ul.ie/~toald/Publications/Imc-11es.pdf, 8 str.

[6] KJELL, O.: An investigation into production scheduling systems, Doctoral dissertation, Karlstad University, 2004, 108 str.

[7] ŽDÁNSKÝ, M., POŽIVIL, J.: Combination genetic/tabu search algorithm for hybride flowshop optimization, Conference on Scientific Computing. Proceedings of ALGORITMY, 2002, str. 230 – 236

[8] MATTHEW, L.: GAlib:A C++ Library of Genetic Algorithm Components, Mechanical Engineering Department, Massachusetts Institute of Tecnology , http://lancet.mit.edu/ga/galib-request@mit.edu, 1996, 101 str.

[9] SANDIKCI, B.: Genetic Algorithms, Department of Industrial Engineering, Bilkent University, http://www.bilkent.edu.tr/~burhan, 2000, 9 str.

[10] RUBÉN, R., CONCEPCIÓN,M.: A comprehensive review and evaluation of permutation flowshop heuristics, European Journal of Operational Research, 165 (2005), str. 479 – 494

LOCAL OPTIMIZATION APPLIED TO A VOLUNTEER TIMETABLING PROBLEM

Matej Borovinšek[1], Janez Žerovnik[1,2]

[1] University of Maribor, Faculty of Mechanical Engineering, Smetanova 17, 2000 Maribor, Slovenija

[2] IMFM/DTCS, Jadranska 19, 1000 Ljubljana, Slovenija

Emails: matej.borovinsek@uni-mb.si, janez.zerovnik@imfm.uni-lj.si

Abstract

The time table assembly is problem in which the feasibility of solutions involves solution of a graph coloring problem while the quality of solution is obtained by optimization of a fitness function. In this paper the time table assembly was done using the method of local optimization. For this purpose a special fitness function was developed with regard to the soft constraints of the particular problem.

Keywords: time table, local optimization

1. Introduction

It is well known that even in its simplest form timetabling is NP-hard. As, on the other hand, there are many practical timetabling and scheduling problems, there is a vast literature on this topic, and many forms of the problems have been studied including course and school timetabling, examination timetabling, sports timetabling, employee timetabling, etc. For references, see the survey [Schaerf, 1999] and a web page [Trick].

The Employee Timetabling Problem consists of a periodical assignment of employees to tasks in shifts with fixed start and end times. A simple example of a real world problem is that of timetabling nurses in a department of a large hospital. Formally, there are *m employees*, *n shifts*, and *t tasks*. We search for an assignment, which can be defined as a 3-dimensional binary matrix X, such that $X(i,j,k)=1$ if employee i is assigned to task j in shift k. Any solution must fulfill additional constraints, which can be briefly given as follows: *Requirements*: Each shift is composed of a number of tasks, some of them multiple times. An employee is needed to be assigned for each task belonging a shift. *Ability*: Each employee has qualifications that enable her/him to fulfill certain types of tasks: that is, each employee has a set of tasks that can be assigned to. *Availability*: There are personal preferences of employees, which restrict them to be assigned only to subsets of the shifts. *Conflicts*: Obviously, an employee cannot be assigned to more than one task in the same shift. In addition, employees cannot be assigned to two shifts that are in *conflict* with each other. Sources of conflicts could be different: overlap in time, consecutive, or combinations that are forbidden by organizational rules. *Workload*: There is an upper and lower limit on the number of tasks that each employee can be assigned to. There are actually a set of limits, because employees can be assigned to a limited total number of tasks per schedule and also to a limited (smaller) number of specific assignments. For more details, see [Meisels and Schaerf, 2003] and the references there.

A feasible solution is *any* assignment that satisfies all of the above constraints. General definitions include also soft constraints that constitute the fitness function (to be minimized). Soft constraints generally regard fair distribution of loads for employees. For example, when a nurse is assigned to two night shifts, it is considered much better to have them spread evenly over the week (e.g., at least two free nights in between). In employee timetabling, a search problem is sometimes considered, meaning the goal is to find any feasible solution because already a feasible solution may be very difficult to find.

In this paper we consider a problem which may be regarded as an example of **volunteer timetabling.** We use the name because of the particular example which motivates this work and because we believe that for volunteer timetabling, the constraints will typically be easy to satisfy while on the other hand various interpretations of fairness are very important. Clearly, a volunteer that feels that he is treated improperly would easily quit the association. Therefore, it is very important to design a fitness function which will force the solutions to satisfy some fairness requirements. In other words, the formal definition of the volunteer timetabling is the same as for employee timetabling, but in the typical instances of volunteer timetabling the difficult part of the problem is expected to be in the optimization of the fitness function.

The paper is organized as follows. A motivation with the practical problem description is given and a simple procedure for generation of feasible solutions is explained. Then the fitness function is defined, which is used for a local search improvement algorithm. Results are briefly discussed.

3. Problem description

In time of ski season on Pohorje there is a mountain rescue service responsible for the safety of ski tourists. The safety is assured with daily duty turns on all ski trails during day and night skiing. The time table for duty turns has to be assembled.

Day and night duty turns have to be distributed along the members of the mountain rescue service. In respect to the month of the ski season and on the specific day of the week there have to be 4, 6, or 8 duty turns per day.

Table 1: Duty turns division and their symbols

Duty turns per day	Daily duty	Nightly duty turns
4	4 (1, 2 ,3, 4)	/
6	6 (1, 2 ,3, 4, A, B)	/
8	6 (1, 2 ,3, 4, A, B)	2 (N,N)

Daily duty turns take place on two different ski areas of Bellevue and Areh. On the Areh ski area there are four rescuers needed to fulfill the safety regulations (symbols from 1 to 4) while on the Bellevue ski area there are only two rescuers needed (symbols A and B). In case of days with night skiing in the area of Bellevue two additional rescuers are needed (both marked with symbol N).

Each member of the mountain rescue service has to fulfill the norm of ten to twelve duty turns per ski season. Because the time table is assembled only for one following month this means the norm is from three to five duty turns per month. In the beginning of the new month each member may choose days of his duty turns (using symbols 1 - 4, A, B, and N) and days when he will be absent and can not do his duty (symbol X). Finally, wishes of members with longer membership are considered of greater importance then wishes of new members.

4. The algorithm for the generation of a feasible solution

The first solution is obtained using the algorithm shown in figure 1. As mentioned before it considers only the condition about the number of demanded duty turns per day. The number of demanded duty turns can be defined separately for each day or the default values are used. The algorithm first determines what kind of day is selected. It can be the day with more, equal or less entered duty turns per day as demanded.

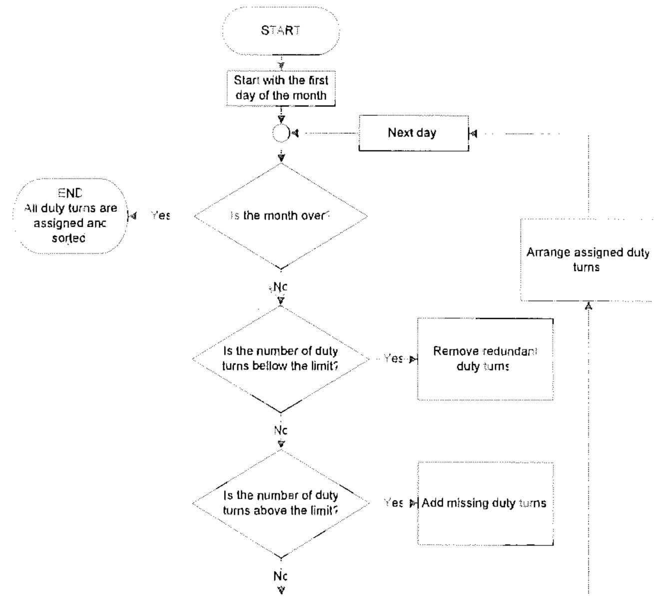


Figure 1: Diagram of the algorithm for the first solution

In case of more entered duty turns per day as demanded it is necessary for redundant duty turns to be removed. Redundant duty turns could be selected randomly but in order for the first solution to be as good as possible the sorted list was used. The list consists only of members whose wish was to do duty on selected day and is sorted using two criteria. The first criterion is the number of monthly duty turns of the member. In case of more than one member with the same number of monthly duty turns they are sorted with regard to their membership time. So when the redundant duty turn is removed it is first removed from the member with the largest number of duty turns per month. If there are more than one such members the duty turn is removed from the member with the shortest membership time. This removal process is then repeated until all redundant duty turns are removed.

In the opposite case the duty turns have to be added. The adding of duty turns is also based on a list sorted similarly as described above but the list consists only of members which do not have a wished duty turn on selected day. The duty turn is first added to the member with the least duty turns per month. The process of adding is repeated until the demanded number of duty turns is reached.

These duty turns have to be subsequently reordered with respect to the ski area and the time of duty (Table 1). This reordering is done using another sorted list which consists only of members whose wish was to perform duty on selected day. In case of two members having the same wish for the selected day the wish of the member with longer membership time is accepted while the other member gets the next available duty turn.

5. The cost function and the optimization algorithm

A cost or fitness function that takes into account the duration of the membership, number of duty turns per month and the number of duty turns changes with respect to expressed wishes. The optimal solution will be the one with the lowest fitness value.

First the evaluation of each mountain rescuer is done by comparison of changes between wished and assigned duty turns. The state of each day in the month is compared and number of removed N_{rem}, added N_{add} and changed N_{cha} duty turns is counted. The changed duty turn represents the change of duty turn area or time. To evaluate these changes three weight parameters were introduced. Parameter b_{rem} to account for removed duty turns, b_{add} to account for added duty turns and b_{cha} for changed duty turns. With those parameters defined the member's fitness s can be written as:

$$s = b_{rem} \cdot N_{rem} + b_{add} \cdot N_{add} + b_{cha} \cdot N_{cha} . \qquad (2.1)$$

Additionally the number of assigned duty turns per month has to be evaluated in respect to prescribed limits. The minimum number of duty turns per month is three while the maximum number of duty turns is five. If N_{dt} is the number of some member's duty turns per month then the difference m between the number of duty turns and it's limits can be written as:

$$m = \begin{cases} 3 - N_{dt}; & N_{dt} \le 3 \\ N_{dt} - 5; & N_{dt} \ge 5 \end{cases} \qquad (2.2)$$

The member's fitness can now be improved considering this difference as:

$$s_p = (b_{rem} \cdot N_{rem} + b_{add} \cdot N_{add} + b_{cha} \cdot N_{cha}) + c^m ,$$
(2.3)

where the c stands for the weight parameter of the difference in duty turns number.

The membership duration is accounted for with the weight parameter a and the position (index) of a member in the membership list. The membership list is a list consisting of all members sorted by their membership duration so that the member with the longest membership is in the first place of the list. The solution fitness value can be calculated as:

$$S = \sum_{i=1}^{N} a^{N-i} \cdot \left[(b_{rem} \cdot N_{i,rem} + b_{add} \cdot N_{i,add} + b_{cha} \cdot N_{i,cha}) + c^m \right], \qquad (2.4)$$

where N is the number of all members and i is an index of a member in the membership list.

Using this function any time table configuration can now be evaluated if the values of weight parameters a, b_{rem}, b_{add}, b_{cha} and c are defined. Based on multiple randomly generated sets of duty turn wishes the values of weight parameters were determined to achieve the convergence as fast as possible. Selected values of weight parameters equal to $a = 1{,}2$, $b_{rem} = 0{,}3$, $b_{add} = 2{,}0$, $b_{cha} = 0{,}1$ and $c = 1000$.

The optimization was done using the method of local optimization. In order to get the optimal solution with this method the existing solution is randomly changed on a small (local) area and these changes are then evaluated using the fitness function. The generated solution is accepted only if it is better than the existing one.

The solution in the described problem is represented by the duty turns of mountain rescue service members. To change such a solution on a local area some members were randomly selected and their duty turn entries completely removed. These duty turns were then arranged back to those members in a random manner. With this process the working algorithm was developed but a lot of generated solutions were necessary to improve the fitness. Because of that a different approach for arranging removed duty turns was developed.

The new approach for arranging removed duty turns included the information about member's wishes. The solution changing was started the same way as before with duty turns removal and then for each day of the month the following procedure was repeated. For all members with duty turns removed it was first checked if any of them expressed a wish for the duty turn on the selected day. In case of existing duty turn wish duty turn was assigned to the members. The membership list was taken into account to give advantage to wishes of members with longer membership time. Lastly the remaining duty turns for the selected day were randomly assigned to the rest of the members. With this approach the increase of quality of generated solutions was achieved (Figure 2).

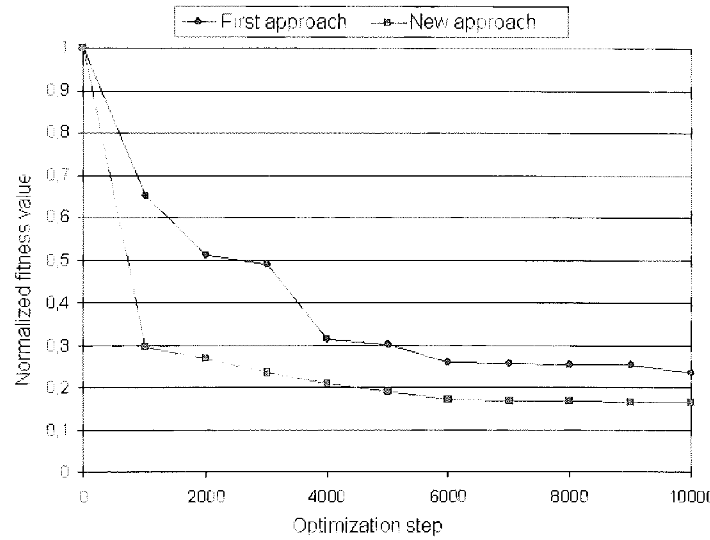


Figure 2: Comparison between the first and improved approach for arranging removed duty turns

The number of members whose duty turns were removed in each optimization step was fixed. The number was based on some optimization time comparison between tests of randomly generated sets of time tables and was set to 10% of all mountain rescue service members. Even faster optimization could be achieved with additional comparisons between optimization times and solution qualities.

6. Results

As a test example the wishes of mountain rescue service members for February were used. The data was entered just for members that expressed their wishes on time. For other members the data was left empty. At the moment there are sixty-seven active members of the mountain rescue service. Because of the high number of expected visitors the number of demanded duty turns per day was set to eight for all days in February.

The first step of the algorithm assures that each day of the month has the demanded number of duty turns. In our case after the first step all days have eight duty turns. But big differences in number of members monthly duty turns are observable. They range from zero to eight duty turns per month and are shown in figure 3. There are thirty-two members with duty turn numbers within limits and the rest of the members is whether under or over the limit. These numbers should be reduced by the optimization step.

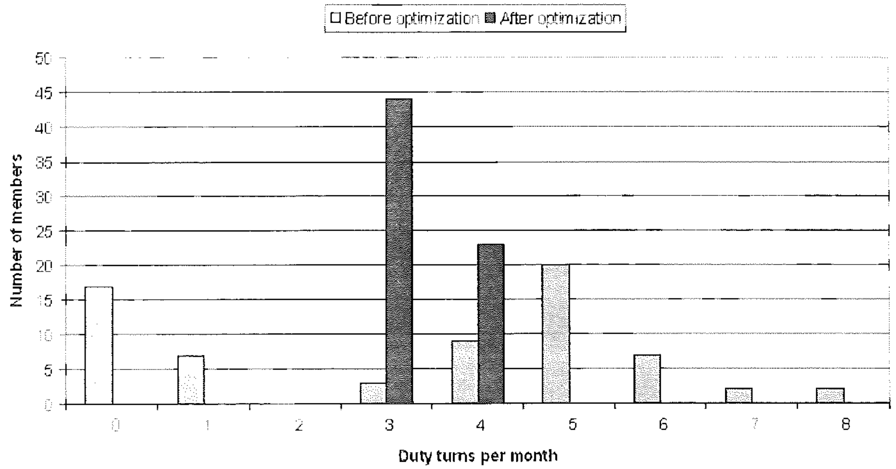


Figure 3: The distribution of the monthly duty turns

To optimize this case approximately 100.000 optimization steps were necessary to fulfill the requirements for the number of duty turns per month of each member. Number of optimization steps is different for each optimization run because of randomness in used optimization method. It can be noticed that the numbers of duty turns of single member per month are all within limits and very evenly distributed. From the final time table configuration high concentration of assigned duty turns is observed by members that entered a larger number of duty turn wishes. This is because the day selection routine in method used for arranging removed duty turns during optimization is not random but straight. The code could be changed to use random day selection but it was not necessary because the reason for entering such large amounts of duty turns was removed with the introduction of computer aided version, namely such large numbers of duty turns were entered only to ease the time table assembly made by hand.

7. The software

As the basis for the computer aided version a Microsoft Office, specifically Microsoft Excel was selected. Excel supports adding and running a code written in software language Visual Basic for Application. Despite the fact that such code runs slower the Excel platform was chosen because of its existing user interface and compatibility. It has a built in interface for work with files, for data entry and for printing. Such solution is included in an Excel file .xls and can be used on every computer with Excel installed.

References

[Schaerf] A.Schaerf, A Survey of Automated Timetabling, Artif. Intell. Rev.13 (1999) 87-127.

[Meisels and Schaerf] A.Meisels and A.Schaerf, Modelling and Solving Employee Timetabling Problems, Annals of Mathematics and Artificial Intellingence 39 (2003) 41-59.

[Trick] Michael Trick's Operations Research Page, http://mat.gsia.cmu.edu/

DYNAMIC PROGRAMMING IN CONTROL

Lado Lenart, Jan Babič
"Jožef Stefan" Institute , Jamova 39, Ljubljana
lado.lenart@ijs.si; jan.babic@ijs.si

Abstract: The use of dynamic programming (DP) in control systems is based on a relation between the Hamilton-Jacoby-Bellman equation (HJBE) and the DP. Without calculating costate variables in canonical HJBE system equations, the iteration algorithm is shown to get solution in minimizing a series of control variables.

Keywords : Hamilton-Jacobi-Bellman equation, optimal control, direct iterative procedure

1.INTRODUCTION

The theoretical basis for the optimal control problem is the Pontryagin's principle. The simplified proof can be found in [1]. The field of non-numerical solutions of the problem is very restricted. Linear systems with quadratic functional already effectively use the analytic solutions of Riccati matrix equation. A broad field of these equation application areas can be found in [2]. The newest development of the general problem theory, connected with HJBE, can be found in [3]. Numerical solutions are often based on two point boundary problems in ordinary differential equations [4] or are presented in the programming environment [5].

Our intention was to prove the behaviour of some algorithms, which work on a direct iteration schema. The method proposed by [6] appeared to be perspective for some problems where the problem of switching surfaces does not appear, or at least is not prevailing. The method is partially modified for the reason of effective programming.

The contribution in the second section presents the connection between DP and HJBE, whereby the connection with optimal principle is systematically avoided. The third, central chapter, shows the possibility to use the direct method of minimizing control sequence for 'smooth' problems.

2. HAMILTON EQUATION BASED ON DYNAMIC PROGRAMMING

In this section, the relation between dynamic programming and optimal control is shown for continuous control systems. The dynamic system is given with (1)

$$\frac{dx_s}{dt} = f_s(x_1,\dots,x_n,u_1,\dots,u_r,t) \qquad (1)$$

or in his vector form (2)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),t) \tag{2}$$

Performance function for terminal time T is :

$$J = h(\mathbf{x}(T),T) + \int_0^T f_0(\mathbf{x}(\tau),\mathbf{u}(\tau),\tau)d\tau \tag{3}$$

Using embedding principle we get:

$$J(\mathbf{x}(t),t,\mathbf{u}(t)) = h(\mathbf{x}(T),T) + \int_t^T f_0(\mathbf{x}(\tau),\mathbf{u}(\tau),\tau)d\tau \tag{4}$$

The construction of the optimal cost functional can be read from fig.1. , formula (5) results (the suffix '*' always means the optimal value - under given restrictions).

$$J^*(\mathbf{x}(t),t) = \min_{\mathbf{u}(\tau);t\le\tau\le t+\Delta t}\left\{ \int_t^{t+\Delta t} f_0 d\tau + J^*(\mathbf{x}(t+\Delta t, t+\Delta t)\right\} \tag{5}$$

Assuming that (5) can be expanded into Taylor series, this is done and after abbreviation follows (6).

$$0 = J_t^*(\mathbf{x}(t),t) + \min_{\mathbf{u}(t)}\left\{ f_0(\mathbf{x}(t),\mathbf{u}(t),t) + \frac{\partial \mathbf{J}^{*T}}{\partial \mathbf{x}}(\mathbf{x}(t),t).[\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),t)]\right\} \tag{6}$$

To find the boundary value for this PDE, at $t=T$, one has

$$J^*(\mathbf{x}(T),T) = h(\mathbf{x}(T),T) \tag{7}$$

If Hamiltonian H is defined as

$$H = f_0(\mathbf{x}(t),\mathbf{u}(t),t) + \frac{\partial \mathbf{J}^{*T}}{\partial \mathbf{x}}(\mathbf{x}(t),t).[\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),t)] \tag{8}$$

(6) is written for optimal control, then one gets the Hamilton-Jacobi-Bellman (HJB) equation in its usual form.

$$0 = J_t^*(\mathbf{x}(t),t) + H(\mathbf{x}(t),\mathbf{u}^*(\mathbf{x}(t),\mathbf{J}_x^*,t),\mathbf{J}_x^*,t) \tag{9}$$

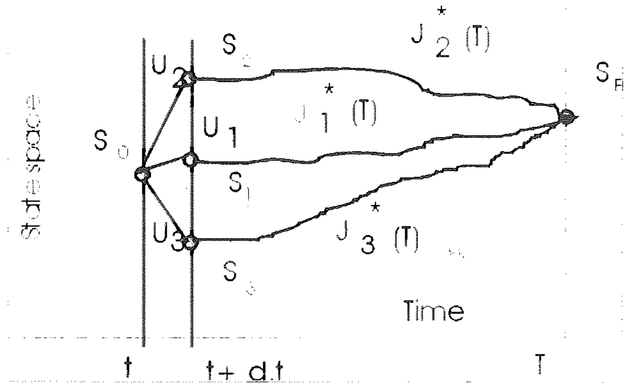


fig.1: Graphic representation of continuous dynamic programming

3. MINIMIZING CONTROL SEQUENCE

Say we simply start with any admissible control $\mathbf{u}_1(\mathbf{x},t)$ transferring the system from state $\mathbf{x}_0$ at time 0 to any state at time T. Then, by definition the sequence $\mathbf{u}_1, \mathbf{u}_2, \ldots$ will be called minimizing sequence, if $J(\mathbf{u}_1), J(\mathbf{u}_2)\ldots$ do not increase and $J(\mathbf{u}_1) \ge J(\mathbf{u}_2)\ldots \ge J(\mathbf{u}^*)$. Now (3) with $\mathbf{u}_1$ is valid and all subsequent equations remain the same with the exception, that minimization cannot be made, as $\mathbf{u}_1$ is not optimal. The most important, dynamic system obeys the HJB equation (9), which shall now be written in the slightly changed form.

$$\frac{\partial V_1}{\partial t} + \sum_{s=1}^{n} \frac{\partial V_1}{\partial x_s} f_s(\mathbf{x},\mathbf{u}_1,t) + f_0 = 0 \tag{10}$$

V_1 in (10) is analogue to J in (6) if V replaces J, the renaming shall make the point, that V_1 is not optimal, but it is built along the trajectory, which is defined with $\mathbf{u}_1$. Now (2) shall run with $\mathbf{u}_1$ and its solution is dependent from initial vector $\mathbf{x}_0$.

$$\mathbf{x} = \mathbf{x}(t,\mathbf{x}_0) \tag{11}$$

From (4) one gets

$$\frac{dV_1}{dt} = -f_0(\mathbf{x},\mathbf{u}_1(t),t) \tag{12}$$

If (12) is integrated, the boundary condition must be again

$$V_1 = h(\mathbf{x}(T,\mathbf{x}_0)) \quad \text{if} \quad t = T \tag{13}$$

It follows for V_1:

$$V_1 = \overline{V}_1(t,\mathbf{x}_0) = h(\mathbf{x}(T,\mathbf{x}_0)) + \int_t^T f_0(\mathbf{x}(\tau,\mathbf{x}_0),\mathbf{u}_1(\tau),\tau)d\tau \tag{14}$$

From (11) one gets the inverse function $\mathbf{x}_0 = \Phi(\mathbf{x},t)$ This function can be inserted into (14) and one gets the identity:

$$V_1(\mathbf{x},t) = \overline{V}_1(t,\Phi(\mathbf{x},t)) = h(\mathbf{x}(T,\Phi(\mathbf{x},t))) + \int_t^T f_0(\mathbf{x}(\tau,\Phi(\mathbf{x},t),\mathbf{u}_1(\tau),\tau)d\tau \quad (15)$$

$\mathbf{x}$ appears as function in (15) and as argument in Φ and on the left side of equation. The recent means, that any solution of (11) can be inserted for $\mathbf{x}$, even with $\mathbf{x}_0$ changed. But, the difference $V_1 - V_1(\mathbf{x},t)$ remains constant for $\mathbf{x}$ as a fixed argument.

Proof:

The solution of (12) is:

$$V_1 = V_1^{(0)} - \int_0^t f_0(\mathbf{x}(\tau,\mathbf{x}_0),\mathbf{u}_1(\tau),\tau)d\tau$$

$$\begin{aligned} V_1 - V_1(\mathbf{x},t) &= V_1^{(0)} - \int_0^t f_0(\mathbf{x}(\tau,\mathbf{x}_0),\mathbf{u}_1(\tau),\tau)d\tau + h(\mathbf{x}(T,\Phi(x,t))) - \\ &- \int_t^T f_0(\mathbf{x}(\tau,\Phi(\mathbf{x},t)),\mathbf{u}_1(\tau),\tau)d\tau = \\ &= V_1^{(0)} - J(\mathbf{u}_1) \end{aligned} \quad (16)$$

The difference above is only functional of $\mathbf{u}_1$, therefore a constant, q.e.d.

Now it is possible to begin with the construction of minimizing control set. Function $V_1(\mathbf{x},t)$ is extended over the whole interval with the new function (17)

$$V_1(\mathbf{x},t) + \int_0^t f_0 d\tau \quad (17)$$

Function above is the 'measure' for the distance between actual point $(\mathbf{x},t)$ and the end surface S at $t=T$. Then the optimal control of transition process (see Annex) can be applied to find the optimal control $\overline{\mathbf{u}}_2(\mathbf{x},t)$. This control (see (28)) will set the function

$$W_1(\mathbf{x},\mathbf{u},t) = \frac{\partial V_1}{\partial t} + \sum_{s=1}^{n} \frac{\partial V_1}{\partial x_s} f_s(\mathbf{x},\mathbf{u},t) + f_0(\mathbf{u},\mathbf{x},t) \quad (18)$$

to its minimum (i.e. the negative maximum) along each point of the trajectory. Surely finding $\overline{\mathbf{u}}_2(\mathbf{x},t)$ is coupled with some numerical work, although when the partial derivations of V_1 are given, the space of $\overline{\mathbf{u}}_2(\mathbf{x},t)$ must be thoroughly checked – this can already be made often in quite direct manner and the procedure like dynamic programming must only be applied occasionally. When obtaining the numerical solution, the fact, that the optimal control of transition process is very similar to the minimal time problem can be used thoroughly. We then

propose, that the generated system has the solution (19). Any proper solution can be found here and is then shifted because $V_1 - V_1(\mathbf{x},t)$ is constant (16).

$$\mathbf{x} = \mathbf{x}_2(t)\,; \qquad \mathbf{x}_2(0) = \mathbf{x}_0\,; \qquad \mathbf{u}_2(t) = \mathbf{u}_2(\mathbf{x}_2(t),t) \quad (19)$$

From $\mathbf{u}_2(\mathbf{x},t)$ the function $V_2(\mathbf{x},t)$ can be constructed in the same manner as V_1 was constructed from $\mathbf{u}_1$ etc. One gets the three sequences : $\mathbf{u}_i, \mathbf{x}_i, V_i$; $i = 1,2,\ldots$, and it remains to prove, that sequence $\mathbf{u}_i$ is the minimizing sequence.

Sketch of the proof:

First from (15) one can see:

$$V_1(\mathbf{x}(0),0) = J(\mathbf{u}_1,\mathbf{x}) \quad (20)$$

Then $W_i(\mathbf{x}_i,\mathbf{u}_i,t)$ is identically equal 0 by (10). It is then

$$W_i(\mathbf{x}_i,\mathbf{u}_{i+1},t) \le 0\,; \quad W_i(\mathbf{x}_{i+1}(t),\mathbf{u}_{i+1}(t),t) \le 0 \quad (21)$$

Before integrating (21) some preparations must be made. The critical step in integration is the term $V_i(\mathbf{x}_{i+1}(0),0)$, where $\mathbf{x}_{i+1}(0)$ is the point in process space and by proposition for constructing functions $\mathbf{u}_i$ (19) is the process initial point $\mathbf{x}_0$. This again is the initial point for the first solution of (15), so one can write:

$$V_i(\mathbf{x}_{i+1}(0),0) = V_i(\mathbf{x}_0,0) = V_i(\mathbf{x}_1(0),0) = V_1(\mathbf{x}_1,0) \quad (22)$$

The last inequality in (21) can be integrated:

$$\begin{aligned} \int_0^T W_i(\mathbf{x}_{i+1},\mathbf{u}_{i+1},t)dt &= V_i(\mathbf{x}_{i+1},(T),T) - V_i(\mathbf{x}_0,0) + \int_0^T f_0(\mathbf{x}_{i+1},\mathbf{u}_{i+1},t)dt = \\ &= h(\mathbf{x}_{i+1}(T)) + \int_0^T f_0(\mathbf{x}_{i+1},\mathbf{u}_{i+1},t)dt - V_i(\mathbf{x}_i,0) \le 0 \end{aligned} \quad (23)$$

From (23) and (22) one can see that:

$$V_i(\mathbf{x}_1,0) \ge J(\mathbf{u}_{i+1},\mathbf{x}_{i+1}) \quad (24)$$

and from (15) it follows directly :

$$V_i(\mathbf{x}_1,0) = J(\mathbf{u}_i,\mathbf{x}_i) \quad (25)$$

Putting together (24) and (25) generates

$$J(\mathbf{u}_l, \mathbf{x}_l) \geq J(\mathbf{u}_{l+1}, \mathbf{x}_l) \qquad (26)$$

Consequently, the sequence $J(\mathbf{u}_l, \mathbf{x}_l)$; $l = 1,2,\ldots$ is not- increasing and the control sequence $\mathbf{u}_1, \mathbf{u}_2, \ldots$ is minimizing q.d.e.

APPENDIX : OPTIMAL CONTROL OF TRANSITION PROCESS

If the dynamic process is given with (1) and initial condition is $\mathbf{x}_0$ at time t=0, then every control $\mathbf{u} \in \mathbf{G}$ induces the dynamics:

$$\mathbf{x} = \mathbf{x}(t, \mathbf{u}, \mathbf{x}_0) \qquad (27)$$

Let **S** be the surface in the process space $(t, x_1, \ldots x_n)$ given by $\mathbf{S}(t, x_1, \ldots, x_n) = 0$. The control task is to choose $\mathbf{u} \in \mathbf{G}$ in the way that $\mathbf{x} = \mathbf{x}(t, \mathbf{u}, \mathbf{x}_0)$ is crossing **S** at time t_1. Additionally the function $V(t, x_1, \ldots, x_n)$ shall exist, as weighing function of the distance between moving point **x** and surface **S**. The task of the control shall be to lessen this distance.

Definition: The control $\mathbf{u}^0 = (u_1^0, \ldots, u_r^0)$ is transition optimal for V, if V decreases along the trajectory $\mathbf{x}(t, \mathbf{u}^0, \mathbf{x}_0)$ in the fastest possible way. One has to find the values for V along the path (27) and the total differential for V can be written in as:

$$\frac{dV}{dt} = \frac{\partial V}{\partial t} + \sum_{s=1}^{n} \frac{\partial V}{\partial x_s} f_s = W(t, \mathbf{x}, \mathbf{u}) \qquad (28)$$

It is obviously, that the optimal transition control gives the function $W(t, \mathbf{x}, \mathbf{u})$ the minimal (negative) value – it means the fastest decreasing of V – for all possible controls **u**, which are all limited of course.

REFERENCES

[1] J.Macki, A.Strauss, Introduction to Optimal Control Theory, Springer , New-York,1983
[2] D.P.Bertsekas, Dynamic Programming, Prentice-Hall,Englewood Cliffs, New-York,1987
[3] M.Bardi,I.C.Dolcetta, Optimal Control and Viscosity Solution of Hamilton-Jacobi-Bellman equations, Birkhauser, Boston, 1997
[4] J.Fox, L.Leslie, The numerical solution of two-point boundary problems in ordinary differential equations, Dover Publications, New-York, 1990
[5] F.Lewis,L.Vassilis,L.Syrmos, Optimal Control, Wiley, New-York, 1995
[6] V.I.Zubov, Lekcii po teorii upravlenija, Nauka, Moskva, 1980

Scheduling Serial - Parallel Processors – A Case Study

Jan Pelikán
Vysoká škola ekonomická v Praze, Czech Republic
pelikan@vse.cz

Abstract
Methods of Operations Research can play an important role in the process of increasing production productivity and decreasing production costs. We show in this case study how discrete optimization models can be used to find the optimal ordering production batches on lathes. The aim of the optimization is to suggest a processing order of production batches, which is connected with the minimum total processing time of the corresponding production order. For solving the problem, an original mathematical model including discrete variables is proposed. The model is solved by using of the system LINGO 8.0. Obtained results are very interesting since they lead to a significant decrease of the total processing time in comparison with the present practice.

Keywords: operations research, mathematical modeling, discrete models, scheduling models

1 Introduction

The subject of this case study is the production scheduling designed for the enterprise K-Bass Brandýs nad Orlicí. The scheduling consists in ordering batches, which are processed on lathes in the order satisfying given technological requirements. The lathes are considered as parallel processors. A production batch is a prescribed amount of a certain product. Each production order consists of a finite number of production batches. The aim of the scheduling is to find such processing schedule of batches, which minimizes the total processing time of the whole production order. It is assumed that before the processing of any batch on a lathe can begin, the corresponding lathe must be adjusted for the batch to be processed. Therefore the adjustment of a machine is the first phase of batch processing and the proper production is its second phase. We shall assume only one worker for adjusting the machines at our disposal. Therefore the adjustment process will be considered as a serial processing of batches on a unique processor. On the other hand, the proper production run, during which each batch is processed on different machines can be considered as the parallel processing of batches on parallel processors. Therefore, delays can occur for the following reasons: (a) all machines are working so that no machine can be adjusted for some next batch (a delay in the adjustment process); (b) a machine must wait for adjustment while another machine is being adjusted, since the adjustment is carried out only by one worker (a delay in the production process). The minimization of the total processing time of all batches of a given production order is carried out in such a way that the resulting schedule leads at the same time to the maximum exploitation of the machines, minimum delays, and minimum production costs.

2 Description of technological conditions of the case study

The production sector consists of ten lathes, six from which are of the same type. The case study will find a schedule of batches for a one-week production. This production order could not be managed within the capacity of the two-shift working time, i.e. within 75 hours. The production order consists of 27 products. Types of products in production batches, numbers of pieces of the corresponding products as well as the corresponding processing and adjustment times are included in the Table 1.

Table 1. The production batches and their assignment to machines

Product number	Machine	Amount in pieces	Adjustment time (minutes)	Processing time (minutes)
1	1	1000	8,16	100,00
2	1	50	8,16	5,50
3	2	1000	10,20	260,00
4	2	50	10,20	16,50
5	2	50	10,20	20,50
6	3*)	1500	61,20	2370,00
7	4*)	2000	91,80	2440,00
8	5*)	2000	61,20	1380,00
9	6*)	667	61,20	1280,64
10	7*)	100	61,20	40,00
11	7*)	300	61,20	54,00
12	7*)	100	61,20	82,00
13	8*)	200	91,80	172,00
14	9	1000	91,80	2800,00
15	9	100	91,80	11,00
16	9	100	91,80	168,00
17	10	100	6,00	14,00
18	10	2000	6,00	460,00
19	10	1500	6,00	795,00
20	10	1000	6,00	140,00
21	10	100	6,00	14,00
22	10	100	6,00	45,00
23	10	300	6,00	18,00
24	10	2000	6,00	820,00
25	10	100	6,00	28,00
26	10	667	6,00	426,88
27	10	200	6,00	58,00

*) the batch can be assigned to any of machines 3 – 8.

3 A model with a priori assignment of batches to machines

First we shall assume that each production batch is assigned to a machine. This assignment mostly follows the fact that certain types of products can be produced on one certain machine (lathes), except the products 6 – 13, which can be produced on an arbitrary from the machines 3 - 8 (see Table 1). These products will be firstly assigned to lathes in such a way that all machines (lathes) are approximately equally loaded as far as the processing time is concerned. The result is given in Table 1.

We can assume that a machine will begin to produce the product immediately after it is adjusted to its production. Otherwise, the worker adjusting the machines will have to wait until the processed product is finished so that a delay in the adjustment process could occur. Further, we will assume that the worker will immediately start to adjust the machine for the next product as soon as the machine is free whenever he is not busy with adjusting some other machine. Since the products cannot be produced without the corresponding adjustment of the machines, which must precede the proper production process, the order, in which the adjustment is carried out, determines the order, in which the products are processed. It means that the products are processed in the order, corresponding to the order of adjusting the

machines for them. The adjustment order is limited by the fact that we have only one worker adjusting the machines and it is assumed that he cannot adjust two or more different machines simultaneously so that he prepares the machines for the production of different products on a sequence similarly as in the one-processor case. In the other words, he can begin to adjust a machine only after the adjustment of a preceding machine had been finished. Another limitation for the adjustment order follows the assumption that the adjustment of any machine can be commenced only after the machine have finished the production and it is idle.

We will introduce binary variables x_{ij}, which will provide the information about the adjustment order in the following sense: $x_{ij} = 1$ if the adjustment for product i precedes the adjustment for product j, $x_{ij} = 0$ otherwise.

Model parameters:
n the amount of products (batches) in pieces;
m – the number of machines;
t_i –the processing time of the i-th batch in minutes;
d_i – the adjustment time for the i-th batch in minutes;
$M >> 0$ – a large positive number;
S_k – the set of indices of those batches, which are assigned to machine k.

Model variables:
T – the time, at which all batches are finished;
x_{ij} – binary variables determining the adjustment order;
c_i – the time, at which the adjustment for product i is finished in minutes and starts the producing.

Model:

$$T \rightarrow \min \qquad (1)$$
$$x_{ij} + x_{ji} = 1 \qquad i, j = 1, \ldots, n,\ i < j \qquad (2)$$
$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) \qquad i,j = 1, \ldots, n,\ i \neq j,\ i \in S_k,\ j \in S_k,\ k = 1, \ldots, m \qquad (3)$$
$$c_j - d_j \geq c_i - M(1 - x_{ij}) \qquad i,j = 1, \ldots, n,\ i \neq j; \qquad (4)$$
$$d_i \leq c_i \qquad i = 1, \ldots, n. \qquad (5)$$
$$c_i + t_i \leq T \qquad i = 1, \ldots, n. \qquad (6)$$
$$x_{ij} \in \{0, 1\} \qquad i,j = 1, \ldots, n,\ i \neq j; \qquad (7)$$
$$c_i \geq 0 \qquad i = 1, \ldots, n. \qquad (8)$$

Equation (2) ensures that for each pair i, j either the adjustment for batch i precedes the adjustment for batch j or the other way round , i.e. the adjustment for batch j precedes the adjustment for batch i.

Inequality (4) ensures that the beginning time of the adjustment for batch j, i.e. $c_j - d_j$, must follow after the moment c_i, at which a preceding adjustment for batch i had been finished under the assumption that the adjustment for batch i precedes the adjustment for batch j (regardless on which machine the proper production process is carried out). This constraint follows the assumption that the worker cannot adjust more than one machine at the same time. Constraint (5) shifts the time moment, at which the adjustment is finished according to the corresponding adjustment. Inequality (6) determines the total processing time of all batches, which is (according to (1)) minimized. The model was solved using the program

LINGO 8.0 (special professional integer programming software) for a one-week production program in Table 1, a number of production batches was $n = 27$ and a number of machines $m = 10$. The sets S_k are given in Table 1, from which it follows that $S_1 = \{1, 2\}$, $S_2 = \{3,4,5\}$, $S_3 = \{6\}$, $S_4 = \{7\}$, $S_5 = \{8\}$, $S_6 = \{9\}$, $S_7 = \{10,11,12\}$, $S_8 = \{13\}$, $S_9 = \{14,15,16\}$, $S_{10} = \{17,18,19,20,21,22,23,24,25,26,27\}$. The mathematical model included 757 variables (729 of them were binary variables) and 1238 constraints. The computations took 5.4 minutes (PC 1.3 GHz). The resulting order, in which the machines are adjusted, is given in Table 2.

4 A model with a partial a priori assignment of batches to machines

Since a priori assignment of products to machines, which was used in the previous section, is not necessary to be optimal in general, we shall try in the sequel to optimize this assignment together with the total processing time of the schedule. As we have already mentioned, the assignment can be chosen only at products 6 – 13, which can be processed on any of machines 3 – 8. The assignment of the other products to machines is fixed.

Table 2. Optimal ordering of production batches

order	batch	machine	Start adj.	End adj.	Finish time
	i	k	c_i-d_i	c_i	c_i+t_i
1	14	9	0	91,8	2891,8
2	20	10	91,8	97,8	237,8
3	4	2	166,4	176,6	193,1
4	8	5	176,6	237,8	1617,8
5	18	10	237,8	243,8	703,8
6	2	1	603,84	612	617,5
7	7	4	612	703,8	3143,8
8	27	10	703,8	709,8	767,8
9	19	10	767,8	773,8	1568,8
10	6	3	823,2	884,4	3254,4
11	26	10	1568,8	1574,8	2001,68
12	5	2	1574,8	1585	1605,5
13	10	7	1585	1646,2	1686,2
14	12	7	1686,2	1747,4	1829,4
15	9	6	1912,56	1973,76	3254,4
16	22	10	2277,4	2283,4	2328,4
17	24	10	2328,4	2334,4	3154,4
18	3	2	2728,6	2738,8	2998,8
19	13	8	2738,8	2830,6	3002,6
20	11	7	2830,6	2891,8	2945,8
21	15	9	2891,8	2983,6	2994,6
22	16	9	2994,6	3086,4	3254,4
23	1	1	3146,24	3154,4	3254,4
24	21	10	3154,4	3160,4	3174,4
25	17	10	3174,4	3180,4	3194,4
26	23	10	3194,4	3200,4	3218,4
27	25	10	3218,4	3224,4	3252,4

Model 1.

To optimize at the same time also the assignment, we shall introduce binary variables y_{ik}, which will determine the assignment in the sense that $y_{ik} = 1$, if batch i is assigned to machine k and $y_{ik} = 0$ otherwise. Further we set $D = \{6,7,\ldots, 13\}$ and $S = \{3, \ldots, 8\}$, i.e. D contains the indices of those products and S the indices of those machines, where the assignment is not fixed. The remaining products are assigned a priori to corresponding machines according to fixed technological requirements. The mathematical model is based on the model (1) – (8) with the following additional constraints :

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ik}) - M(1 - y_{jk}), \quad i, j = 1, \ldots, n, \; i \neq j, \; i \in D, \; j \in D, \; k \in S \quad (9)$$

$$\sum_{k \in S} y_{ik} = 1, \quad i \in D. \quad (10)$$

Constraint (9) replaces constraint (3) for machines $k \in S$ and the corresponding sets of products S_k. It ensures that the adjustment process for a batch can be started only after finishing the production of the preceding batch. Constraint (9) (unlike to (3)) includes variables y_{ik} and y_{jk}, which are both equal to 1, if both products are produced on the same machine k. Equation (10) ensures that each product from D is assigned to exactly one machine from S. This extension of the model (1) – (8) included 805 variables (777 of them were binary variables) and 1576 constraints. The computation time increased to 1 hour and 35 minutes. The resulting total processing time T was equal to 3254.4 minutes, i.e. the total processing time was the same as in the previous model, which means that a priori assignment in the previous model was optimal. The substantial increase of the computational time is remarkable in the case of extended model, which included the optimization of assignment lathes to machines.

Model 2.

The assignment of products to machines have an alternative solution, since the machines, for which the assignment is not fixed are completely identical. Therefore, it is not necessary to assign products to these machines. It is sufficient to find out, which products can be assigned to an identical machine. In this way, we obtain a simpler model with a lower number of variables.

In Model 2, binary variable $y_{ij} = 1$, if product i is assigned to the same machine as product j and $y_{ij} = 0$ otherwise. For the indices of these variables always $i, j \in D$ and $i < j$ holds. Further, we shall replace constraint (9) from Model 1 with the following constraints (11) and (12), which ensure (similarly as constraint (9)) that the adjustment process for any batch can be commenced only after the production of a preceding batch had been finished.

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ij}), \quad i < j, \; i \in D, \; j \in D \quad (11)$$

$$c_j - d_j \geq c_i + t_i - M(1 - x_{ij}) - M(1 - y_{ji}), \quad i > j, \; i \in D, \; j \in D \quad (12)$$

Since we deal with an assignment of 8 products to 6 machines from the set S, it is necessary to assign at least 2 products to the same machine, i.e. at least 2 variables y_{ij} must be equal to 1, which is ensured by the following inequality (13):

$$\sum_{i \in D} \sum_{j \in D, i<j} y_{ij} \geq 2, \quad i \in D. \qquad (13)$$

Therefore inequality (13) replaces relation (10) from Model 1. Model 2 includes, in comparison with Model 1, 24 variables and 280 constraints less than Model 1.

5 Conclusion

The obtained schedule, in comparison with the schedule used in the practice, leads to the total processing time, which is 30% shorter. This result can substantially increase the productivity, exploiting the machines and lead to lower production costs. Nevertheless, it requires on the other hand the usage of sophisticated modeling techniques and a professional optimization software. It seems that an interconnection of these models with the information system would be possible too, because the program system LINGO 8.0 supports this interconnection. On the other side, disadvantage of this approach consists in the fact that we have to solve NP-hard problems and the increase of the size of problems we have to solve may cause a substantial increase of the necessary computational time. This happens also when we passed from Model 1 with a priori assignment to Model 2, which includes also the optimization of the assignment of products to machines. We obtained in this case a larger problem and the necessary computational time was higher too.

References

[1] PELIKÁN, Jan, An Optimization of the Production System - Case Study. New Delhi 08.12.2003-11.12.2003. In: APORS 2003. New Delhi: ORSI,2003,p. 48.

[2] PELIKÁN, Jan, Diskrétní modely v operačním výzkumu. 1. vydání, Praha: PROFESSIONAL PUBLISHING, 2001. 164 p., ISBN 80-86419-17-7.

[3] PELIKÁN, Jan, A minimax parallel processors scheduling problem. Edinburgh 08.07.2002-12.07.2002. In: IFORS 2002. OR in a globalized network world economy. Edinburgh: IFORS, 2002,155 p.

The 8th International Symposium on Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 2:

Stochastic and Combinatorial Optimization

THE KIDNEY EXCHANGE GAME*

Katarína Cechlárová[1], Tamás Fleiner[2] and David Manlove[3]
[1]Institute of Mathematics, P.J. Šafárik University, Jesenná 5, 041 54 Košice, Slovakia
email: cechlarova@science.upjs.sk
[2] Department of Computer Science and Information Theory,
Budapest Univ. of Technology and Economics, Műegyetem rkp. 3-9. H-1111 Budapest Hungary
email: fleiner@cs.bme.hu
[3] Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK
email: davidm@dcs.gla.ac.uk

Abstract. The most effective treatment for kidney failure that is currently known is transplantation. As the number of cadaveric donors is not sufficient and kidneys from living donors are often not suitable for immunological reasons, there are attempts to organize exchanges between patient-donor pairs. In this paper we model this situation as a cooperative game and propose some algorithms for finding a solution.

Keywords. Kidney transplantation, cooperative game, Pareto optimal solution, core, algorithm.

1 Introduction

Renal failure is a very serious illness for which the most effective treatment that is currently known is kidney transplantation. Ideally, a kidney from a deceased donor could be used, but the supply of those in spite of joint efforts of national and even international organisations (for example Eurotransplant Foundation [19] and the United Network for Organ Sharing in the USA [20]) is not sufficient for the growing demand and the waiting time of a patient is unpredictable. As the operation techniques improved and the risk for a living donor of a kidney (a genetic or an emotional relative of the patient) was minimized, the number of live-donor transplantations increased. Moreover, some studies [17] show that grafts from living donors have a higher survival rate.

For a transplantation to be successful, some immunological requirements must be fulfilled. Basically, ABO incompatibility and a positive cross-match are an absolute contraindication, moreover, the greater the number of HLA mismatches between the donor and the recipient, the greater the chance of rejection [9]. Hence, it often happens that a willing donor cannot donate his/her kidney to the intended recipient. Therefore in several countries systematic kidney exchange programs have been established: in Romania [10], the Netherlands [18], USA [13, 14, 15]; in other cases there are isolated examples [7], e.g. in the Middle East.

Kidney exchange (KE for short) is still a controversial issue, however the aim of this paper is not to discuss the ethical and legal aspects of this concept. In spite of some

*This work was supported by the VEGA grant 1/0425/03, Science and Technology Assistance Agency contract No. APVT-20-004104 (Cechlárová), OTKA F 037301 research grant (Fleiner), EPSRC grant GR/R84597/01 (Manlove) and the Royal Society of Edinburgh International Exchange Programme (Fleiner and Manlove).

pessimistic expectations (the British Transplantation Society estimated potential benefits from living donors' exchanges to be around 3% [3]), in an institute of Romania the monthly mean number of transplantations increased from 4.2 to 6.1 since the KE program started [10] and a simulation study using real USA statistical data revealed that the utilization of kidneys from living donors without exchanges was around 55%, whilst with exchanges it increased to 91% when the size of the simulated population was 300 [13].

We follow the approach started in [13] and [14], in that we represent KE as a cooperative game, in which patient-donor pairs seek cyclic exchanges of kidneys. Since all operations on a cycle should be performed simultaneously (to avoid the risk that one of the donors will withdraw his or her commitment after the others have undergone nephrectomy [7]), cycles should be as short as possible for logistical reasons. Therefore we directly incorporate cycle lengths into preference models (notice that in [13] the obtained cycle lengths were just observed after simulations of the algorithm used, and in [15] the number of matched patients was derived as a function of the maximum allowed cycle length). Further, besides Pareto optimal solutions we also consider the core and the strict core of the KE game. We conclude with computational complexity considerations of the proposed algorithms and some open questions.

2 The KE game

An instance of the KE game is represented by a directed graph $G = (V, A)$ where each vertex $v \in V$ corresponds to a patient and his intended (incompatible) donor (or donors). A pair $(i, j) \in A$ if the patient corresponding to vertex i can accept a kidney from a donor corresponding to vertex j. (Hence, if patient i has a compatible donor, $(i, i) \in A$.) Moreover, for each vertex i there is a linear ordering $\leq_i$ on the set of endvertices of arcs incident from i. This ordering is represented by a preference list of i and we assume that i is the last entry in each preference list. If $j \leq_i k$ and simultaneously $k \leq_i j$, we say that i is *indifferent* between j and k, or that j and k are *tied* in i's preference list, and write $j \sim_i k$; if $j \leq_i k$ but not $k \leq_i j$, then i *strictly prefers* j to k, and we write $j <_i k$. Hence, in general, preference lists may contain ties.

Definition 1 *A* kidney exchange *game (KE game for short) is a triple* $\Gamma = (V, G, \mathcal{O})$, *where* V *is the set of players,* G *is a digraph with a vertex set* V *and* $\mathcal{O} = \{\leq_i; i \in V\}$.

Definition 2 *A* solution *of the KE game* $\Gamma = (V, G, \mathcal{O})$ *is a permutation* π *of* V *such that* $i \neq \pi(i)$ *implies* $(i, \pi(i)) \in A$ *for each* $i \in V$. *If* $(i, \pi(i)) \in A$, *we say that* i *is* covered *by* π, *otherwise* i *is* uncovered. *A player* i *is assigned in a solution* π *the pair* $(\pi(i), C^{\pi}(i))$, *where* $C^{\pi}(i)$ *denotes the cycle of* π *containing* i.

A player evaluates a permutation not only according to the player he is assigned to, but he also takes into account the cycle length. This is expressed by the following extension of preferences from $\mathcal{O}$ to preferences over (player,cycle) pairs. The same symbol is used for preferences over players as well as over pairs and permutations.

Definition 3 *A player* i *prefers pair* (j, M) *to pair* (k, N) *if*
(i) $j <_i k$ *or*
(ii) $j \sim_i k$ *and* $|M| \leq |N|$.

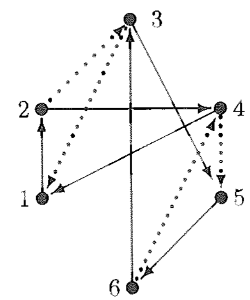
Definition 4 *A coalition* $S \subseteq V$ weakly blocks *a solution* π *if there exists a permutation* σ *of* S *such that*
(i) $(\sigma(i), C^{\sigma}(i)) \leq_i (\pi(i), C^{\pi}(i))$ *for each* $i \in S$ *and*
(ii) $(\sigma(j), C^{\sigma}(j)) <_j (\pi(j), C^{\pi}(j))$ *for at least one* $j \in S$.
A coalition $S \subseteq V$ blocks *a solution* π *if there exists a permutation* σ *of* S *such that* $(\sigma(i), C^{\sigma}(i)) <_i (\pi(i), C^{\pi}(i))$ *for each* $i \in S$.

Now we define the studied solution concepts.

Definition 5 *A permutation* π *is* Pareto optimal *for game* Γ *(belongs to* $PO(\Gamma)$ *for short) if the grand coalition* V *does not block; it is* strongly Pareto optimal *(*$\pi \in SPO(\Gamma)$ *for short) if* V *does not weakly block.*

Definition 6 *A permutation* π *is in the* core $C(\Gamma)$ *of game* Γ *if no coalition blocks* π *and* π *is the* strong core $SC(\Gamma)$ *of* Γ *if no coalition weakly blocks.*

Notice that our concepts of *strongly Pareto optimal* and *Pareto optimal* solutions correspond to what are usually called *Pareto optimal* and *weakly Pareto optimal* solutions respectively [13]. However, we use the first two terms to make the terminology consistent within this paper.



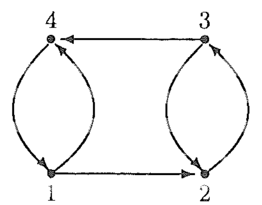
As each blocking coalition is also weakly blocking, we have $SC(\Gamma) \subseteq C(\Gamma) \subseteq PO(\Gamma)$ and $SC(\Gamma) \subseteq SPO(\Gamma) \subseteq PO(\Gamma)$ in general, but the inclusion $C(\Gamma) \subseteq SPO(\Gamma)$ is not always true even in the case without indifferences. Consider the KE game Γ given by the digraph to the left, where the first entries in the preference lists are denoted by solid lines, while the second entries correspond to dotted lines. Permutation $\pi = (1, 2, 3)(4, 5, 6)$ is in $C(\Gamma)$, but $\pi \notin SPO(\Gamma)$, as the grand coalition weakly blocks π by considering permutation $\sigma = (1, 2, 4)(5, 6, 3)$.
However, in the practical kidney exchange application it is not clear which solution concept is the most suitable one.

3 The case with preferences

With no indifferences, the famous Top Trading Cycles (TTC for short) algorithm (Figure 1) can be used for the KE game. The TTC algorithm was originally proposed by Gale in [16] for housing markets, where cycle lengths were not taken into account. It was shown that the TTC algorithm outputs a permutation in the core of the housing market also in the case with indifferences (ties are broken arbitarily). In [12] Roth and Postlewaite proved that if there are no indifferences, the strong core of the housing market is nonempty and contains a unique permutation. Further, Roth [11] proved that the TTC algorithm is strategy-proof. However, a detailed consideration of algorithmic questions connected with the TTC algorithm is quite recent; in [1] its implementation with $O(m)$ time complexity was proposed, where m is the number of arcs in G. In [6], the TTC algorithm is called Algorithm B-stable and it was shown that in the case with no indifferences, its output is in $SC(\Gamma)$.

Input. A KE game $\Gamma = (V, G, \mathcal{O})$. **Output.** A permutation $\pi = C_1 C_2 \ldots C_r$ of V.
Step 0. $N := V$, round $r := 0$.
Step 1. Choose an arbitrary player i_0.
Step 2. Player i_0 points to his favourite i_1 in N, i_1 points to his favourite i_2 in N etc. A cycle arises or player i_k cannot point.
Step 3. $r := r + 1$. If a cycle C was obtained, then $C_r := C$, otherwise $C_r = (i_k)$. $N := N - C_r$.
Step 4. If $N \neq \emptyset$, go to Step 1, otherwise end.

Figure 1: **The Top Trading Cycles (TTC) Algorithm.**



However, the TTC algorithm may fail to find a permutation in the $C(\Gamma)$ and also in $PO(\Gamma)$ if indifferences are present. In the example to the left, the TTC algorithm may, depending on how the ties in vertices 1 and 3 are broken, output the permutation $(1, 2, 3, 4)$, while here $SPO(\Gamma) = PO(\Gamma) = C(\Gamma) = SC(\Gamma)$ and they contain a unique permutation $(1, 4)(2, 3)$.

Moreover, in [5] it was proved that in the case with indifferences it is NP-complete to decide whether $C(\Gamma) \neq \emptyset$, and it is also NP-complete to decide whether $SC(\Gamma) \neq \emptyset$. Notice that it is always the case that $SPO(\Gamma) \neq \emptyset$: if we define a partial order on the set of all permutations of V by setting $\sigma \leq \pi$ if $(\sigma(i), C^\sigma(i)) \leq_i (\pi(i), C^\pi(i))$ for each player, then the minimal permutations in this partial order are strongly Pareto optimal (and hence also Pareto optimal). We show later, however, that it is NP-hard to find a permutation in $SPO(\Gamma)$ (Theorem 5).

4 The Simple KE game

Let us now suppose that all compatible kidneys are equally suitable for transplantation, as suggested in [14] and [15]. We will say that now the players have *dichotomous* preferences (we follow the terminology of [2]), or that all acceptable vertices are tied. Hence patients with a compatible donor are not considered and the KE game is identified simply with a digraph $G = (V, A)$ without loops.

Definition 7 *A simple KE game is the pair $\Gamma = (V, G)$, where V is the set of players and $G = (V, A)$ is a digraph without loops.*

Similarly as for the KE game from Definition 1, the solution of a simple KE game is a permutation π of V such that $i \neq \pi(i)$ implies $(i, \pi(i)) \in A$ for each $i \in V$ and the notation from Definition 2 can be used. Further, preferences of players over (player, cycle) pairs now reduce to saying that a player prefers to be covered to being uncovered, and if covered, he prefers to be in a shorter cycle.

In [14], where dichotomous preferences were considered, the authors allowed only cycles of length 2. Hence the KE game was represented by an undirected graph $G = (V, E)$ with $ij \in E$ if patient i can accept kidney from donor j and also conversely. A strongly Pareto optimal solution was identified as a maximum cardinality matching.

Input: A simple KE game $\Gamma = (V, G)$. **Output:** A permutation π of V.
Step 1. Set $k := 2$ and create the 2-reduced graph $G_2 = (V_2, E_2)$ of digraph G by letting: $V_2 = \{v \in V : v$ lies on a 2-cycle in $G\}$, $E_2 = \{e = ij : (i, j) \in A \ \& \ (j, i) \in A\}$.
Find a maximal matching M_2 in G_2 and set $\pi(i) = j$ for each $ij \in M_2$.
Delete from G all vertices covered by M_2.
Step 2. $k := k + 1$.
Step k. Create the k-reduced digraph $G_k = (V_k, A_k)$ of G by letting:
$V_k = \{v \in V : v$ lies on a k-cycle in $G\}$, $A_k = \{e \in A : e$ lies on a k-cycle in $G\}$.
Find a maximal k-cycle packing M_k of G_k and set $\pi(i) = j$ for each arc $(i, j) \in M_k$.
Delete from G all vertices covered by M_k. If G contains no cycle of length $k + 1$, end. Otherwise go to Step 2.

Figure 2: **Algorithm CoreSimpleGame**

Even without the restriction on cycle lengths, we are able to formulate a necessary and sufficient condition for a permutation to be in $C(\Gamma)$.

Theorem 1 *$\pi \in C(\Gamma)$ if and only if G contains no cycle $C = (i_1, i_2, \ldots, i_k)$ such that $|C^\pi(i_j)| > k$ or $\pi(i_j) = i_j$ for each $j = 1, 2, \ldots, k$.*

Proof. Such a cycle clearly blocks π. On the other hand, if $\pi \notin C(\Gamma)$ then a blocking coalition has a form of C. ∎

Theorem 1 suggests an iterative approach for finding a permutation in the core of the simple KE game. The algorithm is depicted in Figure 2.

Theorem 2 *Algorithm CoreSimpleGame correctly finds a permutation in $C(\Gamma)$ of the simple KE game in polynomial time.*

For the strong core, Theorem 1 can be easily modified:

Theorem 3 *$\pi \in SC(\Gamma)$ if and only if G contains no cycle $C = (i_1, i_2, \ldots, i_k)$ such that either $|C^\pi(i_j)| > k$ or $\pi(i_j) = i_j$ for at least one $j \in \{1, 2, \ldots, k\}$, and either $|C^\pi(i_l)| \geq k$ or $\pi(i_l) = i_l$ for the remaining vertices i_l of cycle C.*

Corollary 1 *If $\pi \in SC(\Gamma)$ then π restricted to the 2-reduced graph G_2 is a perfect matching of V_2 and for each $k > 2$, π restricted to the k-reduced digraph G_k is a partition of V_k into directed k-cycles. Consequently, if either G_2 admits no perfect matching or G_k admits no partition of V_k into directed k-cycles for some k, then $SC(\Gamma) = \emptyset$.*

Hence Algorithm CoreSimpleGame could be used to find a permutation in $SC(\Gamma)$, if in Step 2 a perfect matching and later a partition of V_k into directed k-cycles can be found. However, this problem already becomes difficult.

Theorem 4 *The problem of deciding whether $SC(\Gamma) = \emptyset$ for a simple KE game is NP-complete.*

Proof. Membership in the class NP follows from Theorem 3. In [8, Theorem 3.7] the NP-completeness of UNDIRECTED TRIANGLE PACKING is established. It is easily possible to obtain NP-completeness for DIRECTED TRIANGLE PACKING in graphs with no bidirected arcs, where each vertex lies on a directed triangle, by considering the reduction constructed there and directing appropriately the arcs in the graph shown in [8, Figure 3.8].

An instance $G = (V, A)$ of DIRECTED TRIANGLE PACKING gives rise to an instance Γ of a simple KE game. It is then straightforward to verify that G admits a directed triangle packing if and only if $SC(\Gamma) \neq \emptyset$. ■

However, in spite of the fact that $SPO(\Gamma) \neq \emptyset$ for all KE games, we have:

Theorem 5 *It is NP-hard to find a permutation in $SPO(\Gamma)$ for a simple KE game.*

Proof. Suppose we have a polynomial-time algorithm $\mathcal{A}$ for finding a permutation in $SPO(\Gamma)$. Now let $G = (V, A)$ be a directed graph with no bidirected arcs (given as an instance of DIRECTED TRIANGLE PACKING). Run $\mathcal{A}$ on G to find $\pi \in SPO(\Gamma)$. If π contains only 3-cycles then clearly π gives a directed triangle packing for G. Now suppose that it is not the case that π contains only 3-cycles. Suppose for a contradiction that G admits a directed triangle packing. This defines a permutation σ. Consider any vertex $v \in V$ that does not belong to a 3-cycle in π (there is at least one such vertex). If v is uncovered by π then v improves in σ, since v is covered by σ. Otherwise v belongs to a cycle of length > 3 in π and v also improves in σ, since v belongs to a 3-cycle in σ. Hence V weakly blocks π via σ, contradicting the assumption that $\pi \in SPO(\Gamma)$. ■

5 Conclusion and open questions

The research on efficient algorithms for kidney exchange problems has just started. So there are still many questions to be answered and many problems to be solved. Let us mention just a few directions for possible further research.

- As the basic priority is to treat as many patients as possible, find a solution of the KE game with the maximum number of covered verties.
- As patients enter and leave the waiting lines unpredictably, explore the possibilities of on-line algorithms for kidney exchange.

6 Acknowledgement

We would like to thank to Dr. Ľuboslav Beňa, Transplantcentrum Košice, for detailed explanations of medical and ethical problems connected with transplantations of organs.

References

[1] D. Abraham, K. Cechlárová, D. Manlove and K. Mehlhorn, *Pareto optimality in house allocation problems*, Lecture Notes in Comp. Sci. **3341**, Algorithms and Computation, ISAAC 2004, Hong Kong, December 2004, Eds. R. Fleischer, G. Trippen, 3-15 (2004).

[2] A. Bogomolnaia and H. Moulin, *Random matching under dichotomous preferences*, Econometrica **72**: 257–279, 2004.

[3] J.A. Bradley, British Transplantation Society and Renal Association, *UK Guidelines for Living Donor Kidney Transplantation*, January 2000, http://www.cambridge-transplant.org.uk/program/renal/guidelines.htm

[4] K. Cechlárová and J. Hajduková, *Stability testing in coalition formation games*, Proc. SOR'99, Preddvor, Slovenia, Eds. V. Rupnik, L. Zadnik-Stirn, S. Drobne, 111-116 (1999).

[5] K. Cechlárová and J. Hajduková, *Computational complexity of stable partitions with $\mathcal{B}$–preferences*, Int. J. Game Theory 31 (2002) 3, 353–364.

[6] K. Cechlárová and A. Romero-Medina, *Stability in coalition formation games*, International Journal of Game Theory 29, 2001, 487-494.

[7] F.L. Delmonico, *Exchanging Kidneys Advances in Living-Donor Transplantation*, New England J. Med. 350:1812-1814 (2004).

[8] M.R. Garey and D.S. Johnson, *Computers and Intractability*, Freeman, San Francisco (1979).

[9] T. Kälbe et al., *Guidelines on renal transplantation*, European Association for Urology, February 2003, http://www.uroweb.org/files/uploaded_files/guidelines/renaltransplant.pdf

[10] M. Lucan, P. Rotariu, D. Neculoiu and G. Iacob, *Kidney exchange program: a viable alternative in countries with low rate of cadaver harvesting*, Transplantation Proceedings 35, 933-934 (2003).

[11] A. Roth, *Incentive compatibility in a market with indivisibilities*, Economics Letters **9**:127–132 (1982).

[12] A. E. Roth and A. Postlewaite, *Weak versus strong domination in a market with indivisible goods*, J. Math. Econ. **4**:131–137 (1977).

[13] A. Roth, T. Sönmez and U. Ünver, *Kidney Exchange*, Quarterly J. of Econ. 119: 457–488 (2004).

[14] A. Roth, T. Sönmez and U. Ünver, *Pairwise Kidney Exchange*, J. Econ. Theory, to appear.

[15] A. Roth, T. Sönmez and U. Ünver, *Efficient Kidney Exchange: Coincidence of Wants in a Structured Market*, manuscript, May 2005.

[16] L. Shapley and H. Scarf, *On cores and indivisibility*, J. Math. Econ. **1**:23–37 (1974).

[17] P. I.Terasaki, J. M. Cecka, D. W. Gjertson and S. Takemoto, *High Survival Rates of Kidney Transplants from Spousal and Living Unrelated Donors*, New England J. Med. **333**:333-336, (1995).

[18] M. de Klerk, K. Keizer, W. Weimar, A.J. Goody, A. Spital, F.L. Delmonico, *Donor Exchange for Renal Transplantation*, New England J. Med. 351: 935-937 (2004).

[19] The Eurotransplant International Foundation, http://www.eurotransplant.org/

[20] United Network for Organ Sharing, http://www.optn.org

MODELING STOCK BUYING-SELLING USING A NEW MCDM METHODOLOGY ACCOUNTING FOR UNCERTAINTY

Vladimir I. Kalika
Natural Resources and Environmental Research Center,
Haifa University, Mount Carmel, Haifa 31905, Israel,
E-mail: kalika@econ.haifa.ac.il

Abstract

A new methodology is proposed to select a predetermined number of "reasonable" alternatives from their considerable initial set according to an arbitrary number of optimization criteria. This intuitive methodology, based on accounting for uncertainty factors, inherent in real problems, by performing the multi-level hierarchical system of multi-variant computations and of finding their "stable-optimal" solutions, is applied for modeling the stocks' buying-selling on stock exchange. The "bottleneck" of this modeling - the construction of adequate optimization criteria, is the main object for consideration.

Key words: multi-criteria decision making; six-level hierarchical system of multi-variant computation series; scenarios; statistical data on stock deal sums and stock prices; prognosis trends; sinusoidal character of statistical data on stock prices.

1. Introduction

The paper reflects the attempt of applying the new multi-criteria decision making (MCDM) approach accounting for uncertainty factors to solve a problem of stock buying-selling on stock exchange, where this applied problem is considered mainly as an object suitable for this approach application (for now the practical aspects of behavior on stock exchange are considered by us as less important).

This proposed quite general MCDM approach (its methodology and some applications are reflected in [**1-5**]) is oriented on solving various real problems, connected with decision making, characterized by selecting the "reasonable" (the best in certain sense) alternatives from their considerable (maybe vast) initial set according to many objectives, expressed by multiple criteria including an arbitrary number of optimization criteria. The main specifics of this proposed approach consists in the methodology, developed to take into account uncertainty factors, inherent in such problems, where these factors could have anyone nature. According to this methodology, the following ways are proposed to meet various uncertainty aspects, inherent in such real MCDM problems:

1. *A two-phase process* is considered to reach *a final solution of the whole problem*, where *the first phase of this process* should reflect a computer calculation process to select *a predetermined number of* the resulting *"reasonable" alternatives* (their set is named *RAS*) from their *initial set* (*ISA*) according to the assigned *multiple criteria*. On *the process second phase*, the final solution is found, basing on the derived RAS analysis. It will be a non-formal process to account for uncertainty factors by considering additional qualitative/quantitative criteria and other data, expert estimates, the current observations of last situations, etc.

2. *The methods to reach the RAS*, concerning the first phase of this solution process, may include various ways of accounting for uncertainty factors: (1) considering different versions of conditions and parameters of the problem solved as well as of the appropriate calculation process intended to solve this problem; (2) forming various combinations of such versions to provide performance of *multi-variant computation series* and finding their *"stable-optimal" solutions*; (3) modification of used multi-criteria optimization techniques, other calculation models and procedures to account for

some uncertainty factors; (4) using random variables (e.g., to use *Monte Carlo simulation)*, expert or hypothetic (in situations close to full uncertainty) estimates in various calculation operations, etc.

Thus, *this formalized first phase* of the solution process, intended to solve the suitable real MCDM problems by reaching their resulting RAS, should include the following *basic stages*:

1) *forming a totality of ISA versions*, where such versions should reflect all possible cases of initial alternatives' combinations accounting for availability of appropriate uncertainty factors;

2) *constructing the* versions of *initial set of criteria assessment vectors (ISCAV)*, where each ISCAV version, interrelated with the appropriate ISA version, includes a totality of criteria assessment vectors, determined for each alternative from the ISA using the criteria calculation models developed a priori for all considered criteria. Thus, each criteria assessment vector, corresponding to one initial alternative and conversely, is represented for this alternative by one numerical value, determined for each considered criterion using its calculation model;

3) *multi-criteria optimization on the space of ISCAV (ISA) accounting for uncertainty factors to achieve the resulting RAS* by decreasing the ISA.

Performance of stages 1-2, consisting in constructing (often, simultaneously) the ISA and ISCAV, presents the processes, specific for each considered MCDM problem. However, the approach to perform stage 3 was developed (by us) as a quite universal one. This approach is based on performing *a multi-level hierarchical system of MVC series*, where each level of the system should include *a totality of scenarios*, having the same nature specific only for this level. The scenarios of different levels reflect various versions of: the ISA/ISCAV (they correspond to the first/second levels of this multi-level system, l=1/2), the used multi-criteria optimization techniques, modified to account for uncertainty (the third level, l=3); some conditions and parameter values, involved in the problem (such levels, since with fourth one, l>3, are specific for various MCDM problems as well as for the appropriate types of multi-level hierarchical systems used to solve these problems).

We considered various versions of multi-level hierarchical system of MVC series, differing by various aspects: (a) the number of levels; (b) the specifics of scenarios, accepted for various levels; (c) the peculiarities of calculation process performance, etc.

2. The multi (six)-level hierarchical system, used for the problem of stock buying-selling

We use to solve the considered MCDM problem of stock buying-selling *the six-level hierarchical system* including the following **totalities of scenarios** for each l-level (l=1-6):

l=1/2) ISA/ISCAV versions - 1/2-scenarios, having the numbers k/m (k=1,..,K/m=1,..,M);

l=3) the multi-criteria optimization technique version (n=N=1), presenting our modification [3] of TOPSIS method [6], based on the consideration of the following scalar goal function:

$$\min_{\{i=1,..,I\}} \{c_{i1}\hat{}W_1^* + c_{i2}\hat{}W_2^* + ... + c_{iJ}\hat{}W_J^*\} \qquad (1)$$

where: $c_{ij}\hat{}$ (i=1,..,I; j=1,..,J) is a normalized relative value of j-criterion for i-alternative, calculated for each j-criterion (j=1,..,J) with respect to their maximal/minimal (on i=1,..,I) values $\{c_j^{max}/c_j^{min}\}$; thus, the values $c_{ij}\hat{}$ are calculated according to the formula:

$$c_{ij}\hat{} = (c_{ij} - c_j^{min}) / (c_j^{max} - c_j^{min}), \ i=1,...,I; \ j=1,...,J; \qquad (2)$$

W_j^*(j=1,..,J) is j-criterion random weight, reflected by its possible values interval $[w_j^{min}, w_j^{max}]$;

l=4/5) the intervals $\{[w_j^{min}, w_j^{max}], j=1,...,J\}$ versions (p=1,..,P/r=1,..,R), each of which is a pair of 4/5-scenarios, taken one at a time for each l-level (l=4/5) and corresponding to p-version of "central point" values $\{w_j\hat{}, j=1,...,J\}$ for all these j-intervals (it's 4-scenario) as well as to r- version of all these j-interval bounds (5-scenario), calculated around the values $w_j\hat{}$:

$$w_j^{min} = w_j\hat{}(1-g_j), \quad w_j^{max} = w_j\hat{}(1+g_j), \quad j=1,...,J \qquad (3)$$

where $0<g_j<<1$ is a relative value, assigned a priori to determine these bounds values which might be derived using the equal g_j-values for all j–criteria, e.g., g_j=g, g=0.05, 0.1, 0.15,...).

l=6) All 6-scenarios (q=1,..,Q) reflect various combinations of values $\{w_j, j=1,...,J\}$, corresponding to all j-intervals, formed on the 4/5-levels (l=4,5). Each such q-combination includes J values w_j, taken one at a time inside of each j-interval $[w_j^{min}, w_j^{max}]$ using J Monte Carlo simulations.

The process of forming such **six-level system** to solve the real MCDM problems could be connected with necessity to overcome the "bottlenecks" in forming the 1/2-scenarios, i.e. the ISA/ISCAV creation. In the considered problem we have the simple case of ISA construction (an initial alternative concept is very implicit – a stock buying-selling operation is such an alternative and it's not needed to develop the special calculation procedures to construct a vast ISA, etc,). However, the main "bottleneck" of this problem modeling is connected with the ISCAV creation, where we have the *analytic case*: a necessity *to develop non-implicit artificial criteria*. This case is differed from many other suitable MCDM problems (e.g. **[1,2]**) with the opportunities to form natural optimization criteria (economic, environmental, technical, etc.). This is precisely the main aspect, considered further in this paper.

The calculation process to reach the resulting **RAS** (e.g.[**4,5**]) is performed by fixing the combinations of l-scenarios, where such combination *(full Scenario)* includes l-scenarios, taken on at a time for each l-level (l=1-6). Each *full Scenario* forms a mono-optimization problem by fixing for the random $\{W_1^*,...,W_J^*\}$ in (1) their determined values $\{w_1,...,w_J\}$, using J Monte-Carlo simulations for the appropriate J intervals (3). This problem solution is *a subset of sub-optimal alternatives*. Further, we form a group of *full Scenarios*, changing all 6-scenarios with the unchanged l-scenarios for all upper l–levels (l=1-5), that allows derive *a set of sub-optimal subsets* for this group as well as its *subset of stable-optimal alternatives*, constructed using a special analysis of sub-optimal alternatives' entering quantities in their set. Varying such groups by changing the fixed 5-scenarios only, we derive *a set of stable-optimal subsets* (of 5-level) and its *subset of stable-optimal alternatives* (of 4-level), corresponding to the fixed combination of l-scenarios for the four upper l–levels (l=1-4). Continuing this process with the variation in fixing 4,3,2,1-scenarios, we reach the resulting *RAS* as *a stable-optimal subset* (of 0-level), derived by *the set of stable-optimal subsets* (of 1-level) analysis.

Thus, the calculation process to reach the RAS is a moving from this scenarios' system "bottom" (l=6) to its "top" (l=1), using all its l-scenarios for all its l–levels (l=1-6).

3. Some principles of stock buying-selling modeling using this new MCDM approach

3.1. The contents and purposes of such modeling

The following problem is considered: *to select* (in the current T-day) *a holding of stocks*, including a predetermined number of stocks, *proper for buying and selling them on the stock exchange*. It's proposed to perform stocks' buying for the next prognosis (T+1)-day to sell them for any subsequent (T+S+1)-day, where the integer number S>0 is assigned a priori.

This selection of proper stocks (the operations with them are the "reasonable" alternatives in accordance with our proposed MCDM approach) is performed on a basis of the appropriate statistic data processing, where these data are considered for the period [1,...,T] of T days, as well as of the expert estimates' use. All this concerns two parameters of stock exchange process: *deal sums* (DS), *stock prices* (SP). Thus, *the problem solution purpose* consists in *selecting from a full initial set of stocks* (the ISA in our MCDM approach) *a predetermined number of stocks* (their set is the RAS), *proper to be bought for the prognosis* (T+1)*-day* as well as *to be salt for the prognosis* (T+S+1)-*days*.

The accent on the statistic data processing in comparison with the use of expert estimates for all considered stocks is more convenient for the problem of selecting the proper stocks among their great quantity (selecting the small RAS from the considerable ISA). Usually, the best way to select the proper stocks for their buying-selling is based, first of all, on using the expert estimates of production and financial conditions for the enterprises, whose stocks are bought and salt. However, the principal peculiarity of the considered problem, connected with the competition of great quantity of stocks, seriously limits the possibility to take such expert estimates for all considered stocks. Accounting for it, now we will be oriented on using the statistic data for the quite long period, especially if this period is characterized by "a stable behavior" of the considered stock exchange. In this situation it's possible to expect that these observed statistic data as if accumulate various aspects, affecting on "stock exchange behavior of each considered stock" (among such aspects: the state of appropriate objects and productions, the psychology and interaction of stock exchange buyers and sellers, etc.).

Thus, *the purpose of such our approach with the practical positions* might be treated as to help a stock exchange expert ("broker") or any exchange "player" in the following aspect: to make the preliminary "limitation" for the stocks holding, which this "player" might be analyze further using his regular intuitive methods. For instance, the full totality of stocks in our real sample, see Section 4, includes 97 stocks, but it's possible to select from them, using our approach, a predetermined small number (e.g. 5-10) of "proper" stocks to continue their analysis using the traditional intuitive methods. This case is very convenient for a new unknown situation on the stock exchange considered.

3.2. Peculiarities of this MCDM approach applying to the stock buying-selling modeling

This problem choice to apply this MCDM approach is caused, first of all, by *availability of required initial (statistic) data* as well as of *specific methodological difficulties to realize such approach* for this real situation. Such *difficulties ("bottlenecks") are connected mainly with the criteria modeling and the ISCAV construction.*

The ISA construction has not been of our main methodological interest, since: (a) *an initial alternative concept is very implicit - a stock buying-selling itself is such alternative*; (b) the possible number of stocks isn't vast usually, and the measurable quantity of stocks are possible to be considered on each existing stock exchange. Thus, the ISA (each its version) may be presented as the set $\{i=1,...,I\}$ of i-stocks' (initial alternatives') numbers.

The ISCAV creation is based on using 2 groups of *Criteria calculation models*, reflecting:

Bn) stock buying for the prognosis (T+1)-day, including 6 criteria types (*Criteria 1-5,9*);

Sl) stock selling for the prognosis (T+S+1)-day, including 3 criteria types (*Criteria 6-8*).

These criteria values for each considered i-stock ($i=1,...,I$) are determined on basis of *two principles of statistical data processing*: (a) finding *the trend-prognosis values* (*Criteria 1,2,6-9*), (b) using *the sinusoidal character of these statistical data* (*Criteria 3-5*). The *Criteria 1,2,6,7* reflect the deal sums (DS) prognosis characteristics, the *Criteria 3-5,8,9* – such characteristics for stock prices (SP).

3.3. The Criteria calculation models

The main principle of **Criteria modeling**, based **on** finding **the trend-prognosis values** is as: any such criteria value $\{C_j[i], j=1,2,6\text{-}9;\ i=1,..,I\}$ is determined according to the formula:

$$C_j[i]=C_j^{(1)}[i]w^{(1)}+C_j^{(2)}[i]w^{(2)}+C_j^{(3)}[i]w^{(3)}+C_j^{(4)}[i]w^{(4)}+C_j^{(5)}[i]w^{(5)}+C_j^{(6)}[i]w^{(6)}+C_j^{(Ex)}[i]w^{(Ex)} \quad (4)$$

where: $C_j^{(k)}[i]$ is the *k–trend* ($k=1,..,6$) *prognosis value for i-stock,* corresponding to j-criterion; $C_j^{(Ex)}[i]$ is the *prognosis expert value*; the *weight values* $\{w^{(k)},w^{(2)},w^{(3)},w^{(4)},w^{(5)}, w^{(6)},w^{(Ex)}\}$, assigned by experts or calculated (such method for $w^{(Ex)}=0$ was developed [4,5]).

In the appropriate calculations we use the following 6 *types of trends*: (1) *Linear (Lin)*, (2) *Exponential (Exp)*, (3) *Logarithmic (Log)*, (4) *Polynomial (Pol) 3-rd order*, (5) *Power (Pow)*, (6) *Hyperbolic (Hpr)*. Further, we consider the sense and calculation methods for all Criteria, using the trend-prognosis values and connected with the formula (4) realization.

The i-stocks, having the greatest expected *prognosis* (for (T+1)-day) *absolute values* $\{C_1[i]=A^{pr}(i,T+1),\ i=1,...,I\}$ *for deal sums (DS)*, are preferable with the position of stocks' buying for (T+1)-day, since *the stocks with the greatest values $A^{pr}(i,T+1)$ might be in great demand for* (T+1)-*day and some next days*. Thus, we use *the maximization* (on all i–stocks, $i=1,...,I$) *of the prognosis* (for (T+1)-day) *absolute values* $\{A^{pr}(i,T+1), i=1,...,I\}$ *for deal sums* as the **Criterion 1**:

$$max\ \{C_1[i] = A^{pr}(i,T+1),\ i=1,..,I\} \quad (5)$$

where the prognosis (for any prognosis t-day) values $A^{pr}(i,t)$ are calculated, according to (4), using the trend-prognosis values $\{A^{(1)}(i,t),A^{(2)}(i,t),A^{(3)}(i,t),A^{(4)}(i,t), A^{(5)}(i,t),A^{(6)}(i,t)\}$ as follows:

$$A^{pr}(i,t)=A^{(1)}(i,t)w^{(1)}+A^{(2)}(i,t)w^{(2)}+\ldots +A^{(5)}(i,t)w^{(5)}+A^{(6)}(i,t)w^{(6)}+A^{(Ex)}(i,t)w^{(Ex)} \quad (6)$$

/

Here, all trend-prognosis values $\{A^{(k)}(i,t),\ k=1,...,6\}$ are calculated for all i-stocks ($i=1,..,I$) on a basis of DS statistical data processing for the period $[1,..,t-1]$. For our case t=T+1, the *k-trend-prognosis DS absolute values* $\{A^{(k)}(i,T+1),k=1,..,6\}$ are calculated for the period [1,..,T].

If all these prognosis values are obtained and the *expert values* $A^{(Ex)}(i,T+1)$ are added., the required *prognosis values $A^{pr}(i,t)$ of Criterion 1* are calculated for all i-stocks ($i=1,..,I$) using formula (6), where the weight values $\{w^{(1)},w^{(2)},w^{(3)},w^{(4)},w^{(5)},w^{(6)},w^{(Ex)}\}$ are assigned by experts or calculated. The last case we have [4,5] when the expert values $A^{(Ex)}(t)$ aren't used ($w^{(Ex)}=0$).

The **Criterion 9** model is the same one, but we prognosis the SP values. Thus, we should change the parameters of formulas (5),(6) on the SP parameters: $B^{pr}(i,T+1)$; $\{B^{(k)}(i,t),k=1,..,6\}$, reflecting *the relative* (according to c_{il} ^ in (2)) *trend-prognosis SP values*; $B^{(Ex)}(i,t)$.

The **Criterion 6** realizes the same (5),(6) to define the DS prognosis values $A^{pr}(i,T+S+1)$, but for the (T+S+1)-selling day, using the trend-prognosis values $\{A^{(k)}(i,T+S+1),k=1,..,6\}$. Since our prognosis computations are performed in T-day, we have not the DS fact-statistical data for the days [T+1,...,T+S]. To overcome this "bottleneck", we propose 2 approaches:

A) The statistical k-trend-prognosis, determined for the period [1,...,T], are "as if prolonged" to the (T+S+1)-day, i.e. we accept x= T+S+1 in these k-trend-prognosis formulas.

B) We "extend" the statistical basis for the determination of such k-trend-prognosis to the period [1,..,T+S], using the method: 1) we find the k-trend-prognosis values for the (T+1)-day and add them (as any new "fact" data) to the statistical data for the period [1,..,T]; 2) the k-trend-prognosis values for the (T+2)-day are defined by processing "fact-data" for the period [1,..,T+1], and such "statistical basis" is extended up to the period [1,.., T+S], and for it we determine the required k-trend-prognosis values $\{A^{(k)}(i,T+S+1),\ k=1,..,6;\ i=1,..,I\}$.

Thus, we form **the calculations models for** determining the **Criteria 6A/6B**.

Comparing the DS prognosis values $\{A^{pr}(i,T+1),i=1,...,I\}$ with the DS fact values $\{A^{fact}(i,T),\ i=1,...,I\}$, taken for the last T-day from the statistical data, it's possible to take their differences as the estimates of quality of these DS prognosis values, derived using trends. If the absolute value of such difference is lesser for the considered i^*-stock as compared with other i-stocks, then the DS absolute prognosis value $A^{pr}(i^*,T+1)$ may be considered as more reliable. With such positions, this i^*-stock is considered as more preferable. However, such estimates are actual only for a very short-term period. Thus, we accept the following **Criterion 2**: *the minimization of the relative values* $\{D^{pr}(i,T+1),\ i=1,...,I\}$ *of changing the DS absolute prognosis values in comparison with their fact* (statistic) *values of the preceding* T-*day*, where the *Criterion 2* are presented as follows:

$$min\ \{C_2[i] =D^{pr}(i,T+1)=|(A^{pr}(i,T+1)-A^{fact}(i,T))|\ /\ A^{fact}(i,T),\ i=1,..,I\} \quad (6)$$

The same principle is realized for the **Criteria 7/8** models, reflecting *the maximization of the relative values* $\{D^{pr}(i,S)/P^{pr}(i,S)/,i=1,...,I\}$ *of changing* (for S days) *the prognosis (for the selling (T+S+1)-day) DS absolute / SP relative values in comparison with the same type of values for the buying (T+1)-day* (both a.m. ***A/B*** versions are considered for these *Criteria*):

$$max\ \{C_7[i] = D^{pr}(i,S)=|A^{pr}(i,T+S+1)-A^{pr}(i,T+1)|/\ A^{pr}(i,T+1),\ i=1,..,I\} \quad (7)$$

$$max\ \{C_8[i] = P^{pr}(i,S)=|B^{pr}(i,T+S+1)-B^{pr}(i,T+1)|/\ B^{pr}(i,T+1),\ i=1,..,I\} \quad (8)$$

The **Criteria 3-5 models**, making the prognosis (for the buying (T+1)-day) of SP values, take into account the **sinusoidal character of changing the SP values** during the period (the days [1,..,T]) of these SP statistical data observation. This sinusoidal character is dictated by the nature itself of stock buying-selling on stock exchange: when the SP fall, the volumes of such stocks buying increase, and this tendency is remained up to the days of reaching such SP minimum. After this, another picture is observed, when the SP are growing up, and the holders of these stocks begin to sell such stocks, that leads again to SP fall, and so on. Accounting for it, *it's expediently to buy those stocks, for which are expected the current minimum of their SP sinusoids are closed to the prognosis (T+1)-day.* Using this principle, *the Criteria 3-5* values are determined, based on the calculation of some such SP sinusoids' parameters. Their values for each i-stock, calculated on a basis of constructing its set of "j-sinusoidal Hills" $\{H[i,j],j=1,..,J[i]\}$, allow to select i-stocks with the a.m. minimums [4,5].

3.4. The Solution Models for the problem of "reasonable" i-stocks' selection to buy-sell them

In a framework of the proposed MCDM methodology applying, we form 4 *Solution Models*, based on using various combinations of using the a.m. Criteria calculation models:

Model 1, oriented only on buying *i*-stocks at the (T+1)-day, takes into account the Criteria 1-5. It's proposed the resulting "reasonable" *i*-stocks (RAS), which are found accounting for the prognosis maximal DS/minimal SP values in (T+1)-day, might be freely used (to sell, to accumulate) further by their holders using any additional, mainly non-formal, estimations.

Model 2, oriented on the "reasonable" *i*-stocks' accumulation for the future to sell them for a maximal profit obtaining or their holders' capital increasing, takes into account the Criteria 1-5,9 or 1,2,9 only. Here, finding the prognosis on the buying (T+1)-day, we attempt also to estimate the increasing tendency for SP values (Criterion 9) to maximize the holder capital.

Models 3/4 consider the prognosis operations of *i*-stocks buying at the (T+1)-day jointly with their selling at the (T+S+1)-days, where S are the predetermined natural numbers {e.g., 1,3,...). Both these models account for various combinations of Criteria 1-5,6A/B-8A/B (e.g.,{1-5,6A-8A},{1-5,6B-8B},{1,2,6A-8A},etc.). The distinction between these models consists in the following: **Model 3** *prognosis only one* S^-*day of stocks' selling* with the fixed such (T+S^+1)-day; **Model 4** – *several* (>1) *such* S^-*days* (e.g. S^=1,..,Smax) and its result should reflect the mean resulting estimations, reached for all considered S^-days. Thus, **Model 3** corresponds to a rare situation when it's needed to sell the "reasonable" stocks only in one fixed (T+S^+1)-day (e.g. to accumulate money by the profit obtaining for a duty payment at the next day); **Model 4** corresponds to the more natural situation of speculations on the stock exchange during several nearest S-days to increase the holders' capital. Here, we could use the mean (for all considered S-days) results, when we buy *i*-stocks at (T+1)-day.

4. Testing the proposed MCDM approach and Models of stock buying-selling

We make this testing as a *"debug-regime"*, based on *using the fact data for the past period* [1,..,T,T+1,..,T+1+Smax], that allows: (a) to reach the RAS versions by the ISA/ISCAV decreasing applying *the six-level hierarchical system of MVC series*; (b) to compare the "reasonable" *i*–stocks from the RAS with the *"fact-good" i–stocks*, obtained by analysis of fact data from the whole past period [1,..,T+1+Smax]; this analysis allows to estimate the quality of the prognosis results, obtained by processing the fact-statistical data for the period [1,..,T]. On this way, we met *the difficulties in the exposes of "fact-good" i–stocks*, where these difficulties are initiated by a contradiction between the SP increasing and the DS low levels, observed for the same *i*–stocks. This contradiction might overcome, on our opinion, using any intuitive compromises: the acceptance of limiting levels and adequate weights for the Criteria values, etc.

The real sample, considered early for the testing of Model 1 **[4,5]**, presented the fact (statistic) data, reflecting the "moving" of deal sums (DS) and stock prices (SP) for all "Tel-Aviv-100" *i*-stocks {*i*=1,...,97} of the Israeli stock exchange for the past period (11/11/01-02/01/02) (in all for 33 days). We considered the day 01/01/02 as T-day, and 02/01/02 as the prognosis (T+1)-day. Thus, the period [1,..,T], reflecting the statistic data needed to construct all *Criteria 1-5*, included in the *Model 1*, is 11/11/01-01/01/02 (T=32). The statistic data for the prognosis 33-day were not used in a framework of six-level system calculations, but they were used further to estimate the derived results (in the *debug-process*). To illustrate this six-level system performance, we accepted the *l-scenarios* for all *l*-levels (*l*=1,..,6) of this six-level system (see this paper Section **2** as well as **[5]**):

***l*=1)** the ISA (1 version, K=1), presenting the part of the "Tel-Aviv-100" *i*-stocks {*i*=1,...,20};

***l*=2)** the ISCAV (M=1), including *Criteria 1-5* (according to *Model 1)* values, calculated by the DS/SP processing for the data, taken for the period [1,..,T] (T=32) and for all 20 *i*-stocks;

***l*=3)** the TOPSIS modification (1)-(2) as the multi-criteria optimization technique used (N=1);

***l*=4/5)** 24 versions of the random weight possible value intervals (3) to realize the function (1), that reflects all possible combinations of 6 versions (*p*=1,..,6; P=6) of interval (3) "central points" (4-scenarios) and 4 versions (R=4) of interval (3) bound values (5-scenarios);

***l*=6)** 2 versions (*q*=1,2) of Monte Carlo simulation (MCs) series (6-scenarios). Each MCs series includes 5 MCs, each of which allows fix the interval (3) interior value.

The calculations performed according to the proposed MCDM methodology and the a.m. *l*–scenarios system (*l*=1,..,6) using this real sample data allowed to find a set of RAS versions. Their analysis led to finding a final solution, including 5 best *i*-stocks. Their comparison with the "good-fact" stocks, found using a specific analysis, showed their coincidence on 40-60%.

At present, the testing is performed for the full ISA of this sample (*i*=1,..,97), the extended period of statistical data (up to 40 days with varying values T and Smax) and for Models 3/4.

In the future, such testing will be performed on-line, for the current data of exchange.

Conclusions

The performed work showed the proposed MCDM approach might be applied for the considered problem solution. Further we will accent on practical aspects of such applying.

References

1. Kalika,V., Frant, S. (2000) "Multicriteria Optimization Accounting for Uncertainty in Dynamic Problem of Power Generation Expansion Planning"*Research and Practice in Multiple Criteria Decision Making*. Haimes YY; Steuer RE. (eds.). Lecture Notes in Economics and Mathematical Systems, Springer-Verlag 487,409-420.

2. Kalika, V., Frant, S. (2001) "Multi-criteria Approach for Power Generation Expansion Planning" *Multiple Criteria Decision in the New Millennium*. Koksalan M, Zuionts S. (eds,).Lecture Notes in Economics and Mathematical Systems, Springer, 507, 458-468.

3. Kalika, V. (2001) "Multi-Level Hierarchical System to Take into Account Uncertainty in Multi-Criteria Decision Making Process" *Proceedings of the Sixth International Symposium on the Analytic Hierarchy Process, ISAHP 2001- ISAHP Switzerland 2001*, 173-180.

4 Kalika,V. and Rossinsky, G. (2003) 'Methodology of multi-criteria decision making accounting for uncertainty and some its application', *IJMDM (International Journal of Management and Decision Making)*, 4 , Nos. 2/3, 240-272.

5. Kalika,V. (2004) 'A new methodology of accounting for uncertainty factors in multiple criteria decision making problems', *JTIT (Journal of Telecommunications and Information Technology)*, 3/2004, 44-56.

6. Massam, B.N. (1988). 'Multi-criteria decision making (MCDM): Techniques in planning'. *Progress in Planning* 30, 1-84.

USING GAME THEORY IN ANALYSIS OF MOBILE TELECOMMUNICATIONS MARKET

Lada Rokov
Student of postgraduate study in Operations Research, Faculty of Economics, University of Zagreb
e-mail: rola4@net.hr (or rola4@yahoo.com)

Abstract
This is analysis of telecommunications market where two symmetric mobile operators compete for final prepaid customers but cooperate in setting the price for access to each other's customers. I presented model of four-stages game using Hotelling specification and Bertrand price competition under assumptions of two-way access and termination-based price discrimination.

Keywords: Game theory, Network competition, Telecommunications, Network externalities, Price discrimination.

Introduction

A distinctive feature of the telecommunications industry is the need for competing networks to interconnect in order to provide each customer with access to the entire consumer population (so called two-way access). This leads to an interesting coexistence of vertical and horizontal relationships, where each network is the monopolistic supplier of an intermediate input – the termination of calls directed towards its own subscribers – to its rivals in the retail market, and in turn relies on them to provide access to their customers. It is generally agreed that the outcome will be socially inefficient if each firm sets its termination fee independently, because a unilateral increase by any firm raises its rivals' costs without affecting its own.

Until local competition is developed, two-way access in telecommunications were mainly confined to two types of interconnection (as mentioned by Laffont and Tirole):

- international calls: telecommunications carries have traditionally enjoyed a monopoly position in their respective countries and cooperated in the determination of termination access charges. These settlement charges are determined in private negotiations between the carriers. They are set above the marginal cost of terminating calls. Such high termination charges may be quite inefficient even from the point of view of profit maximization.
- fixed-mobile interconnection: the offering of mobile services by companies other than the incumbent fixed-link operator has given rise to a second set of two-way termination access charges. Charges for termination on a mobile network are often set quite high.

Cooperation among competitors in setting access charges is relatively new in the telecommunications industry and the telecommunications industry and regulators are currently defining principles for interconnection.

Assumptions of two-way access price economic theory are:

- calling company pays termination charge: the call receiver's network charges an amount a per minute to the caller's network for termination. The call receiver pays nothing.
- unregulated retail markets: telecommunications carriers are free to set the prices they wish to final consumers.

Under two-way access, we distinguish between two kinds of calls:
An on-net call is a call that originates and terminates on the same network. That is, the caller and the receiver belong to the same network. In contrast, an off-net call occurs when the two parties belong to different networks.
Retail tariffs offered by operators in most mobile-telephony markets often discriminate between on-net and off-net calls (so called termination-based price discrimination).
A regulator can be interested in this type of pricing because the traffic-dependent costs of on-net and off-net calls may be different due to cost reasons, such as:

- different efficiency levels of the operators' networks;
- terminating access prices set above the traffic-dependent cost of the local loop (so that the traffic-dependent cost of an on-net call is higher than the cost of an off-net call).

Accordingly, the idea is that termination-based price discrimination may result in per-minute prices that reflect the underlying costs and access fees to a better extent than non-discriminatory prices.
To the extent that interconnection fees and costs are publicly observable, it seems highly doubtful that most networks terminate each others' off-net calls at a discount.
When off-net calls are more expensive than on-net calls, then consumers choose the same network as those whom they call the most, so the majority of calls is made at the on-net rate (so called positive network externality). For this reason, the networks tend to prefer a positive markup on interconnection.

Model

Here is presented model based on Hotelling specification and Bertrand competition (which is price competition with the commitment to supply whatever demand is forthcoming at the set price). The model is similar to the one discussed by Peitz, but I focused my analysis on telecommunication operators with prepaid customers.

The point of departure for this model are two operators practising termination-based price discrimination.
Operators i and j offer telephone services to prepaid end users. Both operators use the same technology, so they have symmetric cost structures. Each operator has a full-coverage network, and all consumers (total mass of consumers is n) are subscribed to either one of the two operators. Consumers of both operators do not pay monthly subscriber charge (so called prepaid customers).

The game has four stages, and the order of moves is as follows:
First, in the cooperative stage, the networks jointly determine the access charges. Second, in the competitive stage, they independently and non-cooperatively set their tariffs.
Third, each costumer subscribes to exactly one network.

Finally, consumers place their actual calls according to their calling patterns and tariffs, and payoffs are realized. As usual, the model is solved backwards.

Stage 4 (consumption - demand)

Under termination-based price discrimination (if on-net and off-net calls have different prices), a consumer makes longer or more frequent calls to those who are subscribed to the same network.
Given a price per minute equal to p (p_i^{on} for on-net or p_i^{off} for off-net), each consumer has an individual demand of $x[p]$ call minutes, and derives utility $u[x]$ from calling for x minutes. Consumers derive utility from calls made to consumers subscribed to the same network (their share is s_i) and from calls made to consumers subscribed to the other network.

The indirect utility that a consumer derives from a call at a per-minute price of p can be expressed as:
$\tilde{v}[p] = u[x[p]] - px[p]$
where $x[p_i] = \arg\max_x \{u[x] - xp_i\}$ is the optimal call length given p (individual demand is derived by solving the first-order condition for utility maximization $u'[x] = p_i$).

In addition to the utility from calling, each costumer derives a fixed utility level U_i from subscribing to network i, independent of the number of telephone calls that are made, which may come from, for example:

- brand preference
- services offered in addition to voice telephony (SMSes, VAS)
- the quality of the network and services delivered
- having a telephone connection in the case of unforeseen events
- the possibility to make free calls (ambulance, police,...).

It is assumed that the fixed utility U_i is the same for all consumers of network i.

The total net utility of a subscriber to network i, who optimally chooses its calling time, is composed of a traffic-dependent (utility derived from on-net calls plus the utility derived from off-net calls) and a traffic-independent part U_i, and is written as:

$$v_i[p_i^{on}, p_i^{off}] = U_i + s_i \tilde{v}[p_i^{on}] + (1 - s_i)\tilde{v}[p_i^{off}]. \qquad (1)$$

Traffic dependent part only depends on the calls made and not on the calls received. Clearly, net utility is decreasing in prices.
To guarantee that consumers want to participate in the market, net utility level $v_i[p_i^{on}, p_i^{off}]$ has to be positive.

Stage 3: Subscription

The subscription stage is best described as a market for network membership with unit demand and positive network externalities (positive network externality is when product becomes more valuable as more customers use them) on the consumption side.

It is assumed that networks are horizontally differentiated and Hotelling specification is used.
Consumers are assumed to be uniformly distributed on the $[0,1]$-interval. Operator 1 is located at $l_1 = 0$ and operator 2 at $l_2 = 1$. Consumer $z \in [0,1]$ incurs a disutility $-\theta|l_i - z|$, which is linear in the distance between consumer location and the location of the operator. The parameter θ expresses the substitutability between networks: if $\theta = 0$, networks are perfect substitutes; the larger θ, the more differentiated networks are (larger θ makes it more difficult to gain market share).

Each consumer subscribes to exactly one operator. The consumer who is identified by its location z subscribes to operator 1 if $v_1[p_1^{on}, p_1^{off}] - \theta z > v_2[p_2^{on}, p_2^{off}] - \theta(1-z)$, where $v_i[p_i^{on}, p_i^{off}]$ denotes the net utility of a network at the ideal location z.

The realized market share of operator i is equal to:

$$s_i = \frac{1}{2} + \frac{v_i[p_i^{on}, p_i^{off}] - v_j[p_j^{on}, p_j^{off}]}{2\theta}, \text{ or if using (1):}$$

$$s_i = \frac{1 + \frac{1}{\theta}(U_i - U_j) + \frac{1}{\theta}\left(\tilde{v}[p_i^{off}] - \tilde{v}[p_j^{on}]\right)}{2 - \frac{1}{\theta}\left(\tilde{v}[p_i^{on}] + \tilde{v}[p_j^{on}] - \tilde{v}[p_i^{off}] - \tilde{v}[p_j^{off}]\right)}.$$ Clearly, it is presumed that $s_j = 1 - s_i$.

The market shares resulting from competition and consumers' choices are denoted by $s_1 = s_1[p_1^{on}, p_1^{off}, p_2^{on}, p_2^{off}]$ and $s_2 = s_2[p_1^{on}, p_1^{off}, p_2^{on}, p_2^{off}]$, and as we see, market shares depend on all retail prices.
Consumers base their decision on which network to join on market share because, for instance, at a low on-net price an operator is attractive if it has a large market share. This means that consumers' choices are affected by the market share. Therefore, termination-based price discrimination gives rise to a positive network externality.

Stage 2: Competition in tariffs

In order to specify profit functions and corresponding retail prices, costs must be defined. The fixed costs include all costs that don't depend on traffic (these costs are denoted by f_i). Such costs include, for instance, the costs of building a backbone, the maintenance cost of the local loop, loans repayments, cost of billing, etc. It was implicitly assumed that in the calculation of marginal costs, no traffic-independent costs are included.
Total traffic-dependent costs include an operator's marginal cost (real costs of originating and terminating calls) plus per-minute charges paid to other operator for interconnection and access denoted by a_i (or a_j).

Let c_{ik} denote operator i's traffic-dependent cost per-minute associated with a telephone call of type k and let a_i denote terminating access prices paid to operator i. We distinguish three types of telephone calls under assumption that both operators use the same technology:
- on-net calls: calls that originate and terminate on a single operator's network $(k = 1)$ with associated costs c_1 for operator i;

- off-net calls: calls that terminate on another operator's network $(k = 2)$ with associated costs $c_2 + a_j$ for operator i (on-net costs plus wholesale access price paid to the other operator);
- incoming calls: calls that originate from another operator's network $(k = 3)$ with associated costs c_3 and wholesale access price a_i for operator i.

It is assumed that $c_1 = c_2 + c_3$ so that society's marginal cost of an off-net call is the same as the one of an on-net call.

Before profit functions can be derived, we have to specify calling pattern.
We make the assumption that calling patterns are balanced in the following sense: consumers' share of on-net call minutes corresponds to the market share of the operator to which a consumer is subscribed to, after correcting for differences in per-minute prices. That is, the total volume of call minutes that originates and terminates on network i is $ns_i x[p_i^{on}]$, and the demand for off-net calls is $ns_i x[p_i^{off}]$.

Operator i's profits are:

$$\pi_i[p_i^{on}, p_i^{off}, p_j^{on}, p_j^{off}] = ns_i s_i x[p_i^{on}](p_i^{on} - c_1) + ns_i(1 - s_i)x[p_i^{off}](p_i^{off} - c_2 - a_j)$$
$$+ n(1 - s_i)s_i x[p_j^{off}](a_i - c_3) - ns_i f_i.$$

By Bertrand competition, operators set their on-net per-minute price equal to on-net marginal costs, that is:
$p_i^{on*} = c_1$,
and their off-net per-minute price equal to off-net perceived marginal costs, that is:
$p_i^{off*} = c_2 + a_j$.

This follows directly from the first-order conditions $\frac{\partial \pi_i}{\partial p_i^{on}} = 0$ and $\frac{\partial \pi_i}{\partial p_i^{off}} = 0$.

Consequently, operator i's profits in which operator i chooses its profit-maximizing per-minute prices can be written as:
$\pi_i[p_i^{on*}, p_i^{off*}, p_j^{on}, p_j^{off}] = n(1 - s_i)s_i x[p_j^{off}](a_i - c_3) - ns_i f_i$.
Accordingly, operators have only one source of profit - revenue from incoming calls.

Stage 1: Determination of access charge

This stage can be cooperative - when two networks jointly determine access charges a_i (or a_j) and corresponding access markup (difference between access charge and marginal cost of terminating call per minute), in which case operators usually mutually set reciprocal price $(a_i = a_j)$.
There are some cases when access charges are determined by regulator:
- when operators can not make an agreement
- when regulator wants to stimulate competition – this asymmetric access pricing exists as long as there is one dominant player in the market (especially in the recently liberalized markets)

Operators' cooperative agreement on access charge immediately raises the question of whether an unregulated access charge can be an instrument of collusion in the retail market.

Conclusion

This paper has been focussed on a telecommunications market where two operators compete for retail prepaid customers under two-way access pricing and termination-based price discrimination.
An interesting feature of this new form of competition is that termination-based price discrimination reintroduces network externalities among consumers. Interconnection together with uniform (non termination-based) pricing implies that consumers, when choosing a network, do not take into account the choice of network by the people they want to call. In contrast, if on-net calls are cheaper than off-net calls, consumers are better off if the people they want to call select the same network. More generally, an access markup (access charge above marginal cost) generates positive network externalities among consumers.

Limitation of marginal cost pricing

Social welfare is achieved by setting prices equal to marginal cost. But there are some problems with marginal cost pricing as discussed by Courcoubetis and Weber that follow:
Marginal cost prices can be difficult to compute and they can be close to either zero or infinity. This is a problem since telecommunication networks typically have fixed costs which must somehow be recovered.
Think of a network that is built to carry x calls. The main costs of running the network are fixed costs (such as maintenance, loans repayments, cost of billing, subsidy on handsets and staff salaries), i.e. invariant to the level of usage. Thus when less than x calls are present, the short-run marginal cost that is incurred by carrying another call is near zero. If the network is critically loaded (so that all x circuits are busy), the cost of expanding the network to accommodate another call could be huge.
Another difficulty in basing charges on marginal cost is that even if we know the marginal cost and use it as price, it can be difficult to predict the demand and to dimension the network accordingly. There is a risk that we will build a network that is either too big or too small. Prices are used to signal the need to expand the network.
However, welfare maximization is not the only thing that matters. A firm's prices must ensure that it is profitable, or at least that it covers its costs.
In reality, operators typically impute fixed costs to telephony traffic, enabling them to define a reference point for prices.

References

Courcoubetis, Costas; Weber, Richard, 2003.: "Pricing Communication Networks", Wiley, 130.
Laffont, Jean-Jacques; Tirole, Jean, 2000.: "Competition in Telecommunications", MIT Press, 179-180.
Peitz, Martin, 2004.: "Assymetric access price regulation in telecommunications markets", European Economic Review, 3-8.

AN EXAMPLE OF A GAME AND ITS COMBINATORIAL ANALYSIS

Damir Vukičević[1], Janez Žerovnik[2,3]
[1] Department of Mathematics, University of Split, Nikole Tesle 12, HR-21000 Split, Croatia
[2] University of Maribor, Faculty of Mechanical Engineering, Smetanova 17, 2000 Maribor, Slovenija
[3] IMFM/DTCS, Jadranska 19, 1000 Ljubljana, Slovenija
Emails: vukicevic@pmfst.hr, janez.zerovnik@imfm.uni-lj.si

Abstract. We give an example showing how graph theory may be used to model and analyse practical situations in economics. The example given is a game with three players, perfect information, and known rules of a coalition of two players. An optimal strategy for the third player is given.

Keywords: combinatorial game theory, cooperative game, optimal strategy

1.Introduction

Fifty years old, thanks to the 1994 Nobel Memorial prize to Nash, Selten and Harsanyi, game theory and its application to economics are becoming increasingly popular. Game theory is a distinct and interdisciplinary approach to the study of human behavior. The disciplines most involved in game theory are mathematics, economics and the other social and behavioral sciences. Game theory (like computational theory and so many other contributions) was founded by John von Neumann [Neumann and Morgenstern].

The aim of this paper is to show how combinatorics, in particular graph theory can be used for analysis of a game. Although our example is not very complex, we claim that the analysis given shows the potential of graph theory [Wilson and Watkins] to provide a simple and efficient method for such tasks. We will model the game by a directed graph G. The set of vertices of G denoted by $V(G)$ will represent the set of all possible states of business affairs and the set of arcs $A(G)$ of graph G will represent changes from one state of business affair to the other state of business affair which can, in the example outlined below, be caused solely by one of our actions and predicted reactions from the opposition. There is also the function f which tells how much money we earn during this transition (of course f can take negative values if we spend or loose money in such transition). Our aim is to maximize (in long terms) our profit.

We will define a game, which may be motivated by a practical situation where we have a small number of competing businesses. We assume the players have agreed on some rules and play fair in the sense that they tend to increase their profit but do not attempt to decrease the profit of others. Cooperative games are particularly important in economics. The idea of cooperative games is illustrated by the following example [McCain].

Example. We suppose that Joey has a bicycle. Joey would rather have a game machine than a bicycle, and he could buy a game machine for $80, but Joey doesn't have any money. We express this by saying that Joey values his bicycle at $80. Mikey has $100 and no bicycle, and would rather have a bicycle than anything else he can buy for $100. We express this by saying that Mikey values a bicycle at $100.

The strategies available to Joey and Mikey are to give or to keep. That is, Joey can give his bicycle to Mikey or keep it, and Mikey can give some of this money to Joey or keep it all. it is suggested that Mikey give Joey $90 and that Joey give Mikey the bicycle. This is what we call "exchange." Here are the payoffs:

Table 1

		Joey	
		give	keep
Mikey	give	110, 90	10, 170
	keep	200, 0	100, 80

Explanation: At the upper left, Mikey has a bicycle he values at $100, plus $10 extra, while Joey has a game machine he values at $80, plus an extra $10. At the lower left, Mikey has the bicycle he values at $100, plus $100 extra. At the upper left, Joey has a game machine and a bike, each of which he values at $80, plus $10 extra, and Mikey is left with only $10. At the lower right, they simply have what they begin with -- Mikey $100 and Joey a bike.

If we think of this as a noncooperative game, it is much like a Prisoners' Dilemma. To keep is a dominant strategy and (keep, keep) is a dominant strategy equilibrium. However, (give, give) makes both better off. Being children, they may distrust one another and fail to make the exchange that will make them better off. But market societies have a range of institutions that allow adults to commit themselves to mutually beneficial transactions. Thus, we would expect a cooperative solution, and we suspect that it would be the one in the upper left.

In the next section we give a motivation and a definition of the game that we will analyse. A model using graphs is given in section 3 while section 4 gives an efficient method to find optimal strategies. In the last section the method is applied to the example given in section 2.

2. Example

Let us illustrate our ideas by using a very simple example. Suppose that three people, *A*, *B* and *C* sell ice-cream on the beach and suppose that there are 25 convenient locations for selling ice-cream that are equally distributed along the beach. In order not to fight among themselves the following agreement is obtained:
1) No two sellers will sell ice-cream on the same location.
2) First, *A* chooses any location and sells ice-cream, after a quarter of an our *B* chooses any location different then one chosen by *A*, and after another fifteen minutes *C* chooses any location (different from ones occupied by *A* and *B*).

3) After another fifteen minutes *A* can relocate his vehicle to any location (different form ones occupied by *B* and *C*; *A* can choose to stay at his present location).
4) After another fifteen minutes *B* can relocate his vehicle to any location (different form ones occupied by *A* and *C*; *B* can choose to stay at his present location).
5) After another fifteen minutes *C* can relocate his vehicle to any location (different form ones occupied by *A* and *B*; *C* can choose to stay at his present location).
6) Steps 3) – 5) continuously repeat

Remark. If the Step 2 of the agreement sounds to favour player *A* too strongly, one can prefer to study a variant of the game where the first player and the order is determined randomly.

Further, it can be assumed that people on the beach are equally distributed and that they buy ice-creams from the closest seller of the ice-cream. Let us explain this on one simple example. Suppose that *A* is located on the 4^{th} location, *B* on the 15^{th} location, and *C* on the 23^{rd} location. *A* will sell the ice-creams to the people on the beach near locations 1, 2, 3, 4, 5, 6, 7, 8, and 9, hence he will sell 36% of all ice-creams. *B* will sell the ice-creams to the people on the beach near locations 10, 11, 12, 13, 14, 15, 16, 17, and 18 and also to the half of the people near the location 19 (the other half buys ice-cream from *C*), hence he will sell 38% of all ice-creams. *C* will sell the ice-creams to the people on the beach near locations 20, 21, 22, 23, 24 and 25, and also to the half of the people near the location 19 hence he will sell 26% of all ice-creams.

Suppose that *C*'s strategy is known. He always places his vehicle to such place that he makes the greatest profit in the following fifteen minutes. If there are more than one most profitable location, *C* chooses location that will make the greatest profit of *B* in the next 15 minutes. *B* uses the fact that *B* relocates his vehicle after *C* and that he knows the *C*'s strategy. Hence, he can locate his vehicle in order to optimize his profit in his next 30 minutes. If there are more than one most profitable location, *B* chooses location that will make the greatest profit of *C* in the next 15 minutes. If still the location is not uniquely defined *B* chooses location with the smaller number. Note that *B* and *C* are "in the weak coalition". This increases their profit and reduces the profit of *A*. But since their strategies are known, *A* can adopt his strategy in order to obtain the greatest profit.

Here the state of business affairs is described by the location of the vehicles *B* and *C*. Hence,

$$V(G)=\{(x,y):1\le x,y\le 25, x\ne y\}\cup\{\varnothing\}.$$

The ordered pair (x,y) means that *B* is located at location x and *C* is located at the location y, while $\varnothing$ denotes the position at the very beginning of their business, when no vehicle is located. After *A* relocates his vehicle, it is uniquely defined (since both strategies are known) where vehicles *B* and *C* will be located. Hence, the reaction of *B* and *C* to any action of *A* is uniquely defined. The function f assigns the profit of *A* (i.e. the percentage of the market taken by *A*) in the next 45 minutes. Our aim is to find the cycle with the greatest average value of f on its edges which can be reached from the initial state.

3. Definitions

Denote by $isv(G)$ the vertex of G which represents the initial state vertex. Let $f: A(G) \rightarrow R$ (where R is the set of real numbers) be the function called *profit function*. Let H be the subgraph of G which consists of all vertices that can be reached from the vertex $isv(G)$.

Let C be any oriented cycle in H. Denote

$$av(C) = \frac{\sum_{e \in A(C)} f(e)}{card(A(C))}. \qquad (1)$$

where $card(A(C))$ is the cardinality (or in the finite case the number of elements) of the set $A(C)$.

The aim of this analysis is to find an efficient algorithm that finds one cycle C' such that

$$av(C) \leq av(C') \text{ for each oriented cycle } C \text{ in } G. \qquad (2)$$

Note that cycle C' does not have to be unique. In this case the algorithm (to be explained in the next section) returns one of the possible shortest cycles that satisfies the relation (2). Denote by *SetOpt* the set of all such cycles.

4. The algorithm

One can easily obtain the graph H when G and $isv(G)$ are known. Denote by n the number of vertices in H. Let us inductively define the numbers $a_{i,j,k}$, $1 \leq i, j, k \leq n$ by:

$$a_{1,j,k} = \begin{cases} f(jk), & jk \in A(H) \\ -\infty, & jk \notin A(H). \end{cases}$$

$$a_{i+1,j,k} = \max_{1 \leq x \leq n} \{a_{i,j,x} + a_{1,x,k}\}$$

It can be easily seen that

$$a_{i,j,k} = \max\left\{ \sum_{e \in A(P)} f(e) : P \text{ is path from } j \text{ to } k \text{ of length } i \right\},$$

i.e. $a_{i,j,k}$ is the maximal sum of edge weights over all paths from j to k of length i. In other words, $a_{i,j,k}$ represents the profit of the most profitable path from j to k of length i. Now, denote by $b_{i,j}$, $1 \leq i, j \leq n$ the pair of the numbers $\left(\frac{a_{i,j,j}}{i}, -i\right)$.

Let (i_0, j_0) be the pair of numbers such that $b_{i_0,j_0} \geq_{lex} b_{i,j}$ for each $1 \leq i, j \leq n$, where $\leq_{lex}$ is the lexicographical order. If $b_{i_0,j_0} >_{lex} (-\infty, i_0)$, then there is the cycle $C \in SetOpt$ of length i_0 which passes through the vertex j_0. Denote vertices of this cycle by $v_1, v_2, \ldots, v_{i_0}$ in order of their appearance in such way that $v_1 = j_0$. We can inductively reconstruct the remaining vertices by the following procedure:

$$\text{for each } p = 1, 2, \ldots, i_0 - 1, \text{ let } v_{p+1} = q \text{ such that } a_{p-1,j_0,v_p} + a_{1,v_p,q} + a_{i_0-p,q,j_0} = a_{i_0,j_0,j_0}$$

Note that the choice of q does not necessarily have to be unique, but by choosing different values of q, we always get the cycle with the required properties.

It is clear from the definition that this algorithm can be executed in polynomial time in the number of vertices and that this polynomial is of degree 4. Hence, there is an efficient algorithm that solves this problem.

5. The solution of the example

Now, we use our algorithm to find the solution of the example described in the second section. Recall that $V(G) = \{(x, y): 1 \leq x, y \leq 25, x \neq y\} \cup \{\emptyset\}$. Hence, G has 601 vertices. For this example, we have created graph G by computer. It has 12869 arcs. After that, we have created graph H. It has 59 vertices and 1298 edges. The optimal cycle reads as

$$\overline{\overline{\underline{\underline{(19,6)}}}} \xrightarrow[f=98/3]{a=18} (8,19) \xrightarrow[f=106/3]{a=7} (20,8) \xrightarrow[f=98/3]{a=19} (7,20) \xrightarrow[f=98/3]{a=8} (18,7) \xrightarrow[f=106/3]{a=19} (6,18) \xrightarrow[f=98/3]{a=7} \overline{\overline{\underline{\underline{(19,6)}}}},$$

where a denotes the place where A shall relocate his vehicle. Note that the average profit is $\frac{302}{9} > \frac{100}{3}$, hence using this strategy A earns more than the average, which is interestig bearing in mind that players B and C form a weak coalition against him. On the other hand, it is obviously a major advantage for A that the agreement between players B and C is known to him.

In order to completely demonstrate the strategy for A, one has to explain how to enter the optimal cycle. In fact we can enter in this cycle right from the initial position by the following step

$$\xrightarrow[f=176/3]{a=7} \overline{\overline{\underline{\underline{(19,6)}}}} \xrightarrow[f=98/3]{a=18} (8,19) \xrightarrow[f=106/3]{a=7} (20,8) \xrightarrow[f=98/3]{a=19} (7,20) \xrightarrow[f=98/3]{a=8} (18,7) \xrightarrow[f=106/3]{a=19} (6,18) \xrightarrow[f=98/3]{a=7} \overline{\overline{\underline{\underline{(19,6)}}}}.$$

Remark. If we consider a variant with random order at the start, then things will not change significantly. Analyzing our initial graph with 600 vertices, it can be shown (by exhaustive search, for example) that for each vertex, there is a path that leads from that vertex to the optimal cycle and then the game remains in the same cycle. Since we are interested only in this cycle and not in the first few initial moves everything remains essentially the same. Hence, no matter who starts the game, A will on the long run earn more than average despite the weak coalition of B and C.

References

[Neumann and Morgenstern] John von Neumann and Oskar Morgenstern, Theory of games and economic behavior, 3rd ed., Princeton University Press, Princeton, 1953.

[McCain] Roger A. McCain, An Introductory Sketch of Game Theory, lecture notes at http://william-king.www.drexel.edu/top/eco/game/game.html

[Wilson and Watkins] Robin J. Wilson and John J. Watkins, Graphs, An Introductionary Approach, John Wiley & Sons, New York, 1990.

Section 3:
Algorithms

THE SCALE OF INNOVATIVENESS WITH STRUCTURAL EQUATION MODELING

Majda Bastič
University of Maribor
Faculty of Economics and Business
Razlagova 14, 2000 Maribor
majda.bastic@uni-mb.si

Abstract
The structural equation modeling is the best known methodology for determining the scale for theoretical constructs of interest. The aim of this paper is to present the application of structural equation modeling on determining the scale of organisational innovativeness on the sample of 214 Slovenian organisations. Innovation capability is namely one of the key competitive success factors in the intense global competition.

Keywords: Structural equation modeling, Innovativeness, Dimensions

1. Introduction

Today, the ability to develop new ideas and innovations is one of the top priorities of organisations and the key competitive success factor. Therefore, many researches were focused on factors influencing the innovation capability of firms. In these researches one of the main research questions was associated with the scale of the organisational innovativeness.

The purpose of measurement in theory of testing and development research is to provide an empirical estimate of each theoretical construct of interest [3]. The importance of unidimensionality has been stated succinctly by Hattie: "that a set of items forming an instrument all measure just one thing in common is a most critical and basic assumption of measurement theory" [4].

The scale development process must include an assessment of whether the multiple measures that define a scale can be acceptably regarded as alternative indicators of the same construct. The need for unidimensionality has first been established by Nunnaly by stating, "Items within a measure are useful only to the extent that they share a common core – the attribute which is to be measured."[7].

2. Structural equation model

Structural equation modeling can perhaps best be defined as a class of methodologies that seeks to represent hypotheses about the means, variances, and covariances of observed data in terms of a smaller number of structural parameters defined by a hypothesized underlying model. The general structural equation model consists of two parts: a) the measurement part, linking observed variables to latent variables via a confirmatory factor model and b) the structural part, linking latent variables to each other via systems of simultaneous equations. The estimation of the model utilizes maximum likelihood estimation. In the case where it is assumed that there is no measurement error in the observed variables, the general model reduces to the system of simultaneous equations model developed in econometrics [5].

The structural equations, as represented in a path diagram, are seen as a one-to-one representation of the theory. They are defined by

$$\eta = B\eta + \Gamma\xi + \zeta \tag{1}$$

where η is an $m\times1$ vector of endogenous latent variables, ξ is a $k\times1$ vector of exogenous latent variables, B is an $m\times m$ matrix of regression coefficients relating the latent endogenous variable to each other, Γ is an $m\times k$ matrix of regression coefficients relating endogenous variables to exogenous variables, and ζ is an $m\times1$ vector of disturbance terms.

The latent variables are linked to observable variables via measurement equations for the endogenous variables and exogenous variables. The model used to relate observed measures to factors is the linear factor analysis model defined by the following equations

$$y = \Lambda_y\eta + \varepsilon \tag{2}$$

$$x = \Lambda_x\xi + \delta \tag{3}$$

where y is a $p\times1$ vector of observed responses on p questions assumed to measure an endogenous latent variables, x is a $q\times1$ vector of observed responses on q questions that are assumed to measure exogenous latent variables, Λ_y and Λ_x are $p\times m$ and $q\times j$ matrices of factor loadings, respectively, and ε and δ are a $p\times1$ and a $q\times1$ vectors of uniqueness, respectively. The number of endogenous variables is expressed by p and the number of exogenous variables is expressed by q, k is the number of the important factor's dimensions.

In cases, when the research objective is to find the appropriate indicators to measure a factor of interest then the following model can be applied

$$\eta = \Gamma\xi + \zeta \tag{4}$$

$$x = \Lambda_x\xi + \delta \tag{5}$$

Assessment of model fit refers to the extent to which a hypothesized model is consistent with the data. It is usually done in three stages, involving the a) assessment of the model's overall fit, b) the assessment of the measurement part of the model, and c) the assessment of the structural part of the model.

The purpose of assessing a model's overall fit is to determine the degree to which the model as a whole is consistent with the empirical data in hand. Over the years, a wide range of goodness-of-fit indices have been developed, yet none of them is unequivocally superior to the rest in all circumstances. Particular indices have been shown to operate somewhat differently given the sample size, estimation procedure, model complexity, violation of the underlying assumption of multivariate normality and variable independence. For practical purposes, the results of the chi-square test used in conjunction with root mean square error of approximation (RMSEA), expected cross-validation index (ECVI), standardized root mean square residual (RMR), the goodness-of-fit index (GFI), and the comparative fit index (CFI) should be more than sufficient to reach an informed decision concerning the model's overall fit [2].

The chi-square statistic is the traditional measure for evaluating overall model fit in covariance structure models and provides a test of perfect fit in which the null hypothesis is that the model fits the population data perfectly. A statistically significant chi-square causes rejection of the null hypothesis, implying imperfect model fit and possible rejection of the model. **RMSEA** focuses on the discrepancy between the population covariance matrix and model-based covariance matrix per degree of freedom taking model complexity into account. It is generally regarded as one of the most informative fit indices. Values less than 0.05 are

indicative of good fit, those between 0.05 and under 0.8 show reasonable fit, values between 0.08 and 0.10 are indicative of mediocre fit and values larger than 0.10 are the sign of poor fit. **ECVI** focuses on overall error, i.e. the discrepancy between the population covariance matrix and the model fitted to the sample (implied covariance matrix). It is a useful indicator of a model's overall fit because it measures the discrepancy between fitted covariance matrix in the analysed sample, and the expected covariance matrix that would be obtained in another sample of equivalent size. To assess the model's ECVI, its value must be compared against the ECVI values of other models; the model with the smallest ECVI value is then chosen as representing the greatest potential for replication. **RMR** is a summary measure of fitted residuals. It represents the average value of the element in matrix obtained as the difference between the sample covariance matrix and implied matrix. A problem with interpreting this statistic is that their size varies with the unit of measurement and the latter can vary from variable to variable. This problem can be avoided by using the standardized residuals. A summary measure of standardized residuals is the standardized RMR: values below 0.05 are indicative of acceptable fit. **GFI** is an indicator of the relevant amount of variances and covariances accounted for by the model and thus shows how closely the model comes to perfectly reproducing the observed covariance matrix. Values of GFI should range between 0 and 1 and values greater than 0.9 are usually taken as reflecting acceptable fit. **CFI** is a relative fit index which shows how much better the model fits compared to a baseline model, usually the independence model. It has a range between 0 and 1 and values close to 1 represent good fit.

In evaluating the measurement part of the model, we focus on the relationship between the latent variables and their indicators. The aim is to determine the validity and reliability. Validity reflects the extent to which an indicator actually measures what it is supposed to measure, while reliability refers to the consistency of measurement.

The aim of evaluating the structural part of the model is to determine whether the theoretical relationships specified in the conceptualisation phase are indeed supported by the data. Three issues are of relevance here. The signs of the parameters indicate whether the directions of hypothesized relationship are as hypothesized. The magnitudes of the estimated parameters provide important information on the strength of the hypothesized relationships, and the squared multiple correlations for the structural equations indicate the amount of variance in each endogenous latent variable that is accounted for by the independent latent variables.

3. Dimensions of innovativeness on the case of Slovenian firms

In the past, different dimensions of innovation and their importance were emphasised by different authors [8], [6], [1], [9]. Taking into account their findings, five dimensions and a theoretical model of an organisation's overall innovativeness was established. According to theory organisational innovativeness can be measured by product, marketing, process, behavioural and managerial innovativeness.

Product innovativeness is often referred to as perceived newness, novelty, originality, or uniqueness of products. **Marketing innovativeness** emphasises the novelty of market-oriented approaches. It refers to innovations related to market research, advertising and promotion as well as identification of new market opportunities and entry into new markets [9]. **Process innovativeness** captures the introduction of new production methods and new technology as well as new management approaches that can be applied to improve production and management processes. The primary focus of **managerial innovativeness** is to measure an organisation's ability to manage ambitious organisational objectives and identify a mismatch of those ambitions and existing resources in order to stretch or leverage limited resources creativity. **Behavioural innovativeness** is a fundamental factor that underlies innovative outcomes because it enables the formation of an innovative culture, the overall internal receptivity to new ideas and innovation. It is demonstrated through individuals, teams and management.

Taking into account all five dimensions revealed in previously mentioned researches we tested the following hypotheses:

- H1. The covariance among the items included in the model can be accounted for by one general factor, i.e. organisational innovativeness.
- H2. The covariance among the items included in the model can be accounted for by restricted multi-factor model where each factor presents one of the conceptual components describing the organisational innovativeness. Each item loads only on one factor.
- H3. Responses to each item are reflective of two factors: a general organisational innovativeness factor and a specific component factor.

4. Research methodology

To test a hypothesised factor structure a two-step approach developed by Gerbing and Anderson was used [3]. The measurement model defined by (5) is first developed and evaluated separately from the full structural equation model defined by (4) and (5) which simultaneously models measurement and structural relations. In the approach described, confirmatory factor analysis was applied using computer program AMOS and the Maximum Likelihood estimation method.

A sample of 1000 Slovenian manufacturing organisations was randomly selected from the IPIS database. A total of 254 completed questionnaires were returned, representing a 25.4 per cent response rate which is a normal response rate for most surveys. The rate of usable responses was 21.4 per cent. In the questionnaire, 13 questions referred to the organisational innovativeness. Items V1, V2, V3, V4, V5 referred to behavioural innovativeness, items V6 and V7 to marketing innovativeness, V8 and V9 to process innovativeness, items V10 and V11 to product innovativeness, and items V12 and V13 to managerial innovativeness. The questionnaire used a seven-point Likert scale with verbal anchors of scale. CEOs were chosen as informants as they were most likely to observe and analyse the characteristics of the organization. A pilot study was conducted including 5 experts to aid questionnaire wording and design.

In the first step, all 13 items were included in the first-order measurement model (5) for organisational innovativeness. The initial model fit indices were assessed. They are χ^2=173.765, df=61, p<0.05, GFI = 0.882, RMSEA = 0.093, RMR=0.142 and CFI = 0.895. Their values showed that the original model had to be respecified to obtain better fit with sample data. The pattern of normalized residuals analysis confirmed the need for a respecification. Item V5 had small squared multiple correlation (0.323) and large error variance (2.29) and was therefore removed. Items V12 and V13 had large error covariance (38.754). These two items were thus deleted because each estimated construct is defined by at least two indicators.

Having eliminated 3 items, the modified first-order confirmatory factor analysis model fit indices were: χ^2=33.062, df=28, p>0.1, GFI=0.972, RMSEA=0.029, RMR=0.085, CFI=0.993. All fit indices show that model fits data very well. The standardized regression weights of all variables loadings onto their respective factors were between 0.538 and 0.955, with all critical ratios above 1.96 (which means that all the regressions are statistically significant at the 0.95 per cent confidence level). Their values are given in Table 1. The validation procedure refers to computing reliability for each set of measures. The reliability of the scales using Cronbach's alpha as a measure of internal consistency was encouraging with all the scales adequately meeting standards for such research [7]. The behavioural innovativeness scale achieved an alpha of 0.722, product innovativeness 0.716, process innovativeness 0.716 and marketing innovativeness 0.886.

Table 1. Loadings of the first-order confirmatory analysis

		Standardized first-order loading[a]			
Variable	R^2	Behavioural	Market	Process	Product
Behavioural[b]		-	0.269	0.378	0.284
V1	0.350	0.592[c]			
V2	0.590	0.768 (6.943)			
V3	0.407	0.638 (6.559)			
V4	0.289	0.538 (5.868)			
Marketing			-	0.949	1.152
V6	0.912		0.955		
V7	0.693		0.833 (12.502)		
Process				-	1.206
V8	0.613			0.783	
V9	0.515			0.718 (9.072)	
Product					-
V10	0.580				0.761
V11	0.543				0.737 (9.221)

Note: [a] Standardised first-order loading is the standardised regression weight of the individual variables' loading on to one of the component factor. Numbers in parentheses are critical ratios from the unstandardised solutions; [b] Standardised first-order loading for component factors is the covariance between any two of these component factors; [c] Critical ratio is not available, because the regression weight of the first variable of each component factor is fixed at one.

To assess the multidimensional structure of the organisational innovativeness construct the second-order confirmatory factor analysis using the system of equations (4) and (5) was carried out. The fit indices obtained for this model showed similar results as the first-order confirmatory factor analysis and were: χ^2=37.275, df=30, p>0,1, GFI=0.968, RMSEA=0.034, RMR=0.097, CFI=0.99. The slight difference in the first-order and second-order estimates occurred due to different degrees of freedom. The standardised regression weights of all component factors loadings onto the general factor organisational innovativeness are given in Table 2. They ranged from 0.385 to 0.972.

Table 2. Loadings of the second-order confirmatory factor analysis

Factors	R^2	Standardized Regression Weight	Critical Ratio
Behavioural innovativeness	0.148	0.385	3.977
Product innovativeness	0.945	0.972	8.169
Process innovativeness	0.840	0.916	7.769
Marketing innovativeness	0.478	0.692[a]	

Note: [a] This critical ratio is not available, because the regression weight for the component factor marketing innovativeness is fixed at one.

To assess hypotheses H1 and H2 the hypothesized multidimensional model was compared with a competing unidimensional model. As shown in Table 3, the one-factor model which loaded all 10 indicators to one factor, yielded statistically not significant chi-square of 228.617 while the four-factor model resulted in the statistically significant chi-square of 33.062, suggesting a significant improvement. Furthermore, the improvements in GFI, RMSEA, CFI and NFI were substantial, indicating that the four-factor model presents a better fit to the data. Thus the convergent validity of the constructs is also supported.

Table 3. Results for nested model

Model	Description	χ^2	df	p	GFI	RMSEA	CFI	NFI
1	One general factor	288.617	35	0.000	0.758	0.184	0.664	0.640
2	One general factor + four components	33.062	28	0.233	0.972	0.029	0.993	0.959

5. Conclusions

All hypotheses tested were confirmed. The results obtained showed that innovativeness of Slovenian organisations is also the multidimensional category described by four dimensions. The importance of individual dimensions and their indicators also revealed those features in organizations that may need improvements in order to stimulate innovative capability of organizations.

References

[1] Capon, N., Farley, J.U. Hulbert, J., Lehman, D.R. (1992). Profiles of product innovators among large US manufacturers. *Management Science*, 38 (February): 157-69.

[2] Diamantopoulos, A. and J.A. Siguaw (2000). Introducing LISREL – A Guide for the Uninitiated. SAGE Publications London.

[3] Gerbing, D.W. and J.C. Anderson (1988). An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment. Journal of Marketing Research Vol. XXV (May): 186-92.

[4] Hattie, John (1985), Methodology Review: Assessing Unidimensionality of Tests and Items, Applied Psychological Measurement, 9 (June): 139-64.

[5] Kaplan, D. (2000) Structural Equation Modelling : Foundations and Extensions. Sage Publications, International Educational and Professional Publisher, London.

[6] Miller, D., Friesen, P.H. (1983), Strategy-making and environment: the third link, *Strategic Management Journal*, 4(3): 221-35.

[7] Nunnaly, J. C. (1978), "*Psyhometric Theory, 2nd ed.*" New York: McGraw-Hill Book Company.

[8] Schumpeter, J.A. (1934.), *The Theory of Economic Development*, Harvard University Press, Cambridge, MA.

[9] Wang, C.L., Ahmed, P.K. (2004), The development and validation of the organisational innovativeness construct using confirmatory factor analysis, *European Journal of Innovation Management*, 7(4): 303-13.

Practical bounds in error-free communication

Igor Dukanović
University of Maribor, EPF
Razlagova 14, 2000 Maribor, Slovenia
igor.mat@uni-mb.si

Franz Rendl
Universität Klagenfurt, Institut für Mathematik
Universiätsstraße 65-67, A-9020 Klagenfurt, Austria
franz.rendl@uni-klu.ac.at

Abstract
Celebrated Lovász theta number is an upper bound on the Shannon capacity of a graph defined by $\Theta(G) = \sup_{n\geq 1} \sqrt[n]{\alpha(G^n)}$. In this paper we suggest copositive-programming-motivated upper bounds on numbers in the corresponding sequence.

Keywords: Error-free communication, Shannon capacity.

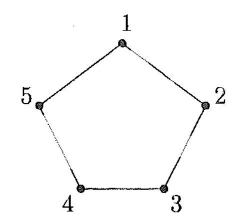
1 Introduction

Consider transmitting data consisting of coding data, transmitting it over a channel and decoding it.

Occasionally errors occur. For example while transmitting binary data digit 1 can change into 0, or vice versa. The standard remedy is to add a parity bit after transmitting every seven (or 8) bits. The added bit equals 1, if the number of ones in the seven bits is odd, and 0 otherwise. This makes it possible to detect, if one error (or an odd number of errors) has been made during transmission. Since double errors are unlikely, this is a very popular approach. Another popular approach which with extremely high probability detects errors is computing checksum consisting of (weighted) sum of the transmitted data.
In this paper we will address a similar but tougher task posed by Claude Shannon, the founder of information theory [24]. Assume that during transmission of a coded message some pairs of symbols occasionally get confused while others do not. What is the maximum rate of transmission such that the receiver may (always) recover the original message without errors?
Such channel is modeled by a so called *confusion graph* $G(V, E)$. The vertex set V consists of all possible symbols. The edge set E consists of all pairs of symbols which can be confused, i.e. symbols B and P in communicating over a phone (voice).
Since we want to communicate error-free we can - if we only send single symbols - use only one letter from each pair that might be confused. In the language of graph theory, this means that the best we can do is to use the symbols from the largest stable subset $S \subseteq V$. Recall that $S \subseteq V$ is *stable*, if

$$i, j \in S \Rightarrow (ij) \notin E.$$

The cardinality of the largest stable set, the stability number, is denoted by $\alpha(G)$.
For example, if the confusion graph is a 5-cycle



we can in error-free communication use only up to two symbols instead of all 5, i.e. $\alpha(C_5) = 2$ and the *information rate* is $\log_2 2$ instead of maximal possible rate for 5 symbols $\log_2 5$.
The obvious idea how to increase information rate is to use larger strings in place of single symbols. Two strings $u_1 u_2 ... u_n$ and $v_1 v_2 ... v_n$ can be confused, if on each position i the two symbols u_i and v_i are either equal either "confusable", i.e. $u_i v_i \in E$. So we define the *n-th confusion graph* G^n with vertex set V^n and edges as defined above. The corresponding rate of information per symbol is so

$$\frac{\log_2 \alpha(G^n)}{n} = \log_2 \sqrt[n]{\alpha(G^n)}.$$

If S is a stable set in G, then obviously S^n is a stable set in G^n. Therefore $\alpha(G^n) \geq \alpha(G)^n$, and

$$\alpha(G^n) \geq \sqrt[n]{\alpha(G)^n}$$

meaning that we never decrease the information rate by using longer strings instead of single symbols. Also replacing short strings of size k with strings with size $ak, a \in N$ can also only increase information rate. This is the basic idea of coding theory (see, i.e. [1]) which motivates the following Shannon's fundamental definition [24].

Definition *The zero-error capacity of a graph G is given by*

$$\Theta(G) := \sup_{n \geq 1} \sqrt[n]{\alpha(G^n)}.$$

Notice that the Shannon capacity of the graph describing transmission of digital data $G(\{0,1\}, \{(0,1)\})$ has Shannon capacity 0, i.e. it is impossible to to be absolutely sure that a transmitted message is correct. Of course, we are more interested in computing the first few numbers in this sequence rather than the "theoretical" limit as the strings in practical applications can not be arbitrarily large.
For example consider the product of two 5-cycles $C_5 \times C_5$. It is easy to check that the set $\{(1,1), (2,3), (3,5), (4,2), (5,4)\}$ is stable. Hence, by using strings of length 2 we have increased the lower bound for the capacity to $\Theta(C_5) \geq \sqrt{5}$.
Can we improve this number by further increasing the number of the symbols in a string? Though already Shannon found an upper bound on his error-free capacity of the graph [24] and thus computed Θ of all graphs on up to 5 vertices, this question remained unsolved up to the break-through paper of Lovász[19].

2 Lovász theta number

Lovász theta number can be formulated as a semidefinite program, so we first introduce some standard notation. Let us denote by $e \in R^n$ the vector of all ones, and by $J := ee^t$ the $n \times n$ matrix of all ones. The trace inner product

$$\langle X, Y \rangle = trace\ XY$$

is a well-known inner product in the space of symmetric matrices. Obviously $\langle X, Y \rangle = \sum_{i,j} x_{ij} y_{ij}$ so $\langle X, J \rangle$ is just the sum of all entries in the matrix J. Notation $A \succeq B$ stands for $A - B \succeq 0$ is a positive semidefinite (symmetric) matrix.
Set S be a stable set in the graph $G(V, E)$. Its characteristic vector χ_S is defined by

$$(\chi_S)_i = \begin{cases} 1 & i \in S \\ 0 & \text{otherwise} \end{cases}$$

Since $|S| = \chi_S^T \chi_S$

$$\langle X, J \rangle = |S|$$

where the matrix X is defined by

$$X := \frac{1}{\chi_S^T \chi_S} \chi_S \chi_S^T. \qquad (1)$$

By construction $X \succeq 0$ and $trace\ X = 1$. Also, if $(ij) \in E$ at least one of the numbers $(\chi_S)_i$ and $(\chi_S)_j$ are zero since such i and j can not both be in the stable set S. Therefore $x_{ij} = 0\ \forall (ij) \in E$.
Consider the maximum over all such matrices

$$\theta(G) = \begin{array}{l} \max \langle X, J \rangle \\ trace\ X = 1 \\ x_{ij} = 0\ \forall (ij) \in E \\ X \succeq 0 \end{array} \qquad (2)$$

Since all matrices (1) are feasible for the above semidefinite program, $\theta(G) \geq \alpha(G)$. Lovász θ number is also well-known upper bound on the Shannon capacity of the graph (see [19, 1] and [7] for an alternative proof).
Goemanns[13] noticed "It seems all roads lead to θ". And indeed $\theta(G)$ is one of the best polynomial bounds sandwiched between two NP-hard problems [11]

$$\alpha(G) \leq \theta(G) \leq \bar{\chi}(G)$$

where $\bar{\chi}(G) := \chi(\bar{G})$ is the chromatic number of the complement graph $\bar{G}$ with vertex set V and edge set $\bar{E}$ defined by

$$\bar{E} = \{(ij) | i \neq j, (ij) \notin E\}.$$

Graphs for which equality $\alpha(H) = \bar{\chi}(H)$ holds for $H = G$ and all induced subgraphs are called *perfect* or *Berge* graphs. Berge conjectured [3] that a graph G is perfect if and only if neither G not $\bar{G}$ contain an odd cycle with length at least 5. This strong perfect graph conjecture has recently been proved [5]. Lovász θ number of any graph can be computed in polynomial time to arbitrary fixed precision since it is a semidefinite program (see survey [25]). But stability number of a perfect graph can be checked in a polynomial time, too [14] as a sequence of computations of θ numbers of induced subgraphs. Therefore the determination of Shannon capacity of perfect graphs sandwiched between two equal numbers $\alpha(G)$ and $\theta(G)$ is trivial, too. So the maximum information rate on a perfect graph is already achieved by transmitting single symbols, since $\alpha(G) = \sqrt{\alpha(G^2)} = \sqrt[3]{\alpha(G^3)} = ...$
While already Lovász showed that the Shannon capacity of C_5 is indeed $\sqrt{5}$ achieved on strings of size two [19, 1] the Shannon capacity of C_7 is still an open problem (see [27]). See also [4] for bounds on complements of odd cycles.

3 Strengthening Lovász θ number toward stability number

In the following we will be interested in bounding a first few numbers in the Shannon's sequence

$$\alpha(G), \sqrt{\alpha(G^2)}, \sqrt[3]{\alpha(G^3)}, ...$$

hoping that the limit $\Theta(G)$ is already achieved, and also since coding very long strings is not practical. Solving semidefinite program (2) does not only give an upper bound on the stability number of the corresponding graph, but also motivates several heuristics for extracting large stable sets (see [15, 2, 26]). All of them are motivated by the belief that the computed near optimal semidefinite matrix variable X of (2) approximates well a matrix X in (1) defined by a single stable set or that X of (2) approximates a convex combination of such solutions. In both cases a large diagonal entry x_{ii} suggests that the vertex i is with a high probability contained in one (or several) large stable sets. So one can order the vertices of G so that

$$x_{i_1 i_1} \geq x_{i_2 i_2} \geq ... \geq x_{i_n i_n}$$

and build large stable sets by a randomized greedy algorithm. A safer (but more expensive) heuristics recompute θ number of the induced graph $G \backslash \{v\}$ after vertex v has been included in a stable set. For both tasks we could benefit by a matrix variable X which better approximates X of (1). To this end notice that the elements of matrices (1) are non-negative. This motivates introducing Schrijver's number [23]

$$\theta^{-}(G) = \max\{\langle X, J\rangle : tr\ X = 1, x_{ij} = 0\,\forall(ij) \in E, x_{ij} \geq 0\,\forall(ij) \in \bar{E}, X \succeq 0\} \quad (3)$$

which is sandwiched between $\alpha(G)$ and $\theta(G)$. See [10, 8] for practical issues in computing these bounds.
A near optimal matrix variable of (3) resembles X of (1) more then does a near optimal solution of (2). Of course, $\theta^{-}(G)$ is also an upper bound on the stability number $\alpha(G)$ by the same reasoning as for $\theta(G)$ while it is at the same time smaller then $\theta(G)$ since the feasible set in (3) is a subset of the feasible set in (2).

Notice further that any matrix (1) is also completely positive. A matrix is *completely positive* if

$$X = \sum_i x_i x_i^T$$

where each vector x_i has only non-negative entries. We denote the cone of all completely positive matrices by $\mathcal{C}^*$ and its dual cone of copositive matrices by $\mathcal{C}$. So Schrijver's number can further be strengthened to

$$\theta^{-\mathcal{C}}(G) = \max\{\langle X, J\rangle : trace\ X = 1, x_{ij} = 0\forall(ij) \in E, X \in \mathcal{C}^*\}. \quad (4)$$

The dual of the (4) is

$$\theta^{-\mathcal{C}^*}(G) = \min\{t : c_{ii} = 1\forall i, c_{ij} = 0\forall(ij) \in \bar{E}, tC - J \in \mathcal{C}\} \quad (5)$$

where the set of copositive matrices $\mathcal{C}$ is characterized by

$$C \in \mathcal{C} \iff \forall x \geq 0^n \quad x^T C x \geq 0$$

Since there are easy to find strictly feasible solutions of (4) and (5) by Slater's theorem there is no duality gap, i.e. $\theta^{-\mathcal{C}^*} = \theta^{-\mathcal{C}}$. So we could apply a primal-dual algorithm. However this number is again NP-hard to compute, so a chain of relaxations

$$\mathcal{P} \subset \mathcal{K}_0 \subset \mathcal{K}_1 \subset \ldots \subset \mathcal{K}_n = \mathcal{C} \quad (6)$$

has been found [6]. Define [6]

$$\theta^{(i)}(G) = \min\{t : c_{ii} = 1\forall i, c_{ij} = 0\forall(ij) \in E, tC - J \in \mathcal{K}_i\} \quad (7)$$

and notice that (6) implies

$$\theta(G) \geq \theta^{(0)}(G) \geq \theta^{(1)}(G) \geq \ldots \geq \theta^{-\mathcal{C}}(G) \geq \alpha(G).$$

It is also known that $\theta^{(0)}(G) = \theta^{-}(G)$ while $\theta^{(1)}(G) = \alpha(G)$ for all odd cycles [6]. For properties of $\theta^{(r)}$ in general see also [16, 22]. Each $\theta^{(r)}$ can be described as a positive semidefinite program of quickly increasing size. For example, in [21] the following characterization of set $\mathcal{K}_1$ is proved

$$X \in \mathcal{K}_1 \iff \forall i \in V \exists M^{(i)} : \begin{array}{l} X - M^{(i)} \succeq 0 \\ m_{ii}^{(i)} = 0 \\ m_{ii}^{(j)} + 2m_{ij}^{(i)} = 0 \quad \forall i \neq j \\ m_{ij}^{(k)} + m_{ik}^{(j)} + m_{jk}^{(i)} \geq 0 \quad \forall i < j < k \end{array} \quad (8)$$

5 Computing $\theta^{(1)}(G)$ of a vertex transitive graph

Definition *Permutation $\alpha : V \to V$ is an automorphism of the graph G, if*

$$(ij) \in E \iff (\alpha(I)\alpha(j)) \in E.$$

The group of all automorphisms is denoted by $\mathrm{Aut}(G)$.
Definition *Graph $G(V, E)$, $V = \{1, \ldots, n\}$ is vertex transitive if there exist automorphisms $\alpha_2, \ldots, \alpha_n$ such that*

$$\alpha_i(i) = 1. \quad (9)$$

All odd-cycles, their complements, all their strong products C_k^n and many other interesting graphs are vertex transitive. So we suggest simplifying (8) to

$$X \in \mathcal{K}_1 \iff \forall i \in V \exists M : \begin{array}{l} X - M \succeq 0 \\ m_{11} = 0 \\ m_{ii} + 2m_{1\alpha_i(1)} = 0 \quad \forall i > 1 \\ m_{ij} + m_{\alpha_j(i)\alpha_j(1)} + m_{\alpha_i(j)\alpha_i(1)} \geq 0 \quad \forall 1 < i < j. \end{array} \quad (10)$$

The matrix M in (10) is defined by $M := \frac{1}{|A|}\sum_{\alpha \in A} P_\alpha^T M^{\alpha(1)} P_\alpha$ where subgroup $A \leq Aut(G)$ and $M^{(1)}, \ldots, M^{(n)}$ satisfy (8), see [9, 7] for details. Notice that in the above characterization the matrix X is "supported" by only one matrix M. We use a primal-dual interior point method to compute $\theta^{(1)}(G)$ by solving (7). Characterization (10) reduces the computational effort to solve one step of the interior point method from $\mathcal{O}(m^9)$ to $\mathcal{O}(m^6)$ where m is the number of vertices of our graph. For example, G^a has $m = n^a$ vertices! This makes even the simplified model hard to compute. However notice that this model has a lot of symmetry. If α is any automorphism which fixes the first vertex, i.e. $\alpha(1) = 1$ (and there are at least $a!$ such automorphisms in G^a, see also [20]) then $P_\alpha^T M P_\alpha$ also satisfies (10). This can be exploited as in a similar bound on chromatic number (see, [9]) to dramatically decrease the number of free variables. This method follows recent successes in exploiting symmetries in semidefinite programs and underlying combinatorial optimization problems [17, 18, 12].

6 Conclusions

The powers of graph G appearing in the definition of the Shannon capacity of a graph have very rich automorphism groups which can be efficiently exploited to compute bounds on the stability numbers $\alpha(G), \alpha(G^2), \ldots, \alpha(G^a)$ for medium powers a.

References

[1] M. Aigner and G. M. Ziegler. Proofs from the book. Springer, 1998.

[2] S. Benson and Y. Ye. Approximating maximum stable set and minimum graph coloring problems with the positive semidefinite relaxation. In *Applications and Algorithms of Complementarity*. Kluwer Academic Publishers, 2000.

[3] C. Berge. Färbung von graphen deren sämtliche beziehungensweise dere ungerade kreise starr sind (zusammenfassung). In *Wissenschafliche Zeitschrift*, Martin Luther Univ. Halle-Wittenberg, Math-Naturwiss Reihe, 114–115, 1961.

[4] T. Bohman and R. Holzman. A nontrivial bound on the Shannon capacities of complements of odd cycles In *IEEE Transactions on Information Theory*, 49:721–722, 2003.

[5] M. Chudnovsky, N. Robertson, P. D. Seymour and R. Thomas. The strong perfect graph theorem. Manuscript, 2003.

[6] E. De Klerk and D. Pasechnik. Approximation of the stability number of a graph via copositive programming In *SIAM J. Optim.*, 12:875–892, 2002.

[7] I. Dukanovic. Semidefinite programming applied to graph coloring problem. PhD thesis. University of Klagenfurt, Austria, 2005 (forthcoming).

[8] I. Đukanović. Zgornja meja za neodvisno število grafa. In *DSI-2003, Ljubljana: Slovenian Society Informatika*, 240-246, 2003.

[9] I. Dukanovic, J. Povh, F. Rendl and A. Wiegele. Reductions in symmetric copositive programming motivated bounds on the clique and chromatic number. Working paper, University of Klagenfurt, Austria, 2005.

[10] I. Dukanovic and F. Rendl. Semidefinite programming relaxations for graph coloring and maximal clique problems. Working paper - *Optimization online*, 2005.

[11] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W.H. Freman and Company, New York, 1979.

[12] K. Gatermann and P.A. Parrilo. Symmetry groups, semidefinite programs, and sums of squares. In *Journal of Pure and Appl. Algebra*, 192:1-3, 95-128, 2004.

[13] M. X. Goemans. Semidefinite programming in combinatorial optimization. *Mathematical Programming*, 79:143-162, 1997.

[14] M. Grötschel, L. Lovász and A. Schrijver. Polynomial algorithms for perfect graphs. In *Annals of Discrete Mathematics*, 325-356, 1984.

[15] G. Gruber and F. Rendl. Computational experience with stable set relaxations. In *SIAM Journal on Optimization*, 13(4):1014-1028, 2003.

[16] N. Gvozdenović and M. Laurent. Semidefinite bounds for the stability number of a graph via sums of squares of polynomials. In *Proceedings of IPCO XI*, Berlin, 2005.

[17] Y. Kanno, M. Ohsaki, K. Murota and N. Katoh. Group symmetry in interior-point methods for semi-definite program. In *Optimization and Engineering*, 2:293-320, 2001.

[18] E.de Klerk, D. V. Pasechnik and A. Schrijver. Reduction of symmetric semidefinite programs using the regular ∗-representation. Priprint, 2005.

[19] L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25:1-7, 1979.

[20] B. D. McKay. Practical graph isomorphism. *Congressus Numerantium*, 30:45-87, 1981.

[21] P. A. Parrilo. *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization.* PhD thesis, California Institute of Technology, 2000.

[22] J. Pena, J. Vera and L. F. Zuluaga. Computing the stability number of a graph via linear and semidefinite programming. Working paper, 2005.

[23] A. Schrijver. A comparison of the Delsarte and Lovász bounds. *IEEE Transactions on Information Theory*, IT-25:425-429, 1979.

[24] C. E. Shannon. The zero-error capacity of a noisy channel. In *IRE Trans. Information Theory* , 3:3-15, 1956.

[25] M.J. Todd. Semidefinite optimization. *Acta Numer.*, 10:515-560, 2001.

[26] E. A. Yildirim and X. Fan. On Extracting Stable Sets from Lovasz's Theta Function. Technical Report, Stony Brook University, 2003.

[27] A. Vesel and Janez Žerovnik. Improved lower bound on the Shannon capacity of C7. *Inf. Process. Lett.*, 81:5, 277-282. Elsevier Nort-Holland, 2002.

CONVERGENCE IN THE CASE OF THE BULLWHIP EFFECT WITH CENTRALIZED DEMAND INFORMATION

Liljana Ferbar, David Čreslovnik, Blaž Mojškerc, Martin Rajgelj
University of Ljubljana, Faculty of Economics
Kardeljeva pl. 17, 1000 Ljubljana, Slovenia
liljana.ferbar@ef.uni-lj.si, david.creslovnik@ef.uni-lj.si,
blaz.mojškerc@ef.uni-lj.si, martin.rajgelj@ef.uni-lj.si

Abstract
Our aim is to improve an algorithm for seeking convergence in the case of the bullwhip effect in a supply chain with centralized demand information. We investigate the conditions needed for production and inventory to converge in the sense that we examine the number of iterations needed for stabilization of these two quantities. Aditionaly we study the influence of the size of perturbation in external demand on total inventory costs of supply chain.

Kewords: supply chain, bullwhip effect, perturbed external demand, centralized information

1 INTRODUCTION

According to Lee et al. (1997b, p. 93), the term "bullwhip effect" was first used by Procter & Gamble, when they experienced extensive demand amplifications for their diaper product "Pampers". Lee et al. (1997a, 1997b) describe the bullwhip effect as the result of information distortion in a supply chain, where companies upstream do not have information on actual consumer demand. Consequently, their ordering decisions are based on the incoming orders from the next downstream company. This may lead to amplified order variability: demand coming in from a downstream company has a lower variability than demand to an upstream company. The bullwhip effect can have dramatic effects on companies that are removed from the end-user (Metters, 1997, p. 89).

The bullwhip effect has been noted and assigned various causes across a range of academic disciplines. The first academic description of the bullwhip phenomenon is usually ascribed to Forrester in 1961 in his seminal book *Industrial Dynamics*. Forrester indicates that it is empirically common for the variance of perceived demand to the manufacturer to far exceed the variance of consumer demand and for seasonality to be larger for manufacturers than for retailers. Furthermore, he notes that the effect is amplified at each stage in the supply chain. He states that the principal cause of this is difficulties involving the information feedback loop between companies and that such systems are too complex for managerial intuition alone to ameliorate. Consequently, Forrester's remedy lies in understanding the system as a whole, and modeling that system with specific "system dynamics" simulation models, so that managers can determine appropriate action. Perhaps the most well-known description of the bullwhip effect emanates directly from Forrester's work and has come from the system dynamics discipline. Sterman (1989) has used a simulation to demonstrate the bullwhip effect: a classroom game known as the "beer distribution game" where pariticipants simulate a supply chain consisting of a beer retailer, wholesaler, distributor and brewery. The decision-making task is straightforward: subjects seek to minimize total costs by managing their inventories appropriately in the face of uncertain demand. As the game proceeds, a small change of a consumer demand invariably is translated into wild swings in both orders and inventory upstream. Sterman (1989, p. 322) concluded that the interaction of individual decisions with the structure of the simulated firm produced aggregate dynamics which diverged significantly and systematically from optimal behavior. He interpreted the phenomenon as a consequence of players' systematic irrational behavior, or "misperceptions of feedback." On the other hand, Lee et al. (1997a, p. 558) claim that demand distortion may arise as a result of optimizing behaviors by players in the supply chain. In a practical study Fransoo and Wouters (2000) have experienced that the theory of measurement of the bullwhip effect in a practical setting has received very limited attention so far. In their paper, they have made a first start with a method to accurately document and define various ways of measuring the bullwhip effect. However, the bullwhip effect can be measured as the ratio of demand-variance at a given supplier to the variance of external demand, assuming that expected values of each stage's demand are the same, i. e., $U_i = \frac{\mathrm{Var}(P_i)}{\mathrm{Var}(D)}$ (Simchi-Levi, Kaminsky and Simchi-Levi, 2000, p. 88).

Many investigations (e.g., Lee et al., 1997a, 1997b; Chen et al., 2000) have shown that providing the supplier upstream with *centralized data* can significantly reduce the bullwhip effect. Such information cuts short all kinds of information distortions in a supply chain. Other improvements that can reduce the bullwhip effect include the reduction of lead times, revising reorder procedures, limiting price fluctuations, and the integration of planning and performance measurement (Lee et al., 1997a, 1997b). Chen et al. (2000) have demonstrated that the bullwhip effect is due, in part, to the effects of demand forecasting. They have drawn the conclusion that centralizing customer demand information can significantly reduce the bullwhip effect. However, centralizing customer demand information does not completely eliminate the bullwhip effect. In addition, they have shown that the difference between the variability in the centralized and decentralized supply chains increases as we move up the supply chain, i.e., as we move from the first stage to the second and third stages of the supply chain.

As a practical matter, the bullwhip effect is a well-documented problem that affects many businesses in serial supply chains across a variety of industries. Although it may seem an obvious inefficiency that is easy to correct, discovery of the bullwhip effect does not automatically lead to its solution: case studies (e.g., Metters, 1997) demonstrate that, despite significant effort, the bullwhip effect can persist. Although the precise causes of the bullwhip effect remain under debate, it is generally agreed that a lack of inter-company communication combined with large time lags between receipt and transmittal of information are at the root of the problem. Simchi-Levi, Kaminsky and Simchi-Levi (2000, p. 91–92) have listed a number of suggeestions for reducing the bullwhip effect or for eliminating its impact. These include reducing uncertainty, reducing the variability of the customer demand process, reducing lead times, and engaging in strategic partnerships.

2 THE MODEL

The focus of our research is the centralized supply chain described above. The assumptions of the model are:

1. Production (P) + Initial Stock (IS) = (Internal) Demand (D) + Final Stock (FS)
2. Final Stock (FS) converges to Balanced Stock (BS)
3. Balanced Stock (BS) = External Demand (ED)

Assuming this we get (i – period, j – stage in a supply chain from retailer toward suppliers):

1. $P_j(i) = ED + D_j(i) - IS_j(i) \ (= D_{j+1}(i))$
2. $IS_j(i+1) = P_j(i) + IS_j(i) - D_j(i)$

We study the model in a case where two more assumptions have to hold:

1. External Demand (x) changes its value by Δx: $ED = x + \Delta x$,
2. The perturbation is negative: $\Delta x < 0$.

These two assumptions tell us that consequences of only one negative change of external demand are examined and that the state following this change remains unchanged. We suppose that these two assumptions can be very real: in many real-world situations we note the stabilization of the external state after single perturbation. The perturbation could, of course, also be positive – in that case the same analysis would follow.

Furthermore if $P_j(i) < 0$, we set $P_j(i) = 0$, since the production can not be negative. The production can be expressed in a form $P = mx + n\Delta x = mx - n|\Delta x|$, so the condition $mx - n|\Delta x| < 0$ implies:

$$\frac{m}{n} < \frac{|\Delta x|}{x} (= \text{relative change}).$$

2.1 Algorithm

Since for any negative perturbation, $|\Delta x| < x$, and any period, i, and supplier, j, we can obtain production, $P_j(i)$, and initial stock, $IS_j(i+1)$, [which equals final stock, $FS_j(i)$] the model can be expressed with the following algorithm:

```
function centralized_model()
// i - period
// j - supplier

   i = 1;
   while TRUE
      for j = 1 to final_j do
         if (2 * D(i, j) - IS(i, j) < 0) then P(i, j) = 0;
         else P(i, j) = 2 * D(i, j) - IS(i, j);
         end if

         D(i, j + 1) = P(i, j);
         IS(i + 1, j) = P(i, j) + IS(i, j) - D(i, j);
      end for

      if ((P(i, final_j) == X + deltaX)) then return i;
      end if

      i = i + 1;
   end while
end function
```

By analysis of the model we can calculate production, $P_j(i)$, and initial stock, $IS_j(i)$, for any supplier at any period in the future with given perturbation, Δx. Using following equations we eliminate time consuming step by step algorithm presented above:

$$P_j(i)=\begin{cases}0 & ; \ \frac{i+j}{i}>\frac{x}{|\Delta x|}\\ i\cdot x+(i+j)\cdot\Delta x & ; \ \frac{i+j}{i}\le\frac{x}{|\Delta x|} \text{ and } \frac{(i-1)+j}{(i-1)}>\frac{x}{|\Delta x|}\\ x+\Delta x & ; \ \frac{(i-1)+j}{(i-1)}\le\frac{x}{|\Delta x|}\end{cases}$$

$$IS_j(i+1)=FS_j(i)=\begin{cases}x+\Delta x & ; \ \frac{i+j}{i}\le\frac{x}{|\Delta x|}\\ x & ; \ \frac{i+(j-1)}{i}>\frac{x}{|\Delta x|}\\ -(i-1)\cdot x-(i+j-1)\cdot\Delta x & ; \ \frac{i+j}{i}>\frac{x}{|\Delta x|} \text{ and } \frac{i+(j-1)}{i}\le\frac{x}{|\Delta x|}\end{cases}$$

From these explicit equations we can get the following table.

Table 1: *Consequences in production, initial and final stock when relative change of external demand lies on an interval* $\frac{3}{7}\le\frac{|\Delta x|}{x}<\frac{4}{9}$

	Supplier 4		Supplier 3		Supplier 2		Supplier 1		Manufacturer		Demand
	Prod	Stock	Prod	Stock	Prod	Stock	Prod	Stock	Prod	Stock	
0	x	x x	x	x x	x	x x	x	x x	x	x x	x
1	0	x x	0	x x	0	x x	0	x $-2\Delta x$	$x+2\Delta x$	x $x+\Delta x$	$x+\Delta x$
2	0	x x	0	x x	0	x $-x-4\Delta x$	$2x+4\Delta x$	$-2\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
3	0	x x	0	x $-2x-6\Delta x$	$3x+6\Delta x$	$-x-4\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
4	$4x+9\Delta x$	x $x+\Delta x$	$4x+8\Delta x$	$-2x-6\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
5	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
	$U\ge 5\frac{1}{2}$		$U\ge 5\frac{1}{10}$		$U\ge 4\frac{9}{20}$		$U\ge 3\frac{11}{20}$		$U=2\frac{2}{5}$		$U=1$

It is obvious that the final stock equals the initial stock in the next period, $FS_j(i)=IS_j(i+1)$, and that the production (order) of supplier j equals the internal demand for supplier $j+1$, $P_j(i)=D_{j+1}(i)$. Another confirming property seen from this table is that the measure of the bullwhip effect, U, increases moving up the supply chain as it was expected.

Further analysis requests the definition of two terms.

Definition 1: *Balance number, k,* is a number of iterations needed for the algorithm to converge.

As it is seen from table above, the balance number in this case is $k=5$.

Definition 2: *Splitting point* is a point on an interval $[0,\ 1]$ such that $P_j(i)$ changes its sign from positive to negative.

Proposition 1: Splitting points are exactly the set $\left\{\frac{i}{i+j}\right\}$ for indices that correspond to condition $\frac{i}{i+j}<\frac{|\Delta x|}{x}$.

Proposition 2: Number of iterations $=\max\left\{i\ ;\ l>\left(\frac{x}{|\Delta x|}-1\right)\cdot i\right\}+2$, where l is the number of links of a chain.

As we can see from Proposition 2, convergence of the algorithm is dependent on the relative change of external demand (and on the number of links). These results of the algorithm can be illustrated by graphs. Figure 1 shows the impact of the relative change of external demand on balance number for two different lengths of supply chain.

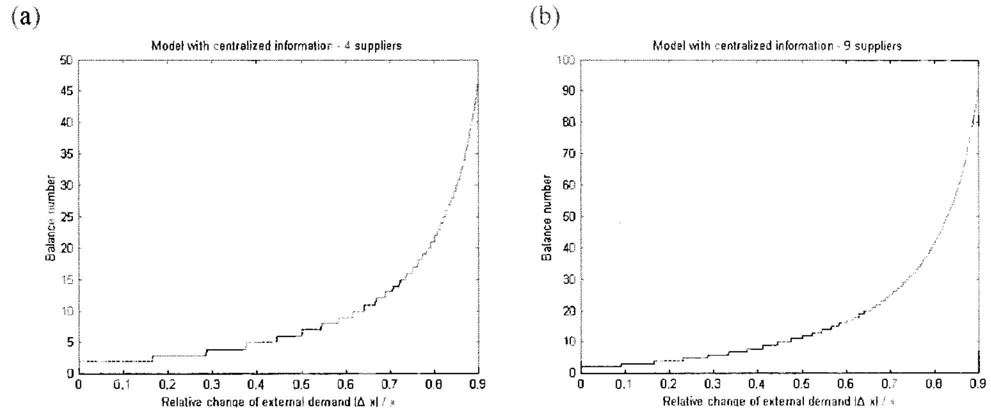


Figure 1: *The dependence of the balance number on the relative change of external demand for two different lengths of supply chain*

Graphs show that up to a certain point the convergence is stable. But as the relative change of external demand approaches towards 1 the balance number rises very quickly. E. g., in the case of a manufacturer and 4 suppliers and at relative change of 0.9 the algorithm converges in 46 iterations.

2.2 Inventory costs

We observe inventory costs of the whole chain from the beginning to the balance number, *k*. *Perturbed costs* are total inventory costs that are attained through our algorithm. *Ideal costs* are total inventory costs that would emerge in the case of immediate adaptation. *Relative costs* are defined as the ratio between the perturbed and ideal costs. Considering this we set:

Perturbed costs: $C_j(i+1)=C_j(i)+IS_j(i+1)\Rightarrow C_p(k)=\sum_{j=1}^{5}C_j(k)$

Ideal costs: $C_j(k)=k\cdot x+(k-1)\cdot\Delta x\Rightarrow C_o(k)=\sum_{j=1}^{5}C_j(k)=5k\cdot x+5(k-1)\cdot\Delta x$

Figure 2 presents the impact of relative change of external demand on relative costs. It is obvious that total inventory costs increase very quickly with increased relative change of external demand, $\frac{|\Delta x|}{x}$. It is also evident that relative changes up to 0.5 do not increase costs so drastically.

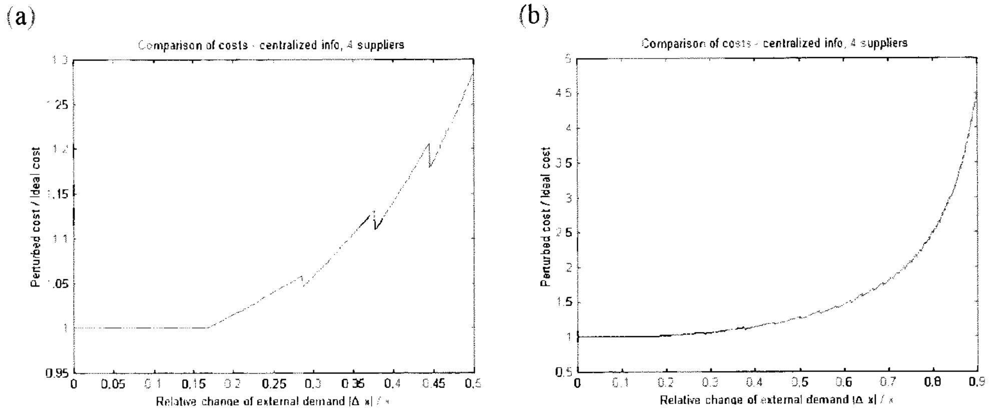


Figure 2: *The dependence of relative costs on the relative change of external demand for two different intervals,* $[0,\ 0.5]$ *and* $[0,\ 0.9]$

3 CONCLUSIONS AND FURTHER RESEARCH

In this paper we have shown that recursion of the algorithm can be replaced by expressing production and stock explicitly. We have also found out that the balance number and relative costs increase very rapidly as relative change of external demand tends to 1.

Further research could include elaboration with positive changes of external demand and also introduce more perturbations.

References

1. Chen et al. (2000): Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times and Information. *Management Science*, Linthicum (MD), Vol. 46, No. 3, pp. 436–443.

2. Fransoo C. Jan, Wouters J. F. Marc (2000): Measuring the Bullwhip Effect in the Supply Chain. *Supply Chain Management*, Bradford, Vol. 5, No. 2, pp. 78–89.

3. Lee L. Hau, Padmanabhan V., Whang Seungjin (1997a): Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, Linthicum (MD), Vol. 43, No. 4, pp. .546–558.

4. Lee L. Hau, Padmanabhan V., Whang Seungjin (1997b): The bullwhip effect in supply chains, *Sloan Management Review*, Cambridge (MA), Vol. 38, No. 3, pp. 93-102..

5. Metters Richard (1997): Quantifying the Bullwhip Effect in Supply Chains. *Journal of Operations Management*, Washington, D.C., Vol. 15, No. 2, pp. 89–100.

6. Simchi-Levi David, Kaminsky Philip, Simchi-Levi Edith (2000): *Designing and Managing the Supply Chain*. Boston : McGraw-Hill.

7. Sterman J. D. (1989): Modelling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment, *Management Science*, Linthicum (MD), Vol. 35, No. 3, pp. 321–339.

CONVERGENCE IN THE CASE OF THE BULLWHIP EFFECT WITH DECENTRALIZED DEMAND INFORMATION

Liljana Ferbar, David Čreslovnik, Blaž Mojškerc, Martin Rajgelj
University of Ljubljana, Faculty of Economics
Kardeljeva pl. 17, 1000 Ljubljana, Slovenia
liljana.ferbar@ef.uni-lj.si, david.creslovnik@ef.uni-lj.si,
blaz.mojskerc@ef.uni-lj.si, martin.rajgelj@ef.uni-lj.si

Abstract

In our contribution we examine consequences of a single perturbation in the case of the bullwhip effect in a supply chain with decentralized demand information. We investigate the conditions needed to stabilize production and inventory levels. The correlation between the size of perturbation in external demand and total inventory costs of supply chain is also studied.

Kewords: supply chains, perturbed external demand, bullwhip effect, decentralized information

1 INTRODUCTION

1.1 The bullwhip effect

Today customers are getting more and more demanding, looking for a customized solution delivered within days, and no longer willing to accept a commodity product in weeks. At the same time, most firms' manufacturing processes are becoming increasingly dispersed and global. These conditions make it difficult for firms to run their supply chains more efficiently and effectively. Why is it, for instance, that Procter & Gamble had to deal with widely fluctuating orders for its nappies, when babies' consumption was generally quite steady? And why was it, that Procter & Gamble's orders to its suppliers fluctuated even more? (Lee et al., 1997b, p. 93) The reason is that each retailer bases his orders on his own, slightly exaggerated, forecast, thus increasingly distorting the information about real consumer demand (Simchi-Levi, Kaminsky, Simchi-Levi, 2000, p. 82). This is one of the most important causes of inefficiency in a supply chain.

This increase in variability as we travel up in the supply chain is referred to as the *bullwhip effect* (Lee et al., 1997a, p. 546). This implies that the variance of the orders becomes larger as we move up the supply chain, so that the orders placed by the second stage of the supply chain are more variable than the orders placed by the first stage (the retailer) and the orders placed by the third stage of the supply chain will be more variable than the orders placed by the second stage, and so on. In other words, the distortion propagates upstream in an amplified form (i.e., variance amplification). Formally; suppose that the expected value of each stage's demand is the same. Then the bullwhip effect occurs when the ratio of demand-variance at a given supplier to the variance of external demand, $U_i = \frac{\mathrm{Var}(P_i)}{\mathrm{Var}(D)}$, increases as we travel up the supply chain (Simchi-Levi, Kaminsky and Simchi-Levi, 2000, p. 87).

The distortion of demand information implies that the manufacturer who only observes its immediate order data will be misled by the amplified demand patterns, and this has serious cost implications. For instance, the manufacturer incurs excess raw materials cost due to unplanned purchases of supplies, additional manufacturing expenses created by excess capacity, inefficient utilization and overtime, excess warehousing expenses and additional transportation costs due to inefficient scheduling and premium shipping rates (Lee et al., 1997a, p. 547).

1.2 Causes of the Bullwhip Effect

Lee et al. (1997a, 1997b) have identified four major causes of the bullwhip effect:

(1) *Demand forecast updating*. Links in the supply chain base their expectations about future demand on orders they receive from the succeeding link. An increase in orders leads to higher demand forecasts, which is transferred to the next link by increased order quantities. That link also sees an increase in demand, updates its forecasts and distorts information for the subsequent link. It works in the reverse way when end customer demand decreases. A possible solution for this is to make data on consumer demand directly available to companies further upstream in the supply chain. Further, a single source of forecasting can be determined for the entire supply chain.

Chen et al. (2000) have shown that if a retailer periodically updates the mean and variance of demand based on observed customer demand data then the variance of the orders placed by the retailer will be greater than the variance of demand. More importantly, they have shown that providing each stage of the supply chain with complete access to customer demand information can significantly reduce this increase in variability. However, they have also shown that the bullwhip effect will exist even when demand information is shared by all stages of the supply chain and all stages use the same forecasting technique and inventory policy.

(2) *Order batching*. Demands come in, depleting inventories but the company may not immediately place an order with its supplier. It often accumulates demands before issuing an order, for instance because of fixed order costs or distribution efficiency. Consider a company that receives daily orders but places orders with its suppliers once a week. Variability of orders placed with the suppliers is higher than the demands the company itself faces. In addition to making consumer demand data available throughout the chain, reducing batch sizes and increasing order frequencies assist in decreasing this effect.

(3) *Price fluctuations*. Because of promotions and trade deals, the price of a product fluctuates, which increases variability of demand. When the price of a product is low, a customer buys in bigger quantities than needed. When the price returns to normal, the customer buys less than needed to deplete its inventory. Stabilizing prices and decreasing the number of promotions is a way of reducing this effect.

(4) *Rationing and shortage gaming*. When product demand exceeds supply, a supplier needs to ration its product to customers. Knowing that, customers may order more than they really need. Later, when there are no shortages, orders disappear. Introducing rationing methods based on past sales rather than on orders placed takes away the incentive for customers to inflate order sizes.

1.3 How to Counteract the Bullwhip Effect

Measuring the total bullwhip effect does not tell which of the different causes contributes most and which solutions are most relevant (Fransoo, Wouters, 2000). For example, to assess the possible benefits of exchanging demand information, it is important to be able to measure which part of the bullwhip effect is due to incomplete demand information in a particular supply chain.

Understanding the causes of the bullwhip effect can help managers find strategies to mitigate it. Indeed, many companies have begun to implement innovative programs that partially address the effect whereby they tackle each of the four causes. Various initiatives and other possible remedies based on the underlying coordination mechanism are:

information sharing, channel alignment, and operational efficiency (Lee et al., 1997b, p. 98). With information sharing, demand information at a downstream site is transmitted upstream in a timely fashion. Channel alignment is the coordination of pricing, transportation, inventory planning, and ownership between the upstream and downstream sites in a supply chain. Operational efficiency refers to activities that improve performance, such as reduced costs and lead time. In his work Metters (1997) has sought to identify the magnitude of the problem by establishing an empirical lower bound on the profitability impact of the bullwhip effect. Results indicate that the importance of the bullwhip effect to a firm differs greatly depending on the specific business environment. Given appropriate conditions, however, eliminating the bullwhip effect can increase product profitability by 10–30%.

2 MODEL AND ALGORITHM

The focus of our research is the decentralized supply chain described above. The assumptions of the model are:

1. Production (P) + Initial Stock (IS) = (Internal) Demand (D) + Final Stock (FS)
2. Final Stock (FS) converges to Balanced Stock (BS)
3. Balanced Stock (BS) = (Internal) Demand (ID)

Assuming this we get (i – period, j – stage in a supply chain from retailer toward suppliers):

1. $P_j(i) = 2D_j(i) - IS_j(i) \quad (= D_{j+1}(i))$
2. $IS_j(i+1) = P_j(i) + IS_j(i) - D_j(i)$

We study the model in a case where two more assumptions have to hold:

1. External Demand (x) changes its value by Δx: $ED = x + \Delta x$,
2. The perturbation is negative: $\Delta x < 0$.

These two assumptions tell us that consequences of only one negative change of external demand are examined and that the state following this change remains unchanged. We suppose that these two assumptions can be very real: in many real-world situations we note the stabilization of the external state after single perturbation. The perturbation could, of course, also be positive – in that case the same analysis would follow.

Furthermore if $P_j(i) < 0$, we set $P_j(i) = 0$, since the production can not be negative. The production can be expressed in a form $P = mx + n\Delta x = mx - n|\Delta x|$, so the condition $mx - n|\Delta x| < 0$ implies:

$$\frac{m}{n} < \frac{|\Delta x|}{x} (= \text{relative change}).$$

Table below shows us consequences in production, initial and final stock when relative change of external demand lies on an interval $\frac{5}{16} \leq \frac{|\Delta x|}{x} < \frac{3}{8}$. At this point it is convenient to define the term *balance number*.

Definition 1: *Balance number*, *k*, is a number of iterations needed for the algorithm to converge.

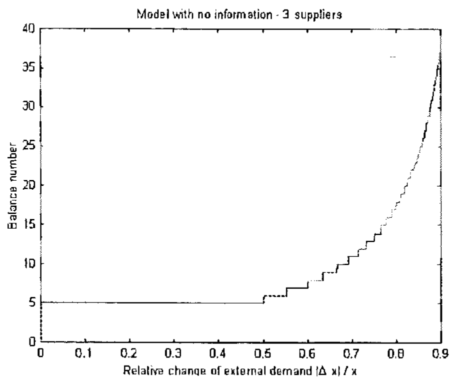
Table 1: *Consequences in production, initial and final stock in the case of relative change of external demand* $\frac{5}{16} \leq \frac{|\Delta x|}{x} < \frac{3}{8}$

	Supplier 3		Supplier 2		Supplier 1		Manufacturer		Demand
	Prod	Stock	Prod	Stock	Prod	Stock	Prod	Stock	
0	x	x x	x	x x	x	x x	x	x x	x
1	0	x x	0	x x	0	x $-2\Delta x$	$x+2\Delta x$	x $x+\Delta x$	$x+\Delta x$
2	0	x $-2x-8\Delta x$	$3x+8\Delta x$	x $2x+4\Delta x$	$2x+4\Delta x$	$-2\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
3	$2x+4\Delta x$	$-2x-8\Delta x$ $-2\Delta x$	$-2\Delta x$	$2x+4\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
4	$2x+4\Delta x$	$-2\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
5	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$	$x+\Delta x$ $x+\Delta x$	$x+\Delta x$
	$U < 11\frac{14}{125}$		$U < 7\frac{1}{5}$		$U < 3\frac{2}{3}$		$U = 2\frac{2}{5}$		$U = 1$

From this table it is obvious that the final stock equals the initial stock in the next period, $FS_j(i) = IS_j(i+1)$, and that the production of supplier j equals the internal demand for supplier $j+1$, $P_j(i) = D_{j+1}(i)$. As it is seen the balance number in this case is $k=5$, whereby that is also the balance number for all cases of relative change of external demand from the interval $[0, \frac{1}{2})$. Another confirming property seen from this table is that the measure of the bullwhip effect, U, increases moving up the supply chain as it was expected.

Convergence of the algorithm is dependent on the relative change of external demand. This correlation can be illustrated by graphs. Figure 1 shows the impact of the relative change of external demand on balance number, for two different lengths of supply chain (for 3 suppliers + a manufacturer, and for 9 suppliers + a manufacturer).

(a)



(b)

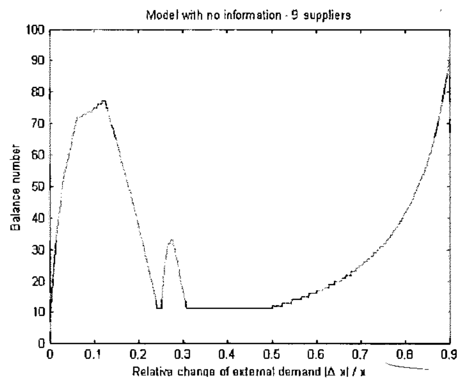


Figure 1: *The dependence of the balance number on the relative change of external demand for two different lengths of supply chain*

The reader can observe an anomaly that occurs on the graph: a "bump" unexpectedly arises at the first third of the interval $[0, 1]$ and from then on the graph has the expected shape. We

conjecture that this bump emerges due to slow transfer of information about small relative changes of external demand to higher-level suppliers.

2.1 Inventory costs

We observe inventory costs of the whole chain from the beginning to the balance number, k. *Perturbed costs* are total inventory costs that are attained through our algorithm. *Ideal costs* are total inventory costs that would emerge in the case of immediate adaptation. *Relative costs* are defined as the ratio between the perturbed and ideal costs. Considering this we set:

Perturbed costs: $C_j(i+1) = C_j(i) + IS_j(i+1) \Rightarrow C_p(k) = \sum_{j=1}^{4} C_j(k)$

Ideal costs: $C_j(k) = k \cdot x + (k-1) \cdot \Delta x \Rightarrow C_o(k) = \sum_{j=1}^{4} C_j(k) = 4k \cdot x + 4(k-1) \cdot \Delta x$

Table 2: *Comparison between perturbed and ideal costs for relative changes of external demand from the interval* $[0, \frac{1}{2})$

	Balance number, k	Ideal costs $C_o(k)$	Perturbed costs $C_p(k)$	Costs increase at least for (in %)	Costs increase at most for (in %)
$\frac{\lvert\Delta x\rvert}{x} < \frac{1}{16}$	5	$20x + 16\Delta x$	$20x + 22\Delta x$	−1.98	0
$\frac{1}{16} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{1}{8}$	5	$20x + 16\Delta x$	$19x + 6\Delta x$	−1.98	1.39
$\frac{1}{8} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{1}{4}$	5	$20x + 16\Delta x$	$20x + 14\Delta x$	1.39	3.125
$\frac{1}{4} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{5}{16}$	5	$20x + 16\Delta x$	$21x + 18\Delta x$	2.5	3.125
$\frac{5}{16} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{3}{8}$	5	$20x + 16\Delta x$	$16x + 2\Delta x$	2.5	8.93
$\frac{3}{8} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{5}{12}$	5	$20x + 16\Delta x$	$19x + 10\Delta x$	8.93	11.25
$\frac{5}{12} \leq \frac{\lvert\Delta x\rvert}{x} < \frac{1}{2}$	5	$20x + 16\Delta x$	$14x - 2\Delta x$	11.25	25

As seen the relative changes of external demand by less than $\frac{1}{16}$ decrease the relative costs but not more than for 1.98%. The decrease of the relative costs occurs again on the interval $[\frac{1}{4}, \frac{5}{16})$. However, other relative perturbations that do not exceed $\frac{1}{2}$ mostly increase relative costs but not more than for 25%. Results presented in table 2 can refer us to expectation of a prevalence of rapid increase of relative costs for the relative changes of external demand bigger than $\frac{1}{2}$.

This expectation is confirmed by figure 2 that presents the impact of relative change of external demand on relative costs for the whole interval used in table 2 and also for a broader interval: figure 2 (a) reflects the relationship on $[0, 0.5]$ and figure 2 (b) shows the same on $[0, 0.9]$. Reader can observe a very rapid increase of total inventory costs as relative change of external demand tends to 1.

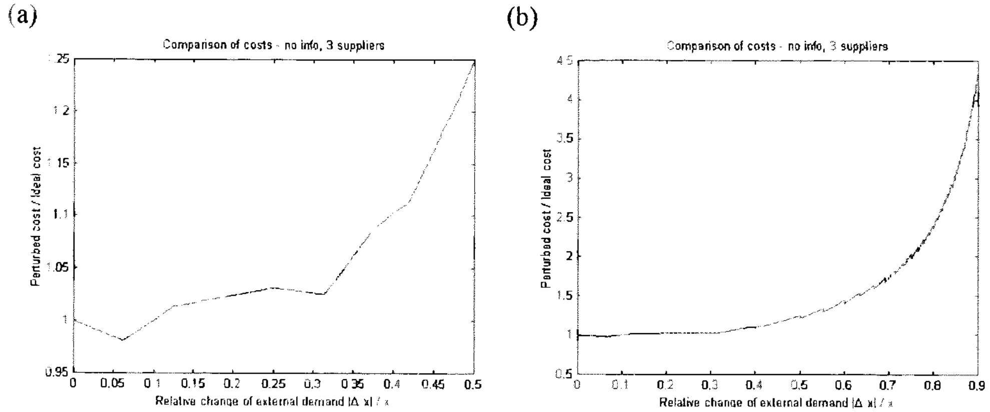


***Figure 2:** The dependence of relative costs on the relative change of external demand for two different intervals,* $[0,\ 0.5]$ *and* $[0,\ 0.9]$

3 CONCLUSIONS AND FURTHER RESEARCH

In this paper we have examined consequences of a single perturbation in the case of the bullwhip effect in a supply chain with decentralized demand information. We have found out that the balance number and relative costs increase very rapidly as relative change of external demand tends to 1.

Further research could include elaboration with positive changes of external demand and also introduce more perturbations.

References

1. Chen et al. (2000): Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times and Information. *Management Science*, Linthicum (MD), Vol. 46, No. 3, pp. 436–443.

2. Fransoo C. Jan, Wouters J. F. Marc (2000): Measuring the Bullwhip Effect in the Supply Chain. *Supply Chain Management*, Bradford, Vol. 5, No. 2, pp. 78–89.

3. Lee L. Hau, Padmanabhan V., Whang Seungjin (1997a): Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, Linthicum (MD), Vol. 43, No. 4, pp. .546–558.

4. Lee L. Hau, Padmanabhan V., Whang Seungjin (1997b): The bullwhip effect in supply chains, *Sloan Management Review*, Cambridge (MA), Vol. 38, No. 3, pp. 93-102..

5. Metters Richard (1997): Quantifying the Bullwhip Effect in Supply Chains. *Journal of Operations Management*, Washington, D.C., Vol. 15, No. 2, pp. 89–100.

6. Simchi-Levi David, Kaminsky Philip, Simchi-Levi Edith (2000): *Designing and Managing the Supply Chain*. Boston : McGraw-Hill.

ON THE EIGENSPACE STRUCTURE OF A MATRIX IN MAX-MIN ALGEBRA

MARTIN GAVALEC, JÁN PLAVKA

ABSTRACT. For a given $n \times n$ matrix A in a max-min algebra, the structure of the eigenspace $\mathcal{F}(A)$ is studied. A necessary and sufficient condition is described, under which the maximum eigenvector is constant. Further results show that $\mathcal{F}(A)$ can contain exponentially many non-empty disjoint intervals of strictly monotone eigenvectors, with distinct endpoints. As a consequence, the eigenspace structure cannot be completely described in polynomial time.

1. INTRODUCTION

Max-min algebras have wide applications in the fuzzy set theory (the max-min algebra on the unit real interval is one of the most important fuzzy algebras). The eigenvectors of max-min matrices are useful in cluster analysis (see [4]), or in fuzzy reasoning (see [7]). The eigenproblem in max-min algebra and its connections to paths in digraphs were investigated in [1, 4, 5, 6]. A procedure for computing the greatest eigenvector of a given max-min matrix was proposed in [7] and an efficient algorithm was described in [2]. The eigenproblem in distributive lattices was studied in [8].

The aim of this paper is to answer several question connected to the structure of the eigenspace $\mathcal{F}(A)$ of a given max-min square matrix A. In [3] the structure of $\mathcal{F}(A)$ was described as a union of intervals containing permuted monotone eigenvectors of various types. As the number of possible permutations is exponentially large, it is natural to ask whether exponentially many of these intervals can simultaneously be nonempty. The positive answer is presented in Theorem 4.2 and it indicates that the structure of the eigenspace $\mathcal{F}(A)$ can be rather complex. Further results concerning constant eigenvectors and permuted monotone eigenspaces are presented in the paper.

2. MONOTONE EIGENVECTORS

By a max-min algebra we understand a linearly ordered set $(\mathcal{B}, \leq)$ with the binary operations of maximum and minimum, denoted by $\oplus$ and $\otimes$. The matrix operations over $\mathcal{B}$ are defined with respect to $\oplus, \otimes$, formally in the same manner as matrix operations over any field. For a given natural $n > 0$, we use the notation $N = \{1, 2, \dots, n\}$, and the notation $\mathcal{B}(n)$ $(\mathcal{B}(n,n))$ for the set of all n-dimensional column vectors (square matrices) over $\mathcal{B}$. We say that a vector $b \in \mathcal{B}(n)$ is increasing, if $b_i \leq b_j$ holds for any $i, j \in N$, $i \leq j$. Vector b is strictly increasing, if $b_i < b_j$ whenever $i < j$. The set of all increasing (strictly increasing) vectors in $\mathcal{B}(n)$ is denoted by $\mathcal{B}^{\leq}(n)$ (by $\mathcal{B}^{<}(n)$). For $x, y \in \mathcal{B}(n)$, we write $x \leq y$, if $x_i \leq y_i$ holds for all $i \in N$, and we write $x < y$, if $x \leq y$ and $x \neq y$. In other words, $x < y$ if $x_i \leq y_i$ for all $i \in N$, but the strong inequality $x_i < y_i$ holds true for at least one $i \in N$.

For given $A \in \mathcal{B}(n,n), h \in \mathcal{B}$, the *threshold digraph* $\mathcal{G}(A,h)$ is the digraph $\mathcal{G} = (N, E)$, with the vertex set N and with the arc set $E = \{(i,j);\ i,j \in N,\ a_{ij} \geq h\}$. The *strict threshold digraph* $\mathcal{G}(A,h^+)$ has the arc set $\{(i,j);\ i,j \in N,\ a_{ij} > h\}$. The set of all permutations on

Date: August 15, 2005.
1991 *Mathematics Subject Classification.* Primary 04A72; Secondary 05C50, 15A33.
Key words and phrases. eigenvector, max-min algebra, fuzzy algebra.
This work was partly supported by Czech Science Foundation, # 406/03/0115.

N will be denoted by P_n. If $A \in \mathcal{B}(n,n)$ and $b \in \mathcal{B}(n)$, $\varphi \in P_n$, then we denote by $A_{\varphi\psi}$ the matrix created by applying permutation φ to the rows and permutation ψ to the columns of A, and by b_φ we denote the vector created by applying the permutation φ to vector b.

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

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

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

Theorem 2.1. [3] *Let $A \in \mathcal{B}(n,n)$, $b \in \mathcal{B}(n)$ and $\varphi \in P_n$. Then $b \in \mathcal{F}(A)$ if and only if $b_\varphi \in \mathcal{F}(A_{\varphi\varphi})$.*

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

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

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

Theorem 2.2. [3] *Let $A \in \mathcal{B}(n,n)$ and let $b \in \mathcal{B}(n)$ be a strictly increasing vector. Then $b \in \mathcal{F}(A)$ if and only if $m^\star(A) \leq b \leq M^\star(A)$. In formal notation we can write*

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

3. Constant eigenvectors

It was stated in [1] that the greatest eigenvector $x^\star$ exists for every matrix A, and its entries are given by the formula

$$x_i^\star = \max_j \min\{a_{ij}^\star, a_{jj}^\star\}$$

i.e., the ith entry of the maximum eigenvector equals the maximum capacity $c(p)$ of a path beginning at i and ending by a cycle at j, where the capacity $c(p)$ is assigned to each path p as

$$c(p) = \min_{(k,l) \in p} a_{kl}$$

In the following theorem, an interval for constant eigenvectors is described. For $A \in \mathcal{B}(n,n)$, we define the value $M(A) \in \mathcal{B}$ as $M(A) := \min_{i \in N} \max_{j \in N} a_{ij}$.

Theorem 3.1. [3] *Let $A \in \mathcal{B}(n,n)$ and let $b \in \mathcal{B}(n)$ be a constant vector. Then $b \in \mathcal{F}(A)$ if and only if $O \leq b_1 \leq M(A)$.*

Corollary 3.2. *The maximum eigenvector $x^\star$ is constant if and only if $x_i^\star = M(A)$ holds for all $i \in N$.*

Theorem 3.3. *The maximum eigenvector $x^\star$ is constant if and only if the digraph $G_{>M(A)}$ is acyclic.*

Proof. Suppose that $x_i^\star = M(A)$ holds true for all $i \in N$, and $\mathcal{G} = G_{>M(A)}$ contains a cycle $\sigma = i_1, i_2, \ldots, i_k$. By the definition of the maximum eigenvector $x^\star$ we have $x_i^\star > M(A)$ for every $i \in \{i_1, i_2, \ldots, i_k\}$, which is a contradiction.

For the proof of the converse implication, let us suppose that there exists $i \in N$ such that $x_i^\star \neq M(A)$. As $x^\star$ is the maximum eigenvector, in view of Theorem 3.1 we have $x_i^\star > M(A)$. The formula

$$x_i^\star = \max_j \min\{a_{ij}^\star, a_{jj}^\star\}$$

implies that there is $j \in N$ such that $a_{jj}^\star > M(A)$. Hence, there exists a closed path p from j to j such that $c(p) > M(A)$, i.e. p is a cycle in $\mathcal{G}$. □

4. Permuted eigenspace structure

Theorem 4.1. *Let $\mathcal{B}$ be a dense max-min algebra and let $A \in \mathcal{B}(n,n)$. Then the following statements are equivalent*

(i) *$a_{ij} < a_{kk}$ for every $i, j, k \in N$, $i \neq j$, $i \neq k$*
(ii) *$\mathcal{F}^{<}(A_{\varphi\varphi}) \neq \emptyset$ for every $\varphi \in P_n$*

Proof. (i) $\Rightarrow$ (ii) Let statement (i) be fulfilled, and let $\varphi \in P_n$ be an arbitrary permutation on N. Clearly, the matrices A, $A_{\varphi\varphi}$ have the same sets of diagonal and non-diagonal elements. Moreover, any two elements belonging to a common row (column) in A, lie in a common row (column) in $A_{\varphi\varphi}$ as well. Thus, statement (i) is invariant with respect to permutations on N, and it is sufficient to consider only the case $\varphi = \mathrm{id}_N$, i.e. $A = A_{\varphi\varphi}$.

Clearly, the inequality

$$m^{(i)}(A) = \max_{l>i} a_{il} \leq \max_{l \geq i} a_{il} = M^{(i)}(A) \tag{4.1}$$

holds true for every $i \in N$. Further, by the assumption (i) we get a strict inequality

$$m^{(j)}(A) = \max_{l>j} a_{jl} < a_{kk} \leq \max_{l \geq k} a_{kl} = M^{(k)}(A) \tag{4.2}$$

for every $j, k \in N$, $j < k$.

By the inequalities (4.1), (4.2) we get

$$m_i(A) = \max_{j \leq i} m^{(j)}(A) \leq \min_{k \geq i} M^{(k)}(A) = M_i(A) \tag{4.3}$$

and, for $i < l$ we have

$$m_i(A) = \max_{j \leq i} m^{(j)}(A) < \min_{k \geq l} M^{(k)}(A) = M_l(A) \tag{4.4}$$

According to Theorem 5.3 in [3], the inequalities (4.3), (4.4) imply that $\mathcal{F}^{<}(A) \neq \emptyset$.

$\neg$(ii) $\Rightarrow$ $\neg$(i) Let us assume that statement (i) is not satisfied, i.e. there exist indices $i, j, k \in N$, $i \neq j$, $i \neq k$ such that

$$a_{ij} \geq a_{kk} \tag{4.5}$$

We shall construct a permutation $\varphi \in P_n$ with $\mathcal{F}^{<}(A_{\varphi\varphi}) = \emptyset$ in the following way. We start with a permutation which interchanges indices $k, n \in N$, i.e. we put $\varphi(k) = n$, $\varphi(n) = k$ and $\varphi(l) = l$ for all $l \neq k, n$. By assumption $i \neq k$ we have $\varphi(i) \neq \varphi(k)$, i.e. $\varphi(i) < n$.

Further, the assumption $i \neq j$ implies $\varphi(i) \neq \varphi(j)$. If $\varphi(j) < \varphi(i)$, then $\varphi(j) < n = \varphi(k)$ and the mutual interchanging of indices $\varphi(i), \varphi(j)$ does not influence the value $\varphi(k) = n$. Hence, without any loss of generality, we may assume that $\varphi(i) < \varphi(j)$ and $\varphi(k) = n$. Let us denote the matrix $A_{\varphi\varphi}$ by A'. The assumption (4.5) can now be written in the form

$$a'_{\varphi(i)\varphi(j)} \geq a'_{nn} \tag{4.6}$$

Then we have

$$m^{(\varphi(i))}(A') = \max_{l > \varphi(i)} a'_{\varphi(i)l} \geq a'_{\varphi(i)\varphi(j)} \geq a'_{nn} = M_n(A') \tag{4.7}$$

what implies

$$m_{\varphi(i)}(A') \geq m^{(\varphi(i))}(A') \geq M_n(A') \tag{4.8}$$

In view of Theorem 5.3 in [3], the inequality (4.8), gives

$$\mathcal{F}^<(A') = \mathcal{F}^<(A_{\varphi\varphi}) = \emptyset$$

i.e. the statement (ii) does not hold true. □

Theorem 4.2. *Let $\mathcal{B}$ be a dense max-min algebra. If $|\mathcal{B}| > 1$, then there is a matrix $A \in \mathcal{B}(n,n)$ such that $\mathcal{F}^<(A_{\varphi\varphi}) \neq \emptyset$ for every $\varphi \in P_n$.*

Proof. Let us choose $x, y \in \mathcal{B}$ such that $x < y$. We define a matrix $A \in \mathcal{B}(n,n)$ by putting

$$a_{ij} = \begin{cases} x & \text{if } i \neq j \\ y & \text{if } i = j \end{cases}$$

It is easy to verify that matrix A satisfies the condition (i) in Theorem 4.1. Hence $\mathcal{F}^<(A_{\varphi\varphi}) \neq \emptyset$ for every $\varphi \in P_n$. □

Remark 4.1. The matrix A in Theorem 4.2 is invariant with respect to all permutations in P_n. As a consequence, we have

$$m^\star(A) = m^\star(A_{\varphi\varphi}) = (x, x, \ldots, x)$$
$$M^\star(A) = M^\star(A_{\varphi\varphi}) = (y, y, \ldots, y)$$

Thus, all intervals $\mathcal{F}^<(A_{\varphi\varphi})$, $\varphi \in P_n$ have the same endpoints.

Theorem 4.3. *Let $\mathcal{B}$ be a dense max-min algebra with $|\mathcal{B}| > 1$, let $n > 2$. Then there is a matrix $A \in \mathcal{B}(n,n)$ with at least 2^{n-2} distinct intervals of the form $\mathcal{F}^<(A_{\varphi\varphi}) \neq \emptyset$, $\varphi \in P_n$.*

Proof. Let the assumptions of the theorem be fulfilled. As $\mathcal{B}$ is dense, there is an increasing sequence of length $2n+1$ in $\mathcal{B}$. The elements of the sequence will be denoted by numbers $0, 1, 2, \ldots, 2n$. The matrix $A \in \mathcal{B}(n,n)$ is defined by

$$a_{ij} = \begin{cases} n+i & \text{if } i = j \\ i & \text{if } i < n = j \\ j & \text{if } j < n = i \\ 0 & \text{otherwise} \end{cases}$$

It is easy to see that A satisfies the statement (i) in Theorem 4.1, hence $\mathcal{F}^<(A_{\varphi\varphi}) \neq \emptyset$ holds true for every $\varphi \in P_n$.

Now, let ε be an arbitrary mapping $\varepsilon : N \to \{0,1\}$ with the two last values $\varepsilon(n-1) = \varepsilon(n) = 0$. The unit values in the sequence $(\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(n))$ occur in m ($m \geq 0$) disjoint intervals $(\varepsilon(i_k), \ldots, \varepsilon(j_k))$, $k = 1, 2, \ldots, m$, separated by zeros. Each interval is of the form $(1, 1, \ldots, 1, 0)$ with indices $1 \leq i_k < j_k \leq n-1$ and $j_k < i_{k+1}$. Thus, for every $i \in N$ we have

$$\varepsilon(i) = \begin{cases} 1 & \text{if } i_k \leq i < j_k \text{ for some } k = 1, 2, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

We assign to ε a permutation $\varphi_\varepsilon = \varphi \in P_n$, defined for $i \in N$, as follows

$$\varphi(i) = \begin{cases} i+1 & \text{if } i_k \leq i < j_k,\ k = 1, 2, \ldots, m \\ i_k & \text{if } i = j_k \\ i & \text{otherwise} \end{cases}$$

By the above definition, the permutation preserves the value $n = \varphi(n)$. Therefore, when applied to the rows and the columns of the matrix A, φ permutes only the elements in the last row and in the last column, and the elements on the diagonal, while the remaining entries of A remain equal to zero. Further, it is clear that φ consists of m disjoint cycles (cyclic permutations) $(i_k, i_k+1, \ldots, j_k)$. Having this in mind, we can easily verify that

$$m_i^\star(A_{\varphi\varphi})) = \begin{cases} i + \varepsilon(i) & \text{if } 1 \leq i < n \\ n-1 & \text{if } i = n \end{cases}$$

If ε runs over all 2^{n-2} of possible mappings $\varepsilon : N \to \{0,1\}$ with fixed values $\varepsilon(n-1) = \varepsilon(n) = 0$, then the corresponding permutations φ create 2^{n-2} of permuted matrices $A_{\varphi\varphi}$ with distinct lower bounds in intervals

$$\langle m^\star(A_{\varphi\varphi}), M^\star(A_{\varphi\varphi})\rangle \cap \mathcal{B}^<(n) = \mathcal{F}^<(A_{\varphi\varphi}) \neq \emptyset$$

□

Corollary 4.4. *Let $\mathcal{B}$ be a dense max-min algebra with $|\mathcal{B}| > 1$, let $n > 2$. Then the eigenspace structure of a matrix in $\mathcal{B}(n,n)$ cannot be, in general, completely described in polynomial time, with respect to n.*

References

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

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

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

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

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

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

[7] E. Sanchez, *Resolution of eigen fuzzy sets equations*, Fuzzy Sets and Systems **1** (1978), 69-74.

[8] Yi-Jia Tan, *Eigenvalues and eigenvectors for matrices over distributive lattices*, Lin. Algebra Appl. **283** (1998), 257-272.

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

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

Preservation of Efficiency and Inefficiency Classification in Data Envelopment Analysis*

Luka Neralić, University of Zagreb, Faculty of Economics
Trg J. F. Kennedya 6, 10000 Zagreb, Croatia
E-mail: lneralic@efzg.hr

Extended Abstract. Sensitivity Analysis in Data Envelopment Analysis (abbreviation: DEA) for the Additive model (see Charnes *et al.* [1]) was studied by Charnes and Neralić [2] for the case of the simultaneous change of all inputs or/and of all outputs of an arbitrary efficient Decision Making Unit (abbreviation: DMU) preserving its efficiency. Sufficient conditions for preserving efficiency of DMU under these changes were obtained. Sensitivity in DEA for arbitrary perturbations (and for non-negative perturbations) of all data in the Additive model was studied by Neralić [4]. For some recent developments of sensitivity and stability analysis in DEA see Cooper *et al.* [3].

The aim of this paper is firstly to study the region of efficiency around an efficient DMU_o according to the Additive model in DEA, which is projection of an inefficient DMU_g to the efficiency frontier, under the simultaneous non-negative perturbations of all data of all DMUs preserving efficiency of DMU_o. Based on sufficient conditions for preserving efficiency of DMU_o under these changes (see Theorem 1, p. 54) region around DMU_o and the corresponding region around DMU_g are obtained. Considering relationship between these regions, conditions for simultaneous preservation of efficiency of DMU_o and inefficiency of DMU_g are given. Secondly, using the result on simultaneous preservation of efficiency of DMU_o and inefficiency of DMU_g (see Theorem 3, p. 55) the case of region of joint efficiency around every efficient DMU with the corresponding region around each inefficient DMU is studied under the non-negative perturbations of all data of all DMUs. Sufficient conditions for preserving efficiency of all efficient DMUs and inefficiency of all inefficient DMUs simultaneously are obtained (see Theorem 4, p. 56). An illustrative example with data from Seiford and Thrall [5] is provided.

*This is an extended abstract of the paper L. NERALIĆ, *Preservation of Efficiency and Inefficiency Classification in Data Envelopment Analysis*, Mathematical Communications **9** (2004) 51-62. This research was partly supported by the MZT grant 0067010 of the Republic of Croatia.

Because conditions in Theorem 1 are sufficient but not necessary, so are conditions in Theorem 3 and Theorem 4. An open question is to find necessary and sufficient conditions for these cases. In order to apply the results in practice it is necessary to make an algorithm and computer code, which is a challenge for the research in the future.

An open question is also to find (necessary and) sufficient conditions for simultaneous preservation of efficiency of efficient DMU(s) and inefficiency of inefficient DMU(s) according to the Charnes-Cooper-Rhodes (CCR) model and the Banker-Chanes-Cooper (BCC) model under the non-negative (or arbitrary) changes of outputs or/and inputs of all DMUs. The same holds for the proportionate change of all data for the Additive, CCR and BCC models. The results for these cases will be discussed elsewhere.

Key words: Data Envelopment Analysis, Efficiency, Additive Model, Preservation of Efficiency and Inefficiency Clasification, Sensitivity Analysis, Linear Programming.

References

[1] A. CHARNES, W. W. COOPER, B. GOLANY, L. M. SEIFORD, J. STUTZ, *Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions*, Journal of Econometrics **30** (1985) 91-107.

[2] A. CHARNES, L. NERALIĆ, *Sensitivity Analysis of the Additive Model in Data Envelopment Analysis*, European Journal of Operational Research **48** (1990) 332-341.

[3] W. W. COOPER, S. LI, L. M. SEIFORD, K. TONE, R. M. THRALL, J. ZHU, *Sensitivity and Stability Analysis in DEA: Some Recent Developments*, Journal of Productivity Analysis **15** (2001) 217-246.

[4] L. NERALIĆ, *Sensitivity in Data Envelopment Analysis for Arbitrary Perturbations of Data*, Glasnik Matematički **32** (1997) 315-335.

[5] L. M. SEIFORD, R. M. THRALL, *Recent Developments in DEA: The Mathematical Programming Approach to Frontier Analysis*, Journal of Econometrics **46** (1990) 7-38.

RECENT APPROACHES TO THE QUADRATIC ASSIGNMENT PROBLEM

Janez Povh[a], Renata Sotirov[b]

[a]School for business and management, Na Loko 2, 8000 Novo mesto, Slovenia

[b]The University of Melbourne, Department of Mathematics and Statistics, Parkville, Australia

email: janez.povh@guest.arnes.si, rsotirov@ms.unimelb.edu.au

Abstract In this paper we present several recent approaches for solving the Quadratic Assignment Problem (QAP). We first recall several semidefinite relaxations of QAP, and then present new QAP relaxations based on copositive programming. We discuss on the strength of the new relaxations and compare them with SDP ones.

Key words: quadratic assignment problem, semidefinite programming, copositive programming.

1 Introduction

The Quadratic Assignment Problem (QAP) can be stated in the following way. Let Π be the set of $n \times n$ permutation matrices. For given A and B real symmetric $n \times n$ matrices and C real $n \times n$ matrix, find a permutation matrix $X \in \Pi$ which minimizes

$$\text{(QAP)} \quad \text{OPT}_{\text{QAP}} = \min_{X \in \Pi} \text{trace}\,(AXB^T + C)X^T. \tag{1}$$

This formulation is called the trace formulation of QAP. The QAP is nowadays widely considered as a classical combinatorial optimization problem, but it is also known as a generic model for various real-life problems.

The QAP is well known to be NP–hard, and even finding an ϵ–approximation of QAP is a NP–hard problem. The computational effort to solve the QAP is very likely to grow exponentially with the problem size, and problems of size $n \geq 25$ are currently considered as huge problems. The recent developments in algorithms as well as in computational platforms have resulted in a large improvement in the capability to solve QAPs exactly. In 2002, Anstreicher et al. [2] made a break–through by solving a number of previously–unsolved large QAPs. Their computations are considered to be among the most extensive computations ever performed to solve discrete optimization problems. Anstreicher et al. [2] incorporated the Branch & Bound (B&B) algorithm on a supercomputer. For the summary of recent advances in the solution of QAP by B&B see [1].

The study of lower bounds for QAP is very important for the development of B&B algorithms. Namely, the performance of B&B algorithms depends on the computational quality and efficiency of the lower bounds. The most recent and promising trends of research for the bounding methods for QAP are based on semidefinite programming [8, 10] and lift–and–project techniques [4]. Bounding procedure that is used in [2] is known as a convex quadratic programming bound (QPB), and it combines orthogonal, semidefinite and convex quadratic relaxations in a nontrivial way. The bounds computed in [4] and [8] are stronger than QPBs. SDP bounds were also tested in B&B settings and the promising results are reported, see [8].

In this paper, we first recall semidefinite programming (SDP) relaxations of QAP. Semidefinite programming is an extension of linear programming where the nonnegativity constraints are replaced by positive semidefiniteness constraints on matrix variables. SDP has been shown to be a very powerful tool for providing tight relaxations for hard combinatorial problems, notably QAP. In Section 2, we recall and summarize the approach from [8, 10] to derive SDP relaxations for QAP. For deriving SDP relaxations we use new matrix representations derived in [9]. All SDP relaxations are formulated in the space of symmetric matrices of order $(n-1)^2+1$. The strongest SDP relaxation of QAP contains $O(n^3)$ equations and $O(n^4)$ sign constraints.

Rendl and Sotirov [8] solve that relaxation using the Bundle Method and report very strong bounds. In Section 3 we present several new QAP relaxations based on the copositive programming. We give first a reformulation of the QAP as linear problem over the cone of completely positive matrices. Since copositive programming is not tractable, we approximate the cone of the symmetric completely positive matrices with the cone of symmetric doubly nonnegative matrices, and obtain a new relaxation. The relaxation obtained in this way is relaxed further by omitting the nonnegativity constraints and adding extra $O(n^3)$ constraints. In Section 4 we discuss on the strength of the new relaxations and compare them with SDP relaxations.

1.1 Notation

In this paper we consider following convex cones: symmetric $n \times n$ matrices are denoted by $\mathcal{S}_n = \{X \in \mathbb{R}^{n\times n} : X = X^T\}$, the $n \times n$ positive semidefinite matrices $\mathcal{S}_n^+ = \{X \in \mathcal{S}_n : y^TXy \geq 0, \forall y \in \mathbb{R}^n\}$, the $n \times n$ copositive matrices $\mathcal{C}_n = \{X \in \mathcal{S}_n : y^TXy \geq 0, \forall y \in \mathbb{R}^n_+\}$, the $n \times n$ completely positive matrices $\mathcal{C}_n^* = \{X = \sum_{i=1}^k y_iy_i^T, k \geq 1, y_i \in \mathbb{R}^n_+, 1 \leq i \leq k\}$, and the $n \times n$ symmetric nonnegative matrices $\mathcal{N}_n = \{X \in \mathcal{S}_n : X_{ij} \geq 0, \forall i, j\}$.

For $X, Y \in \mathcal{S}_n$, $X \succeq 0$ (resp. $X \succ 0$) denotes positive semidefiniteness (resp. positive definiteness). For two matrices $X, Y \in \mathbb{R}^{m\times n}$, $X \geq Y$, $(X > Y)$ means $x_{ij} \geq y_{ij}$, $(x_{ij} > y_{ij})$ for all i, j. The Kronecker product of matrices $X = (x_{ij})$ and $Y = (y_{kl})$ is $X \otimes Y = (x_{ij}y_{kl}) = (x_{ij}Y)$, $\forall i, j, k, l$. We denote by $\langle \cdot, \cdot \rangle$ the standard scalar product. For $u, v \in \mathbb{R}^n$ is therefore $\langle u, v \rangle = u^Tv$ and for $X, Y \in \mathbb{R}^{m\times n}$ is $\langle X, Y \rangle = \text{trace}(X^TY)$. For $X \in \mathbb{R}^{m\times n}$, $\text{vec}(X)$ denotes the vector in $\mathbb{R}^{mn}$ that is formed from the columns of the matrix X.

We denote the ith standard unit vector with e_i, the vectors of all ones is $u_n \in \mathbb{R}^n$, a unit matrix is $E_{ij} = e_ie_j^T$. The square matrix of all ones we denote by J_n (or simply J when n is obvious) and the identity matrix by I_n or I. For matrix columns and rows we use MATLAB notation, hence $X(:, i)$ and $X(i, :)$ stands for ith column and row, respectively.

A linear program over the cone $\mathcal{S}_n^+$ is called a semidefinite program, while a linear program over $\mathcal{C}_n$ or $\mathcal{C}_n^*$ is called a copositive program.

2 SDP Relaxations of QAP

In order to derive a SDP relaxation of QAP we rewrite the objective function in the following way

$$\text{trace}\ (AXB + C)X^T = \langle x, \text{vec}(AXB + C)\rangle = x^T(B \otimes A)x + x^Tc,$$

where $x = \text{vec}(X)$ and $c = \text{vec}(C)$. Therefore QAP becomes

$$\min\{x^T(B \otimes A)x + x^Tc : \ x = \text{vec}(X), X \in \Pi\}. \tag{2}$$

Since $c^Tx = \text{trace}(\text{Diag}(c)(xx^T))$, for $X \in \Pi$, the equivalent formulation to (2) is

$$\min\{\text{trace}(B \otimes A + \text{Diag}(c))xx^T : \ x = \text{vec}(X), X \in \Pi\}. \tag{3}$$

Thus, the feasible set of the last formulation of QAP is

$$\mathcal{P} := \text{conv}\{xx^T : x = \text{vec}(X), X \in \Pi\}.$$

In order to obtain tractable relaxations for QAP which will give good lower bounds for QAP, we need to approximate the set $\mathcal{P}$ by larger sets containing $\mathcal{P}$. We first impose a semidefiniteness constraint on elements $Y \in \mathcal{P}$, i.e., $Y - \text{diag}(Y)\text{diag}(Y)^T \succeq 0$. This condition is well known to be equivalent to the convex constraint

$$\begin{pmatrix} 1 & \tilde{y}^T \\ \tilde{y} & Y \end{pmatrix} \succeq 0, \ \tilde{y} = \text{diag}(Y). \tag{4}$$

In order to derive the SDP relaxations we exploit the fact that the row and column sums of permutation matrices are one. In [9] is proved that the set of $n \times n$ matrices with row and column sum equal to 1 is equivalent to the following set:

$$\left\{X \in \mathbb{R}^{n\times n} : X = \frac{1}{n}(J_n - VJ_{n-1}V^T) + VX(1 : n-1, 1 : n-1)V^T\right\};$$

where

$$V := \begin{pmatrix} I_{n-1} \\ \hline -u_{n-1}^T \end{pmatrix}.$$

Now is easy to prove (see [8, 9]) that for any $Y \in \mathcal{P}$ exists a symmetric matrix R of order $(n-1)^2 + 1$, indexed from 0 to $(n-1)^2$, such that

$$R \succeq 0, \ R_{00} = 1 \text{ and } Y = WRW^T, \ \text{ where } \ W = \left(\frac{1}{n}(e_{n^2} - (V \otimes V)e_{(n-1)^2}), V \otimes V\right). \tag{5}$$

If Y is an extreme point of $\mathcal{P}$ then $R_{ij} \in \{0, 1\}$, otherwise $R_{ij} \in [0, 1]$ for $i, j \in \{0, \ldots, (n-1)^2\}$, see [9]. By exploring that fact, we can formulate SDP relaxation that impose constraints on the set of smaller dimensional space than $\mathcal{P}$ and explore sparsity efficiently, see [9].

From conditions (4) and (5) we derive the following set $\hat{\mathcal{P}}$ containing $\mathcal{P}$.

$$\hat{\mathcal{P}} := \left\{Y \in \mathcal{S}_{n^2} : \exists R \ \text{ s. t. } \ R_{00} = 1, Y = WRW^T, \tilde{y} = \text{diag}(Y), \begin{pmatrix} 1 & \tilde{y}^T \\ \tilde{y} & Y \end{pmatrix} \succeq 0\right\}.$$

We can show that the interior of $\mathcal{P}$ is not empty by finding a barycenter of $\mathcal{P}$. In what it follows, we explore further a structure of the elements from $\mathcal{P}$. To express the zero pattern, we index the elements of the matrix $Y \in \mathcal{P}$ by $y_{r,s} = Y_{(i,j)(k,l)}$ for $r, s \in \{1, \ldots, n\} \times \{1, \ldots, n\}$, $i, j, k, l \in \{1, \ldots, n\}$. The zero pattern is covered by the following equalities:

$$y_{rs} = 0 \ \text{ for } \ r = (i, j), s = (i, k), \ \text{ or } \ r = (j, i), s = (k, i), \ j \neq k. \tag{6}$$

We collect all these equalities in the constraint $G(Y) = 0$ and arrive to the following relaxation, see [8]

$$(\text{QAP}_{R_2}) \quad \min\{\text{trace}(B \otimes A + \text{Diag}(c))Y) : G(Y) = 0, Y \in \hat{\mathcal{P}}\}.$$

Note that the relaxation is formulated in the space of symmetric matrices of order $(n-1)^2 + 1$. Relaxation QAP_{R_2} contains $O(n^3)$ equations ($n^3 - n^2$ to be precise). Model QAP_{R_2} is introduced in [10] as the *Gangster model*. The results in [8] show that QAP_{R_2} provides very tight approximations of OPT_{QAP}. Because of the large number of the constraints in QAP_{R_2} the interior-point method is not appropriate for solving this relaxation, and therefore Rendl and Sotirov [9] solve it by the Bundle Method.

The relaxation QAP_{R_2} can be further tightened by adding nonnegativity constraints

$$(WRW^T)_{rs} \geq 0, \quad \forall r, s = 1, \ldots, n^2. \tag{7}$$

We collect the inequalities (7) which are not yet covered by $G(WRW^T) = 0$ in the constraint $N(WRW^T) \geq 0$ and arrive at the final relaxation, see [8, 10]

$$(\text{QAP}_{R_3}) \quad \min\{\text{trace}(B \otimes A + \text{Diag}(c))Y) : G(Y) = 0, N(Y) \geq 0, Y \in \hat{\mathcal{P}}\}.$$

The resulting SDP has $O(n^4)$ sign constraints and $O(n^3)$ equality constraints. The relaxation QAP_{R_3} can not be solved straightforward by interior-point methods for interesting instances ($n \geq 15$). In [8] the QAP_{R_3} relaxation is solved successfully (up to $n = 32$) by the Bundle Method.

3 QAP Models Based on Copositive Relaxation

In [3] Anstreicher and Wolkowicz show that the quadratically constrained quadratic programs where the quadratic constraints correspond to the matrix orthogonality condition $X^TX = I$ can equivalently be expressed through semidefinite programming. We follow their idea and find a new Lagrangian relaxation of the QAP. Using the following characterization of the set of permutation matrices $\Pi = \{X \geq 0 \colon X^TX = I\}$ we first rewrite the QAP as

$$\min\{\ \mathrm{trace}(AXB^T + C)X^T : X^TX = XX^T = I,\ X \geq 0\}.$$

Note that we added the (redundant) constraint $XX^T = I$. In the sequel we dualize the orthogonality constraints and write $x = \mathrm{vec}(X)$, $c = \mathrm{vec}(C)$.

$$\begin{aligned}
\mathrm{OPT}_{\mathrm{QAP}} &= \min_{X\geq 0} \mathrm{trace}(AXB + C)X^T + \max_{S,T\in\mathcal{S}_n} \langle S, I - XX^T\rangle + \langle T, I - X^TX\rangle \\
&= \min_{X\geq 0}\max_{S,T\in\mathcal{S}_n} \mathrm{trace}(S) + \mathrm{trace}(T) + x^T(B\otimes A + \mathrm{Diag}(c) - I\otimes S - T\otimes I)x \\
&\geq \max_{S,T\in\mathcal{S}_n}\min_{X\geq 0} \mathrm{trace}(S) + \mathrm{trace}(T) + x^T(B\otimes A + \mathrm{Diag}(c) - I\otimes S - T\otimes I)x \\
&= \max\ \mathrm{trace}(S) + \mathrm{trace}(T) \\
&\quad\ \text{s. t. } S, T \in \mathcal{S}_n \\
&\quad\quad\ B\otimes A + \mathrm{Diag}(c) - I\otimes S - T\otimes I \in \mathcal{C}_{n^2} \\
&= \min\ \langle B\otimes A + \mathrm{Diag}(c), Y\rangle \\
&\quad\ \text{s. t. } Y \in \mathcal{C}^*_{n^2} \\
&\quad\quad\ \sum_i Y^{ii} = I \qquad (\mathrm{QAP}_{LD}) \\
&\quad\quad\ \mathrm{trace}(Y^{ij}) = \delta_{ij}.
\end{aligned}$$

In the last formulation Y^{ij} is (i,j)th block of Y, i. e. $Y^{ij} = [Y_{pq}]$, $p = (i-1)n+1, \ldots, in$, $q = (j-1)n+1, \ldots, jn$ and δ_{ij} is the Kronecker delta. The last equality above follows from strict feasibility of the last but one problem, i.e., for $T = S = -\alpha I$ the matrix $B\otimes A + \mathrm{Diag}(c) - I\otimes S - T\otimes I$ is positive definite for α sufficiently large, hence in the interior of $\mathcal{C}_{n^2}$ and therefore strictly feasible.

There may exist arbitrary large duality gap between $\mathrm{OPT}_{\mathrm{QAP}}$ and solution of QAP_{LD}. Let $B = \alpha(J - I)$, $A = \beta J$ and $C = 0$, where $\alpha, \beta > 0$. For every permutation matrix X is $\mathrm{trace}(AXB^T + C)X^T = \alpha\beta n(n-1)$, hence $\mathrm{OPT}_{\mathrm{QAP}} = \alpha\beta n(n-1)$. In the relaxed problem matrix $B\otimes A$ and every feasible matrix Y have only nonnegative entries, hence $\mathrm{OPT}_{\mathrm{LD}} \geq 0$. If we take $Y = 1/n\, I_{n^2}$, then Y is feasible for the relaxed problem and $\langle B\otimes A, Y\rangle = 0$. This means that Y is optimal and $\mathrm{OPT}_{\mathrm{LD}} = 0$. The duality gap is $\alpha\beta n(n-1)$ and may be arbitrary large.

In the sequel we add to the original problem redundant constraints $\mathrm{trace}(JXD_{ij}X^T) = 1$, where $D_{ij} = \frac{1}{2}(E_{ij} + E_{ji})$, $1 \leq i < j \leq n$. By adding new constraints to the (QAP_{LD}) we obtain the following copositive program

$$(\mathrm{QAP}_{CP}) \qquad \begin{array}{lrcl}
\min & \multicolumn{3}{l}{\langle B\otimes A + \mathrm{Diag}(c), Y\rangle} \\
\text{s. t.} & Y & \in & \mathcal{C}^*_{n^2} \\
& \sum_i Y^{ii} & = & I \\
& \mathrm{trace}(Y^{ij}) & = & \delta_{ij},\ 1 \leq i \leq j \leq n \\
& \mathrm{trace}(JY^{ij}) & = & 1,\ 1 \leq i < j \leq n.
\end{array}$$

The new relaxation is tight. In [6] Povh shows that $\mathrm{OPT}_{\mathrm{QAP}} = \mathrm{OPT}_{\mathrm{CP}}$. This result leads to several new approximation models for the QAP. If we replace the intractable constraint $Y \in \mathcal{C}^*_{n^2}$

with some good tractable constraint, then we may get a good lower or upper bound for the $\mathrm{OPT}_{\mathrm{QAP}}$. Recently Parillo has suggested in [7] a hierarchy of cones, which approximates the copositive cone from the inside arbitrary close. The first cone in this hierarchy (we will respect the usual notation and denote it as $\mathcal{K}_0$) is the sum $\mathcal{N}_n + \mathcal{S}^+_n$, and the $(r+1)$-th cone (denoted by $\mathcal{K}_r$) is the cone of those matrices Y for which the polynomial $(\sum_{ij} Y_{ij}x_i^2x_j^2)(\sum_i x_i^2)^r$ is a sum of squares. The hierarchy of dual cones approximates the cone $\mathcal{C}^*_n$ from the outside. We will focus on these cones. For the first member we have a cone of symmetric doubly nonnegative matrices $\mathcal{K}^*_0 = \mathcal{N}_n \cap \mathcal{S}^+_n$, while the $\mathcal{K}^*_1$ is already quite complicated and it's description requires n matrices with n positive semidefinite constraints, see [7].

Our next model will be therefore:

$$(\mathrm{QAP}_{K0^*}) \qquad \begin{array}{lrcl}
\min & \multicolumn{3}{l}{\langle B\otimes A + \mathrm{Diag}(c), Y\rangle} \\
\text{s. t.} & Y & \in & \mathcal{N}_{n^2} \cap \mathcal{S}^+_{n^2} \\
& \sum_i Y^{ii} & = & I \\
& \langle I, Y^{ij}\rangle & = & \delta_{ij},\ 1 \leq i \leq j \leq n \\
& \langle J, Y^{ij}\rangle & = & 1,\ 1 \leq i < j \leq n.
\end{array}$$

We have to emphasize that this is already expensive model since the constraint $Y \in \mathcal{N}_{n^2}$ implies $O(n^4)$ linear inequalities. We can get a cheaper model by demanding only $Y \in \mathcal{S}_{n^2}$, but this leads to a poor lower bound, which is only slightly better than Hoffman-Wielandt lower bound. A good compromise seems to be the Gangster constraint, see the zero pattern in (6). If we combine the sign constraint $Y \geq 0$ with second and the third constraint from QAP_{K0^*}, we get that the off-diagonal elements in diagonal blocks and the diagonal elements in the off-diagonal blocks of Y must be zero. This is exactly the Gangster constraint. Our final relaxation is

$$(\mathrm{QAP}_{GM}) \qquad \begin{array}{lrcl}
\min & \multicolumn{3}{l}{\langle B\otimes A + \mathrm{Diag}(c), Y\rangle} \\
\text{s. t.} & Y & \in & \mathcal{S}^+_{n^2} \\
& G(Y) & = & 0 \\
& \sum_i Y^{ii}_{jj} & = & 1 \\
& \langle I, Y^{ii}\rangle & = & 1,\ 1 \leq i \leq n \\
& \langle J, Y^{ij}\rangle & = & 1,\ 1 \leq i < j \leq n.
\end{array}$$

In the following section we compare copositive relaxations with SDP models.

4 Numerical results

In this section we compare the lower bounds for the previously derived relaxations. Table 1 reads as follows. The first column gives the problem instances and their sizes. In the second columns we provide the optimal value for each instance, and the remaining columns are bounds for the relaxations derived in this paper. The QAP_{R_2} and QAP_{R_3} bounds are computed by the Bundle Method, see [9]. The QAP_{GM} bounds for all listed instances, except Nug20 and Nug21, we compute by SDPT3 solver (url: `http://www.math.nus.edu.sg/~mattohkc/sdpt3.html`). Difficulties arising solving problems with IPMs for $n > 18$ are due to the hardware limitations. For computing approximate solutions for Nug20 and Nug21 we use a Boundary Point Method (BPM) introduced in [5], which seems to be an efficient method for solving large problems (see Table 1). Our preliminary results in computing QAP_{K0^*} bounds with IPM are promising, i.e., the obtained bound for Nug12 is 568, which is better bound than any computed SDP bound for the same problem.

5 Conclusions

In this paper we recall SDP relaxations for QAP and derive new relaxations based on copositive programming for QAP. Both approaches turn out to be strong tools for developing approximate

problem	OPT	QAP_{R_2}	QAP_{R_3}	QAP_{GM}
Nug12	578	528	557	529
Nug14	1014	958	992	959
Nug15	1150	1069	1122	1070
Nug16a	1610	1526	1570	1528
Nug16b	1240	1136	1188	1138
Nug18	1930	1798	1852	1801
Nug20	2570	2380	2451	2380
Nug21	2438	2244	2323	2244

Table 1: Bounds for SDP and copositive relaxations

models for the QAP. The resulted relaxations contain up to $O(n^4)$ constraints and therefore IPMs are not appropriate solving methods for larger instances. Therefore, for SDP and larger copositive relaxations ($n > 18$) we use the Bundle Method and the Boundary Point Method that has been recently introduced by Dukanovic et al. [5]. The preliminary results show that the BPM is a promising method for solving copositive relaxations. Our results also show that QAP_{KO^*} provides strong lower bounds for QAP. Thus, our future work will be solving QAP_{KO}-relaxation for large instances with the new Boundary Point Method.

References

[1] Anstreicher, K.: Recent Advances in the Solution of Quadratic Assignment Problems. *Math. Program. B*, 97:27–42, 2003.

[2] Anstreicher, K., Brixius, N., Goux, J.–P., Linderoth, J.: Solving Large Quadratic Assignment Problems on Computational Grids. *Math. Program. B*, 91:563–588, 2002.

[3] Anstreicher, K., Wolkowicz, H.: On Lagrangian Relaxation of Quadratic Matrix Constraints. *SIAM J. Matrix Anal. Appl.*, 22(1):41–55, 2000.

[4] Burer, S., Vandenbussche, D.: Solving Lift-and-Project Relaxations of Binary Integer Programs. Technical report, June 2004.

[5] Dukanovic, I., Povh, J., Rendl, F., Wiegele, A.: A Boundary Point Method for Semidefinite Programs. University of Klagenfurt, Austria, Working paper, 2005.

[6] J. Povh: Application of semidefinite and copositive programming in combinatorial optimization. PhD Thesis, University of Ljubljana, Faculty of mathematics and physics, Ljubljana, 2005 (in preparation).

[7] Parrilo, P. A.: Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization. Dissertation, California Institute of Technology, 2000.

[8] Rendl, F., Sotirov, R.: Bounds for the Quadratic Assignment Problem Using the Bundle Method. Technical report, University of Klagenfurt, Austria, 2003.

[9] Rendl, F., Sotirov, R.: Representations of Matrices with Sum of Row and/or Columns Equal to One. Technical report, The University of Melbourne, Australia, 2005.

[10] Zhao, Q., Karisch, S. E., Rendl, F., and Wolkowicz, H.: Semidefinite Programming Relaxations for the Quadratic Assignment Problem. *J. Comb. Optim.*, 2:71-109, 1998.

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 4
Environment and Human Resource Management

CROSS - TABULATION ANALYSES OF THE SURVEY RESEARCH ON THE MORAL VALUES

Josip Arnerić, Elza Jurun
University of Split, Faculty of Economics
Matice hrvatske 31
21000 Split, Croatia
jarneric@efst.hr, elza@efst.hr

Abstract

This paper is a part of the extensive on - going scientific research on contemporary family and moral values. Among a wide spectrum of results the authors have selected those pertaining to the state of moral values from the point of view of individuals 18 years and older living in the Split-Dalmatian County. This paper pays particular attention towards numerous cross-tabulation analyses of the survey research based on sample of 1000 units. The purpose of this analysis was to determine which factors have a significant and which ones a slight influence on forming of moral values.

Keywords: survey research on moral values, cross-tabulation analysis, bribe, euthanasia

1. INTRODUCTION

Motivated by previous research on values in the world and in Croatia at the beginning of the third millennium, the authors have decided to carry out this study on the state of family and values in the Split-Dalmatian County. On one hand, this paper is a part of the on - going extensive research, and on the other hand it is a continuation of the paper "Family Values Today - Case Study of Split-Dalmatian County" with which it forms a coherent unity. The paper was published in "Proceedings of the 7th International Symposium on Operation Research" and the cross-tabulation analyses were announced in the conclusion of the paper.

Based on the important data on the state of the contemporary family and moral values, the authors have chosen the parts of the complex scientific research, carried out in the area of Split-Dalmatian County. A sample of 1000 interviewees 18 years and older with current residence in the Split-Dalmatian County meets all the criteria of the contemporary statistics and makes a representative and random sample. Consequently, according to sex, age, employment, profession, qualifications and marital status the structure of the interviewees gives a precise picture of the structure of the population of the Split-Dalmatian County according to 2001 census and indicates no significant statistical deviations.

The given answers, which could at first be defined as easy-going, do not remain on the level of description of forms of behaviour, but help us reveal the essence of interpersonal relations within a marriage and a family as well as relations of an individual to different spheres of his social environment.

2. AN ABSTRACTION ON SOME MORAL VALUES IN SPLIT-DALMATIAN COUNTY

For the purpose of this paper parts of the analysis of attitudes towards certain moral values as a part of the complex output of the scientific research have been prepared in the adequate form. Statistical analyses have been carried out on the influence of sex and age of the interviewees, their upbringing, place of the childhood residence until the age of 15, marital status, qualifications, employment, current residence place and financial status on his opinion on abortion, euthanasia, homosexuality and bribe.

According to the authors, these attitudes are symptomatic because they go to the very heart of the attitudes towards life. Therefore, this abstraction focuses on the attitude of the interviewees towards abortion, euthanasia, homosexuality and towards giving and taking a bribe. So, the authors analyse the attitude which is not abstract one, but the one which manages daily decisions of the human being as the person. These decisions have recognisable forms and far-reaching influence on the contemporary intimate, family, religious, business and social life and therefore create an atmosphere of the "culture of life".

Due to the limited size of this paper it has been necessary to extract even a smaller output which will be presented. Therefore, the analysis focuses on the cross-tabulation analysis of the attitudes of the interviewees towards bribe and euthanasia as towards the so called representatives of the attitudes towards "business moral" and "life ethics".

2.1. CROSSTABULATION ANALYSES OF THE SURVEY RESEARCH ON THE BRIBE

Among many attitudes in the context of the world of business it has been necessary to choose one which will act as a representative of all others. The authors have chosen the attitude of the interviewee towards bribe believing that it mirrors symptomatically all other attitudes within the context of business moral.

Hence, a very detailed cross-tabulation analysis has been carried out aiming at the analysis of the factors which statistically significant influence the attitude of the interviewee towards bribe. We will give the results of the crosstabulation analysis of sex and age of interviewees, place of childhood residence until the age of 15, current residence, employment, qualifications, financial status, marital status and their attitudes towards bribe.

Table 1. *Influence of the sex of interviewees on their attitudes towards bribe*

YOU WOULD OFFER A BRIBE	SEX: HAVEN T ANSWER	SEX: FEMALE	SEX: MALE	TOTAL
HAVEN T ANSWERED	0	7	4	11
	.0%	63.6%	36.4%	100.0%
	.0%	1.1%	1.1%	1.1%
NEVER	0	145	102	247
	.0%	58.7%	41.3%	100.0%
	.0%	23.3%	27.3%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	2	281	161	444
	.5%	63.3%	36.3%	100.0%
	50.0%	45.1%	43.2%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT, JOB)	2	166	71	239
	.8%	69.5%	29.7%	100.0%
	50.0%	26.6%	19.0%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	0	24	35	59
	.0%	40.7%	59.3%	100.0%
	.0%	3.9%	9.4%	5.9%
TOTAL	4	623	373	1000
	.4%	62.3%	37.3%	100.0%
	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	21.967a	8	.005
Likelihood Ratio	22.589	8	.004
N of Valid Cases	1000		

a. 6 cells (40.0%) have expected count less than 5. The minimum expected count is .04.

As it can be seen in table 1. sex in general has statistically significant influence on attitude towards bribe ($Sig.\chi^2 = 0,05\%$). For example 19% males would offer a bribe under special circumstances (to solve a vital issue like housing, job, education...), but even 26,6% females would do the same under the above mentioned conditions. Just opposite, while 9,4% males would offer a bribe always, if that is the safest and easiest way of achieving goal, only 3,9% females would do the same.

Table 2. *Influence of the age of interviewees on their attitudes towards bribe*

YOU WOULD OFFER A BRIBE	AGE: HAVEN T ANSWERED	AGE: 18 - 30	AGE: 31 - 40	AGE: 41 - 50	AGE: 51 - 60	AGE: 61 OR MORE	TOTAL
HAVEN T ANSWERED	0	5	1	3	1	1	11
	0%	45.5%	9.1%	27.3%	9.1%	9.1%	100.0%
	.0%	1.5%	.4%	1.1%	1.0%	2.1%	1.1%
NEVER	2	85	45	67	31	17	247
	.8%	34.4%	18.2%	27.1%	12.6%	6.9%	100.0%
	28.6%	25.1%	19.9%	24.2%	29.5%	36.2%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	3	124	106	139	52	20	444
	.7%	27.9%	23.9%	31.3%	11.7%	4.5%	100.0%
	42.9%	36.7%	46.9%	50.2%	49.5%	42.6%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT, JOB)	2	90	65	57	18	7	239
	.8%	37.7%	27.2%	23.8%	7.5%	2.9%	100.0%
	28.6%	26.6%	28.8%	20.6%	17.1%	14.9%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	0	34	9	11	3	2	59
	.0%	57.6%	15.3%	18.6%	5.1%	3.4%	100.0%
	.0%	10.1%	4.0%	4.0%	2.9%	4.3%	5.9%
TOTAL	7	338	226	277	105	47	1000
	.7%	33.8%	22.6%	27.7%	10.5%	4.7%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	38.736a	20	.007
Likelihood Ratio	38.733	20	.007
Linear-by-Linear Association	14.205	1	.000
N of Valid Cases	1000		

a. 11 cells (36.7%) have expected count less than 5. The minimum expected count is .08.

In this case general conclusion is the same as in previous cross-analysis. As illustration of statistically significant influence of the age to the attitude towards bribe it can be underlined that 3.5 times (10.1%) more of the interviewees in the age 18-30 would give bribe always as opposed to those who are in the age 51-60 (2.9%).

Table 3. *Influence of residence of the childhood until the age of 15 on attitude towards bribe*

YOU WOULD OFFER A BRIBE	RESIDENCE OF THE CHILDHOOD UNTIL THE AGE OF 15: HAVEN T ANSWERED	VILLAGE	TOWN WITH LESS THAN 10.000 INHABITANTS	TOWN WITH UP TO 30.000 INHABITANTS	TOWN WITH OVER 30.000 INHABITANTS	TOTAL
HAVEN T ANSWERED	0	5	0	1	5	11
	.0%	45.5%	.0%	9.1%	45.5%	100.0%
	.0%	1.0%	.0%	.9%	1.9%	1.1%
NEVER	0	127	31	26	63	247
	.0%	51.4%	12.6%	10.5%	25.5%	100.0%
	.0%	26.5%	20.7%	24.1%	24.0%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	1	236	63	51	93	444
	.2%	53.2%	14.2%	11.5%	20.9%	100.0%
	100.0%	49.3%	42.0%	47.2%	35.5%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT. JOB)	0	93	44	21	81	239
	.0%	38.9%	18.4%	8.8%	33.9%	100.0%
	.0%	19.4%	29.3%	19.4%	30.9%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	0	18	12	9	20	59
	.0%	30.5%	20.3%	15.3%	33.9%	100.0%
	.0%	3.8%	8.0%	8.3%	7.6%	5.9%
TOTAL	1	479	150	108	262	1000
	.1%	47.9%	15.0%	10.8%	26.2%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	33.257a	16	.007
Likelihood Ratio	35.257	16	.004
Linear-by-Linear Association	8.433	1	.004
N of Valid Cases	1000		

a. 8 cells (32.0%) have expected count less than 5. The minimum expected count is .01.

In comparison to table 3. where residence of the childhood until the age of 15 statistically significant influences on attitudes towards bribe , current residence place in table 4. has statistically influence at 6% level of significance.

Table 4. *Influence of current residence place on attitude towards bribe*

YOU WOULD OFFER A BRIBE	CURRENT RESIDENCE PLACE: HAVEN T ANSWERED	VILLAGE	TOWN WITH LESS THAN 10.000 INHABITANTS	TOWN WITH UP TO 30.000 INHABITANTS	TOWN WITH OVER 30.000 INHABITANTS	TOTAL
HAVEN T ANSWERED	0	3	1	1	6	11
	.0%	27.3%	9.1%	9.1%	54.5%	100.0%
	.0%	1.0%	6%	7%	1.5%	1.1%
NEVER	3	83	32	34	95	247
	1.2%	33.6%	13.0%	13.8%	38.5%	100.0%
	42.9%	28.1%	20.8%	24.8%	23.3%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	1	137	74	67	165	444
	.2%	30.9%	16.7%	15.1%	37.2%	100.0%
	14.3%	46.4%	48.1%	48.9%	40.5%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT JOB)	2	59	31	27	120	239
	.8%	24.7%	13.0%	11.3%	50.2%	100.0%
	28.6%	20.0%	20.1%	19.7%	29.5%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	1	13	16	8	21	59
	1.7%	22.0%	27.1%	13.6%	35.6%	100.0%
	14.3%	4.4%	10.4%	5.8%	5.2%	5.9%
TOTAL	7	295	154	137	407	1000
	.7%	29.5%	15.4%	13.7%	40.7%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

In the cross-tabulations represented in table 5. there are less significant influences ($\text{Sig.}\chi^2 = 8{,}8\%$) than in table 6., where qualifications have significant influences ($\text{Sig.}\chi^2 = 0{,}03\%$) on attitudes towards bribe.

Table 5. *Influence of employment on attitude towards bribe*

YOU WOULD OFFER A BRIBE	EMPLOYMENT: HAVEN T ANSWERED	PERMANENT EMPLOYEE	TEMPORARY EMPLOYEE	UNEMPLOYED WHO RECEIVES AN UNEPLOYMENT BENEFIT	UNEPLOYED WHO DOES NOT RECEIVES AN UNEMPLOYMENT BENEFIT	RETIRED	TOTAL
HAVEN T ANSWERED	1	3	0	2	4	1	11
	9.1%	27.3%	.0%	18.2%	36.4%	9.1%	100.0%
	3.1%	.7%	.0%	2.4%	1.5%	1.1%	1.1%
NEVER	7	101	13	20	75	31	247
	2.8%	40.9%	5.3%	8.1%	30.4%	12.6%	100.0%
	21.9%	22.2%	20.6%	23.8%	27.4%	33.3%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	14	223	28	36	100	43	444
	3.2%	50.2%	6.3%	8.1%	22.5%	9.7%	100.0%
	43.8%	49.1%	44.4%	42.9%	36.5%	46.2%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT JOB)	6	107	15	20	76	15	239
	2.5%	44.8%	6.3%	8.4%	31.8%	6.3%	100.0%
	18.8%	23.6%	23.8%	23.8%	27.7%	16.1%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	4	20	7	6	19	3	59
	6.8%	33.9%	11.9%	10.2%	32.2%	5.1%	100.0%
	12.5%	4.4%	11.1%	7.1%	6.9%	3.2%	5.9%
TOTAL	32	454	63	84	274	93	1000
	3.2%	45.4%	6.3%	8.4%	27.4%	9.3%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 6. *Influence of qualifications on attitude towards bribe*

YOU WOULD OFFER A BRIBE	QUALIFICATIONS: HAVEN T ANSWERED	DIDN T GO TO SCHOOL	NOT GRADUATED FROM PRIMARY SCHOOL	GRADUATED FROM PRIMARY SCHOOL	GRADUATED FROM VOCATIONAL SCHOOL	GRADUATED FROM SECONDARY SCHOOL	GRADUATED FROM A TWO-YEAR COLLEGE	GRADUATED FROM FACULTY	MASTERS DEGREE	DOCTORS DEGREE	TOTAL
HAVEN T ANSWERED	0	0	0	1	0	6	2	2	0	0	11
	0%	0%	0%	9.1%	0%	54.5%	18.2%	18.2%	0%	0%	100.0%
	.0%	.0%	.0%	1.6%	.0%	1.0%	1.6%	1.3%	.0%	.0%	1.1%
NEVER	1	0	5	22	16	117	28	48	3	7	247
	4%	0%	2.0%	8.9%	6.5%	47.4%	11.3%	19.4%	1.2%	2.8%	100.0%
	100.0%	.0%	38.5%	36.1%	30.8%	20.4%	22.2%	30.6%	50.0%	100.0%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	0	2	5	26	28	258	62	60	3	0	444
	.0%	5%	1.1%	5.9%	6.3%	58.1%	14.0%	13.5%	.7%	.0%	100.0%
	.0%	66.7%	38.5%	42.6%	53.8%	44.9%	49.2%	38.2%	50.0%	.0%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT, JOB)	0	0	2	8	5	157	25	42	0	0	239
	0%	0%	8%	3.3%	2.1%	65.7%	10.5%	17.6%	.0%	.0%	100.0%
	.0%	.0%	15.4%	13.1%	9.6%	27.4%	19.8%	26.8%	.0%	.0%	23.9%
ALWAYS IF THAT S THE SAFEST, EASIEST WAY OF ACHIEVING A AIM	0	1	1	4	3	36	9	5	0	0	59
	0%	1.7%	1.7%	6.8%	5.1%	61.0%	15.3%	8.5%	.0%	.0%	100.0%
	.0%	33.3%	7.7%	6.6%	5.8%	6.3%	7.1%	3.2%	.0%	.0%	5.9%
TOTAL	1	3	13	61	52	574	126	157	6	7	1000
	.1%	.3%	1.3%	6.1%	5.2%	57.4%	12.6%	15.7%	.6%	.7%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 7. *Influence of financial status on attitude towards bribe*

YOU WOULD OFFER A BRIBE	FINANCIAL STATUS: HAVEN T ANSWER	VERY GOOD	GOOD	MODEST BUT SUFFICIENT	LOW	VERY LOW	TOTAL
HAVEN T ANSWERED	1	0	3	4	3	0	11
	9.1%	.0%	27.3%	36.4%	27.3%	.0%	100.0%
	12.5%	.0%	.7%	1.0%	3.4%	.0%	1.1%
NEVER	3	13	100	101	24	6	247
	1.2%	5.3%	40.5%	40.9%	9.7%	2.4%	100.0%
	37.5%	24.5%	22.8%	26.0%	27.3%	26.1%	24.7%
UNDER SPECIAL CIRCUMSTANCES IF I HAD SERIOUS HEALTH PROBLEM	3	24	194	181	33	9	444
	.7%	5.4%	43.7%	40.8%	7.4%	2.0%	100.0%
	37.5%	45.3%	44.2%	46.5%	37.5%	39.1%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT JOB)	0	10	120	84	21	4	239
	.0%	4.2%	50.2%	35.1%	8.8%	1.7%	100.0%
	.0%	18.9%	27.3%	21.6%	23.9%	17.4%	23.9%
ALWAYS IF THAT S THE SAFEST EASIEST WAY OF ACHIEVING A AIM	1	6	22	19	7	4	59
	1.7%	10.2%	37.3%	32.2%	11.9%	6.8%	100.0%
	12.5%	11.3%	5.0%	4.9%	8.0%	17.4%	5.9%
TOTAL	8	53	439	389	88	23	1000
	.8%	5.3%	43.9%	38.9%	8.8%	2.3%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Financial status also significantly influences on attitude towards bribe ($\text{Sig.}\chi^2 = 2{,}2\%$). For example, those who evaluate their financial status as modest but sufficient would offer a bribe 3.5 times less (4,9%) than the interviewees whose financial status is very low (17,4%).

Table 8. *Influence of marriage status on attitude towards bribe*

YOU WOULD OFFER A BRIBE	MARRIAGE STATUS: HAVEN T ANSWERED	SINGLE	MARRIED	DIVORCED	WIDOW(ER)	TOTAL
HAVEN T ANSWERED	0	4	5	0	2	11
	.0%	36.4%	45.5%	.0%	18.2%	100.0%
	.0%	1.2%	.8%	.0%	5.4%	1.1%
NEVER	0	94	137	6	10	247
	.0%	38.1%	55.5%	2.4%	4.0%	100.0%
	.0%	27.1%	23.1%	28.6%	27.0%	24.7%
UNDER SPECIAL CIRCUMSTANCES, IF I HAD SERIOUS HEALTH PROBLEM	1	134	281	9	19	444
	.2%	30.2%	63.3%	2.0%	4.3%	100.0%
	33.3%	38.6%	47.5%	42.9%	51.4%	44.4%
IF THAT WOULD CONTRIBUTE TO SOLVING VITAL ISSUE (FLAT JOB)	1	83	147	5	3	239
	.4%	34.7%	61.5%	2.1%	1.3%	100.0%
	33.3%	23.9%	24.8%	23.8%	8.1%	23.9%
ALWAYS IF THAT S THE SAFEST EASIEST WAY OF ACHIEVING A AIM	1	32	22	1	3	59
	1.7%	54.2%	37.3%	1.7%	5.1%	100.0%
	33.3%	9.2%	3.7%	4.8%	8.1%	5.9%
TOTAL	3	347	592	21	37	1000
	.3%	34.7%	59.2%	2.1%	3.7%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Marriage status is among the factors of strong influence on attitude towards bribe ($\text{Sig.}\chi^2 = 0{,}07\%$).

In comparison to widow(er) who would offer bribe to solve vital problems in 8.1% cases, married people would do the same generally in 3 times more cases (24,8%).

2.2. CROSSTABULATION ANALYSES OF THE SURVEY RESEARCH ON THE EUTHANASIA

It has been difficult to choose the attitude which will in a certain way unite the overall attitude of the interviewee towards "culture of life" in general. Nevertheless, the authors have chosen the attitude towards euthanasia, because it reflects the respect to one's own life as well as to the life of his beloved which has been primary given to us without our contribution or our consent. Recently, we all witnessed the reports in the media on the "top matters" which referred to the situations when the closest relatives, or husbands and wives needed to make an irrevocable decision not on their own lives but on lives of the persons under their tutorage. Therefore, the authors have chosen euthanasia as a particular representative of the complex of attitudes towards life and they have carried out the cross tabulation analyses of the same size and structure as in the case of the attitudes towards bribe.

Table 9. *Influence of residence of the childhood on the opinion towards euthanasia*

EUTHANASIA IS	RESIDENCE OF THE CHILDHOOD UNTIL THE AGE OF 15					
	HAVEN'T ANSWERED	VILLAGE	TOWN WITH LESS THAN 10.000 INHABITANTS	TOWN WITH UP TO 30.000 INHABITANTS	TOWN WITH OVER 30.000 INHABITANTS	TOTAL
HAVEN'T ANSWERED	0	12	1	3	5	21
	.0%	57.1%	4.8%	14.3%	23.8%	100.0%
	.0%	**2.5%**	**.7%**	**2.8%**	**1.9%**	**2.1%**
MURDER	0	254	51	46	75	426
	.0%	59.6%	12.0%	10.8%	17.6%	100.0%
	.0%	**53.0%**	**34.0%**	**42.6%**	**28.6%**	**42.6%**
ALLOWED IN SPECIAL CASES, E.G. WHEN A PATIENT IS KEPT ALIVE	0	100	46	26	107	279
	.0%	35.8%	16.5%	9.3%	38.4%	100.0%
	.0%	**20.9%**	**30.7%**	**24.1%**	**40.8%**	27.9%
ACT OF MERCY TO A PERSON IN GREAT PAIN WAITING TO DIE	1	113	52	33	75	274
	.4%	41.2%	19.0%	12.0%	27.4%	100.0%
	100.0%	**23.6%**	**34.7%**	**30.6%**	**28.6%**	27.4%
TOTAL	1	479	150	108	262	1000
	1%	47.9%	15.0%	10.8%	26.2%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 10. *Influence of current residence place on the opinion towards euthanasia*

EUTHANASIA IS	CURRENT RESIDENCE PLACE					
	HAVEN'T ANSWERED	VILLAGE	TOWN WITH LESS THAN 10.000 INHABITANTS	TOWN WITH UP TO 30.000 INHABITANTS	TOWN WITH OVER 30.000 INHABITANTS	TOTAL
HAVEN'T ANSWERED	0	6	5	4	6	21
	.0%	28.6%	23.8%	19.0%	28.6%	100.0%
	.0%	**2.0%**	**3.2%**	**2.9%**	**1.5%**	**2.1%**
MURDER	4	158	53	60	151	426
	9%	37.1%	12.4%	14.1%	35.4%	100.0%
	57.1%	**53.6%**	**34.4%**	**43.8%**	37.1%	**42.6%**
ALLOWED IN SPECIAL CASES, E.G. WHEN A PATIENT IS KEPT ALIVE	1	55	44	32	147	279
	.4%	19.7%	15.8%	11.5%	52.7%	100.0%
	14.3%	**18.6%**	**28.6%**	**23.4%**	**36.1%**	**27.9%**
ACT OF MERCY TO A PERSON IN GREAT PAIN, WAITING TO DIE	2	76	52	41	103	274
	.7%	27.7%	19.0%	15.0%	37.6%	100.0%
	28.6%	**25.8%**	**33.8%**	**29.9%**	**25.3%**	27.4%
TOTAL	7	295	154	137	407	1000
	7%	29.5%	15.4%	13.7%	40.7%	100.0%
	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

3. FINAL CONSIDERATIONS

This paper is only an abstraction on the parts of extensive on-going research related to the cross-tabulation analyses towards a few moral values. Strictly defined frame of the paper limited presentation and analysis of the results. So, in the case of euthanasia only attitudes of residence of the childhood until the age of 15and current residence place of interviewees have been presented. They are the only factors of statistically significant influence, because all others cross-tabulation analyses resulted at significant level more than 10%. Contrary to bribe which is under strong influence of numerous factors, the opinion on euthanasia is much more stable and only surroundings during upbringing as well as current residence environment could significantly destabilize "culture of life". This is only one of numerous considerations which originates from the cross-tabulation analyses and the authors hope to publish all them in next research paper.

4. REFERENCES

1. E. Jurun; L. Tomašević (2003): "*Family Values Today - Case Study of Split-Dalmatian County*", Proceedings of the 7th International SOR 2003. Podčetrtek, Slovenia
2. H.F. Weisberg; J.A. Krosnick; B.D. Bowen: "*Survey research, Polling and Data Analysis*" *Third Edition*, (2005) Ohio State University, Assessment Systems Corporation, USA

COINTEGRATION AND PURCHASING POWER PARITY IN SELECTED NEW MEMBERS OF EU

Darja Boršič, PhD
University of Maribor, Slovenia
E-mail: darja.borsic@uni-mb.si

Abstract:

The paper deals with empirical testing of purchasing power parity (PPP) theory. Using Johansen cointegration test, the most relaxed PPP version is investigated for Slovenia, Czech Republic, Hungary and Poland in comparison to Austria, Germany, France and Italy in the period of January 1992 to December 2001. The results show no evidence in favour of PPP theory.

Key words : Johansen cointegration test, stationarity, prices, exchange rates

INTRODUCTION

The theory of purchasing power parity (PPP) is a rather simple theory, to which a lot of research has been devoted in the last couple of decades[1]. It was first invented in 16^{th} century in Spanish School of Salamanca. During its history it was invented and forgotten for a couple of times. This was mainly due to authors not knowing previous research at that time. Thanks to Gustav Cassel, a Swedish economist in the beginning of 20^{th} century, the theory became very well known and has been a topic of many discussions during the 20^{th} century (Boršič 2004).

The purchasing power parity theory suggests that exchange rate system should provide a mechanism, which would enable a basket of goods being purchased in both analysed countries to cost the same amount of money when recalculated in one currency.

This paper analyses the validity of purchasing power parity in Slovenia, Czech Republic and Hungary in comparison with selected members of European Union: Austria, Germany, France and Italy, which are also main EU trading partners of the Central European countries in question. The observed period ranges from January 1992 (1993 for Czech Republic) to December 2000. That is from the beginning of transition till the end of the individual European currencies and the introduction of Euro.

The general model of testing for purchasing power parity (Cheung and Lai 1993) is the following:

$$e_t = \alpha_0 + \alpha_1 P_t - \alpha_2 P_t^* + \xi_t \quad (1),$$

where e_t stands for nominal exchange rates, presented as the price of foreign currency in the units of domestic currency, P are domestic prices and P* are foreign prices. All the variables are in the logarithmic form. In the most restrictive form, there are the following restrictions: $\alpha_0 = 0$, $\alpha_1 = \alpha_2 = 1$. The symmetry restriction applies that α_1 and α_2 are equal, while the

[1] Review articles in this field are: Officer (1976), Froot and Rogoff (1995), Rogoff (1996), Sarno and Taylor (2002) and Taylor and Taylor (2004).

limitation of α_1 and α_2 being equal to one is called the proportionality restriction (Froot in Rogoff 1995).

The first empirical analysis started off with the most restrictive version of the model ($\alpha_1 = \alpha_2 = 1$), that is testing the real exchange rates. In the context of the relative PPP the movements in exchange rates are expected to compensate for price level shifts. Thus, real exchange rates should be constant over a long run and their time series should be stationary. Example of such empirical analysis contain Boršič (2003, 2005), Holmes (2001), Parikh and Wakerly (2000) and May (1999).

Relaxing the proportionality condition in equation (1) allows us to test if nominal exchange rates and relative prices are cointegrated. PPP holds if the presence of long-run equilibrium relation is confirmed. The test used in this step is usually the Engle-Granger test of cointegration. Taylor (1988), Kim (1990), Mark (1990), Pufnik (2002) and Boršič (2003, 2005) are some of the examples of this approach.

When all restrictions in equation (1) are omitted, it becomes the least restrictive version of PPP. The only restriction that remains is the signs of the coefficients. This implies that we are looking for any linear relation ship among the observed variables. Taking into account the unstable characteristics of non-stationary time series, the existence of stationary relation ship among them are more important than deviations of coefficients from the strict PPP theory (Liu 1992). If a cointegration among nominal exchange rates, domestic consumer prices and foreign consumer prices is found and it is presented by cointegrating vector of $(1, \alpha_1, -\alpha_2)$, the validity of PPP theory is proven (Chen 1995).

JOHANSEN COINTEGRATION TEST

Since we are looking for a stationary linear combination of three variables, Johansen co-integration test is appropriate to use (Maddala in Kim 1998). This method is base on a VAR and can be briefly described as follows (Johansen 1991):

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + Bx_t + e_t, \tag{2}$$

where t ranges from 1 to T, Y_t is a vector of i variables, which are integrated of the first order. VAR in equation (2) can be also written as:

$$\Delta Y_t = \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta Y_{t-i} + Bx_t + e_t \tag{3}$$

$$\text{where } \Pi = \sum_{i=1}^{p} A_i - I \text{ in } \Gamma_i = -\sum_{j=i+1}^{p} A_j \tag{4}$$

Matrix Π contains information about long run variation of time series. According to Granger representation theorem (Engle in Granger 1987, Johansen 1991) matrix Π can be divided to k x r matrices ρ in α with rank of r ($r \leq k-1$), so that $\Pi = \rho\,\alpha'$, if also Π has reduced rank $r < k$. Matrix α contains r linear cointegrating vectors, while matrix ρ presents adjustment coefficients of error correction model.

Number of cointegrating vectors are tested by two statistics. Trace statistic (LR_{tr}) tests H_0: number of cointegrating vectors is less or equal to r. The trace statistic can be written as:

$$LR_{tr}(r \mid k) = -T \sum_{i=r+1}^{k} \log(1-\lambda_i) \tag{5}$$

where λ_i is i-th maximum eigen value of matrix Π in equation (3). The maximum eigen value statistic (LR_{max}) tests H_0: number of cointegrating vectors is equal to r and H_1: number of cointegrating vectors is equal to r + 1. LR_{max} can be calculated as follows:

$$LR_{max}(r \mid r+1) = -T \log(1-\lambda_{r+1}) = LR_{tr}(r \mid k) - LR_{tr}(r+1 \mid k) \tag{6}$$

Table 1: Results of Johansen cointegration tests

Number of coint. equations	Eigen value	Slovenia	Czech Republic	Hungary	Poland
Austria					
r=0		**61,4312	23,3980	**44,0881	**37,2447
r≤1	LR_{tr}	13,6185	9,0559	12,5636	13,2543
r≤2		*4,7704	0,7835	0,8093	0,1168
r=0		**47,8127	14,3421	**31,5245	*23,9904
r=1	LR_{max}	8,8481	8,2724	11,7543	13,1375
r=2		*4,7704	0,7835	0,8093	0,1168
Germany					
r=0		**42,2441	23,9747	**48,2695	**56,5127
r≤1	LR_{tr}	**20,7690	8,9509	*19,1047	*17,9221
r≤2		2,7287	0,4842	*4,5854	0,8039
r=0		*21,4751	15,0238	**29,1648	**38,5906
r=1	LR_{max}	*18,0402	8,4667	*14,5197	*17,1181
r=2		2,7287	0,4842	*4,5854	0,8039
France					
r=0		25,9711	29,5547	*33,6738	**72,8139
r≤1	LR_{tr}	6,7105	8,2806	4,6155	**30,5526
r≤2		0,7913	0,7824	0,0275	0,7848
r=0		19,2606	*21,2741	**29,0583	**42,2614
r=1	LR_{max}	5,9192	7,4982	4,5880	**29,76774
r=2		0,7913	0,7824	0,0275	0,7848
Italy					
r=0		**37,7476	*34,0048	**58,0352	**49,8470
r≤1	LR_{tr}	12,3176	*15,4926	*18,5722	**21,1667
r≤2		*4,4377	2,2730	2,0778	0,2165
r=0		*25,4300	18,5122	**39,4629	**28,6803
r=1	LR_{max}	7,8798	13,2196	*16,4944	**20,9502
r=2		*4,4377	2,2730	2,0778	0,2165

Note: ** (*) presents 1% (5%) degrees of freedom.

Critical values for Johansen cointegration test are stated in Johansen (1988) and Johansen and Juselius (1990). However, this study applies improven critical values of Osterwald-

Lenum (1992), which are also used by EViews. To undertake the Johansen cointegration test an appropriate lag had to be found. Lags estimated on the basis VAR's Akaike Information Criteria (AIC) and Final Prediction Error (FPE) are presented next to the individual countries names in table 2.

In order to be cointegrated, time series have to be integrated of the same order. That's why, ADF tests were conducted, in order to find the order of integration of individual time series (nominal exchange rates, domestic consumer prices and foreign consumer prices). The tests[2] showed that the time series are mostly I(1). With the exception of Czech exchange rates and consumer prices. Thus, the results of Czech cointegration test are doubtful but are still presented as a matter of comparison. However, MacDonald (1993) claims that even in the event of different order of integration of time series, it is possible for their linear combination to be cointegrated. This is impossible in the case of three time series being integrated of three different orders (Granger 1986).

Table 2. Cointegration and adjustment coefficients

Slovenia	α_1	α_2	ρ_e	ρ_1	ρ_2
$Austria_3$	-0,7075	-1,8258	-0,0424	-1,98E-05	-0,0020
$Germany_6$	-0,7090	2,2350	-0,0228	-0,0006	-0,0050
$France_5$	-0,3307	-4,4198	-0,0703	2,24E-05	0,0025
$Italy_2$	-1,1967	0,9779	-0,0832	-0,0006	-0,0046
Czech Republic					
$Austria_1$	-3,2421	12,3977	-0,0872	0,0133	-0,0030
$Germany_2$	-1,2001	4,8959	-0,0746	0,0406	-0,0033
$France_1$	-2,8237	12,9935	-0,0942	0,0313	-0,0039
$Italy_2$	-1,5897	4,1891	-0,0874	0,0151	-0,0099
Hungary					
$Austria_6$	-3,2617	9,8276	0,0092	0,0253	-0,0035
$Germany_6$	-1,5443	4,8563	0,0316	0,0644	-0,0051
$France_6$	-2,8604	15,3784	-0,0178	0,0380	-0,0012
$Italy_6$	-2,1690	6,9434	-0,1171	0,0363	4,67E-05
Poland					
$Austria_3$	-116,0942	130,8193	0,0056	0,0001	-7,92E-06
$Germany_2$	-3,4839	10,9507	0,0457	0,0130	-0,0009
$France_2$	-6,7725	36,3192	0,0871	0,0137	-0,0007
$Italy_3$	-6,3788	-2,9567	0,0126	0,0027	7,76E-05

Tables 1 and 2 present the results of Johansen cointegration tests. Table 1 shows the estimated test statistics for different null hypothesis and table 2 presents cointegrating and adjustment coefficients. In the case of Slovenia, the results show that there is cointegration among the three time series in comparison to Austria, Germany and Italy, but the signs of cointegrating coefficients are wrong to confirm PPP theory. In comparison to Germany, there is no proof of cointegration eighter. In the case of Czech Republic there is no evidence of cointegration in comparison to Austria and Germany, while for Italy and France there is a conflict in results among the two test statistics. Taking into account the different orders of integration of individual time series, this is not surprising.

Analysing Hungarian time series the results are the following. In comparison to Austria, Italy and France, there is evidence of cointegration but the signs of cointegrating coefficients are again not in favour of PPP validity. In comparison to Germany, there is evidence of three cointegrating equations, which is according to cointegration theory impossible (among three time series there can be at most two cointegrating equations). For Poland the results are similar. In all four cases H_0 (no cointegration) is rejected but the coefficients in cointegrating equations are of the opposite signs as proposed by PPP theory.

The last three columns of table 2 present adjustment coefficients estimated from error correction model, which is also a product of Johansen cointegration test. Adjustment coefficients show the magnitude of correction of deviation from long term equilibrium in the curent time period. In every case it can be seen that deviations are not diminishing, they are rather getting larger. In analysing PPP theory, variations of exchange rate deserve a special attention. Their adjustment coefficients (ρ_e) are presented in the third coloumn of table 2.

CONCLUSION

Relaxing the assumption of proportionality and symmetry in empirical testing of PPP theory allows to test for cointegration of nominal exchange rates, domestic prices and foreign prices. In order to find evidence of PPP theory in Slovenia, Czech Republic, Hungary and Poland, Johansen cointegration tests were conducted. Though some cointegration among the observed variables is proven, the coefficients of cointegrating vectors are not of an appropriate sign to confirm the validity of PPP. The results show that none of the observed currencies and prices exhibit such a long run equilibrium relationship that the purchasing power parity would hold. These results are compatible with most of the rare studies of these countries in the field of purchasing power parity.

Further effort in searching for validity of PPP theory should concentrate on panel data. Namely, panel unit roots test are well-known to be of a higher power that time series unit root test. Thus, it would be interested to see the results of panel unit root tests in the case of real exchange rates. However, further analysis is out of the scope of this paper.

REFERENCES

1.Boršič, Darja. 2003. Empirično preverjanje teorije paritete kupne moči v Sloveniji. *Naše gospodarstvo* 5/6: 418-435.
2.Boršič, Darja. 2004. Cassel in njegovi predhodniki: razvoj teorije paritete kupne moči. *Naše gospodarstvo* 3/4: 68-79.
3.Boršič, Darja. 2005. Cointegration Among Prices and Exchange Rates in New Members of EU. *ERE Conference 2005*. Luxembourg.
4.Chen, Baizhu. 1995. Long-Run Purchasing Power Parity: Evidence from some European Monetary System Countries. *Applied Economics* 27: 377-383.
5.Cheung, Yin-Wong, and Kon S. Lai. 1993. Long-Run Purchasing Power Parity During the Recent Float. *Journal of International Economics* 1/2: 181-192.

[2] Due to limited size of the paper, the results of ADF tests are available upon request.

6 Cheung, Yin-Wong and Kon S. Lai. 2000. On the Purchasing Power Parity Puzzle. *Journal of International Economics* 52(2): 321-330.
7 Christev, Atanas and Abbas Noorbakhsh. 2000. Long-Run Purchsing Power Parity, Prices and Exchange Rates in Transition. The Case of Six Central and East European Countries. *Global Finance Journal* 11: 87-108.
8 Engle, Robert F., in C. W. J. Granger. 1987. Co-integration and Error Correction: Representation, Estimation and Testing. *Econometrica* 2: 251-276.
9 Froot Kenneth A. and Kenneth Rogoff. 1995. Perspectives on PPP and Long run Real Exchange Rates. V: Grossman G. in K. Rogoff. *Handbook of International Economics. Vol. III*. Elsevier Science: 1647-1688.
10. Granger, C.W.J. 1986. Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics* 3: 213-228.
11. Holmes, Mark J. 2001. New Evidence on Real Exchange Rate Stationarity and Purchasing Power Parity in Less developed Countries. *Journal of Macroeconomics* 23(4): 601-614.
12. Johansen, Søren. 1988. Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control* 12: 231-254.
13. Johansen, Søren. 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica* 6: 1551-1580.
14. Johansen, Søren, in Katarina Juselius. 1990. Maximum Likelihood Estimation and Inference on Cointegration – with Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics* 2: 169-210.
15. Kim, Yoonbai. 1990. Purchasing Power Parity in the Long Run: A Cointegration Approach. *Journal of Money, Credit and Banking* 22: 491-503.
16. Liu, Peter C. 1992. Purchasing Power Parity in Latin America: A Co-Integration Analysis. *Weltwirtschaftliches Archiv* 4: 662-679.
17. MacDonald, Ronald. 1993. Long-Run Purchasing Power Parity: is It for real? *The Review of Economics and Statistics*: 690-695.
18. Maddala, G.S., and In-Moo Kim. 1998. *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.
19. Mark, Nelson C. 1990. Real and Nominal Exchange Rates in the Long Run: An Empirical Investigation. *Journal of International Economics* 28: 115-136.
20. May, Roger D. (1999). An econometric Evaluation of Purchasing Power Parity [on line]. East Caroline University. Economics Department. Available: http://www.ecu.edu/econ/ecer/honors/may.pdf [17.02.2001].
21. Officer, Lawrence H. 1976. The Purchasing Power Parity Theory of Exchange Rates: A Review Article. *IMF Staff Paper* 1: 1-60.
22. Osterwald-Lenum, Michael.1992. A Note with Quantiles of Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statististics* 3: 461-472.
23. Parikh, Ashok and Elizabeth Wakerly. 2000. Real Exchange Rates and Unit Root Tests. *Weltwirtschaftliches Archiv* 136(3): 478-490.
24. Pufnik, Andreja. 2002. Purchasing Power Parity As a Long-Run Equilibrium: Co-Integration Test in the Case of Croatia (1991-1996). *Croatian Economic Survey 1996-1999*: 29-54.
25. Rogoff, Keneth. 1996. The Purchasing Power Parity Puzzle. *Journal of Economic Literature* 2: 647-668.
26. Sarno, Luciano and Mark P. Taylor. 2002. Purchasing Power Parity and the Real Exchange Rate. *IMF Staff Papers* 1: 65-105.
27. Taylor, Alan M., in Mark P. Taylor. 2004. The Purchasing Power Parity Debate. *NBER Working Paper* 10607.
28. Taylor, Mark P. 1988. An Empirical Examination of Long Run Purchasing Power Parity using Cointegration Techniques. *Applied Economics* 20: 1369-1381.

SIMULATION OF THE BUSINESS PROCESS SALES_CLAIM

Nadja Damij[1], Janez Grad[2]

[1]Faculty of Economics
University of Ljubljana
Kardeljeva ploscad 17
SI-1000 Ljubljana

[2]Faculty of Administration
University of Ljubljana
Gosarjeva ulica 5
SI-1000 Ljubljana

Abstract

The objective of this work was to run the simulation of the business process Sales_Claim in order to improve it to become a more effective and efficient process. To do that, a business process Sales_Claim was first modeled using the flowchart technique and then the simulation process was run using iGrafx software. The simulation results showed that the business process Sales_Claim is well modeled, effective and at the present time does not need further improvement.

Keywords: Business process, Process simulation, Flowchart diagram

1. Introduction

The majority of business processes were developed over time or badly modeled. Such business processes became ineffective and need to be improved. For this reason, business process modeling and simulation became an important way of ensuring changes in an organization's functioning in order to create a better, more competitive and successful enterprise.

The aim of this work is to show that business process simulation could be used as an essential tool to carry out business process improvements such as better resource utilization, shorter transaction times, etc. The process simulation is done using iGrafx software.

The paper consists of five sections. Sections 2 and 3 discuss the simulation of business processes and introduce the flowchart technique as a tool used by different software packages to run the simulation. Section 4 represents the results of simulation of a business process called Sales_Claim. The final section contains some useful remarks and conclusions.

2. Simulation

Modeling, analysis, improvement and simulation of business processes are on the increase as only a thorough comprehension of the business processes within an organization can lead to effective, efficient and value-adding systems. According to Aguilar-Saven (2003), it is the business processes that are the key element when integrating an enterprise. Furthermore, conceptual modeling of business processes is deployed on a large scale to facilitate the development of software that supports the business processes, and to permit the analysis and re-engineering or improvement of them (Aguilar-Saven, Olhager, 2002). Business processes are modeled with the aim of analyzing their current states within the organization, as well as improving them through the execution of potential "what-if" simulation scenarios.

Currently, few organizations maintain a formal model of their business process network (Weyland, Engiles, 2003). Even so, these formal models were usually the product of a series of interviews, reviews, examinations, etc. conducted at a time when the organization encountered apparent and serious problems. Construction of a business process model makes use of several accepted modeling constructs: delays associated with performing work, statistical distribution of these delays to represent observed variability, dependence of processes on completion of earlier processes, queuing of input entities waiting to be "processed", decision logic that directs entities into alternate flow paths depending on their characteristics, application of resources to the work of a business process, and the costs associated with these resources (Weyland, Engiles, 2003). Consequently, an "as-is" model encompasses the above constructs with the aim of imitating the real business process under inspection within the organization.

Simulation modeling according to Pidd (1998) is based on very simple principles: the analyst builds a model of the system of interest, writes a computer program which embodies the model and uses a computer to initiate the system's behaviour when subject to a variety of operating policies. Furthermore, it is thought that simulation modeling extends the potential of business process modeling and analysis.

According to Weyland and Engiles (2003), modeling and simulation of a business process network serves three immediate purposes: organizing the results of interviews and research, identifying the cause of observed performance issues, and exploring alternate process network configurations that improve performance. The aim is to efficiently image the process network. Weyland and Engiles (2003) argue that hierarchical modeling tools are the most useful, as they contain features like simultaneous understanding of high-level and detailed views, as well as the manageability of the model.

Simulation is the imitation of the operation of a real-world process or system over time (Banks et al., 2001). A simulation model enables the analyst to observe and study the system's behaviour as it advances through time.

Simulation represents a powerful approach for analysis and quantitative evaluation of business processes (Laguna, Marklund, 2005). Furthermore, they classify simulation models into three groups depending on their attributes:

- Static or dynamic. A static model is a model where time within the real business process is insignificant, and a dynamic model incorporates changes over a time period.
- Deterministic or stochastic. A deterministic model is defined by a sequence of events such as input recognition enabling the output definition, whereas a stochastic simulation model has, according to Banks et al. (2001), one or more random variables as inputs.
- Discrete or continuous. A discrete model is a model that consists of discrete events which are events that happen at particular times, and a continuous one consists of variables changing continuously over time.

Consequently, Laguna and Marklund (2005) state that business processes are in general represented as computer-based dynamic, stochastic, and discrete simulation models which are defined as abstractions of the actual business processes, represented in the computer as a network of connected activities and buffers through which jobs or customers flow, and must also capture the resources and various inputs needed to perform the activities. Discrete-event

simulation describes how a system with discrete flow units or jobs evolves over time (Laguna, Marklund, 2005). Therefore, according to Banks et al. (2001) discrete-event simulation examines the modeling of systems in which the state variable changes only at a discrete set of points in time. What differentiates a discrete-event model from a continuous one is the fact that it deals with the attribute time only when the event actually happens.

Consequently, according to Laguna and Marklund (2005), such a perspective of the events and time enables significant time compression because it makes it possible to skip through all time segments between events when the state of the system remains unchanged. Simulation packages enable simulation runs of vast and various numbers of events that may in reality happen over a long period of time. A discrete-event simulation model focuses on the state of the business process at specific time points when the events occur. Hence, when executing the simulation run, the simulation clock jumps between the events and regards the system as staying the same in the meanwhile.

A simulation model is normally based on a set of assumptions regarding the system's operation. These assumptions are expressed in mathematical, logical, and symbolic relationships between the entities, or objects of interest, of the system (Banks et al., 2001). After a simulation model has gained form and been validated, it is deployed to examine various "what-if" questions regarding the real-world system, so that any future alterations of the system are first simulated and as a result it provides forecasts about the impact of the alterations on systems effectiveness.

Laguna and Marklund (2005) summarized some of the main attributes that make simulation powerful:

- Simulation, like analytical modeling, provides a quantitative measure of performance.
- Simulation, unlike analytical and symbolic models, is able to take into consideration any kind of complex system variation and statistical interdependencies.
- Simulation is capable of uncovering inefficiencies that usually go undetected until the system is in operation.

The availability of special-purpose simulation languages, massive computing capabilities at a decreasing cost per operation, and advances in simulation methodologies have made simulation one of the most widely used and accepted tools in operations research and system analysis (Banks et al., 2001).

3. Flowchart

A flowchart is a simple diagram used by different software packages such as iGrafx to model and run the simulation of a business process under discussion. iGrafx software was used in this work to run the simulation of the business process "Sales_Claim".

A flowchart is defined as a formalized graphical representation of a program logic sequence, work or manufacturing process, organization chart, or similar formalized structure (Lakin et al., 1996). A flowchart is commonly used to show the flow of a process from its start to its end. It usually consists of different symbols connected by lines, arranged in such a way to lead us in correct sequence order through a series of steps. Process flow is traced by following the connecting lines between the symbols drawn. These symbols include start and end, activity, input and output, decision, and department.

A flowchart, according to Arlow and Neustadt (2002) begins with a starting point and finishes with an ending point. The terminus symbol is commonly used in flowcharting to designate the beginning and the end.

An activity is represented by a rectangle and means an elementary task or a subprocess. The path by which processes flow through the diagram consists of connecting lines between activities. A set of activities could be contained by a container called a department. An input is indicated by an arrow, which enters an activity. An output is shown by an arrow, which leaves an activity. An arrow connects one activity to another, showing the movement of the diagram.

If an activity is a decision, it specifies alternative paths based on some Boolean expression and is shown by a diamond. There can be only one input path to a decision, but there can be many output paths (Arlow, Neustadt, 2002). A decision is a point at which the process flow can take one of several possible paths based on a defined criterion.

To model a task performed simultaneously by different departments or to model parallel activities, we define different outputs from an activity as split outputs. A split is made by defining multiple paths from a single activity to a set of activities. After parallel tasks have been performed, outputs of those activities which performed the parallel tasks could be modeled to enter a single activity; this is called a joint input.

According to Aguilar-Saven (2003) flowcharts are built to offer an enhanced comprehension of the process, which is a requirement for process improvement. By grouping tasks into logical areas of activity (processes) and drawing flowcharts of the events which occur, it is possible to get a concise picture of the way particular processes are completed within the organization[1]. The flexibility of the flowchart technique is argued by some authors to be its advantage as it allows each modeler to unite various pieces of the process together to obtain the overall picture as he/she feels they fit best. On the other hand, other authors argue that the technique is too flexible, describing large models without illustrating the hierarchy of different layers.

4. Results

Business process Sales_Claim represents a difficult problem in a large trading company. Many customers are not satisfied with the solution obtained considering their claim applications. The purpose of this paper was to model, simulate and improve the business process of Sales_Claim. For this reason, in this section we are simulating the process considered to find if there are possibilities for its improvement.

Figure 1 shows the flowchart of a reduced business process Sales_Claim. As it is evident from Figure 1, the process is performed in two departments; these are Sales and Warehouse. In the framework of the Sales department 15 activities are performed, while 7 activities are executed in the framework of the Warehouse department. The real Sales_Claim business process consists of more than 200 activities.

[1] http://www.hci.com.au/hcisite2/toolkit/flowchar.htm

We ran the simulation of the process Sales_Claim shown in Figure 1 taking into consideration 600 claims, 80 claims of them were already in different phases of the process, and we postulated that 30 claims were received by the organization every day.

To do that, a standard calendar was used, that is, 8 hours/day, 5 days/week and 22 days/month. And the following resources were defined: 6 Sales_Claim_Clerk, 4 Sales_Clerk, Stock_Keeper and Warehouse_Claim_Clerk.

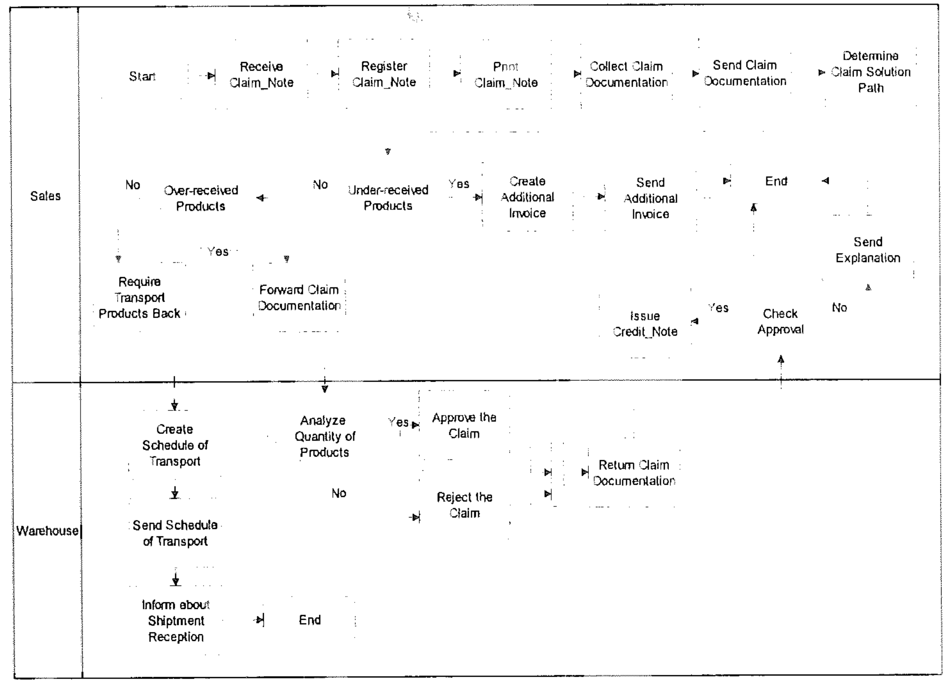


Figure 1. Flowchart of Sales_Claim business process

The results of running the simulation of the business process Sales_Claim were as follows:
- Average cycle time for one claim is 3.74 days;
- Elapsed time for considering 600 claims is 22.73 days.

For this reason, it seems that the process was improved earlier and therefore there is no reason or possibility to carry out further improvement of the process.

5. Conclusions

The objective of this work was to run the simulation of the business process Sales_Claim in order to improve it to become a more effective and efficient process. The results of running the simulation of the reduced business process Sales_Claim show that the process considered is well modeled and effective. Nevertheless, these results also confirm the necessity of using the simulation as a tool for improving business processes.

References

1. Aguilar-Saven R. Business Process Modeling - Review and Framework. International Journal of Production Economics. 2003;90(2):129-49.
2. Aguilar-Saven R, Olhager J. Integration of Product, Process and Functional Orientations - Principles and a Case Study. Preprints of the International Conference on Advanced Production Management Systems, APMS 2002, IFIP, The Netherlands. 2002.
3. Arlow, J., Neustadt, I. UML and the Unified Process - Practical Object-Oriented Analysis & Design. Pearson Education, Boston. 2002.
4. Banks J, Carson II JS, Nelson BL, Nicol DM. Discrete-Event System Simulation. 3rd Edition. Prentice Hall International Series in Industrial and Systems Engineering. Prentice Hall, Inc., Upper Saddle River, New Jersey. 2001.
5. Laguna M, Marklund J. Business Process Modeling, Simulation, and Design. Pearson Education, Inc. New Jersey. 2005.
6. Lakin R, et al. BPR Enabling Software for the Financial Service Industry. Management Services, ISSN: 0307-6768. 1996.
7. Miller AR. Medical Diagnostic Decision Support Systems-Past, Present, and Future. J Am Med Inform Assoc. 1994;1(1):8-25.
8. Pidd M. Computer Simulation in Management Science. Chichester: John Wiley. 1998;118-32.
9. Weyland JH, Engiles M. Towards Simulation-Based Business Process Management. Proceedings of the 2003 Winter Simulation Conference, (eds) Chick S, Sanchez PJ, Ferrin D, Morrice DJ., 2003;225-27.

Boom and bust phenomena in electricity market*

Andreas Ehrenmann[1], Marek Kočan[2] and Yves Smeers[3]

[1]CORE, Université catholique de Louvain, Louvain La Neuve, Belgium

ehrenmann@core.ucl.ac.be

[2]Institute of Mathematics, P.J.Šafárik University, Košice, Slovakia

kocan@science.upjs.sk

[3]CORE, Université catholique de Louvain, Louvain La Neuve, Belgium

smeers@core.ucl.ac.be

Abstract: This paper studies a possibility of boom and bust cycles in electricity market. A single-node model is presented, where the boom and bust cycles emerge. The model is simulated on the real data from electricity market represented by Germany, France, Netherlands and Belgium.

Keywords: Electricity market, deregulation, boom and bust cycles, Cournot game.

1 Introduction

The deregulation of electricity markets should help to decrease the prices and increase the competition. But as the crisis in California in 2000 and 2001 convinced, along with these benefits, instability in prices can arise. This problem was discussed in Ford [8], where the author studies boom and bust cycles in electricity market.

In general, the boom and bust cycles are characterized by periods of underinvestment, rising prices and profitability (the boom), followed by overinvestment and falling prices (the bust) [1]. One can observe the boom and bust cycles for example in markets with commodities (even if the products of those markets can be stored in inventory as a buffer between production and consumption) and in real estate markets [9]. Because of the special properties of electricity (impossibility to store significant amounts of it) the booms and busts in electricity market can have even more severe impact on electricity producing industry than on any other. This was the case of the crisis in California that started two years after the deregulation of the electricity market in 1998 and continued until the spring of 2001 [8]. During this time unprecedented outages appeared. On the other hand, prices started to increase sharply along with marginal costs (the price of gas for CCGT turbines increased), but since many end customers had no incentive to reduce demand because of contractual protection from price increases, several electricity producers went bankrupt.

The research was carried out during a placement in Electrabel, Belgium within the project SK/04/A/F/PL-401 286 of the Leonadro da Vinci program.

*This is a preliminary version of paper appearing in Studies in Nonlinear Dynamics and Econometrics, please do not quote. For quoting please write to kocan@science.upjs.sk.

According to *Electricity directive 2003/54/EC* [4], electricity markets shall become open markets not later than July 2007 in all EU countries. Countries in western Europe, Scandinavia and also Great Britain have already implemented this directive and electricity markets in these countries are deregulated and connected with each other. Also the new EU countries like Slovakia, Czech republic etc. are on their way to deregulation (the process of privatization of transmission and production capacities has already started) so a natural question arises: "Shall Europe and its electricity markets expect similar problems like those in California at the end of the 20-th century?"

There are several articles dealing with behavior of electricity producers. These articles present an idea of what quantities the producers shall offer to the market when they know the demand [10], or what prices the producers shall offer to the market and consequently what quantities to maximize their profits [13]. But the models presented in that articles are static and they do not focus on price evolution and the factors that affect the price stability.

In this article we present a simple dynamic model, where the boom and bust cycles in electricity market can be studied. The electricity market will be represented by one node, where the consumption and production will be concentrated. We assume two producers of electricity. To increase the capacities, the producers have to make some investment decisions in each period (e.g. year) based on their expectations of future prices. The model consists of two simultaneous processes. In the first one we calculate the price of 1 *unit* of electricity for each period and in the second one the investments of each player will be predicted. To model the investment behavior, the Cournot game [12] will be used. Then real data from the market represented by France, Germany, Netherlands and Belgium are applied and the price behavior is studied.

2 Modeling price expectations

Let us assume a single-node electricity market. In this market two electricity producers (players, $i = 1, 2$) are operating. These players are myopic (i.e. they are not able to predict the future accurately). We suppose that the market evolves over periods (e.g. months, years,...), a generic period denoted by t. Let t_0 be the initial period. Suppose, that the power plants used to generate electricity do not retire and to build a new power plant takes L periods. Let l_i represent the new capacity obtained by investing 1 *monetary unit*. Denoting by $C_i(t)$ the capacity of player i in period t, $I_i(t)$ the new investment of player i in period t, we can express the capacity of player i in period $t+L$ by:

$$C_i(t+L) = C_i(t+L-1) + l_i I_i(t) \tag{1}$$

This implies that the total capacity of electricity production in period $t+L$ is:

$$C(t+L) = C_1(t+L) + C_2(t+L) \tag{2}$$

We further suppose that the demanded quantity of electricity increases linearly over periods. Let r denote the rate of this increase (% per period) and $D(t)$ the quantity demanded in period t, we can therefore write

$$D(t+L) = (1+r)^L D(t) \tag{3}$$

We assume that in each period the players produce their full capacity. Using this full capacity and quantity demanded, the particular price will be determined for that period. We further suppose a linear demand function with a constant slope $k > 0$, where the increase of quantity demanded corresponds to the shift of the demand function. The demand function for period $t+1$ has the form:

$$p(t+1) = a(t+1) - kQ_D(t+1) \tag{4}$$

where $Q_D(t+1)$ is the variable representing quantity demanded in period $t+1$ and $a(t+1)$ is a constant.

In period t, the players observe the total capacity $C(t)$ and the price $p(t)$ for which the electricity is sold in that period and demanded quantity $D(t)$. If the price in the following period $t+1$ stayed at the same value as in the previous period, i.e. $p(t+1) = p(t)$, then the increase in demand would set the demanded quantity in period $t+1$ to $(1+r)D(t)$. Having the point $[p(t), (1+r)D(t)]$ on the demand function (4) we get:

$$a(t+1) = p(t) + k(1+r)D(t)$$

After substituting this expression into (4), the demand function in period $t+1$ becomes:

$$p(t+1) = p(t) + k(1+r)D(t) - kQ_D(t+1). \tag{5}$$

On the other hand, the players produce their full capacity in each period, so the quantity supplied in period $t+1$ equals the total capacity, i.e. $Q_S(t+1) = C(t+1)$. The intersection of function (5) with $Q_S(t+1) = C(t+1)$ will result in the price for which electricity in period $t+1$ will be sold:

$$p(t+1) = p(t) + k(1+r)D(t) - kC(t+1) \tag{6}$$

Using a similar reasoning for L periods ahead, we derive that players expect the price in period $t+L$ to be:

$$p(t+L) = p(t) + k((1+r)^L D(t) - C(t+L)) \tag{7}$$

In this equation, expression $(1+r)^L D(t) - C(t+L)$ has a direct impact on price change. When the expression is positive (undercapacity), the price increases and when the expression is negative (overcapacity) the price decreases.

Equation (7) creates a mistake in predicting prices. We can see that when $L > 1$ the players do not take into consideration possible changes in price between periods t and $t+L$. This is caused by their myopic behavior. That mistake will adversely influence the value of investments of both players.

3 The Cournot game

The second step of the model studies the investment behavior of players and it will be represented by a Cournot game [12].

When playing a Cournot game, the players try to maximize their profits by choosing quantities to supply. In our case the quantities that will be chosen are the investments in each period. Since it takes L periods to build a power plant, to find the optimal investments in period t, the players maximize their profits in period $t+L$. Denoting by

m the marginal costs for producing 1 *unit* of electricity (we suppose that marginal costs also include the discounted past investment expenditures), the profit function of player $i = 1, 2$ for period $t + L$ is following:

$$\Pi_i(I_1(t), I_2(t)) = p(t+L)C_i(t+L) - mC_i(t+L) \quad (8)$$

Substituting (1) and (7) we get from (8) for $i = 1, 2$:

$$\Pi_i(I_1(t), I_2(t)) = \Big[p(t) + k\big((1+r)^L D(t) - C(t+L-1) - l_1 I_1(t) - \\ - l_2 I_2(t)\big)\Big]\Big[C_i(t+L-1) + l_i I_i(t)\Big] - m\Big[C_i(t+L-1) + l_i I_i(t)\Big]$$

denoting $A = p(t) + k\big((1+r)^L D(t) - C(t+L-1)\big)$ we have that for $i = 1, 2$:

$$\Pi_i(I_1(t), I_2(t)) = AC_i(t+L-1) - kl_1 I_1(t) C_i(t+L-1) - kl_2 I_2(t) C_i(t+L-1) + \\ + Al_i I_i(t) - kl_1 l_i I_1(t) I_i(t) - kl_2 l_i I_2(t) I_i(t) - mC_i(t+L-1) - ml_i I_i(t)$$

To calculate the investments, both players solve the following optimization problem:

$$\max_{I_i(t)\geq 0} \Pi_i(I_1(t), I_2(t)) \qquad i = 1, 2 \quad (9)$$

Being in period t both players know $p(t)$, $D(t)$ and $C(s)$ for $s = t_0, \ldots, t+L-1$. So the only unknowns in the profit functions Π_i are the investments themselves. This implies that the best response functions are:

$$\frac{\partial \Pi_1(I_1(t), I_2(t))}{\partial I_1(t)} = -kl_1 C_1(t+L-1) + Al_1 - 2kl_1^2 I_1(t) - kl_1 l_2 I_2(t) - ml_1 = 0 \quad (10)$$

$$\frac{\partial \Pi_2(I_1(t), I_2(t))}{\partial I_2(t)} = -kl_2 C_2(t+L-1) + Al_2 - 2kl_2^2 I_2(t) - kl_1 l_2 I_1(t) - ml_2 = 0. \quad (11)$$

Using (10) and (11) we get that if $I_1(t) > 0$, $I_2(t) > 0$ the formulas for optimal investments for period t are

$$I_1(t) = \frac{p(t) + k(1+r)^L D(t) - 3kC_1(t+L-1)}{3kl_1}$$

$$I_2(t) = \frac{p(t) + k(1+r)^L D(t) - 3kC_2(t+L-1)}{3kl_2}.$$

If $I_1(t) = 0$ and $I_2(t) > 0$, only the second player influences the profit of both players and then the formulas are:

$$I_1(t) = 0$$

$$I_2(t) = \frac{p(t) + k(1+r)^L D(t) - 2kC_2(t+L-1) - kC_1(t+L-1)}{2kl_2}.$$

Because of symmetry, the formulas for $I_1(t) > 0$, $I_2(t) = 0$ are

$$I_1(t) = \frac{p(t) + k(1+r)^L D(t) - 2kC_1(t+L-1) - kC_2(t+L-1)}{2kl_1}$$

$$I_2(t) = 0.$$

Using the formulas derived, the players decide in each period how much to invest. According to their investment decisions, actual price, demand and capacity available, the price for next period is set using equation (6). When repeating the process of finding optimal investments and prices over periods, we get the price evolution scenario and we can search for the triggers of boom and bust cycles.

4 Simulation

We performed simulations using real data from electricity market represented by France and Belgium (the first player), and Netherlands and Germany (second player).

The capacity of the first player in initial period t_0 is the joined capacity available in France and Belgium at the end of 2004 and its value is $C_1(t_0) = 84725.18$ MWh, similarly, the capacity of second player is $C_2(t_0) = 95151.87$ MWh [3].

The initial price taken is the average week-day price from the beginning of 2005, announced by *European Enery Exchange* [5] and it is $p(t_0) = 37.1$ *Euro/MWh*. Quantity demanded is taken from [5] as a possible demand scenario, for electricity market presented, for 2005. The value is $D(t_0) = 168915$ MWh.

We assume that both players operate and build the CCGT turbines [9]. The expectations for 2005 are that $m = 36.4$ *Euro/MWh* when running CCGT turbines. The investment costs for building 1MWh of new capacity are 51900 *Euros* and they are fixed over the run of the simulation. So $l_1 = l_2 = 1/51900$ [3].

In general, the growth rate of the demand for electricity is highly correlated with the growth rate of GDP. So for r we take the average of forecast of GDP of France, Begium, Germany and Netherlands for 2005 [7], $r = 2\%$.

The long-term price elasticity of the electricity market in western Europe is approximately 0,1 [11]. Using this elasticity and the data presented before we set $k = 0.002$.

Having all the values necessary we can start a simulation by calculating the optimal investments of both players using formulas derived in previous section and by calculating the prices for each period using equation (6).

For the simulation a spreadsheet (Microsoft Excel) was used. Results of running the simulation of the model over 100 periods for $L = 1$, $L = 2$, $L = 3$ are presented on Figures 1 to 3. On the horizontal axis are the periods, on the primary vertical axis we show price (in Euros/MWh) and on the secondary vertical axis the values of total investments (in millions of Euros). The light curve represents the price evolution, the dark one represents the joined new investments of both players.

All three Figures show the same behavior in the first four periods. Price decreases and players do not invest. This is caused by the initial overcapacity. Then a sharp increase of price in all three Figures appears until it is optimal for players to invest. The first investments come in the 10-th period but with the difference that after that period for $L = 1$ the players invest in all succeeding periods and the price increases in a steady way until the end of the simulation. With $L = 2$ and $L = 3$ the evolution is different. After the 10-th period in both cases the players are investing over next five periods. But with $L = 2$ they invest one and a half more and with $L = 3$ almost twice as much as with $L = 1$. This overinvestment forces the price in next periods to decrease in both cases, but for $L = 2$ it decreases only for one period, and then the price increases in a steady way until the end of simulation. With $L = 3$, the decrease of the price is sharper, but after the price starts to increase, this increase is not permanent. The periods of decreasing prices and overinvestment alternate with increase of prices over the run of the simulation.

5 Conclusion

As we have seen in the simulation, when the players decide about the investments, they have to make some price expectations. The more they have to predict into future (the

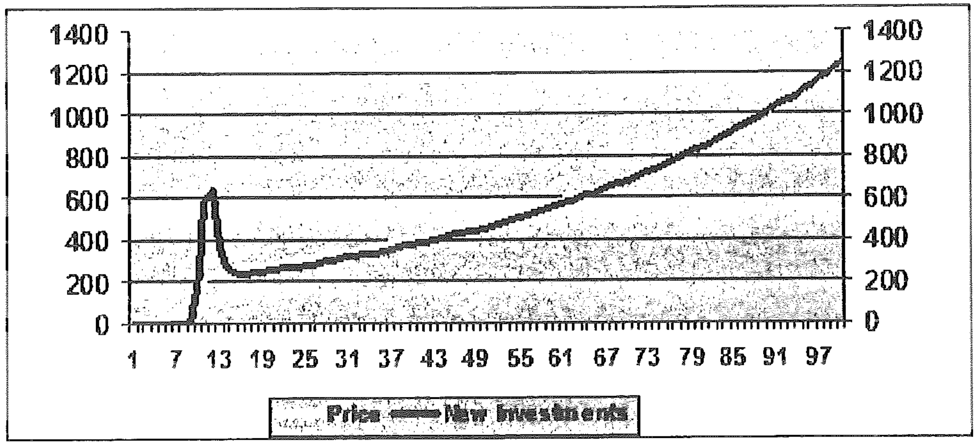


Figure 1: Price evolution and the joined new investments for $L = 1$

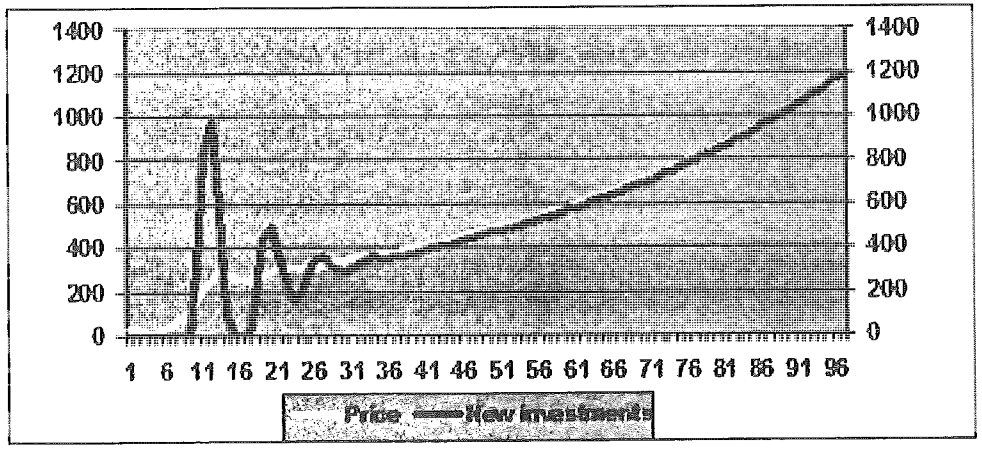


Figure 2: Price evolution and the joined new investments for $L = 2$

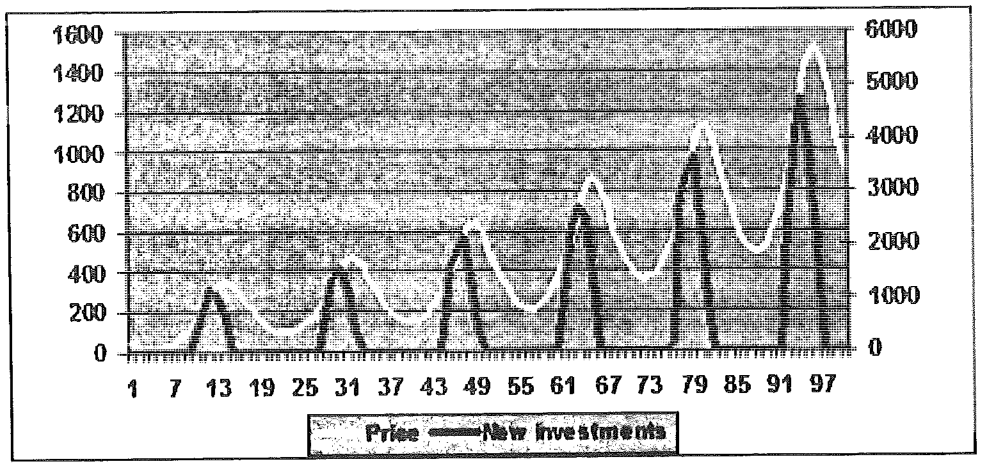


Figure 3: Price evolution and the joined new investments for $L = 3$

bigger is L in the models) the bigger mistake they make in predicting the prices and so the bigger is the mistake when calculating the investments. When we have applied the real data we have seen that this mistake had a huge impact on price stability along with appearance of boom and bust cycles.

The model presented is a single-node model where two electricity producers are concentrated. As a topic for further research it would be interesting to see the evolution of the prices in models with a greater number of players, or a multi-node network.

References

[1] Boston Consulting Group report: Keeping the lights on, *May 2003.*

[2] A. Ehrenmann K. Neuhoff: A comparison of electricity market designs in networks, *Cambridge Working Papers in Economics CWPE 0341, August 2003.*

[3] Electrabel database: Electrabel S.A., Belgium, *2005.*

[4] Electricity directive 2003/54/EC: *http://europa.eu.int/comm/energy/electricity/legislation/legislation_en.htm*

[5] Electricity markets: *http://www.electricitymarkets.info/*

[6] European Energy Exchange: *http://www.eex.de/*

[7] Eurostat: *http://epp.eurostat.cec.eu.int/portal/page?_pageid=1090,30070682,1090_33076576&_dad=portal&_schema=PORTAL*

[8] A. Ford: Simulation Scenarios for the Western Electricity Market, *A Discussion Paper for the California Energy Commission Workshop on Alternative Market Structures for California, November 2001.*

[9] A. Ford: Waiting for the Boom: A simulation study of power plant construction in California, *March 2001.*

[10] D. Moitre: Nash equilibria in competitive electric energy markets, *Electric Power Systems Research 60, 2000.*

[11] F. Murphy, Y. Smeers: Forward Markets May not Decrease Market Power when Capacities are Endogenous, *CORE DISCUSSION PAPER 2005/28, Louvain-la-Neuve University, Belgium, 2005.*

[12] Hal A. Varian: Intermediate Microeconomics, 3-rd Edition, *W.W.Norton & Company, USA, 1993.*

[13] H. Xu, E. J. Anderson: Nash equilibria in electricity markets with discrete prices, *Mathematical methods of operational research, 2004.*

EVALUATION OF PARAMETERS WHICH INFLUENCE THE ASSOCIATION OF PRIVATE FOREST OWNERS BY THE USE OF AHP AND DEXi

Špela Malovrh
University of Ljubljana, Biotechnical Faculty,
1000 Ljubljana, Slovenia
spela.malovrh@bf.uni-lj.si

Abstract
In this article the two multi-criteria decision making methods AHP and DEXi are compared with regard to establishing which factors were more important for private forest owners in making their decision whether to become members of the private forest owners association of the Mirna Valley. Both methods rely on hierarchical decomposition of criteria. The comparison was carried out on a real-life example of private forest owners becoming associated.

Keywords: association of private forest owners, multi-criteria decision making, Analytic Hierarchy Process (AHP), DEXi

1. INTRODUCTION

With social development, the demands of people towards nature have changed. For quite a long while, the forest has not only been a producer of wood, but has had to assure a variety of non-material functions which are becoming more and more important for society [2].

Small and fragmented forest property is one of the main reasons for the low efficiency and non-competitive position of private forest management. Small and fragmented forest property distributed among a large number of forest owners and co-owners is a typical characteristic of Slovenian forestry, creating limitations for private forest owners. This is why many people have decided to establish associations of private forest owners. But before a forest owner becomes a member of an association he/she always wishes to know which factors influence his/her decision to join the association.

People have to make decisions in every day life over and over again. Some of these decisions are easy and cause no damage if the decision is wrong; others can be complicated and exert a strong influence if the decision is wrong.

When one must pick out of several alternatives (decisions) and when each alternative consists of several criteria (attributes) we speak about multi-criteria decision making. One of the problems of multi-criteria decision making is how to define the importance of the alternatives and attributes [5]. Undoubtedly, one alternative is more important than another for private forest owners, but it is difficult to ascertain how much important one alternative is than another.

The goal of this paper is to establish which factors can influence a private forest owner to become a member of the private forest owners association of the Mirna Valley. With the help of the decision making methods AHP and DEXi we attempted to find out which alternatives and attributes were more important for the private forest owner and had a decisive influence on his joining the association. The forest owner's property size was taken into account.

2. MULTI-CRITERIA DECISION MAKING

Multi-criteria decision making is based on the fact that the choice of a solution is affected by numerous criteria, the importance of which varies. A decision making problem is thus broken down into smaller subordinate problems (parameters, criteria, attributes), and these are then assessed separately for each parameter. Final assessment is obtained by means of a specific combining procedure.

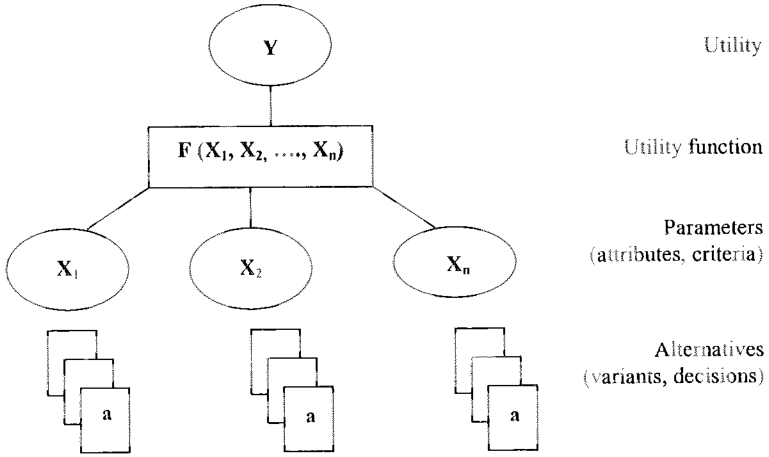


Figure 1: The multi-criteria decision making model

As evident from Figure 1, the model is formed by means of parameters (attributes), X_i. These are variables which represent subordinate problems of a decision making process (attributes which define the quality of alternatives). Utility function F is a rule according to which the values of individual parameters are combined to form variable Y, which represents a final assessment of usefulness of alternatives [3]. Alternatives are described according to basic parameters by values a_i. On the basis of these values, the utility function will provide a final assessment of each alternative.

In the case of multi-criteria decision making, two notions are encountered:

- preferential relation S; $x_1 S x_2$; i.e. x_1 is preferred to x_2
- utility function - $F(x_i)$, which determines the degree of suitability (preference, priority) of parameter x_i.

In actual cases we know what we prefer, but we are unable to assign a certain value to a solution (the utility function is not known). What is needed, then, is a procedure which converts preference relation into utility function. One of the procedures which make this possible is the AHP method [1], [6].

3. BRIEF DESCRIPTION OF DECISION SUPPORT METHODS

3.1 AHP METHOD

The analytic hierarchy process (AHP) is a well known method for a numerical evaluation and analysis of options [1]. The AHP method was developed in the early seventies by Thomas L. Saaty. The method is based on step by step mutual comparison of two parameters at the same level. For comparison, a scale of 1 to 9 is used, where 1 means equal importance, so two activities contribute equally to the objective, and where 9 means extreme importance of one activity over another. The comparisons between elements are then entered into a pair-wise comparison matrix, matrix $A = a_{ij}$ (i = 1...m, j = 1...m), if there are m parameters.

A is a square matrix with ones placed diagonally. Symmetrical values are inverse: $a_{ij} = \frac{1}{a_{ji}}$.

In the process of assessment, the problem of consistency of assessments is encountered: $a_{ik} \cdot a_{kj} = a_{ij}$. If assessment values are inconsistent, they should be corrected, or the results will be of no use [1],[4].

$$A = [A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \vdots & & & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n} \end{bmatrix} \quad (1)$$

Utility vector w is obtained by searching for eigenvalues λ of matrix A: $Aw = \lambda_{max} w$, where λ_{max} is the maximum eigenvalue of matrix A and w is the corresponding eigenvector. In practical cases, this tends to be a complex calculation procedure. The eigenvector which belongs to the maximum eigenvalues of positive reciprocal matrices can be obtained in various ways:

- method of powers – the matrix is put to a sufficiently high power, the values are then added and normalized by lines
- normalization – the matrix is normalized so that the sum according to columns is 1, and the average element in the line is calculated [6].

The eigenvalues which correspond to the eigenvector are obtained as:

$$\lambda_{max} = \frac{1}{m} \sum_{i=1} \frac{(Aw)_i}{wi} \quad (2)$$

The measure of inconsistency is defined by the difference (λ_{max} - m). It is expressed by the consistency index CI: $CI = (\lambda_{max} - m) / (m - 1)$. (3)

A random index is then introduced, which is given in tabular form [6]

m	1	2	3	4	5	6	7	8	9	10
RI	0	0	0,58	0,90	1,12	1,24	1,32	1,41	1,45	1,51

and inconsistency (CR quotient) is calculated as: CR = CI / RI. If CR < 0.1, the matrix is sufficiently consistent. In the opposite case, the matrix should be corrected; otherwise the results will not be correct.

For the purpose of this paper we used AHP as implemented in the program Expert Choice.

3.2 DEXi

DEXi is a multi-criteria decision method which is based on the construction of a decision problem in a hierarchic structure of attributes. In contrast with AHP, DEXi uses qualitative attributes. Each attribute in hierarchy is defined as a discrete variable which can take its values from a set of symbols. These symbols need to be defined for each attribute separately and typically consist of words like not important, very important ... In DEXi, the use of qualitative attribute requires a different approach to aggregation than AHP. As opposed to using weights, the aggregation is carried out by decision rules. These are simple "if-then" rules defined by the designer, with which the utility function is expressed point-by-point for all possible combinations of alternative values [5].

4. PRACTICAL CASE

The comparison and evaluation of AHP and DEXi was carried out on a real-life case of private forest owners joining the association of private forest owners of the Mirna Valley. The problem was complex and simple enough to enable a sound comparison of the both methodologies.

Because of the problems that forest owners meet, they decided to found the association of private forest owners of the Mirna Valley, aiming to improve the competitiveness of the private forestry sector. But every new member had to decide which alternatives were important for his choice to join to the association. A poll was carried out on the pattern of 40 members [2]. With the information obtained from the interviewed members of the association, mutual comparisons were assessed and a pair-wise comparison matrix was formed for each level of the decision tree.

We distributed private forest owners into four groups – corresponding to criteria (attributes) with regard to the size of estate: private forest owners with 1-5 ha of forest, private forest owners with 5-10 ha of forest, private forest owners with 10-25 ha of forest and private forest owners with more then 25 ha of forest, because we expected the size of the property would influence the reasons for joining the owners' association. Each of these criteria was then supported by alternatives - reasons why private forest owners became members of the association: education, counselling, cooperation in elaboration of forest management plans, conclusion of timber sale contracts, joint purchase of machinery, equipment and literature (Figure 2).

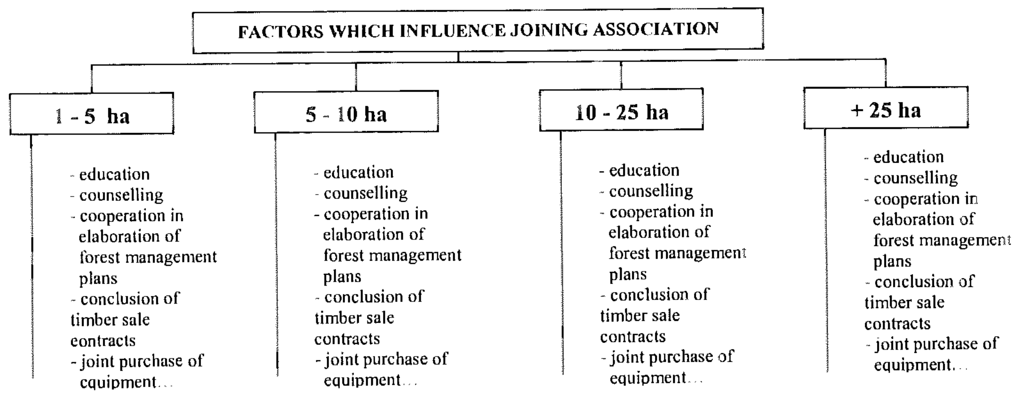
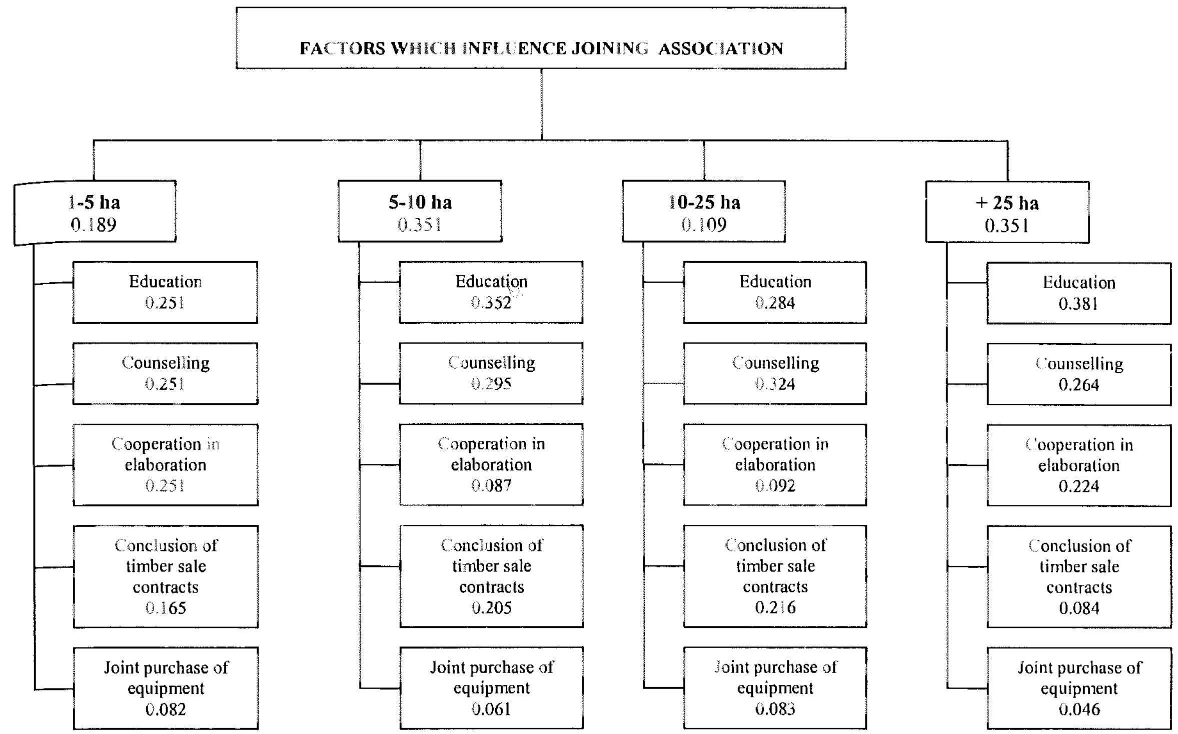


Figure 2: Decision tree for private forest owners in the Mirna Valley

The analysis which we made with the AHP method in the program Expert Choice showed that for the private forest owners the most rewarding alternative regarding the criteria was education; counselling took second place, followed by cooperation in elaboration of forest management plans, then came conclusion of timber sale contracts, while joint purchase of machinery, equipment and literature were stated as least important for the owner's decision to join the association (Figure 3).

When we compared the private forest owners by property size we discovered that there were only minor differences between them. The most important factor why owners became members of the association was education, since private forest owners wish to improve their knowledge on safe work in the forest and on silviculture, and become acquainted with new technologies. The second factor that was important for all of the owners was receiving counselling about the timber market and being informing about the best contractors.

The difference between owners with regard to the size of property was obvious with respect to the following two factors: cooperation in elaboration of forest management plans and conclusion of timber sale contracts. For small (1-5 ha) and for big (+25ha) forest owners cooperation in the elaboration of forest management plans was more important than for the other owners.



	1-5 ha	5-10 ha	10-25 ha	+ 25 ha		Attributes		Result
Education	0.251	0.352	0.284	0.381		0.189		**0.336**
Counselling	0.251	0.295	0.324	0.264		0.351		0.279
Cooperation in elaboration	0.251	0.087	0.092	0.224	·	0.109	=	0.167
Conclusion of timber sale contracts	0.165	0.205	0.216	0.084		0.351		0.156
Joint purchase of equipment	0.082	0.061	0.083	0.046				0.062

Figure 3: Results of evaluation with AHP

Concluding timber sale contracts was more important to all the small proprietors that to the biggest ones (+25 ha). Joint purchase of machinery, equipment and literature was least important for all proprietors without regard of size.

The analysis that we made in DEXi shows that the most important alternatives for private forest owners are education and counselling, less important is conclusion of timber sale contracts and not important for the owners are cooperation in elaboration of forest management plans and joint purchase of machinery, equipment and literature (Figure 4).

	Education	**Counselling**	**Cooperation in elaboration**	**Conclusion of timber sale contracts**	**Joint purchase of equipment**
Factors	very important	very important	not important	less important	not important
1- 5 ha	less important	very important	not important	less important	not important
5 - 10 ha	very important	very important	not important	not important	not important
10 - 25 ha	less important	very important	not important	not important	not important
+ 25 ha	very important	less important	not important	very important	not important

Figure 4: Results of evaluation with DEXi

If we compare private forest owners with regard to property size we see that education and counselling are important for all owners. Conclusion of timber sale contracts is only important for owners with property sized 1-5 ha and for those with the largest property (+ 25 ha). For small proprietors the conclusion of timber sale contracts is important because they carry out tree felling in longer time intervals and thus have only scant information about the timber market. But this factor is also important for big proprietors (+ 25 ha) because they sell timber during the whole year and need to get a fair price during this whole period. Cooperation in elaboration of forest management plans and joint purchase of machinery, equipment and literature are not important for private forest owners.

5. CONCLUSION

Both AHP and DEXi use hierarchical decomposition for developing decision models. The structure of attributes is identical, but the methodologies use different techniques for describing the options and aggregating the values of input attributes into the final result. The AHP uses comparison matrices, while DEXi is based on decision rules that aggregate qualitative values. When comparing the results of evaluation, DEXi gives very comprehensible symbolic results, while AHP produces a numeric evaluation. So the DEXi method evaluates several factors with the same values, while AHP determines a unique value for every project. The results of evaluation with both methodologies were almost the same, excluding cooperation in elaboration of forest management plans. AHP evaluated it to be important (in a third place of five), slightly better than conclusion of timber sale contracts and joint purchase of machinery, equipment and literature. DEXi evaluated it to be not important.

DEXi with its qualitative values is suitable for modelling problems that are by their nature qualitative and would be difficult to describe with numbers. In our case all the attributes were qualitative, so for this kind of problem DEXi is more suitable than the AHP method [5].

Interesting future work could be implementing a combination of DEXi and AHP, where DEXi would classify an option into a certain class and AHP would be used for detailed evaluation within a certain class. The most important advantage of AHP is the ability to distinguish between similar options, for which creating a comparison matrix is easier than for complex problems. Such two-step models usually give better results than models based on a single methodology [3].

6. REFERENCES

[1] LIPUŠČEK, I.: Assessment of ecologically most favourable manufacture of wooden chairs by means of AHP method. Zbornik posvetovanja, SOR 2001, SDI, Ljubljana, pp.307-312

[2] MALOVRH, Š.: Prispevek društva lastnikov gozdov na razvoj podeželja (študija primera: Društvo lastnikov gozdov mirenske doline), seminarska naloga, Ljubljana, 2005

[3] MANRAI, A.K.: Mathematical models of brand choice behaviour, European Journal of operational research 82, Elsevier Science B.V., 1-17, 1995

[4] SAATY, T.L.: Fundamentals of decision making and priority theory. RWS Publication, Pittsburgh, 1994

[5] ŠPENDL, R., RAJKOVIČ, V., BOHANEC, M.: Primerjava kvalitativnih in kvantitativnih odločitvenih metod: DEX in AHP pri ocenjevanju projektov. Zbornik posvetovanja, Organizacija in management, XV. posvetovanje organizatorjev dela 1996, Univerza v Mariboru, Fakulteta za organizacijske vede, Kranj, pp.190-199

[6] WINSTON, W.L: Operations research: Applications and algorithms. Duxbury Press, Bellmont, CA, 1994

Simulation Analysis during Law Adoption

Bojan Peček, M.Sc.
University of Ljubljana, Faculty of Administration
Gosarjeva 5, 1000 Ljubljana, Slovenia
bojan.pecek@fu.uni-lj.si

Abstract

Simulations are commonly used in technical systems. This paper demonstrates their applicability to systems where the human factor is more important. It discusses an example of a procedure in public administration where the effects of reorganization are simulated. The intention of the reorganization is positive – but easing the load on clients means that the procedure takes longer. With simulations it is possible to monitor the procedure and obtain data which are otherwise impossible to obtain.

Keywords: simulations, discrete event simulations, iGrafx

1 INTRODUCTION

Public administration is almost entirely a service activity, for which it is characteristic that the quality of the product (service) cannot be measured. It is probably the most expensive human activity for which we have no, even approximate, data on the costs of the services.

Up to the present, simulations have been used mainly in technical systems. Now they are becoming increasingly applied to processes which are unpredictable. "Many services are stochastic complex processes operating in resource-constrained environments. Most of them are discrete processes" [4]. In this paper we demonstrate the utility of the method of discrete simulation in public administration procedures. The hypothese which we wish to prove through our research is the following: *Simulations can be used during the adoption of a new law.*

In the research we analyzed the procedure of granting social assistance benefits. With the help of the discrete simulation method we compared the course of the existing manner of performing the procedure with the procedure prescribed by the new Law on General Administrative Procedure.

2 METHODOLOGY AND SOFTWARE

Software packages for the simulation of processes are as old as the first computers [8]. Today we can find several types of software for performing discrete simulations: software which operates in one of the standard computer languages (Fortran, Pascal, C, etc.), special languages for simulation, block-structured systems such as GPSS, and visual interactive modelling systems (VIMS) [5].

One of the interactive systems for modelling and simulation is the iGrafx Process 2000 program from the Micrografx Company. Since the program does not have a strict drawing methodology like some other programs (cf. [3], or [7]), one of the existing techniques must be applied. We used the technique of the swimlane diagram from the methodology of the Unified Modelling Language (UML), which some people refer to as the "de facto industrial standard object-oriented language for modelling" [2], [10]. The symbols which we use are shown in Table 1. Solutions for more complex situations are represented in Table 2. It is the notation as suggested by Scott [9], and the methodology shown by Lunn [6].

Table 1 Swimlane diagram symbols

Symbol	Meaning
Org. unit1 / Org. unit2	The organizational unit is the performer of the activity. It can be any kind of unit, such as a department, client, ministry, secretary, department head, etc.
activity	Activity is an individual step which must be taken within the process
alternative	Alternative has two possible meanings: a node where a decision is made about the continuation of the procedure, or a condition for the beginning of the performance of an activity
Start	Start/end – each diagram has at least one start and one end

Table2 Denoting parallels and alternatives in swimlane diagrams

Situation	UML notation following Lunn
Alternative starts: Activity is performed when transaction arrives along any branch	
Joining of partial transactions: Activity is performed only when both flows arrive	
Alternative continuation: Procedure continues along one of the alternatives	
Splitting: Procedure continues along both branches simultaneously	

3 THE EXISTING PRACTICE OF THE PROCEDURE

The procedure begins at a Center for Social Work when a client files an application along with the necessary certificates and documents which they have previously obtained from the competent institutions – the Tax Office, Employment Service and Registry Office, or already has in his or her possession. For the rest of the simulation it is important that the clients themselves obtain the documentation. In the case of picking them up in person they are issued on the spot and registered in the logbook. The issuing of certificates does not require the opening of new files or the starting of new cases, and therefore no additional time is required.

After the filing of the completed application at the Center for Social Work, the case is entered into the records. When the matter reaches the clerk who will process it, it waits its turn to be processed. From the point of view of work organization this phase is the most controversial, since it cannot be excused.

Occasionally (estimated at one case in twenty), the client is called in for a discussion. When this is completed, a positive or negative decision is made. In the case of an unjustified (i.e. unreasonable) procedure, the applicant is notified that the procedure has been terminated. In the case that the procedure is reasonable but the applicant is not entitled to receive social assistance benefits, a decision is drafted. This is usually written more precisely than a decision in favour of the applicant, since there is a greater chance of an appeal, and the legal findings must be accurately defined. The percentage of rejections and unfavourable decisions is surprisingly small, around 15%. The model of the procedure is represented in Figure 1.

The results of the simulation for an individual transaction show:

- It takes 357 days to process 2000 applications. This is essentially in keeping with expectations, as we received data for one year of operations.

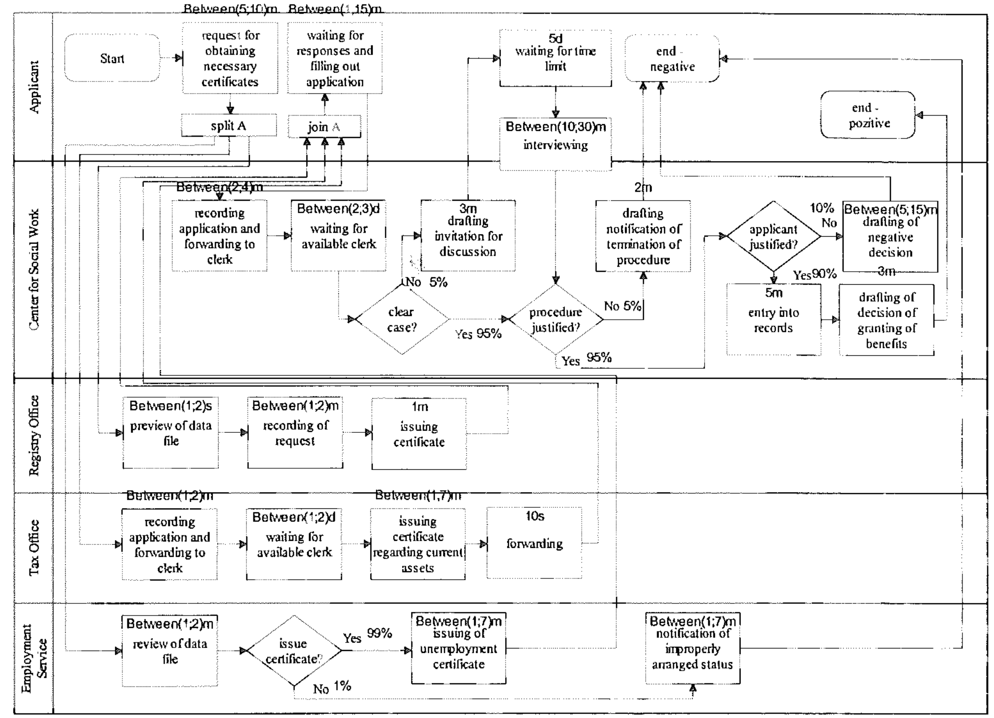


Figure 1: *Existing practice in performing the procedure of granting social assistance benefits*

- The average cycle per transaction is 7.14 days. This is the time from the beginning of collecting the documents to receiving the final decision. The time for waiting for decisions to become binding as prescribed by law was not considered.
- The average work time (recording, processing, deciding, etc.) was 40.21 min. That is 0.41% (less than one percent!) of the total average time the transaction was in the system. The rest of the time (99.59%) was spent waiting, in delivery, etc.
- For calculating the efficiency of the process, the number of approved applications is unimportant. But we should still remember that 1693 applications were approved and 307 rejected. This data is important because the process will be compared with other models. Through the number of accepted and rejected applications we will convince ourselves that the new model is of equal value to the previous one.

The formal correctness of the model was confirmed, since the simulation ended normally. For verification we repeated the questioning of the operative performers of the procedure. The average transaction cycle of around 7 days seemed much too long to them. When we set in the activity 'clear case?' the probability of continuing along the 'Yes' branch at 100%, the average cycle was reduced to 6.5 days. Another change was made to a sub-procedure at the Tax Office, as if they issued their confirmations on the spot when clients go there in person. This reduced the average transaction cycle to 3.44 days. Taking into consideration the fact that clerks are not aware of the time clients lose collecting their documents, we quickly come to an acceptable approximation of the response we got from the operators to the question: "How much time does it take for a client to receive a processed application?" "It's fast: in two to three days, four at most." In short, we can be quite satisfied with the estimate of the transaction cycle for an individual application.

4 COURSE OF THE PROCEDURE PURSUANT TO THE NEW LAW ON GENERAL ADMINISTRATIVE PROCEDURE

The Law on General Administrative Procedure (LGAP) has brought major changes to the operations of official bodies. In point three of Article 139 it prescribes that bodies shall have the obligation that "the clerk who is in charge of the procedure shall obtain information ex officio on the facts about which the body competent for decision making keeps official records". Similar is prescribed also in the article 175 of the same Law.

Obviously clients no longer have to obtain the documents themselves, but can file their applications without enclosures. The necessary certificates must be obtained by the clerk in charge of processing the matter. But what is apparently an elegant and user-friendly solution contains a number of hidden traps.

A significant change which the new approach introduces has to be taken into consideration. In the former case the client went to the body in person to obtain the desired document and the body only recorded the issuing of the document. In the case of an official inquiry by the body the mail is recorded and filed pursuant to the rules on office operations. In this case the course of the procedure for obtaining missing documentation is as follows: The clerk in the first body (the addressing body) drafts a request for the second body (the addressed body), where they explain their request. This request is sent to the clerk's supervisor for signing. The signed request is sent to the forwarding office, the matter is entered into a logbook. The addressed body receives the request, and pursuant to the rules on office operations records and classifies and partially marks it. The clerk in charge of the matter at the addressed body puts the request into their stack of things to be done. When the matter comes to the top of the pile, it is processed and sent to the department head for signing. The signed document is then sent to the forwarding office, which sends it to the addressing body. The head office of the first body matches the document with the case. After the time limit from the logbook expires the entire matter with an enclosed opinion is returned to the first clerk. The course of the entire procedure of obtaining information from other organizational units is shown graphically in Figure 2.

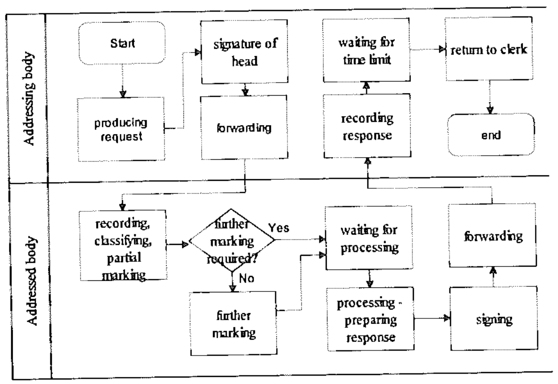


Figure 2: *Course of official request for issuing an opinion between administrative bodies*

In the case of the procedure in question, an application for social assistance benefits, the addressing body is the Center for Social Work. There are three addressed bodies: the Registry Office, the Tax Office and the Employment Service.

The new procedure differs from the old one in that the applicant simply files his application and his part of the procedure is then over. He is no longer required to obtain the necessary documents; this is now the responsibility of the clerk who is in charge of the procedure.

The Center for Social Work still first records classifies and marks the application. Then it is put into the stack of matters which are waiting for an available clerk. The clerk first determines whether the procedure is reasonable. If it is reasonable, he prepares requests for the issuing of opinions. We suppose that he will have a program prepared for preparing the

necessary forms and that it will be sufficient if he enters the information once and the computer prints them automatically. This is a very optimistic scenario, that the centers will be well organized and equipped. Therefore we have predicted a very optimistic time for this operation – a total of 2 minutes with the supervisor's signature and dispatching.

The activity 'drafting requests for reviewing and forwarding office' are followed by a gate which indicates the splitting of the transaction into a new family of partial transactions. With this we simulate that each document travels through the procedure independently of the other two. The gate opens and the transactions are performed each day at the end of the day.

The transactions, i.e. the opinions issued by the Registry Office or the Tax Office are joined before the activity 'justified?'. If the applicant is justified, the joined partial transaction is further joined by the certificate of active participation in the employment program. This is followed by an entry into the data base and the drafting of a positive decision. Procedure is presented in Figure 3. After execution of the simulation the following is concluded:

- First we check whether the modelled process is still comparable to the previous one: as in the previous process 2000 transactions were processed. 308 were rejected, the rest were approved. Therefore we can conclude that the models are equal.
- The average cycle time grew to 10.27 days, which is a 43.8% increase in the duration of the procedure! Interestingly, the effective amount of work per transaction is reduced by over 15% - from 40.24 to 33.81 minutes. This decrease is most of all due to the amount of time the applicant is busy, since in this model he only fills out his application once and no longer has to wait to obtain his documents – from 16.41 to 8.55 minutes. The amount of time the clerk is busy is slightly increased – from 19.40 to 20.43 minutes.

The conclusion of the experiment demonstrates that the time needed for processing a request is significantly increased if Article 139 of the Law on General Administrative Procedure is strictly adhered to.

We can therefore conclude that: if we take into account the recommendations of the Law on General Administrative Procedure, the average time for processing an application is increased from 7 to at least 10 days. We have to be aware that these applications are filed by people who are in fact in serious trouble. For them the reaction time of the body is most important. As the saying goes: "He who gives quickly gives twice".

5 CONCLUSION

The research confirms the initial hypothese that using simulation we can monitor not just technical systems but also difficult to predict systems which are almost entirely based on people and their stochastic behaviour. The calculations showed and proved the exceptional utility of simulation.

At the end of the research we are justified in asking whether there is any other method of obtaining similar data from which we could make calculations on the data which we obtained from the described system of simulation. How could we calculate the 'cost of the product', in this case the cost of processing one application? Definitely not with the existing information system used in public administration. There is simply no data on how many minutes and seconds a person spends on an individual matter or an individual document, in this case an application for the granting of social assistance benefits. We could (over the course of a year) monitor the processing of the applications using dozens of stopwatches. Even if this could be done, the clerks would perform their work completely differently if they knew that they were being watched. From this we can conclude that simulation is more than just experimenting with data – it represents the augmentation of an information system. Therefore the sentence written by Bosilj Vukšić and Kovačič, "the use of simulation modelling will have to become standard business practice" [1], should be expanded to "the use of simulation will have to become a component part of information systems".

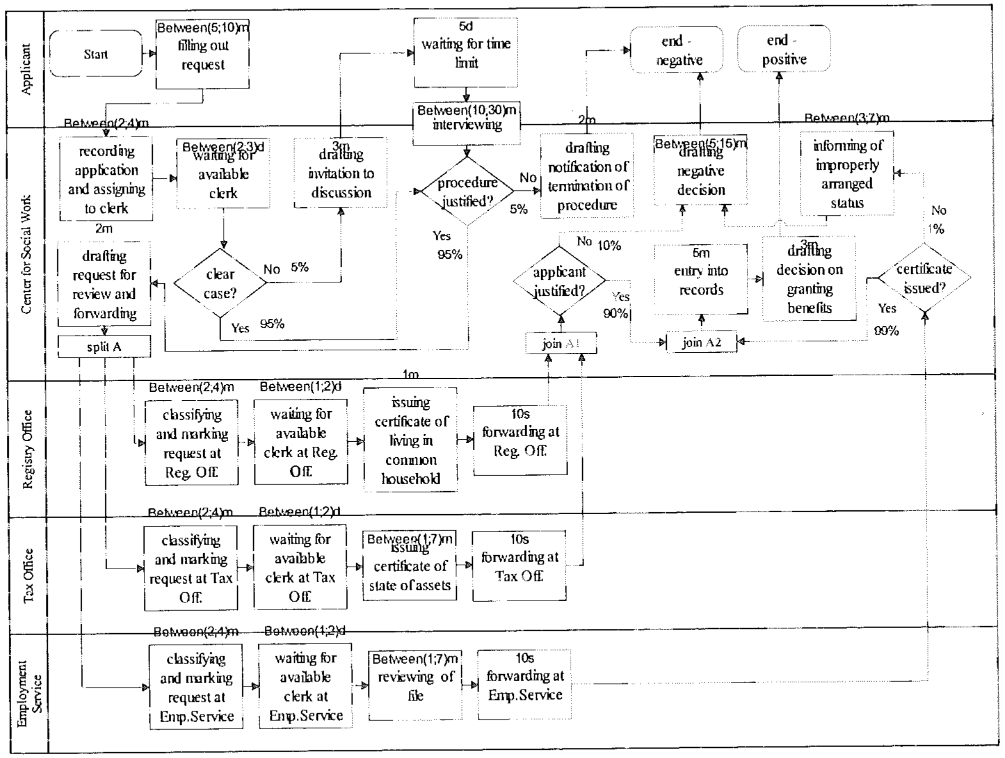


Figure.3. *Diagram of the procedure of obtaining social assistance benefits pursuant to Article 139 of the LGAP*

The simulations have shown that otherwise good intentions – the reduction of the burden on people requesting necessary documents – can turn into an even more time-consuming procedure if new laws are not followed by appropriate and effective changes to information and communications equipment.

REFERENCES

[1] Bosilj Vukšić, V., Kovačič, A. (2004): Upravljanje poslovnim procesima, Sinergija, Zagreb
[2] Engels, G., Heckel, R., Sauer, S. (2000): UML – A Universal Modeling Language?, ICATPN 2000, (Ed. Nielsen M., Simpson D), Springer Verlag, Berlin
[3] Kelton, D. W., Sadowski, R. P., Sadowski, D. A. (1998): Simulation with Arena, McGraw-Hill
[4] Laughery, R., Plott, B., Scott-Nash, S. (1998): Simulation of service systems, Handbook of Simulations (ed. Banks, J.), John Wiley & Sons
[5] Law, M. A., Kelton, W. D. (2000): Simulation modelling and analysis, McGraw-Hill
[6] Lunn, K. (2003): Software development with UML, Macmillan
[7] Pegden, D. C., Shannon, R. E., Sadowski, R. P. (1995): Introduction to simulation using Siman, McGraw Hill
[8] Pidd, M. (1998): Computer simulation in management science, John Wiley & Sons
[9] Scott, K. (2001): UML explained, Addison-Wesley
[10] Siau, K., Halpin, T. (2001): Unified Modeling Language: System analysis, design and development issues, Idea Group Publishing

DISTRIBUTION CHANNELS FOR ORGANIC PRODUCED AGRICULTURAL PRODUCTS

Petra Ulamec, ARSO, Vojkova 1b, 1000 Ljubljana
Lidija Zadnik Stirn, Biotehniška fakulteta, Jamnikarjeva 101, 1000 Ljubljana
Katja Vadnal, Biotehniška fakulteta, Jamnikarjeva 101, 1000 Ljubljana
petra.ulamec@gov.si, lidija.zadnik@bf.uni-lj.si, katja.vadnal@bf.uni-lj.si

Abstract

The organic farmers, who intend to appear on the market, have to decide what strategy to choose to enter the organic market in the optimal way. On basis of benchmarking and qualitative analysis, surveys and linear programming we have generated the models for determining the optimal strategy of distribution channels for Slovenian organic products. The goal of determined strategies is to reduce the recognizable, physical and economic distances between the producers and consumers of the organic products. With linear programming we have established that organic markets, which are at the emerging stage, do not exploit all given possibilities.

Keywords: organic farming, organic products, marketing, distribution channels, benchmarking, linear programming

Introduction

Formation of new distribution channels and selection of the existent ones belong to the more important strategic decisions of every organic farm that produces the products with intention of marketing the products (Ulamec, 2005). Slovenian organic farms market their products through one or two distribution channels. They use the distribution channels that were typical for the start of organic farming in Western Europe. The prevailing distribution channels in European countries today are of lesser importance for the Slovenian farmers; this may threaten their competitive position on the organic market (Bratuša, 2000, Vadnal and Bratuša, 2000). When programming distribution channels the producers have to decide between what is ideal and possible and what is applicable (Kotler, 1998). When choosing distribution channels for organic products produced in Slovenia the models for optimising strategies can drastically reduce the recognisable, physical and economic distance between the producers and real and potential consumers of those products (Zadnik Stirn, 1995 and Ulamec, 2005).

Material and methods

With benchmarking we have analysed and compared the stage and development of the European organic market with the situation in Slovenia. According to Watson (1993) we have with the purpose to identify the best solutions on the whole European organic market determined the following spheres (Ulamec, 2005):

- impact of agricultural political measure on the development of the European organic markets
- development of European organic markets
- evaluation of the importance of individual distribution channels for organic products
- impact of the media
- constraints in the development of supply with organic products
- and constraints in the development of demand for organic products.

In our research we paid a special attention to organic farms from Gorenjska region. The goal of research was to find out whether the used methods and ways of distributing the organic products from farms of Gorenjska region are different from the methods used when distributing conventional products. On the basis of analysis we established what part of production the organic farmers from Gorenjska region use for:

- direct sale on their own farms
- sales through wholesalers
- their own use.

The questionnaire was distributed to 50 % of organic farmers from Gorenjska region in July 2002 (Ulamec, 2005). We made the qualitative analysis of sales of organic products with focus group in October 2004. We considered the principle of homogeneity of the group. As the goal of the survey was known to the group members beforehand, we used the method of the direct interview (Vadnal, 2002). The goal of implementing the quantitative analysis was to find out that the marketing strategy of organic products based on price difference is supported by strategy of distribution channels (Ulamec, 2005).

With linear programming we have defined (Ulamec, 2005):

- market share that can be achieved on differently developed organic markets by organic agricultural products whereby evaluation of importance of different distribution channels is to be considered,
- market shares, that can be achieved on differently developed markets by organic products whereby evaluation of importance of different distribution channels is to be considered, as well as share of organic farms and share of land area under organic management,
- maximal income, that can be reached by organic producers by using different distribution channels on differently developed organic markets and considering the interest of consumers,
- market share by using different distribution channels on differently developed organic markets whereby considering achieved market share on individual distribution channel and the interest of the consumers.

With linear programming we also have considered the difference in development of organic markets. We used the data of evaluation of the importance of individual distribution channels (Padel et al., 2003). We defined the quoted data of achieved market shares on differently developed organic markets as the average of the countries taken into consideration by their share in the total food sales in the year 2002 (International Trade Centre, 2003). We summarized the data on income share achieved by producers using different distribution channels, by the results of survey distributed in Slovenia in the year 2000. The survey included all Slovenian organic farms (Bratuša, 2000). Survey results gave us data on the interests of consumers (Kvas, 1999). We obtained the data about the shares of organic farms and share of land area under organic management by doing sums of average value of countries taken into consideration, which consist of individual groups with varied developed markets (Willer and Yussefi, 2004). We also applied our own calculations with separate linear programs on achieved market shares on particular distribution channels, on differently developed organic markets, whereby considering the share of organic farms, share of land area under organic management and evaluation of different distribution channels (Ulamec, 2005). For a solution of problems with more parameters where the restrictions are passed in form of linear non-equations we used the method of benchmarking (Winston, 1994 and Zadnik Stirn, 2001).

Results

We only present here the calculation of market share for different distribution channels in urban regions; other calculations are in (Ulamec, 2005). Table 1 indicates the data of the maximal market share of organic products in total food trade, which can be achieved by different distribution channels on differently developed urban organic markets.
Here we took into account the value of different distribution channels applying the current situation. The only restriction is mark 5, which represents the highest possible value of importance of separate distribution channel.

Table 1: Importance of different distribution paths for reaching maximal market share on differently developed urban organic markets

Distribution channels /Stage of organic market development	Established market	Growing market	Emerging market	Constraints
Multiple retailers	3,8	3,9	2,9	5
Direct marketing	1,9	2	1,9	5
Specialist organic shops	2,4	2,4	2,7	5
Other shops	1,7	1,6	2,1	5
Catering/public services	0,6	2	0,3	5
Market share	2,5	1,1	0,6	max

Mathematical formulation of the problem in the form of LP:

$$x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0, \qquad (1)$$

$$\begin{array}{lllll} 2{,}5\,x_1 + & 1{,}1x_2 + & 0{,}6\,x_3 & & \max \\ 3{,}8\,x_1 + & 3{,}9\,x_2 + & 2{,}9\,x_3 & \leq & 5 \\ 1{,}9\,x_1 + & 2\,x_2 + & 1{,}9\,x_3 & \leq & 5 \\ 2{,}4\,x_1 + & 2{,}4\,x_2 + & 2{,}7\,x_3 & \leq & 5 \\ 1{,}7\,x_1 + & 1{,}6\,x_2 + & 2{,}1\,x_3 & \leq & 5 \\ 0{,}6\,x_1 + & 2\,x_2 + & 0{,}3\,x_3 & \leq & 5 \end{array}$$

The solution of LP (1):

Maximum value of objective function = 3,29

variable	value
X 1	1,31
X 2	,00
X 3	,00
slack	value
S 1	,00
S 2	2,50
S 3	1,84
S 4	2,76
S 5	4,21
constraint	shadow price
C 1	,65
C 2	,00
C 3	,00
C 4	,00
C 5	,00

We introduced the unknowns x_1, x_2, x_3. The x_1 represents the maximal increase in market share of organic products on developed organic market, the x_2 represents the maximal growth of market share on organic market, which is in stage of progress and growth, and the

x_3 represents maximal increase in market share on organic markets, which are based on starting development level.

The solution of problem (1) tells that market share of organic products produced through different distribution channels on urban organic markets is 3,29 % of total food trade in European countries. The result given by linear program (1) also shows that it is possible to reach the maximal market share on developed organic market and this is in accordance with expectations. On the developed organic market we could reach the maximal market share by growth of 1,31 (value x_1). The increase of the developed organic market depends on development of organic markets that are currently in stage of progress and growth. In optimal solution marketing the condition is fulfilled with multiple retailers. At direct marketing we consider only half of the value of importance of distribution channel. Optimal solution of linear programme increases the value of importance on this distribution channel in comparison with the value given by experts. To reach the maximal market share within specialized organic shops the value considered is 3,16 of that particular distribution channel. Available is still 1,84 of the value. At other shops there is 2,76 available value, at catering the value is 4,21 of importance of distribution channel.
The dual programme indicates how the maximum value of objective function could change if we change constrains – distribution channels: $u_1 = 0,65$, u_2 , u_3 , u_4, $u_5= 0$. If the marketing importance through multiple retailers increased for one more unit, then the market share would increase for the additional 0,65 %. In case of increase of importance of other distribution channels there would not come to increase of market share.

In countries with developed organic market the market share of organic products represents 0,8 % - 2,5 % of the total food market. We believe that the calculated market share, considering predicted growth of 15 – 30 % per annum, will be reached within few years. With linear programming we calculated that the market share that depends on marketing through multiple retailers could reach 3,29 % on total food market in urban regions. The increase of developed organic market is possible to reach with the development of other lesser-developed organic markets. Within the optimal solution the condition for the importance of marketing through multiple retailers is fulfilled. The optimal solution considers only 4,6 % of the included organic farms. There is only 2,6 % of land under organic management spare (Ulamec, 2005). On developed organic markets there are 26 % - 90 % of all organic products marketed by multiple retailers (McDonald, 2004).

With linear programming which we do not present here due to the limited space we have calculated that the market share within direct marketing can reach 4,3% of the total food market. The only condition to reach the market share is that it increases for 1,72. In the optimal solution we get partially unrealised condition of importance for direct marketing.
Included in the solution is 3,11 worth of value. Optimal solution does not include the total value of organic farms shares. Included are 6,04 %, to be used there are still 1,96 % of organic farms (Ulamec, 2005). Direct way of marketing is characteristic for the beginning of development of organic markets. In the developed organic markets' countries the direct sales is of the lesser spread distribution channels, as it has from 7 – 19 % share. In Slovenia there are 85 % organic farms that use this distribution channel (McDonald, 2004).
According to the optimal solution the current market share of organic products within marketing them in specialized organic shops could reach 4,3 %. If the developed organic markets increased for 1,72, then maximal market share would be achieved. Optimal solution of the linear programming implements the condition of the average share of land under organic management in total. In the optimal solution there is partly unaccomplished

condition on importance of marketing within specialised organic shops. In the solution there are 4,14 worth of value, unavailable are still 0,86 % of this distribution channel (Ulamec, 2005). At the revue of marketing data within the specialised organic shops we have found out that in several countries with developed markets this particular distribution channel has developed differently. In some countries with developed organic markets it has not got a major effect. The difference can be even up to 4 5% (McDonald, 2004).
With optimal solution the market share of organic products can reach 9,048 % with regard to the total food market in urban regions, inclusive the use of different distribution channels on differently developed organic markets, as well as considering the created income share on separate distribution channels and buyers' interests. In linear programming we have considered the maximal value of the individual distribution channels, which we calculated. This is also the reason for reaching such a high market share, as all three most expanded distribution channels are considered. The obtained results present us the final market share, which the organic products could reach (Ulamec, 2005).

Wishing to consider the importance of distribution channels from consumers' point of view we have used the models of the goal programming (Zadnik Stirn, 2001). In the programme we included all parameters, which we had used or calculated in the previous linear programming. In the optimal solutions we came very close to the consumers' wishes on two from three distribution channels. The results of goal programming reveal that in urban regions we can consider 48,65 % of consumers' wishes, in rural regions but up to 60,2 %. This is the way to enable fulfilling the expectations of the consumers, who wish to buy the products in specialised shops as well as through the direct marketing. In models we did not wish to limit the achievement of the maximal market shares neither with shares of inclusive organic farms nor with land under organic management. The data demonstrate that both used parameters are higher than the current situation in Slovenia. We have to emphasise that random effects are not considered in calculations, although we are aware that their influence can be very important (Ulamec, 2005).

Conclusion

With benchmarking of European organic markets we have discovered that with forming the measures for stimulation of organic farming in Slovenia we need to pay more attention to organic farms with market production. With benchmarking we confirmed that for the development of Slovenian organic market the insufficient supply of products as well as their production resources are very limited. The surveyed farmers use the same means of marketing, which were used before turning into organic farming. For the further market the surveyed farms offer very small part of their production. Organic farmers will have to stand on the market far more organised if they want to come closer to the potential consumers. The consumers often do not distinguish between organic products from other products, which appear on Slovenian market. In Slovenia the acknowledged selling channel are multiple retailers. The selling channels are expanding within the development of the organic market. However it depends on developing market which selling channel is the most important. Organic markets at their emerging stage do not use all their given possibilities.
We have also discovered that on developed organic markets new challenges are to be looked for, as within the investigated parameters the further growth is not possible. The results directed towards the future and which are based on development of all discussed organic markets prove this statement. Consecutively the differences between them are becoming smaller.

The models which are based on modified units of more linear programmes prove that more developed markets currently grow mainly with extension into the undeveloped regions. To prepare the strategy for the adequate distribution channels we have to consider the potential consumers as well as the current situation on the field of organic farming. The forming of the strategy for the distribution channels is dependent on more parameters. The generated model is based on consideration of optimal parameters of European organic markets with the emphasis on Slovenian consumers. Based on goal programming we have proved how important it is to consider the interest of the consumers. They represent the most important factor in development of organic market and herewith they help in promotion of organic farming, which is directed to marketing production.

References

1. Bratuša A. 2000. Načini trženja ekoloških pridelkov. Diplomska naloga. Ljubljana, Biotehniška fakulteta, Oddelek za agronomijo:129 str.
2. International Trade center. 2003. Overview world markets for organic food & beverages (estimates):1 str. http://www.intracen.org/mds/sectors/organic/overview.pdf (27 junij 2004)
3. Kotler P. 1998. Marketing, management, trženjsko upravljanje, analiza, načrtovanje, izvajanje in kontrola. Ljubljana, Slovenska knjiga: 831 str.
4. Kvas K. 1999. Modna muha ali preudarna izbira? Ona - tedenska priloga Dela, 1,19: 18 - 19
5. McDonald P. 2004. The organic food market in Europe. Queens land government trade & investment office Europe: 46 str.http://www.dpi.qld.gov.au/extra/pdf/organiceurope.pdf (18 maj 2004)
6. Padel S., Seymour, C., Foster C. 2003. Organic marketing initiatives and rural development QLK5-2000-01124 SWP 5.1: Report of all three rounds of the Delphi inquiry on the European market for organic food*: 64 str. (23.maj 2004) http://www.irs.aber.ac.uk/omiard/publications-pdf/Delphi_Final_Report_Oct_2003_D.pdf
7. Ulamec P. 2005. Optimiranje tržnih poti ekološko pridelanih kmetijskih pridelkov. Magistrsko delo. Ljubljana, Biotehniška fakulteta, Oddelek za agronomijo: 121 str.
8. Vadnal K., Bratuša A. 2000. Tržne poti slovenskih ekoloških kmetij. V: Zbornik simpozija (Moravske Toplice 14 in 15 december 2000) Novi izzivi v poljedelstvu 2000. Ljubljana, Slovensko agronomsko društvo: 284- 288
9. Vadnal K. 2002. Trženjski model oživitve etnične jedi na primeru ajdovega močnika. V: Zbornik Biotehniške fakultete Univerze v Ljubljani. Ljubljana, Biotehniška fakulteta, Oddelek za agronomijo: 47-52
10. Watson G.1993. Strategic Benchmarking. New York, John Wiley and Sons Inc.: 269 str.
11. Willer H.,Yussefi M. 2004. The world of organic agriculture statistics and emerging trends. Bonn, Germany, IFOAM -International Federation of Organic Agriculture Movements: 167 str. http://www.ifoam.org (24 junij 2004)
12. Winston W.L., 1994. Operations research: applications and algorithms. Duxbury Press, Belmont, CA: 1318 str.
13. Zadnik Stirn, L. 1995. Razvoj tržnih potencialov slovenskih kmetijskih pridelkov (1.osnutek kvantitativnega modela – poročilo raziskovalne naloge). Ljubljana, Univerza v Ljubljani, Biotehniška fakulteta: 18 str.
14. Zadnik Stirn L. 2001. Metode operacijskih raziskav za poslovno odločanje. Novo mesto, Visoka šola za upravljanje in poslovanje: 182 str.

Section 5
Location Theory and Transport

HIDDEN LOGISTIC POTENTIALS IN THE COMPANY

T. Berlec[1], J. Kušar[1], L. Lenart[2], M. Starbek[1]

[1]University of Ljubljana, Faculty of Mechanical Engineering, Ljubljana, Slovenia

[2]Institut Josef Stefan, Ljubljana, Slovenia

Abstract

The companies operate in a turbulent environment. In these circumstances only those companies are successful which can respond to market changes as fast as possible. The article explains how the tools for modeling the flow of orders through working systems can be used in order to discover hidden logistic potentials of working systems in a company and how these potentials can be exploited. Presented are the results of finding hidden logistic potentials of working systems in the tool-production department company and the measures proposed for exploiting these logistic potentials.

Keywords:

working system, lead time, inventory, performance, logistic potentials

1 INTRODUCTION

Analysis of research results regarding inventory of orders, performance, range and lead time of working systems [1], [2], [3], [4], [5] has shown that the tools presented in Figure 1 should be used to model the flow of orders through the company working systems.

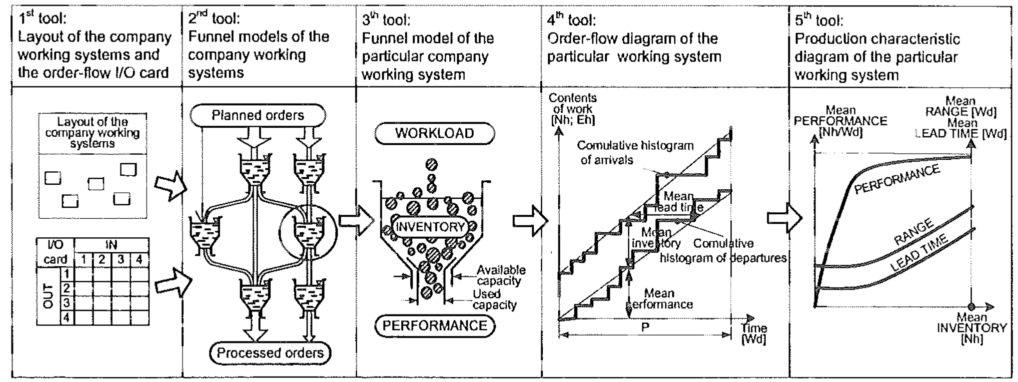


Figure 1: Tools for modelling the flow of orders through working systems

2 FINDING THE LOGISTIC POTENTIALS OF WORKING SYSTEMS IN THE COMPANY

The term "logistic potentials" of working systems in a company refers to hidden logistic potentials of mean inventory (ΔIm), mean range (ΔRm) and mean lead time (ΔTLm) of all working systems in a company.

Research has shown that hidden logistic potentials of working systems in a company can be found in the following ways:

1) On the basis of the measurements of the order-flow through working systems in the company in the interval P, an order-flow diagram is designed for each working system [1], [2], [5], [6] and actual operating point of a working system is defined.
2) Target operating point of a working system is defined [6].

Logistic potentials of the j-th working system are shown in Figure 2.

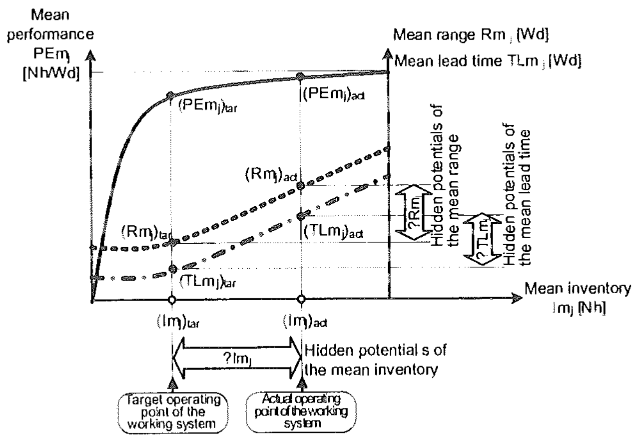


Figure 2: Logistic potentials of the j-th working system

Definition of target operating point of a working system, i.e. the target mean inventory of orders $(Im_j)_{tar}$ target mean performance $(PEm_j)_{tar}$, target mean range $(Rm_j)_{tar}$ and target mean lead time $(TLm_j)_{tar}$ of the working system is based on equations, developed by P. Nyhuis and H. P. Wiendahl [6].

2.1 Procedure for finding and exploiting the hidden logistic potentials of working systems in the company

Procedure for finding the hidden logistic potentials of working systems can be described by the following steps:

Step 1: Order-flow measurements through working systems

Step 2: Design of order-flow diagrams for working systems and definition of actual operating points of working systems

Step 3: Calculation of actual mean range and mean lead time for processing all orders in the working system

Step 4: Calculation of actual mean inventory, mean range and mean lead time for processing all orders in all working systems

Step 5: Designing production-characteristic diagrams of working systems and definition of target operating points of working systems

Step 6: Calculation of target mean range and target mean lead time for processing all orders in working systems

Step 7: Calculation of target mean inventory, mean range and mean lead time for processing all orders in all working systems

Step 8: Calculation of logistic potentials in the working system

Step 9: Calculation of logistic potentials of all working systems

Step 10: Selection of measures for exploiting logistic potentials of working systems

2.2 Computer-aided finding of hidden logistic potentials of working systems in the company

Manual calculation and drawing of flow diagrams and production characteristics and determining the hidden logistic potentials of working systems is rather time consuming, so the "DIAGRAM" computer software was developed for this purpose.

The software was written in Java programming language, it runs on a PC and it can be added to any ERP system (figure 3).

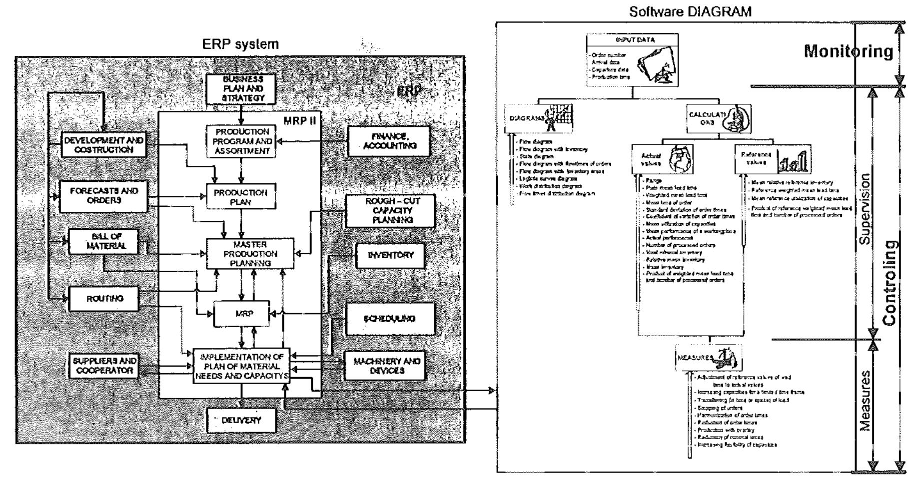


Figure 3: Linking the ERP system with the DIAGRAM software [7]

Using the Diagram software it is possible to draw or calculate for any working system of the company: order-flow diagram of any working system, actual mean inventory performance, utilization, range and lead time of the working system, production characteristic diagram of the working system, target mean inventory, performance and lead time of the working system, logistic potentials of the working system.

For all working systems of the company it is possible to draw or calculate logistic potentials of all working systems in a company.

3 CASE STUDY

Procedure for determining the logistic potentials of working systems was tested in a tool-production department of ETI d.d. company from Izlake, Slovenia.

It produces tools for transforming and cutting, jet machines for thermoplastic materials, jet and press machines for duroplastic materials, and automated assembly appliances.

There are 38 working systems in the tool-production department.

The management decided that hidden logistic potentials of working systems in this department should be found out in the interval from Wd 291 till 350 and that these potentials should be exploited.

Steps for finding and exploiting the hidden logistic potentials of working systems in the tool-production department:

Step 1: Measurement of the flow of orders through working systems in the tool-production department

In each working system the company employees recorded:

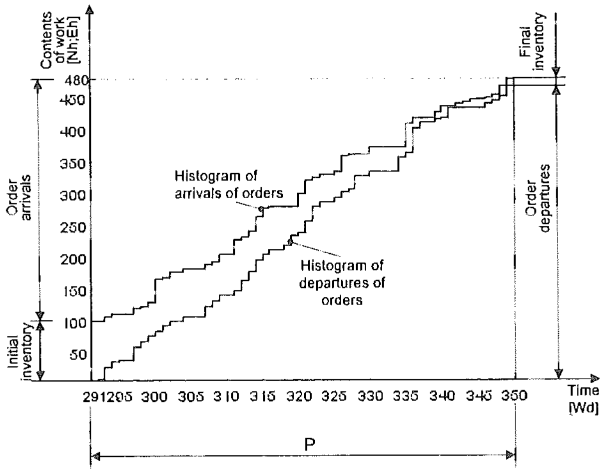
- order numbers and dates of order arrival at and departure from the working systems,
- contents of work of orders which arrived at the working system or departed from it and
- number of orders which departed from the working systems.

There have been 2011 orders with 3883 working operations processed.

Step 2: Design of order-flow diagram for working systems in the tool-production department and definition of actual operating points of working systems

The Diagram software was used to draw the flow diagram and to calculate actual mean inventory of orders, relative actual mean inventory of orders, actual mean performance, actual mean range and actual mean lead time for each working system of the tool-production department.

Figure 4 presents results of designing the order-flow diagram for the "Milling machine ALG 200" and calculation of the actual operating point.



Actual operating point for the »Milling machine ALG 200«
1. $(Im_j)_{act}$ = 48.45 Nh
2. $(Im,r_j)_{act}$ = 604.42 %
3. $(PEm_j)_{act}$ = 8 Nh/Wd
4. $(Rm_j)_{act}$ = 6.06 Wd
5. $(TLm_j)_{act}$ = 6.06 Wd

Figure 4: Results of designing the order-flow diagram for the "Milling machine ALG 200" and calculation of the actual operating point

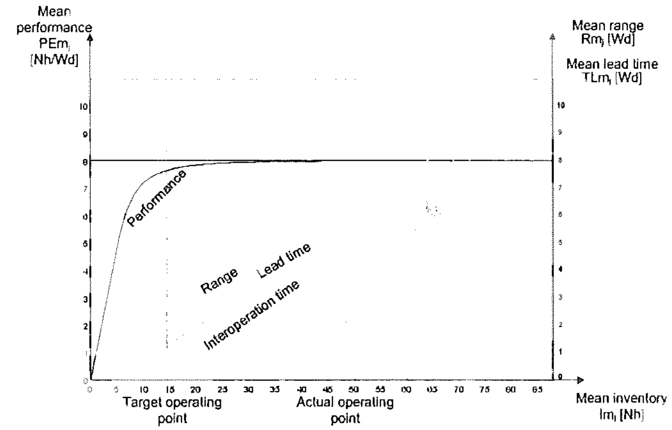
Steps 3 and 4: Calculation of actual mean range and mean lead time for processing all orders in a particular working system of the tool-production department, and definition of actual mean inventory, mean range and mean lead time for processing all orders in all working systems.

These results are used in step 8 for calculation of hidden logistic potentials in a particular working system in the tool-production department.

Step 5: Definition of target points and drawing production-characteristic diagrams of working systems in the tool-production department

The Diagram software was used to draw the flow diagram and to calculate relative target mean inventory of orders, target mean inventory of orders, target mean range and target mean lead time for each working system of the tool-production department.

Figure 5 presents the results of calculation of the target operating point and drawing the production characteristic diagram for the "Milling machine ALG 200".



Target operating point for the »Milling machine ALG 200«
$(Im,r_j)_{tar}$ =180 %
$(Im_j)_{tar}$ = 14.43 Nh
$(Rm_j)_{tar}$ = 1.90 Wd
$(TPm_j)_{tar}$ = 1.27 Wd

Figure 5: Results of calculation of the target operating point and drawing the production characteristic diagram for the "Milling machine ALG 200"

Steps 6 and 7: Calculation of target mean range and target mean lead time for processing all orders in a particular working system, target mean inventory, target mean range and target mean lead time for processing all orders in all working systems of the tool-production department.

These results are used in step 9 for calculation of hidden logistic potentials of all working systems in the tool-production department.

Step 8: Calculation of hidden logistic potentials in a particular working system in the tool-production department

Hidden logistic potentials of the "Milling machine ALG 200" (working system No. 17) reveal that by transition from actual to target operating points it would be possible to reduce the mean inventory of orders by 70.22 %, to reduce mean range of the working systems by 68.65 %, and to reduce the mean lead time of the working systems by 79.04 %.

Step 9: Calculation of hidden logistic potentials of all working systems in the tool-production department

Hidden logistic potentials of working systems in the tool-production department reveal that by transition from actual to target operating points of working systems it would be possible to reduce the mean inventory of orders by 69.69 %, to reduce mean range of the working systems by 68.55 %, and to reduce the mean lead time of the working systems by 75.36 %.

Step 10: Selection of measures for exploiting the hidden logistic potentials of working systems in the tool-production department

If it was found in a particular working system that the relative actual mean inventory of orders is larger than the target mean inventory, then it is possible to exploit the hidden logistic potentials of the working system either by temporarily increasing capacities of the working system, or by temporarily stopping arrival of orders, or by a time- or space transfer of a part of a working-system load to a system of similar technology.

If it was found in a particular working system that the relative actual mean inventory of orders is less than or equal to the target mean inventory, then it is necessary (by harmonizing or reducing the order times or by reducing interoperation times) to find new logistic potentials of the working system.

4 CONCLUSION

Because of ever fiercer domestic and international competition of companies, and transition from the market of sellers to the market of customers, the companies have to continuously increase their economy which largely depends on inventory of orders, range and lead times of the companies' working systems [8],[9].

The article proposes a procedure for finding and measures for exploiting hidden logistic potentials of working systems in a company; these can be found only by continuously monitoring inventory of orders, range and lead times of working systems. Controlling allows for a continuous monitoring of actual values of inventory, range and lead times of working systems; if these values deviate from the target values, measures can be taken in order to bring the actual values closer to the target ones. Only after logistic potentials of working systems have been exhausted it is necessary to remove the barriers (by additional harmonization of order times, or reduction of interoperation times, or manufacturing by overlapping, or increasing flexibility of working systems) that prevent further reduction of mean inventory, range and lead time of working systems.

Case study of finding and exploiting hidden logistic potentials of working systems in the tool-production department of the ETI d.d. company from Izlake has shown that there are considerable logistic potentials which should be exploited in future.

5 REFERENCES

[1] Wiendahl HP.: Load-Oriented Manufacturing Control. Springer, Verlag, Berlin, 1995, pp. 41-199.

[2] Starbek M., Grum J.: Operation lead time control. Robotics and Computer Integrated Manufacturing 16, 2000, pp. 443-450.

[3] Starbek M., Grum J.: Control of the state of orders on machines. Computers & Industrial Engineering 40, 2001, pp. 35-49.

[4] Kušar J., Brezovar A., Grum J., Starbek M.: Realistic lead time scheduling of operations of orders. Machine Tools & manufacture, Vol. 44, 2004, pp. 1037-1046.

[5] Starbek M., Grum J., Kušar J.: Realistic lead times of Operations and Efficiency of the PPC System. Journal of Mechanical Engineering, Vol.48, 2002, pp. 98-111.

[6] Nyhuis P., Wiendahl HP.: Logistische Kennlinien. Springer Verlag Berlin, 1999, pp. 81-94.

[7] Berlec T.: Diagnostika pretočnega časa operacij, Magistrsko delo, Univerza v ljubljani, Fakulteta za strojništvo, Ljubljana 2003

[8] Starbek M., Menart D.: The optimization of material flow in production. International Journal of Machine Tools & Manufacture, Vol. 40, 2000, pp. 1299-1310.

[9] Lenart L., Kušar J.: Separable programming in optimal control. Proceeding of the 7th International Symposium on Operational Research in Slovenia, Podčetrtek, Slovenia, Slovenian Society Informatika, Section for Operational Research, 2003, pp. 235-240

REVERSE LOGISTICS IN VALUE CHAIN

Ludvik Bogataj and Marija Bogataj
University of Ljubljana
E-mail: Marija.Bogataj@guest.arnes.si

Abstract: To control the flows in total supply chain the analysis in frequency domain is more compact and assures better visibility of total chain. Location decision influence lead time in any activity cell of chain and often simultaneously also the contribution of reverse logistics. It is the reason that the approach to the analysis of location decisions in frequency domain is suggested here indirectly through changes of lead time. How to choose the best location of activity cells in reverse logistics particularly? How the (simultaneous changes) of different locations along the whole supply chain influence financial flow which reflects on the Net Present Value (NPV) can be studied using extended Material Requirements Planning (MRP) model. It is extended, including not only distribution but also recycling (reverse logistics) formalization. The recycling rate α is embedded in the model. The suggestion how to consider the perturbations of lead times in reverse logistics is given. The sensitivity analysis approach to the lead time perturbations is suggested in case when recycling is introduced in a supply chain.

Key words: Reverse logistics, Supply network, Material Requirements Planning, Laplace Transforms, Input – Output modelling, Lead time.

1. Introduction to the reverse logistics, studied in integrated supply chain

Optimal decisions (i) where to produce, (ii) how to distribute the product and (iii) at what time to order or deliver the items in integrated supply chain can be successfully discussed and evaluated in transformed environment, where lead times and other time delays can be consider in linear form. The site and capacity selection, as for instance the problems where it is best to locate a facility and what capacity is needed to achieve the most rapid response, are discussed more easily in transformed environment. Lead times in the entire supply chain can be analysed in compact form using MRP and I-O analysis in Laplace transformed space.

We can consider the activities in integrated supply chain as three distinct parts:

- materials management,
- physical distribution and
- recycling part,

All three parts can be defined in time or in frequency space.

Various operating stages in the logistic chain (nodes of the chain) can be represented by a simple model of a material – transformations or location - transformations processing cell. In each processing cell a value is added and some costs are incurred. At each processing cell there is a supply and a demand and often both are stochastic by nature.

Inventories are insurance against the risk of shortage of goods in each cell of logistic chain. They are limited by the capacity of each processing node of the chain and transportation capability of input and output flows and influence all kind of inventory costs.. Ordering goods (input flow) in distribution centres can be studied as a multi-period dynamic problem. The demand (output flow) during each period has to be considered as a stochastic variable. The distribution of this variable is often described with a certain probability function (see Bogataj & Bogataj /6/ or Tang /18 /).

The flows of items in supply chains influence transportation costs and costs of activities in logistic nodes of global economy and consequently the net present value (NPV) of all activities that have to be performed in such logistic networks. The costs of item flows between two cells depend on their location which can have advantages or disadvantages considered in detail in /5/, /7/, /8/ and /10/.

An integrated supply chain includes the purchasing of raw materials, manufacturing with assembly and the distribution of produced goods to the final clients. A third part which is more and more often included in the analysis of such integrated system, especially because environmental restrictions, is the reverse logistics, having the same formal properties in the networks. In a supply chain the key variables that have to be considered at each activity cell are activity level and timing, inventor level and lead times between the order time and the moment of the arrival of items in the required activity cell. The managers of a supply chain have two main goals: (a) to keep the level of inventory in the supply chain as low as possible, to reduce inventory costs; (b)To move the inventory in its continually changing form from raw material to final product and its delivery through the physical distribution part of the supply chain to the final consumer as fast as possible.

The final goal is mostly to achieve the maximal net present value NPV of all activities in the supply chain and not only to reduce the costs of operations.

Raw materials or parts of supply are delivered and stored in the raw material warehouses at the production centres. From there, raw material and parts are withdrawn as much as they are needed by the production centres. There the raw materials are transformed into subassemblies. The subassemblies are transformed in semi- finished goods and so on to the last production centre where final products are produced. From there finished goods are put into final product storage. From storage the finished goods are delivered to the distribution centres, or to their own warehouses or directly from there to retail outlets. There the goods are available to the final consumer. The third part, which can be considered in the same way as the first (material management) and the second (business logistics) part of the logistic chain, is reverse logistics part. Here the question appear when to include activities of the reverse logistics in a supply chain and how the increased percentage of recycling items in production influence NPV of total integrated chain. This is particularly difficult to evaluate when lead time is varying The compact analysis in frequency domain, where lead time perturbations can be studied in linear form, can help to give also direct evaluations of EU directives about environmental protection by recycling and remanufacturing activities.

A subject closely tied with production and inventory control is material requirements planning (MRP). It consists of a set of logically related procedures, decision rules and records designed to translate a master production schedule into time phased net requirements. To fulfil the planned coverage of all these requirements, a schedule needs to be implemented for each activity cell and its component of inventory items. A modern approach using Input – Output analysis and Laplace transforms has been introduced by Grubbström (1967) and his Linköping School (see: 1989, 1990, 1994, 1998...) The list of papers considering this problem is available at http://www.ipe.liu.se/rwg/rwg.htm . Such an approach gives us good theoretical and practical results also for the extension of the analysis to the distribution and especially to the reverse logistics part of supply network.

The concept of all three parts of an integrated logistic chain can be illustrated as in Fig.1.

From one stage to another, physical characteristics of the items and their locations are changing step by step. In the process of globalization, the distances between the pairs of production cells or between production and distribution cells are increasing, which influences the transportation costs and a lead time for the delivery of items.

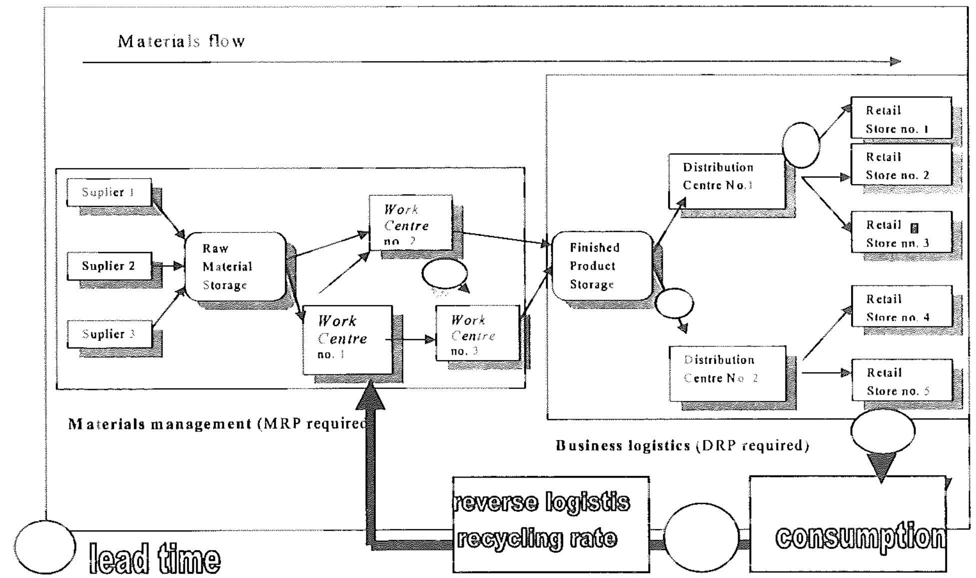


Fig. 1: The presentation of the inner structure in logistic chain, having production part, distribution part and reverse logistics.

2. Numerical presentation of a supply chain containing reverse logistics

Fig.2 presents a part of distribution and reverse logistic chain where 100 units of product A are packed in one unit (let it be the palette) of A_1 and transported to the certain destination Ω_1 . At the destination Ω_1 the palette A_1 is opened and two different products are made A_{21} and A_{22}. From one palette of products A $p\%$ of pallet A_{21} is filled with p good products and $(100-p)\% = q\%$ pallet of defected products A_{22} are collected.

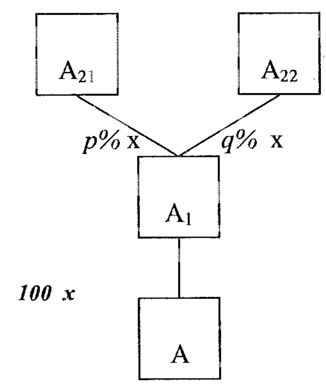


Fig. 2: An example of a simple arborescent product structure. 100 units of product A are packed in one unite of A_1. Unit A_1 is repacked into p% of pallet containing good products A_{21} and q% of pallet containing bad products collected in pallet A_{22}.

Let us suppose that 3 components of C are needed for production of 1 component of B and 4 components of C and 2 components of B are needed for 1 product A. This is described by Input – Output matrix H . In case when B in this production is covered

100α % by recycling of A_{22}, the extended Input – Output matrix (I_O) in this supply chain is:

$$H^{tot}=\begin{bmatrix} H & H_M^D \\ R(\alpha) & H_D^D \end{bmatrix}=\begin{array}{c} \begin{array}{cccccc} A & B & C & A_1 & A_{21} & A_{22} \end{array} \\ \left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 100 & 0 & 0 \\ 2(1-\alpha) & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & p/100 & q/100 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha/q & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{c} A \\ B \\ C \\ A_1 \\ A_{21} \\ A_{22} \end{array} \tag{1}$$

3. Supply chain lead times in fundamental equations of integrated supply chain containing reverse logistics

For describing time developments and time lags of the relevant production, demand and inventory variables in a compact way, the use of Laplace transforms for the production part of the supply system and input- output approach has been suggested by Grubbström /11-16/ and could be extended in total chain by introduction of I_O matrix described above. He introduced the fundamental equations which could be extended as follows. We have N items (produced, distributed and recycled in N activity cells) in the system. Demand D^{tot}, stock S^{tot} and production P^{tot} are presented by N dimensional column vectors. These vectors are transformed to frequency space and are denoted by tildes or by $\pounds\{\cdot\}$. Cumulative values are denoted by bars and Laplace inverse by $\pounds^{-1}\{\}$. The lead times $\tau_j, j=1,2,...N$ create the dynamics of internal demands and are changing especially because of changing the location of activity cells. They are represented by a diagonal lead time matrix $\widetilde{\tau}$, having $e^{s\tau_j}$ in the j-th diagonal position, where s is Laplace frequency. We denote the generalized input matrix by $\widetilde{H}^{tot}=H^{tot}\widetilde{\tau}$ (being disturbed by $\widetilde{\tau}^P(s)=\widetilde{\tau}(s)+\Delta\widetilde{\tau}(s)s$), which describes component requirements in each production cell together with their required timing. Grubbström /11-16/ introduced and developed the following notations for the production part of the logistic chain, which are here extended for entire supply chain including reverse logistics and therefore marked by $(\ )^{tot}$:

Total inventory	$\widetilde{S}^{tot}$	Cumulative deliveries	$\widetilde{F}^{tot}(s)/s$
Cumulative net production	$(I-H^{tot}(q,\alpha)\widetilde{P}/s$	Initial inventory	$S^{tot}(0)$
Generalised input matrix	$\widetilde{H}^{tot}(q,\alpha)= = H^{tot}(q,\alpha)\widetilde{\tau}$	Cumulative internal demand together with its required timing	$\widetilde{H}^{tot}(q,\alpha)\widetilde{P}(s)/s$
Available inventory	$\widetilde{R}^{tot}(s)$	Allocated component stock	$\widetilde{A}^{tot}(s)$
Cumulative gross production	$\widetilde{P}^{tot}(s)/s$	Backlogs	$\widetilde{B}^{tot}(s)$
Cumulative internal demand	$H^{tot}(q,\alpha)\widetilde{P}^{tot}(s)/s$	External demand in the frequency space	$\widetilde{D}^{tot}(s)$

The total inventory in time domain can be written as the sum of the available inventory and the allocated component stock:

$$S^{tot}(t)=\left[\pounds^{-1}\left\{\frac{R^{tot}(0)-B^{tot}(0)+(I-H^{tot}\widetilde{\tau}^{tot})\widetilde{P}^{tot}(s)-\widetilde{D}^{tot}(s)}{s}\right\}\right]^{+}+\pounds^{-1}\left\{\frac{A^{tot}(0)+H^{tot}(\widetilde{\tau}^{tot}-I)\widetilde{P}^{tot}(s)}{s}\right\}=$$
$$=Max\left\{\pounds^{-1}\left\{\frac{A^{tot}(0)+H^{tot}(\widetilde{\tau}^{tot}-I)\widetilde{P}^{tot}(s)}{s}\right\},\pounds^{-1}\left\{\frac{S^{tot}(0)-B^{tot}(0)+(I-H^{tot})\widetilde{P}^{tot}(s)-\widetilde{D}^{tot}(s)}{s}\right\}\right\} \tag{2}$$

Here production and distribution levels as control variables are collected into a vector $\widetilde{P}^{tot}(s)$:

$$\widetilde{P}^{tot}(s)=\left[x_1^M,x_2^M,\ldots,x_{m1}^M,x_{m1+1}^D,\ldots\ldots,x_{m1+m2}^D\right]^T \tag{3}$$

where x_i^M, $i=1,2,...m1$, denotes the activity level for $i-th$ production process, x_i^D $i=m1+1,m1+2,......m1+m2$, denotes the activity level for $i-th$ distribution process but the levels of recycling activities are determined through the submatrix $R(\alpha)$ presented in (1).

The expected net present value of the production part derived from /16/ has got the extended form on the time horizon $\hat{T}$ as:

$$E(NPV)=r(E(\widetilde{D}^{tot}(\rho))-\rho E(\widetilde{B}^{tot}(\rho))-E(B^{tot}(\hat{T}))e^{-\rho\hat{T}}-\vec{c}\widetilde{P}^{tot}(\rho)-K^{tot}\widetilde{\tau}^{tot}(\rho)\widetilde{\upsilon}^{tot}(\rho) \tag{4}$$

where $E(.)$ denotes expected value, r_j as components of r are sales prices of the products at different production levels and different locations which appear at different production and distribution cells in the supply chain, K^{tot} is vector of set- up costs at production cells and distribution cells for each activity (for example, in transportation there are always some fixed costs, which do not depend on the transportation volume). Here $\vec{c}$ is the row vector of unit incremental production costs, $\widetilde{\upsilon}^{tot}(\rho)$ is the vector of all completion times of production or distribution activities in the case when complex frequency equals to ρ, $\widetilde{\tau}^{tot}(\rho)$ is the lead time matrix, $\widetilde{\tau}^{tot}(\rho)=diag(\ e^{\tau_1 s}\ldots e^{\tau_N s})$, which depends on the time distance between two activities in production or distribution cells and on input or output capacities which define setup, production or distribution times, which can be changed due ti changing location and capacities by $\Delta\widetilde{\tau}^{tot}(\rho)=diag(\ e^{\tau_1 s}\Delta\tau_1,\ldots,e^{\tau_N s}\Delta\tau_N)$. For details see analogous production part /16/. A sensitivity study of the influence of long term changing locations of activity cells (the matter of strategic logistics and strategic alliances) on the lead times and consequently on the expected net present value of the activities in the logistic chain can be carried out in the presented compact form from /16/ directly:

$$\begin{aligned} E^P(NPV)=r(E(\widetilde{D}^{tot}(\rho))-\rho E(\widetilde{B}^{tot}(\rho))-E(B^{tot}(\hat{T}))e^{-\rho\hat{T}}-\vec{c}\widetilde{P}^{tot}(\rho)- \\ -K^{tot}\widetilde{\tau}^{tot}(\rho)\widetilde{\upsilon}^{tot}(\rho)+\Delta E(NPV(\Delta\tau,\alpha)) \end{aligned} \tag{5}$$

Let us note that in this case we can split the perturbed value of expected net present value $E^P(NPV)$ into unperturbed part $E(NPV)$ plus the effect of the perturbations $\Delta E(NPV)$, which is particularly useful in the sensitivity study to get a simple, quick and easily achievable estimate of the simultaneous effect of response time perturbation in the case of non-robust perturbations ($\Delta\tau_k \to 0, \forall k, k = 1,2,...N$.)and reverse logistic ratio α in $R(\alpha)$. In this case the perturbation of the net present value is:

$$\Delta E(NPV(\Delta\tau,\alpha) = \tilde{c}\begin{bmatrix} I \\ H''(\alpha)\tilde{\tau}(\rho) \end{bmatrix}(\sum_{k\geq 1} k(H'(\alpha)\tilde{\tau}(\rho))^{k-1}\tilde{\Delta}\tilde{\tau}(\rho)\tilde{P}'(\rho))\rho +$$
$$+\tilde{c}\begin{bmatrix} I \\ H''(\alpha)\tilde{\Delta}\tilde{\tau}(\rho)\rho \end{bmatrix}((I - H'(\alpha)\tilde{\tau}(\rho)\ )^{-1}\tilde{P}'(\rho) -$$
$$-K\tilde{\Delta}\tilde{\tau}(\rho)\tilde{v}(\rho)\rho + \Gamma(o(\tilde{\Delta}\tilde{\tau}(\rho)\rho)) \tag{6}$$

In (6) the amounts of end items and the timing of their production in unperturbed distribution environment are all captured in the vector $\tilde{P}'(\rho)$. The original input matrix in (6) is denoted by H'. In the matrix H'' the lower M rows of the extended matrix of H with capacity requirements are included. In (6) the term $\Gamma(o(\tilde{\Delta}\tilde{\tau}(\rho)\rho))$ is of lower order than $o(\tilde{\Delta}\tilde{\tau}(\rho)\rho)$.

4. Conclusion

Using (6) offers us the opportunity to study the influences of delay perturbations and reverse logistics ratio simultaneously. As we can see, a similar procedure can also be introduced to study more robust perturbations. In this case the changes in (6), which are included in the perturbation part of expression, have to be calculated more accurately, considering the sufficient number of terms in the appropriate expression of Taylor series in $\Delta E(NPV)$.

The obtained sensitivity results for non-robust perturbations are particularly useful in practice when it is necessary to make a quick estimate of the risk value owing to the unexpected perturbations of lead times which appear in a certain activity cell at a certain location and compensation of it with changing reverse logistics ratio α. This approach helps to balance the costs of a certain level of inventories as the insurance of lean and/or smooth production and supply.

References

1. Baldiwala, Q.: "Developing a Global Supply Chain ", *Logistics Spectrum.* Vol.35/4, 2001, pp. 25-29.
2. Bogataj, D., Bogataj, M. "Capacity planning in stochastic programming of logistic chains ". Lipičnik, M. (Ed.). *Transport, promet, logistika*, Portorož, Slovenija, *2. Annual proceedings,* 2000, pp. 11-16.
3. Bogataj, L., Horvat, L.,"Stochastic considerations of the Grubbström -Molinder model of MRP, Input-Output analysis and multi-echelon inventory systems", *Int. J. Prod. Econ.*, Vol.45, 1996, pp 329-336.
4. Bogataj, L., Bogataj, M., " Minimization of distribution costs and response time in global supply chain defined in the frequency space ", *Annual proceedings.* Orlando: The International Society of Logistics, 2001, pp 1-12.
5. Bogataj, M., "Inventories in spatial models", *Int.J.Prod.Econ.*, Vol. 45, 1996, pp 337-342.
6. Bogataj, M., Bogataj, L." Compound distribution of demand in location - inventory problems". in: Papachristos, S. (Ed.), Ganas, I. (Ed.). Third ISIR summer school, Ioannina, 1998. *Inventory modelling in production and supply chains : research papers.* Ioannina: International society for inventory research: University of Ioannina, 1998, pp. 15-46.
7. Bogataj, M., "Uncertainty of delivery determining the market area ". In Grubbström, R.W., Bogataj,L. (Eds).Input-Output Analysis and Laplace Transforms in Material Requirements Planning. Storlien 1997. FPP Portorož 1998, pp 115-122.
8. Bogataj, M. "Inventory allocation and customer travelling problem in spatial duopoly ". *Int. J. Prod. Econ.* , 1999, vol. 59, no. 1/3, pp. 271-279,
9. Edwards, P., Sharman,G., "The Effectivenes of Information Systems in Supporting the Extended Supply Chain" presented at the *Council of Logistics Management educators conference*, New Orleans, Florida, 2000.
10. Girlich, H.J., "On the Metric Transportation Problems and Their Solution", in Bogataj, L.,(Ed.), *Inventory Modelling. Lecture Notes of the International Postgraduate Summer School.* Vol.2, ISIR, Budapest- Portorož, 1995, pp13-24.
11. Grubbström, R.W., Molinder, A., Further theoretical consideration on the relationship between MRP, Input-Output analysis and multiechelon inventory systems", *Int.J.Prod.Econ.*, Vol. 35, 1994, pp 299-311
12. Grubbström, R.W. "On the Application of the Laplace Transform to Certain Economic Problems", Management Science, Vol.13, No.7, 1967, pp. 558-567
13. Grubbström, R.W., "Stochastic Properties of a Production-Inventory Process With Planned Production Using Transform Methodology", *Int.J.Prod.Econ*, Vol. 45, 1996, pp 407-419.
14. Grubbström, R.W., "Material Requirements Planning and Manufacturing Resources Planning", in Warner, M., (Ed), *International Encyclopedia of Business and Management,* Routladge, London, 1996.
15. Grubbström, R.W., Bogataj, L. (Eds).*Input-Output Analysis and Laplace Transforms in Material Requirements Planning.* Storlien 1997. FPP Portorož 1998.
16. Grubbström, R.W., "A Net Present Value Approach to Safety Stocks in Planned Production ", *Int.J.Prod.Econ*, Vol. 56-57, 1998, pp 213-229.
17. Segerstedt, A., "A Capacity- Constrained Multi-level Inventory and Production Control Problem" *Int.J.Prod.Econ*, Vol. 45, 1996, pp 449-461.
18. Tang, Ou, "Application of transforms in a compound demand process" *Promet-Traffic-Traffico,* Vol. 13/5, 2001, pp. 355-364.
19. Waller, D.L.,"Operations Management, A Supply Chain Approach, ITP, London, 1999.

INTERMUNICIPAL GRAVITY MODEL OF SLOVENIA

Samo Drobne* and Marija Bogataj**

* University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia, sdrobne@fgg.uni-lj.si

** University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia, marija.bogataj@guest.arnes.si

Abstract

In this paper, an extended Lowry-like intermunicipal model of daily commuting in Slovenia is presented. The regression model has been already implemented as a web-based application, where citizens could participate and give an opinion to policy-makers about the forecasting role and results of the model.

Keywords: gravity model, spatial model, commute, municipality, Slovenia.

1 INTRODUCTION

As a cause and a consequence of social change, human migration and periodical (daily or weekly) commuting to job, school or for other activities are regarded as one of the most important factors underlaying the demographic and socioeconomic composition of regions of different sizes (varying from states and administrative regions to municipalities and/or communities). Thus, for anyone attempting to analyze the general process of regional change, an understanding of interregional migration is vital [1]. Cadwallader [9] has pointed out, that 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 and daily commuting. The migration between regions can be slowed down by daily commuting, which is becoming a surrogate for migration, if the commuting is bringing higher social well-being than any migration. If the contacts between regions, because of improved transportation abilities and removed barriers, are becoming less expensive and easier, the inhabitants often prefer daily commuting [13].

In our previous investigations [2,3,4,5,6], we proved that daily commuting also has an important role in the context of a socio-economic issue in Slovenia. In [7], we summarized all results: we investigated the main factors of interregional migratory and daily commuting flows in Slovenia, previously predicted in above mentioned papers. For most regional analysis, interregional extended gravity model should be precise enough: it can even predict how regionalization and changes of regionalization influence gravity. However, more precise forecasting results about daily commuting can be calculated only using smaller spatial units than regions. For this purpose, we calculated an extended Lowry-like intermunicipal model of daily commuting in Slovenia for 2005. The model has been already implemented in web-based application, where citizens could participate and give an opinion to policy-makers about the forecasting role and results of the model.

2 PROBLEM FORMULATION

The gravity models, which belong to the family of spatial interaction models, offer a framework for building integrated models of land use compared to the econometric models. The Lowry model, designed in 1964 for the Pittsburgh metropolitan region [12] and revised several times later, is the basic in this group of models. However, Lowry-like models miss many other aspects of integration; the impact of the transportation network on the land use, which were particularly emphasized in [7], where we calculated an Lowry-like interregional model of daily commuting. However, for more precise forecasting results, we studied intermunicipal daily commuting flows for persons in employment and implemented them in the gravity model.

Let us study 192 municipalities of Slovenia with a total of population $P = 1\,998\,079$ inhabitants. We denote with i the living municipality and with j the destination municipality. The number and structure of external daily commuters, persons in employment, were obtained from Census 2002 data and was $\sum_i \sum_j DC_{i,j} = DC = 440\,299 \ (i \neq j)$. Data on Slovenian population as well as the number of external daily commuters were obtained from statistical data (Census 2002 and Statistical news [14]). Figure 1 shows population and daily commuters – persons in employment in Slovenian municipalities in 2005.

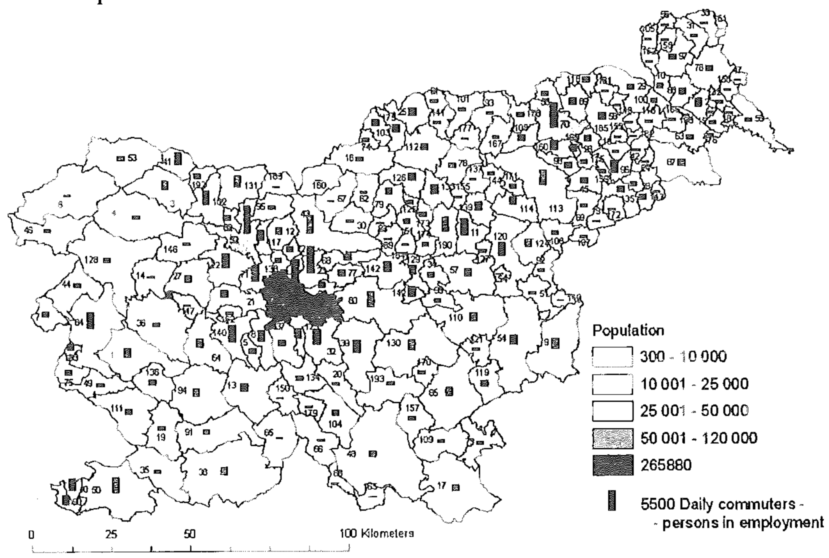


Figure 1: Population and daily commuters – persons in employment in Slovenian municipalities in 2005

The simple gravity model helps us determine the expected number of daily commuters which originate in municipality i and terminate in municipality j under the assumption that no costs and no time are involved in undertaking a trip from one municipal to another (for the starting analysis, the friction of distance is zero). If there are no mentioned differences among the municipal, we may expect for a particular person in employment, living in municipality i, the probability to travel to municipality j is P_j/P, where P_j is the total population of the j-th municipality. By homogeneity assumption we can estimate the number of daily commuters between pairs of municipalities.

Now, we can define the daily commuting coefficient between the municipalities k_{DC}^{MUN}:

$$k_{DC}^{MUN} = \frac{DC}{P} = 0.13 \tag{1}$$

Following the principles of simple gravity model and assumption mentioned above, the hypothetical number of daily commuters – persons in employment between the pairs of municipalities (originating in i and terminating in j, $i \neq j$), is then:

$$T_{i,j}^{DCs} = \frac{k_{DC}^{MUN} P_i P_j}{P} \tag{2}$$

Following to the principles of the gravity models, the distances between studied pairs of municipalities should be included in the model. In [4,7], we proved that a new number of daily commuters $DC_{i,j}^{*}$ is:

$$DC_{i,j}^{*} = \frac{T_{i,j}^{DCs}}{d_{i,j}^{*}(\cdot)^{\kappa}}, \tag{3}$$

where $d_{i,j}^{*}(\cdot)$ is a new, expected (improved) distance between municipalities after investments in the network, $(\cdot)$ denotes that a distance can be treated in different ways (Euclidian, road or time-spending distance), and κ is an exponent – usually defined in the regression analysis between actual distance $d_{i,j}(\cdot)$ and ratio of actual to hypothetical daily commuters

$$\rho_{i,j}^{DCs} = \frac{DC_{i,j}}{T_{i,j}^{DCs}}. \tag{4}$$

The model (3) enables the forecasting of number of daily commuters between pairs of analyzed spatial units. However, to extend gravity model some additional economic coefficients were added to the model. To be included in the model, the following coefficients was investigated:

$$K_{GDP,i} = \frac{GDP(i)}{GDP(SI)}, \ K_{GEAR,i} = \frac{GEAR(i)}{GEAR(SI)}, \ K_{EMP,i} = \frac{EMP(i)}{EMP(SI)}, K_{UEMP,i} = \frac{UEMP(i)}{UEMP(SI)},$$

where $GDP(i)$ is Gross Domestic Product per capita in municipality i[1], and in Slovenia (SI) respectively, $GEAR(i)$ is an average gross earning per person in municipality i and in Slovenia (SI) respectively, $EMP(i)$ is the number of persons in employed in the municipality i and in Slovenia (SI) respectively, and $UEMP(i)$ is the level of registered unemployment in the municipality i divided by the level of registered unemployment in the country, and $UEMP(SI)$ is the level of registered unemployment in Slovenia, and $i = 1,2,\ldots,192$. The model was tested for Euclidian, the shortest road distances, as well as for the quickest time-spending distances.

3 PROBLEM SOLUTION

The Census 2002 data enable the study of the volume of daily commuting to job and home between pairs of municipalities in dependence of time-distances between the municipal centres. Time-spending distances were calculated in transport planning software OmniTRANS. It is impossible to show the intensity of all flows on a map; Figure 2 represents the intensity of daily commuting flows in 2002 for the workers employed out of their municipality of living – only for Pomurje region.

[1] The smallest spatial unit, that Gross Domestic Product was defined in Slovenia in 2005, is statistical region. So, the presumption was, that all municipalities in the same region have the same GDP.

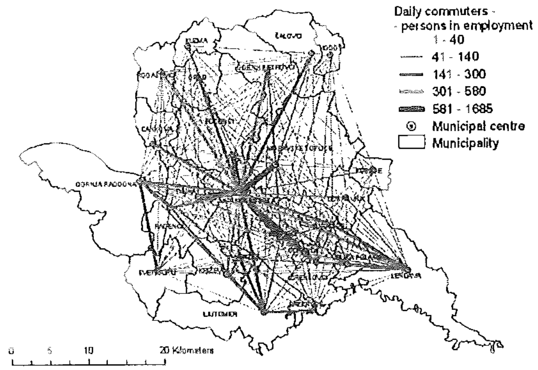


Figure 2: Intermunicipal flows of daily commuters - persons in employment in Pomurje region in 2005

In the regression analysis of daily intermunicipal commuters – persons in employment, who travel by car, only the time-spending distance and following coefficients gave the best results (R^2 is the highest and P-values are the lowest):

$$DC^{*}_{i,j} = \frac{aP_i^{\alpha_i} P_j^{\alpha_j}}{d(t)^{\beta}_{i,j}} K^{\gamma_1}_{GDP,i} K^{\gamma_2}_{GDP,j} K^{\gamma_3}_{GEAR,j} K^{\gamma_4}_{UEMP,i} K^{\gamma_5}_{UEMP,j} \quad (5)$$

We got the regression parameters for the intermunicipal daily commuting flow equation, where R-square is 58 % for 7099 observations and where $d(t)$ is the road time-spending distance needed for travelling by car; see Table 1.

Table 1: Extended gravity model coefficients and summary output for intermunicipal daily commuting ($DC^{*}_{i,j}$) – persons in employment in 2005.

Name of Coefficients	*Value of Coefficients in (5)*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Value of Coefficients without γ_4*
$Ln(a)$	-2.0567	-11.7022	0.0000	-2.4012	-1.7121	-2.0676
α_i	0.4146	32.6054	0.0000	0.3897	0.4395	0.4163
α_j	0.7177	52.2147	0.0000	0.6907	0.7446	0.7164
$-\beta$	-1.7307	-92.4236	0.0000	-1.7674	-1.6940	-1.7303
γ_1	-0.8986	-10.2788	0.0000	-1.0699	-0.7272	-0.9526
γ_2	-0.2255	-2.6514	0.0080	-0.3923	-0.0588	-0.2229
γ_3	0.6484	5.9864	0.0000	0.4361	0.8607	0.6488
γ_4	0.0391	1.0367	0.2999	-0.0348	0.1129	
γ_5	-0.1538	-3.9385	0.0001	-0.2303	-0.0772	-0.1438
Multiple R = 0.76 R Square = 0.58						

The intermunicipal regression model of daily commuters – persons in employment is then:

$$DC^{*}_{i,j} = \frac{0.13P_i^{0.41} P_j^{0.72}}{d(t)^{1.73}_{i,j}} K^{-0.90}_{GDP,i} K^{-0.23}_{GDP,j} K^{0.65}_{GEAR,j} K^{0.04}_{UEMP,i} K^{-0.15}_{UEMP,j} \quad (5a)$$

Using data from Census 2002, the results of regression analysis show that coefficient γ_4 is not statistically significant (see also Table 1). It was expected that unemployment in the municipality of origin was one of the most important factors pushing inhabitants to commute to their work. However, excluding this coefficient from the analysis does not change a lot the values of other coefficients (compare values in second and last column in Table 1).

4 MODEL APPLICATION

Using the regression model of daily commuters – persons in employment, who travel between Slovene municipalities (5a) and data on new, improved time-spending distances after highways will be finished in 2013, new volumes of daily commuters were evaluated; a model of calculation of time-spending distances in a raster-based GIS environment is discussed in [10]. Figures 3 and 4 shows the growth of daily commuters - persons in employment due to shortening of time-spending distances (new highways) in municipalities of origin in Slovenia from 2005 to 2013, and in municipalities of destination respectively.

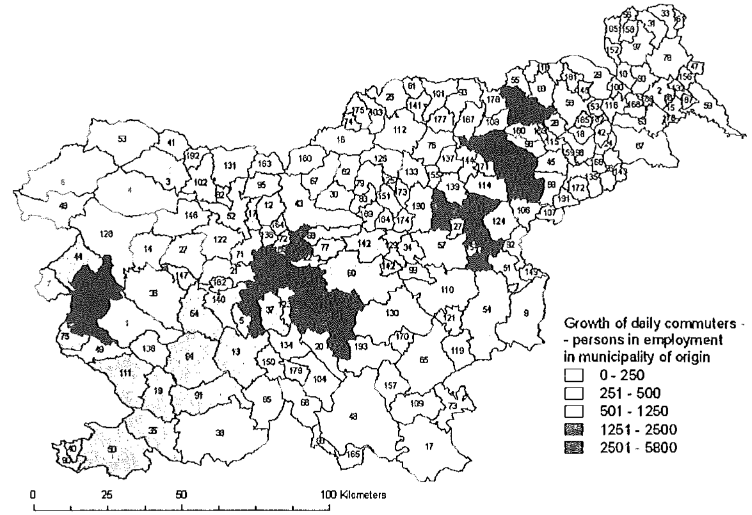


Figure 3: Growth of daily commuters - persons in employment due to shortening of time-spending distances because of new highways in municipalities of origin in Slovenia from 2005 to 2013

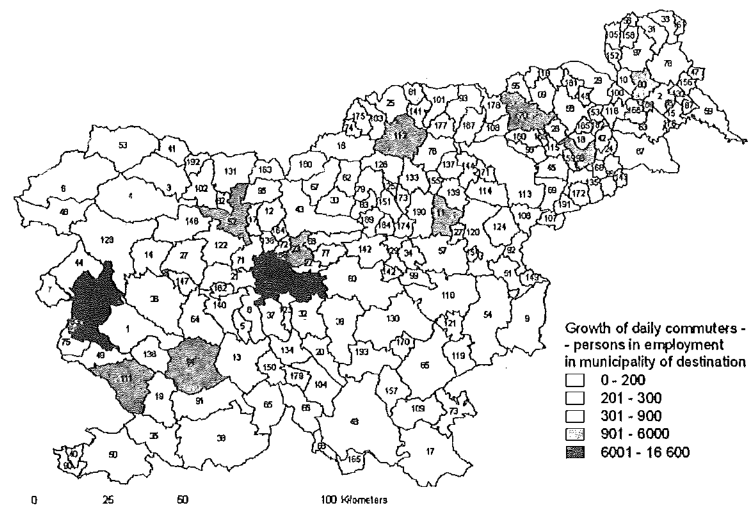


Figure 4: Growth of daily commuters - persons in employment due to shortening of time-spending distances because of new highways in municipalities of destination in Slovenia from 2005 to 2013

The model (5a) is implemented also as a web-based application of intermunicipal gravity model of Slovenia [11] – interested reader is invited to test it at http://193.2.92.34/ – where decision-makers and citizens can work together to find constructive solutions in the decision making about investments in the infrastructure, municipality growth and development.

5 CONCLUSIONS

The results of intermunicipal gravity model of daily commuters – persons in employment shows that the reduction of time-distances between Slovene municipalities increases the daily commuting flow. Though it is expected that high time delays appear between the spatial investment and the reaction of the population, this significance is already recognized and proved in [7]. However, these results require a further study of the influence of the newly built motorways in terms of employment and commuting between the spatial units of different size in Slovenia. Finally, we need to ask ourselves how will the new investments in roads, which will induce the transit traffic in Slovenia, influence the demand of land and land use in the central places of Slovenia. This is one of the main parameters needed for successful spatial planning.

References

[1] Anjomani, A., 2002: Regional growth and interstate migration, *Social-Economic Planning Science*, (36):239–265.

[2] Bogataj, M., Drobne S., Bogataj L., 1995a: System approach to urban structure and growth. In: Rupnik, V. and M. Bogataj (ed.), SOR '95 proceedings, International Symposium on Operational Research, Portorož, 37–46.

[3] Bogataj, M., Drobne S., Bogataj L., 1995b: The influence of investment and fiscal policy on growth of spatial structure, Suvremeni promet, Zagreb, 15(5):239–245.

[4] Bogataj, M., Drobne S., 1997: The influence of investments in highways on gravity and interaction in Slovenia. In: Rupnik, V., L. Zadnik Stirn and S. Drobne (ed.), SOR'97 proceedings, International Symposium on Operational Research, Preddvor, 55–60.

[5] Bogataj, L., Bogataj, M., Drobne, S., Vodopivec, R., 2003: Management of investments in roads and in capacities of border regime to induce the flow of human resources in and out of region. In: Zadnik Stirn L., M. Bastič and S. Drobne (ed.), SOR '03 proceedings. International Symposium on Operational Research, Podčetrtek, 43–46.

[6] Bogataj, L. Bogataj, M., Drobne, S., Vodopivec, R., 2004: The Management of Investments in Roads and in Capacities of Border Regime to Improve the Human Resources in Regional Development. Suvremeni promet, Special issue, Mostar, Vol. 30, to appear.

[7] Bogataj, M., Drobne S., 2005: Does the improvement of roads increase the daily commuting? Numerical analysis of Slovenian interregional flows. In: Zadnik Stirn, L., Indihar Štemberger, M., Ferbar, L., Drobne, S., Selected Decision Support Models for Production and Public Policy Problems, Slovenian Society Informatika, Ljubljana.

[8] Briassoulis, H., 2000: Analysis of Land Use Change: Theoretical and Modeling Approaches. Regional Research Institute, West Virginia University.

[9] Cadwallader, M., 1992: Migration and residential mobility: macro and micro approaches. University of Wisconsin Press, Wisconsin.

[10] Drobne, S., Bogataj, M., Paliska, D., Fabjan, D., 2005: Will the future motorway network improve the accessibility to administrative centres in Slovenia?. In Zadnik Stirn, L., Drobne, S., Proceedings of SOR'05, Nova Gorica, here.

[11] Gaiduchenko, O., 2005: Development of WEB-based gravity model for forecasting the commuters' flows in Slovenia. Diploma work, Technical University of Riga, Latvia.

[12] Lowry, I. S., 1966, Migration and metropolitan growth: two analytical models. Chandler Publishing Company, San Francisco.

[13] Nijkamp, P., 1987: Handbook of Regional and Urban Economics. Vol. 1, Regional Economics, North – Holland.

[14] SURS 2005: Statistical Office of the Republic of Slovenia, URL: http://www.stat.si/eng/index.asp, date accessed: 03-May-2005.

WILL THE FUTURE MOTORWAY NETWORK IMPROVE THE ACCESSIBILITY TO ADMINISTRATIVE CENTRES IN SLOVENIA?

Samo Drobne[*], Marija Bogataj[**], Dejan Paliska[***] and Daša Fabjan[***]

[*] University of Ljubljana, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia, sdrobne@fgg.uni-lj.si

[**] University of Ljubljana, Faculty of Economics, Ljubljana, Slovenia, marija.bogataj@guest.arnes.si

[***] University of Ljubljana, Faculty of Maritime Studies and Transportation, Portorož, Slovenia, dejan.paliska@fpp.edu, dasa.fabjan@fpp.edu

Abstract

A two-stage raster-based GIS approach for evaluation of continuous accessibility fields has been presented. The proposed methodology can be essential for a complete integration of accessibility in many different studies - fully integrated in GIS environment – at macro or mezzo level. In the model applied, the improvement of time-spending distances to administrative centres due to new motorways has been studied for the years 2005 and 2013.

Keywords: accessibility, travel time, administrative centre, GIS, Slovenia.

1 INTRODUCTION

The concept of accessibility can be used in many contexts and in different ways; for instance, as a goal in transportation policy, as measures in rural development policy, as an indicator of rural deprivation, and as a variable in location analysis. Accessibility is one of the main issues in development and physical planning policies, which are mostly concerned with equity and a better distribution of people and activities in the territory. So, accessibility is considered, by planners and other technicians, to be a key variable for territorial development and planning.

Accessibility is the ability to control the transportation facilities that are necessary to reach the desired location at a suitable time [7]. However, the concept of accessibility incorporates not only the transport link between origin and destination, and the ability for travelling by the target group, but also characteristics of the destination and the objective of the trip.

The concept of accessibility can be applied to many spatial problems. Thus, the analysis based on this concept is therefore suitable to be integrated within geographic information systems (GIS).

Accessibility can be measured in several different ways, already described in [8]. However, every measure of accessibility must consist of two elements. The first is distance between the person or place and the destination; the second is the activity in the various destinations. These two elements are combined by using comparative or composite measures. In the context of GIS, both measures are easy to handle – although the composite measures are more useful, than the comparative measures.

The strength of GIS lies in their flexibility; they can represent irregular spatial shapes and patterns, as well as the spatial relationships between them. This characteristic gives the possibilities for modelling, where proximity and coincidence of irregularly spaced phenomena are very important. The relations between these phenomena can be modelled with standard GIS functions - such as recode, buffer, and overlay – implemented in cartographic modelling.

This paper expands our work in modelling accessibility fields [3, 4, 5 and 6], originated from Donnay and Ledent [2] and tested by Julião [9]. In the model applied we tested, whether the future approved motorway network is going to improve the accessibility to administrative centres in Slovenia or not.

2 PROBLEM FORMULATION

The assessment of accessibility can be solved using various algorithms, the so called shortest path algorithms from graph theory. The research of the shortest path in a graph constitutes an application frequently implemented in GIS software. The issue fits in with geographical data available in vector format. However, Donnay and Ledent [2] stressed two major restrictions that feature the vector solution of accessibility. First, accessibility is computed only at the nodes of the network. The accessibility of intermediate points on one segment can only be assessed with the interpolation between its two ending nodes, which is valid only for the segments lying radially from the origin. And second, the accessibility of the points located away from the network is completely ignored by the vector approach. Consequently, the drawing of contour lines (isodistances or isochronous lines), which is often an important goal of the application, is subjective and not reliable.

An overall solution would be to consider accessibility as a field; i.e. a spatial continuous geographic variable. Then the accessibility can represent an actual value at every point in the region of interest, a minimum value at the origin, and a maximum value inside the network grid when the accessibility is expressed in terms of distance or time-distance. The handling of the accessibility field presupposes a transcription of the problem in the raster format. This can introduce some restrictions on data, but it offers new processing possibilities concerning surface modelling.

The raster-based GIS methodology for accessibility evaluation, proposed in this paper, required a two-stage modelling. Layers describing the public road network, administrative centres, and dwelling houses with data on residents were used. The image resolution obviously influences the accuracy of accessibility evaluation. In our application, the vector layers were rasterized with the resolution of 100 m, which is accurate enough when working at the administrative level.

3 PROBLEM SOLUTION

The key issue in modelling accessibility fields is the definition of cost surfaces. Cost surfaces are defined by data on distances from origin (features) in terms of costs measures (e.g. cost distances). Evaluation of cost surfaces requires a friction surface that indicates the relative cost of moving through each cell. In our application, costs of movement across the region were expressed as travel time. These represent the time necessary to move through areas with certain attributes.

The friction surface was calculated using a simple model that fixes the value of the cell crossing time (see also [3]):

$$CCT = \frac{PS \cdot 60}{TS \cdot 1000} \cdot \vartheta \qquad (1)$$

where CCT is Cell Crossing Time (in minutes), PS is Pixel Size (in meters), TS is Travel Speed (in kilometres per hour), and ϑ is the correction coefficient explained below.

To use properly the raster-based approach, stereological sampling has to be considered where the road distance between two points is determined by number of the intersections between the road and the raster grid. In accessibility analysis, the category of road determines the average travel speed between two nodes in the vector-based approach, and the average travel speed between two intersections of that road with the raster grid. The expected number of intersections of each road category has the same structure as is the structure of road categories in reality. Using the solution of Buffoon's Needle Problem [1], we can conclude the following: "In the structure of intersections of roads of different categories in the analyzed area, the ratio of intersections of category X compared to the total number of intersections is equal to the ratio between the length of all roads of category X and the total length of roads in the analyzed area". Thus, in [5] we calculated the correction coefficient for CCT, which is $\vartheta = \pi/4$.

While in [3] we proposed a one-stage raster-based GIS approach for modelling of accessibility fields, cartographic modelling of accessibility fields at higher levels (administrative, regional, interregional, state or interstate level) makes it necessary to work out specific methodologies. In this case, two different cost surfaces are needed: one indicating the travel time considering the whole road network, and another one excluding the motorways. Two-stage modelling of accessibility fields as a spatial continuous geographic variable is necessary because one can not get in and out of a motorway at any point – which is the case for all other road categories.

Two-stage modelling of accessibility fields – applied in our cartographic model – follows the next major six steps:

1. Calculation of travel time to the centres outside (without) the motorways and major roads.
2. Calculation of travel time from the centres using motorway or major road to motorway or major road connections.
3. Calculation of travel time from hinterland (from each location, defined by pixel size) to the motorway or major road connections. And, territorial allocation for each connection according to the travel time to the connections.
4. Adding up travel time from hinterland to connections (step 3) with the travel time from the centres to the connections (step 2).
5. Determining the minimum travel time from the administrative centre (comparing travel times from the centres without motorways or major roads (step 1) with travel times using motorways and major roads (step 4)).

Figure 1 shows the cartographic model designed and operating in GIS tool Idrisi, where violet rectangles denote input data and intermediate and final results, white rectangles attribute data for values assignment, green rectangles vector intermediate results, and red rhombuses denote analytical modules in Idrisi.

4 MODEL APPLICATION

The calculations of travel time to administrative centres in Slovenia were performed for the base year 2005 and for the year 2013, when the approved motorway network is to be finished. To calculate the friction surface using the simple model (1) and to later add the correction coefficient ϑ, the average speed was defined for every category of road network as well as for the hinterland. The overall data and results are presented in Table 1.

Table 1: Average travel speed and cell crossing time according to the road category

Road category	*Average speed (in km/h)*	*CCT (Cell Crossing Time; in minutes)*	*CCT * ϑ (in minutes)*
Motorway (MW)	110	0.0545	0.0428
Major road (MR)	90	0.0667	0.0524
Main road 1 (MR_1)	70	0.0857	0.0673
Main road 2 (MR_2)	65	0.0923	0.0725
Regional road 1 (RR_1)	65	0.0923	0.0725
Regional road 2 (RR_2)	50	0.1200	0.0942
Regional road 3 (RR_3)	40	0.1500	0.1178
Regional road 3 - tourist road (RT)	35	0.1714	0.1346
Local road (LR)	40	0.1500	0.1178
Other road (OR)	30	0.2000	0.1571
Outside the road network (OUT)	*15*	*0.4000*	*0.3142*

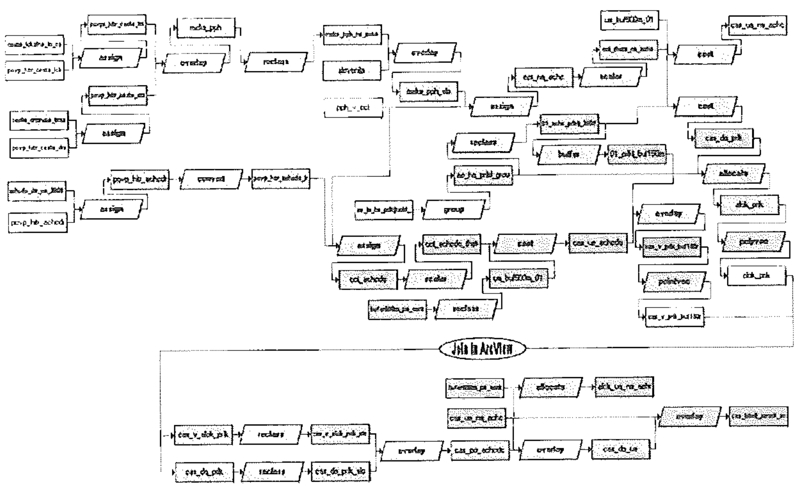


Figure 1: Raster-based two-stage cartographic model for calculation of time-spending distances to selected objects

In our application, we analyzed the accessibility of every location in Slovenia to the nearest administrative centre. Therefore, the cost value (time distance) was calculated as the less cumulative costs starting from the origins (administrative centres) and moving through a friction surface. Note, that the time-spending distance was calculated for the ideal circumstances without consideration of traffic flows by day time schedules, traffic restrictions on road segments, etc.

Comparing time-spending distances to the selected centres in analyzed years 2005 and 2013, when approved motorway network will be finished according to the plans, the improvement of time-spending distances due to new motorways was calculated. Figure 2 shows the improvement of time-spending distances due to new motorways from 2005 to 2013.

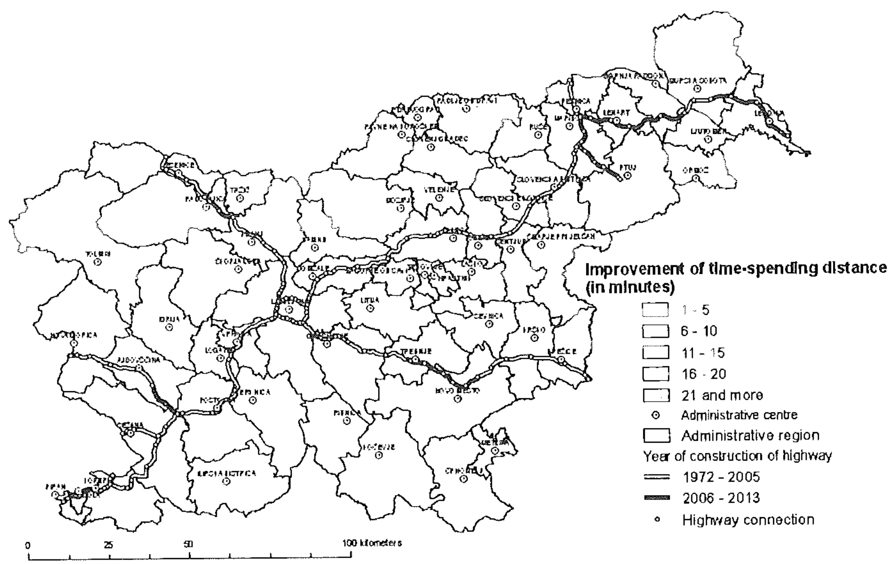


Figure 2: Improvement of time-spending distances to administrative centres due to new motorways from the year 2005 to the year 2013

Once the cost surfaces were determined, different accessibility analyses could be performed. In our application, we compare the location of dwelling houses in Slovenia with the improvement of time-spending distances to administrative centres. In 2005, there were about 1 997 000 inhabitants living in almost 483 000 dwelling houses in Slovenia. Figure 3 shows dwelling houses of 2005 defined according to the improved time-spending distance to administrative centres due to new motorways from the year 2005 to the year 2013.

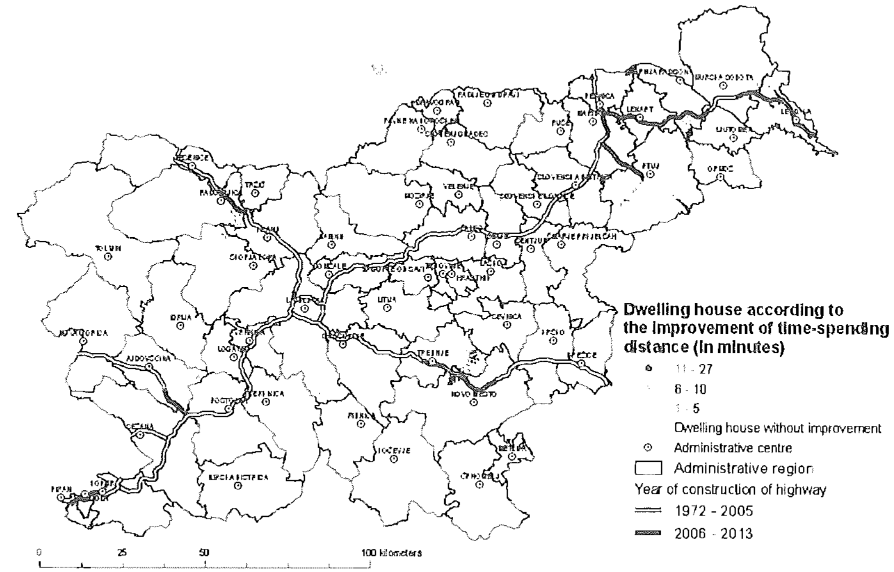


Figure 3: Dwelling houses in Slovenia in 2005 according to the improved time-spending distance to administrative centres due to new motorways from 2005 to 2013

Table 2: Number of dwelling houses and population in Slovenia in 2005 according to the improved time-spending distance to administrative centres due to new motorways from 2005 to 2013

Improved time-spending distance (in minutes)	*Number of dwelling houses*	*Population*	*Percentage of population*
1 - 5	28 389	88 448	4.4 %
6 - 10	5602	16 333	0.8 %
11 - 27	1454	3717	0.2 %
together:	*35 445*	*108 498*	*5.4 %*

When talking about the improvement of accessibility to administrative centres in Slovenia due to the new motorways from the year 2005 to the year 2013, we can realize that this will affect only 5.4 % of all population in Slovenia (app. 108 500 inhabitants), which can be seen in Table 2 above. For most of them (4.4 %; app. 28 500 inhabitants), the time-spending distance will be improved for only 5 minutes or less. However, there are only 2 ‰ of Slovene population, whose accessibility to the administrative centres will be improved for more than 10 minutes, when all approved motorways will be completed in 2013.

5 CONCLUSIONS

In this paper, a two-stage raster-based GIS approach for evaluation of continuous accessibility fields has been presented. The proposed methodology can be essential for a complete integration of accessibility in many different studies. The application made here showed that accessibility analyses fully integrate in GIS environment in order to obtain a dynamic and full territorial coverage. However, to make the accessibility analysis more complex, some additional methodologies have to be considered in the model regarding the barriers and traffic data. Physical barriers and other information regarding land surface are important to get more accurate results of territorial accessibility, especially that considering off-road travelling. In this way, for simulation of travelling outside road network, especially data on terrain slope has to be considered. And, instead of calculating a simple cost surface based on a fixed travelling time, it is possible to integrate in the cartographic model an accessibility behaviour that can often be expressed with traffic flows by day time schedules and traffic restrictions on road segments. All discussed improvements of calculating the accessibility fields should be incorporated as weighted layers in the model.

A raster-based cartographic model suggested here should be used only in accessibility analysis at mezzo or macro level (at a country level, or statistical or administrative regions' levels; conditionally also at municipal levels) where the high accuracy of the results is not necessary. But, when calculating accessibility fields at micro level (city or town level) or when the high accuracy of results is of a big importance, one can use different methodologies implemented in vector-based GIS approach [10].

References

[1] Buffon G. L. L. 1777: Comte de. Essai d'Arithmétique Morale. In: Supplément à l'Histoire Naturelle, 4. Paris: Imprimerie Royale.

[2] Donnay J. P., Ph. Ledent, 1995: Modelling of Accessibility Fields. *Proceedings JEC-GI '95*, vol. I, pp. 489-494.

[3] Drobne S., 2003: Modelling Accessibility Fields in Slovene Municipalities. Proceedings of the 7th Symposium on Operation Research in Slovenia (SOR'03), Ljubljana, Slovenia, p. 89-96.

[4] Drobne S., 2005: Do Administrative Boundaries Fit Accessibility Fields in Slovenia? Proceedings of the 6th international conference Environmental engineering, Vilnius, Lithuania, pp. 537-542.

[5] Drobne S. and M. Bogataj, 2005: Korekcija (časovne) razdalje v postopku modeliranja dostopnosti v rasterskem GIS-u = Correction of (time-spending) distance in accessibility modells using raster-based. Proceedings DSI - Dnevi slovenske informatike 2005 (DSI 2005), pp. 484-490.

[6] Drobne S., D. Paliskan and D. Fabjan 2004: Rastrski pristop dvostopenjskega modeliranja dostopnosti v GIS-u = Raster-based two-stage approach of modelling accessibility in GIS. *Geografski informacijski sistemi v Sloveniji 2003-2004*. Ljubljana, ZRC, pp. 265-276.

[7] Geertman S.C and J. R. Ritseman van Eck, 1995: GIS and Models of Accessibility: An Application in Planning. *Int. J. Geographical Information Systems*. 9(1): 67-80.

[8] De Jong T., J. R. Ritseman van Eck, 1997: Location Profile-Based Measures as an Improvement on Accessibility Modelling in GIS. *Computers, Environment and Urban Systems*, 20(3): 181–190.

[9] Julião R. P., 1999: Measuring Accessibility Using GIS. *GeoComputation Proceedings*, http://www.geovista.psu.edu/sites/geocomp99/Gc99/010/gc_010.htm, date accessed: 18-04-2003.

[10] Paliska D., S. Drobne and D. Fabjan, 2004: Uporaba GIS-a za proučevanje prostorske dostopnosti v analizi povpraševanja po storitvi JPP = Using GIS to study the transit accessibility in public transportation service demand analysis. *Geografski informacijski sistemi v Sloveniji 2003-2004*. Ljubljana, ZRC, pp. 79-88.

SOME CONSIDERATIONS ON OPTIMISATION IN LOGISTIC SYSTEMS

Peter Köchel

Chemnitz University of Technology, Department of Informatics, D-09107 Chemnitz

pko@informatik.tu-chemnitz.de

Abstract

The paper investigates two decision problems for multi-location logistic systems. The first one is the fleet-sizing-and-allocation problem, which is dealing with questions about the size of a vehicle fleet and the empty vehicle allocation. We briefly discuss the simulation optimisation approach. Next we consider so-called hub-and-spoke systems. Such a system consists of a single central, the hub, and M locations or spokes. The spokes generate a random demand for transportation orders. For a defined cost structure the decision problem is to find such a number of transportation units that minimizes for the steady state the expected cost per time unit. We formulate a solution algorithm and demonstrate it by an example.

Keywords: fleet sizing and allocation problem, simulation optimisation, hub and spoke systems

1. Introduction

Optimal design and control of dynamic stochastic systems is an actual and interesting field for theorists as well as practitioners. Among all application domains logistics is one of the most important. With respect to logistic systems exist very different formulations and models for optimisation purposes. Here we will concentrate on some aspects of how to use transportation resources in an optimal way. More explicitly, we will discuss solution approaches to the problem of the optimal size and allocation of transportation resources. In logistic literature we can find multitudinous models and solution methods dealing with that problem. As a rule optimisation in logistic systems leads to very complex decision problems. Thus it is not astonishing that corresponding mathematical models suffer from various simplification assumptions. For instance, most models assume deterministic known demand for transportation services. Common approaches are then linear and non-linear network programming models. More sophisticated models and solution methods investigate the case of stochastic demand. In that case inventory and queueing theory models are applied. However, in general no closed-form solutions are available. One can say that the majority of papers investigate algorithms that define an approximate solution for the discrete-time case with known demand and infinite transportation capacity. To come to more realistic models and solutions the broadest applicable approach is to build a simulation model. Simulation however is not an optimisation tool. What can we do?

In our paper we are dealing with this question. We will demonstrate and discuss by means of two decision problems the potential possibilities of both the simulation and the queueing theory approaches. Simulation we combine with an appropriate optimisation tool. Through such a simulation based optimisation we get an approach to find reasonable solutions for the fleet-sizing-and-allocation problem. This will be discussed in the second section. In section three we use queueing theory to investigate a multi-location system with hub-and-spoke structure. A brief resume finishes the paper.

2. The Fleet-Sizing-and-Allocation problem

Practical needs of modern transportation networks lead to a control problem, which in the literature is known as the fleet-sizing-and-allocation problem (FSAP). In fact the FSAP itself combines two closely connected problems – the fleet sizing problem, which is dealing with the question about the size of a vehicle fleet, and the empty vehicle allocation problem, i.e., how to redistribute empty vehicles among the locations of the network. The term vehicle applies to a single reusable unit, which realises a given kind of transportation service. Such units may be cars, trucks, railcars, airplanes, containers, material handling equipments and so on. A long time the two problems were investigated independently from each other (see e.g. /1/, /2/, /3/). Investigations, which consider the FSAP as a whole, can be found for instance in /4/ and /8/. Common approaches to solve the FSAP or at least one of its partial problems are linear and non-linear network programming models (e.g. /13/) as well as inventory and queueing theory models (cp. /4/, /6/). Dynamic programming models (see for instance /2/, /7/, /9/) are also used. From the complexity of the problem follows that we cannot expect to get closed-form solutions. Therefore most papers are dealing with algorithms for approximate solutions in the discrete-time case with known demand and infinite transportation capacity. Simulation optimisation is a promising approach to overcome these restrictions. It is applied in /8/, where the FSAP for M locations is investigated under the assumption of continuous time, Poisson arrivals of clients, given queueing capacities for waiting clients, homogeneous vehicles without failures and aging, infinite capacity to tranship vehicles between locations, and random transhipment times. The decision problem is to choose a fleet size $K\in\{1, 2, \dots\}$ and a reallocation policy $\pi\in\Pi$ such that for the steady state the expected cost per time unit will be minimised, i.e.

$$\text{(P)} \qquad C(K,\pi) \quad \to \quad \underset{(K,\pi)\in N\times\Pi}{\text{MIN}} \, !$$

To get an impression on goal function $C(K,\pi)$ we introduce the following notions:

- wv_i - the waiting cost per time unit and vehicle available (not used) in location i;
- wc_i - the waiting cost per time unit and client waiting for a vehicle in location i;
- r_i - the rejection cost per rejected (lost) client in location i;
- W_i - the mean number of waiting vehicles in location i;
- Q_i - the mean number of waiting clients in location i;
- R_i - the mean number of per time unit rejected clients in location i.

Then we can express the goal function as

$$C(K,\pi) = \sum_{i=1}^{M}\left[wv_i \cdot W_i + wc_i \cdot Q_i + r_i \cdot R_i\right] + T(K,\pi), \tag{2.1}$$

where $T(K, \pi)$ denotes for fleet size K and reallocation policy π the expected transhipment cost per time unit in steady state. However, we have no analytical expression for $T(K, \pi)$. The same holds for the performance measures W_i, Q_i and R_i, which also depend on K and π. Thus we can try either to simplify the problem and to get an approximate solution or to use a non-classical approach like simulation optimisation. A variant of the second way is realised in /8/. In a first step policy set Π is restricted to a single element π_0, which realises no reallocations at all. In this case term $T(K, \pi_0)$ can be omitted in (2.1) and we get a *fleet sizing problem without vehicle reallocation*, which is analysed by queueing network theory. On the basis of that analysis we restrict policy set Π in a second step to policies with heuristic ingenious structure. Thus let us consider the case $\Pi=\{\pi_0\}$ for the *no-backorder case*, i.e., demand which cannot be satisfied by an available vehicle will be lost and all Q_i vanish. We further assume that a vehicle rented in location i will be returned in location j with probability p_{ij} after a random using time, which has a Coxian distribution with expected value μ_{ij}, $i, j = 1..M$ (see e.g. /10/). These assumptions allow us to model the M-location system as a closed BCMP queuing network with K jobs, where the vehicles are the jobs (cp. /10/). That part of the network, which is related to location i, is shown in Figure1. With each node i are connected M artificial nodes (i, j) representing that a vehicle is rented at location i and will be returned in location $j = 1..M$. Figure 1 contains also the corresponding transition probabilities between nodes.

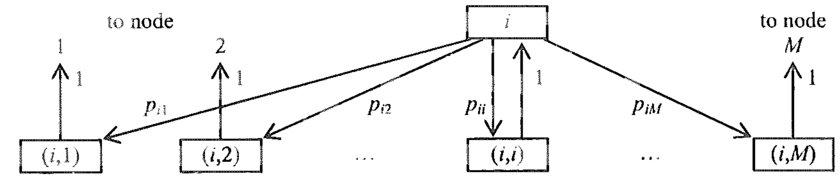


Figure 1. Part of the BCMP queueing network for the fleet sizing problem without reallocation

We notice that for closed queueing networks the steady state probabilities for being k vehicles in a node, $p_i(k)$ and $p_{(i,j)}(k)$, $k = 0..K$, $i, j = 1..M$, exist. With these probabilities the needed steady state performance measures can be calculated as $W_i = \sum_{k=1}^{K} k \cdot p_i(k)$ for the mean number of waiting vehicles in location i and $R_i = \lambda_i \cdot p_i(0)$ as the mean number of per time unit in location i rejected clients, $i = 1..M$, where λ_i denotes the arrival rate of clients to location i. Thus goal function (2.1) reduces to

$$C(K) := C(K,\pi_0) = \sum_{i=1}^{M}\left[wv_i \cdot W_i + r_i \cdot R_i\right] \tag{2.2}$$

and the initial optimisation problem (P) to problem

(P') {Define $K^* \in N$: $C(K^*) \le C(K)$ for all $K \in N$.}

Similar as in /8/ to solve (P') we can use the following greedy algorithm:

Initialisation: Set $K = 0$; compute $C(0) = \sum_{i=1}^{M} r_i \cdot \lambda_i$.

Iteration: REPEAT
$K := K+1$;
Compute $C(K)$
UNTIL $C(K) > C(K-1)$.

Output: $K^* := K-1$; $C(K^*) = C(K-1)$.

To compute for given K goal function $C(K)$ we need the performance measures W_i and R_i. For that we adapt the well-known *Mean Value Algorithm* (see /8/ and /10/). Regardless of all simplifications in (P') results on the fleet sizing problem without reallocations may be helpful for the FSAP. First, solution K^* of (P') is a certain indication for the search of the optimal fleet size for the FSAP. Next, $C(K^*)$ is an upper bound of the minimal cost for the FSAP. And third, realising the mean value algorithm we get the information which locations are surplus and which are shortage locations. That knowledge we can exploit for simulation optimisation. As already mentioned simulation optimisation combines simulation and optimisation. To realise that very promising idea we have to go through the following three stages:

1. Parameterisation of the optimising problem.
2. Choice of one or more optimisation approaches.
3. Realisation of simulation optimisation, i.e., iterate proposal of a solution by the optimiser with performance estimation of that solution by the simulator until a defined stopping criteria is fulfilled.

Stage 1 we accomplish through a restriction to reallocation policies with heuristic ingenious structure. Since we have a continuous-time model we concentrate on *situation-dependent* policies (see /8/). Situation-dependence means that a reallocation order will be created as a result of the occurrence of a defined event in the system, similar to an (s, Q)-policy in inventory systems. Two groups of such events are of interest – a *shortage*-event when a location has a shortage of vehicles, and a *surplus*-event when a location has too many vehicles. Therefore we define three policy sets: Π_1 for shortage-based, Π_2 for surplus-based, and Π_3 for shortage/surplus-based reallocation policies. A second attribute for a reallocation policy is the number of locations, which are taking part in a given reallocation. If all random variables in the system are assumed to be continuous then only following cases are relevant: single-supplier-single-receiver, single-supplier-multi-receiver, and multi-supplier-single-receiver case. As in /8/ we introduce now the notion of *location balance* B_i as the sum of vehicles waiting in location i or being in transport to location i minus the number of clients waiting in location i. The parameterisation of the optimising problem is realised in /8/ through the following parameterisation of the policies:

(1) Location i applies a (s_i, Q_i)-policy if it observes the following rule:
IF $B_i < s_i$ THEN place a demand order for $Q_i \geq 1$ vehicles.

(2) Location i applies a (S_i, q_i)-policy if it observes the following rule:
IF $B_i > S_i$ THEN place a delivery offer for $q_i \geq B_i - S_i$ vehicles.

(3) Location i applies a (s_i, Q_i, S_i, q_i)-policy if it observes the following rule:
IF $B_i < s_i$
THEN place a demand order for $Q_i \geq 1$ vehicles
ELSE IF $B_i > S_i$ THEN place a delivery offer for $Q_i \geq B_i - S_i$ vehicles.

In the currently simulator, available at Chemnitz University of Technology in a laboratory version, these policies are implemented for all three supplier-receiver cases.
As optimisation approaches corresponding to Stage 2 we prefer heuristic search methods like Genetic Algorithms or Simulated Annealing. Especially for the FSAP we have used a Genetic Algorithm (see /8/ for more details). In /8/ you will also find some results for a system with five locations. From these empirical investigations follow at least two conclusions, i.e., wise reallocation policies decrease the cost in orders of magnitude and the optimal solution for policy π_0 can be a very bad approximation for the FSAP solution.
To finish our excursion to simulation optimisation we want to point to some main pros and cons of that approach. Advantageous is that

1. problems respectively systems with practically arbitrary complexity can be handled;
2. the influence of a policy on the dynamics of the whole system as well as on particular elements of the system are observable;
3. estimation of individual cost or gain components is enabled.

The main disadvantages are the need for an appropriate simulator, which increases cost and time of the development process, and a considerable increase of computing time.
However, regardless of the great potential of simulation optimisation we believe generally that introducing some structure into an optimisation problem will lead to considerable improvements in the optimisation process. This means for instance to move from the FSAP for an arbitrary multi-location system to one with a defined and in reality existent structure. An example will be given in the following section.

3. Multi-location systems with hub-and-spoke structure

In the previous section we have seen that the general FSAP is too complex for an analytical solution. Now we want to investigate, what is possible if we assume some structure for the multi-location system, i.e., we concentrate on so-called hub-and-spoke systems (HS-systems). Let us assume that our HS-system consists of M+1 locations. Location 0 denotes a single hub and locations 1 to M the spokes. For instance, we may think location 0 as a central storage for a given product, whereas the locations 1 to M have to meet a random demand for that product. At the hub is stored an ample amount of product. Transports of product from the hub into the spokes have to be realised by transportation units (TU's) with given capacity. Altogether K

identical TU's are available, i.e. the fleet includes K vehicles. Spoke i generates a demand for a single TU in accordance with a Poisson process with parameter $\lambda_i > 0$, $i = 1..M$. TU's move from the hub to a spoke and return to the hub. The time for a trip from the hub to spoke i, for unloading and the return trip to the hub is assumed to be an exponentially distributed random variable with parameter $\mu_i > 0$. Let all random variables be independent. Further, let the hub serve transportation orders in accordance with first-come-first-served (FCFS) policy. If all TU's are on the trip arriving transportation orders will be queued. Following cost parts are considered:

$c > 0$ – fixed cost per TU and per time unit,
$w > 0$ – waiting cost per time unit and waiting transportation order,
$c_s > 0$ – cost per time unit per TU in use.

Now the decision problem is to choose such a number K^* of TU's, i.e., a fleet size K^*, which minimises for the steady state the average long-run cost per time unit.

Remark 3.1
With the above stated assumptions we can model the HS-system as an M/M/K/∞ queueing system[1] with arrival rate λ and service rate μ for a single server, where

$$\lambda = \sum_{i=1}^{M} \lambda_i \qquad \text{and} \qquad E(servicetime) = 1/\mu = \sum_{i=1}^{M} \frac{\lambda_i}{\lambda} \times \frac{1}{\mu_i}.$$

Subject to Remark 3.1 we consider now the HS-system as an M/M/K/∞ queue. From queueing theory we know that for the existence of the steady state regime the number K of servers must be greater than $a = \lambda / \mu$. For all $K > a$ we have following formulas for the different performance measures (see e.g. /10/):

(i) For the steady-state probabilities $p_n(K)$ that n jobs resp. clients are in the system it holds

$$p_0(K) = 1 / \left\{ \sum_{n=0}^{K} a^n / n! + a^{K+1} / [K!(K-a)] \right\} \tag{3.1}$$

and

$$p_n(K) = p_0(K) \cdot a^n \cdot \begin{cases} 1/n! & , n \leq K, \\ 1/(K! K^{n-K}), & n \geq K, \end{cases} \qquad n = 0, 1, 2, \ldots$$

(ii) The average number $Q(K)$ of waiting clients is given by

$$Q(K) = p_0(K) \times a^K / K! \times K \times a / (K-a)^2. \tag{3.2}$$

(iii) For the average waiting time W of a client holds

$$W = \lambda \cdot Q = \frac{a^K}{\mu \cdot (K-1)!(K-a)} \times \frac{1}{(K-a) \sum_{n=0}^{K-1} a^n / n! + a^K / (K-1)!}.$$

(iv) The average number of clients in the system can be computed as $L = Q + a$.
(v) The average number of busy servers is equal to E(number of busy servers) $= a = \lambda/\mu$.

[1] M/M/K/∞ queueing system means Poisson arrivals, exponentially distributed service times, K identical servers, and infinite waiting capacity.

To get a mathematical formulation of the decision problem above all we have to define the criterion function. Therefore let $C_{HS}(K)$ denote the average cost per time unit in the steady-state regime if there are K servers. With the assumed cost structure it holds that

$$C_{HS}(K) = K \cdot c + Q(K) \cdot w + a \cdot c_s \,. \tag{3.3}$$

Having in mind that for the existence of steady-state regime K should be greater than a (see Remark 3.1) we formalize the decision problem as optimisation problem

$$\text{(P-HS)} \qquad \begin{cases} C_{HS}(K) \rightarrow MIN! \\ K > \lambda / \mu \\ K \in N \end{cases}$$

To define an optimisation algorithm properties of $C_{HS}(K)$ with respect to K will be useful. If we could prove that $C_{HS}(K)$ is convex with respect to K, the design of an optimisation algorithm is straightforward. Since $K{\cdot}c$ is a linear function and a is a constant we have to consider $Q(K)$ only. In /5/ is proven the following

Theorem 1:
For the M/M/K/∞ system the performance measure $Q(\cdot)$ is a decreasing convex function of K for $K \geq a$.

With Theorem 1 the criterion function $C_{HS}(K)$ is convex with respect to the number K of servers. Now the validity of the following optimisation algorithm lies on hand.

Algorithm *optimal number of servers*

1. Initialisation: $K := \lfloor a \rfloor + 1$; C1 := $C_{HS}(K)$; C0 := MAX.

2. WHILE (C1 < C0) DO
   BEGIN
   $K := K+1$;
   C0:= C1;
   C1:= $C_{HS}(K)$
   END.

3. RETURN $K^* = K$-1 and $C_{HS}(K^*)$ = C0.

Within the algorithm we have to calculate for different server numbers K (fleet size values) function $C_{HS}(K)$ from (3.3). The most time consuming step is thereby the calculation of $p_0(K)$ and $Q(K)$ from (3.1) respectively (3.2). Its complexity is O(K). We remark that the notion $\lfloor a \rfloor$ means the greatest integer not exceeding a.

To demonstrate the algorithm we consider
Example 3.1:
Let $M = 5$, $c = 20$ EUR/day, $w = 500$ EUR/day, and $c_s = 100$ EUR/day. The arrival and service rates with measuring unit day^{-1} are given in Table 1. From the data in Table 1 and the formulas in Remark 3.1 we calculate $\lambda = 6$ day^{-1}, $\mu = 1$ day^{-1}. Since $a = \lambda/\mu = 6$, we need at

least 7 servers respectively TU's. Table 2 contains the results of the numerical computations. These results show the following:

1. The probability $p_0(K)$ that the system is empty is an increasing function of the server number K. From the formula for $p_0(K)$ we see that $\lim_{K\to\infty} p_0(K) = e^{-a} = e^{-6} = 0.00247875$.
2. The average number $Q(K)$ of waiting transportation orders in the steady-state regime is, as stated in Theorem 3.1, a convex function of the number K of TU's or servers.
3. The average cost per time unit in the steady-state regime if there are K servers, $C_{HS}(K)$, is a convex function of the number K of TU's or servers.
4. The optimal number of TU's or servers is equal to $K^* = 11$. For the given values of the cost parameters and rates underestimation of K^* is more dangerous than overestimation.
5. With increasing K function $C_{HS}(K)$ becomes a linear function with grade $c = 20$ EUR/day. This follows from the fact that $Q(K)$ vanishes for $K \to \infty$.

n	1	2	3	4	5
λ_n	0.3	1.2	0.6	2.4	1.5
μ_n	0.5	1.0	0.3	2.0	1.25

Table 1. Arrival and service rates for Example 3.1

K	$p_0(K)$	$Q(K)$	$C_{HS}(K)$ [EUR/day]
7	0.00157878	3.682981	1 841.4904
8	0.00214238	1.070945	1 295.4726
9	0.00235231	0.391962	975.9810
10	0.00243174	0.151949	875.9744
11	0.00246166	0.059066	849.5332
12	0.00247273	0.022474	851.2371
13	0.00247670	0.008269	864.1346
14	0.00247808	0.002924	881.4618

Table 2. Numerical results for Example 3.1

Obviously the presented model can be extended in various directions.

A) General distributions.

For the M/M/K/∞ system the interarrival times as well as the service times are exponentially distributed random variables. In case of arbitrary distribution functions we get a G/GI/K/∞ system. In /12/ is proven that for the G/GI/K/∞ system the performance measure $Q(\cdot)$ is a decreasing convex function of K for $K \geq a$. Thus the above formulated algorithm works again. However, for G/GI/K/∞ systems we have for Q only approximate formulas. Now let $A_i(\cdot)$ and $B_i(\cdot)$ denote the distribution function of the generation time for transportation orders in location i and the service time by the centre, respectively. We assume only that the first moments $m_1(A_i)$ and $m_1(B_i)$ are finite for all locations. If $\lambda_i = 1 / m_1(A_i)$ denotes the arrival intensity of transportation orders from i then we get a G/GI/K/∞ system by setting $\lambda = \sum_{i=1}^{M} \lambda_i$,

$A(t) = \sum_{i=1}^{M} \lambda_i / \lambda \cdot A_i(t),\ t \geq 0$, and $B(t) = \sum_{i=1}^{M} \lambda_i / \lambda \cdot B_i(t),\ t \geq 0$.

B) Finite capacity of the order queue.

We can assume that no backorders or only a finite number of backorders is possible. Then we get the lost-case models M/M/K/0 or G/GI/K/0 respectively the finite models M/M/S/L or G/GI/K/L.

C) Variations of the decision problem.
Other criterion functions as well as other constraints are possible.

D) Different from FCFS service disciplines.
The hub can serve waiting transportation orders in accordance with other service disciplines, e.g., in accordance with some priority rules.

E) Existence of extern transportation resources.
If it is possible to rent from outside the HS-system TU's we have to answer two questions – how much own TU's and how much rented TU's is optimal.

4. Resume

We have investigated two different formulations of the so-called fleet-sizing-and-allocation problem. The main conclusion is that in general only a combination of several approaches will lead to satisfying solutions. Thus for instance analytical approaches can give some information on properties of the goal function as well as on the structure of optimal or at least good solutions. On the other hand simulation optimisation can exploit that information for searching optimal values of the parameters of those policies. It is more promising to investigate problem classes with well-defined structure instead to deayl with general problem formulations. Finally, we want to point to the fact that various tools from computer science in connection with powerful computer technique open new horizons for solving realistic problems.

Acknowledgements

The author wants to thank an anonymous referee for very helpful comments on a first version of this paper.

References

/1/ Beaujon, G.J.; Turnquist, M.A.: A Model for Fleet Sizing and Vehicle Allocation. *Transportation Sci.*, 25, 19-45, 1991

/2/ Cheung, R.K.; Powell, W.B.: An Algorithm for Multistage Dynamic Networks with Random Arc Capacities, with an Application to Dynamic Fleet Management. *Oper. Res.*, 44, 951-963, 1996

/3/ Crainic, T.G.: Service network design in freight transportation. *European Journal of Operational Research*, 122 , 272-288, 2000

/4/ Du, Y.; Hall, R.: Fleet Sizing and Empty Equipment Redistribution for Center-Terminal Transportation Networks. *Man. Sci.*, 43, 145-157, 1997

/5/ Dyer, M.E.; Proll, L.G.: On the validity of marginal analysis for allocating servers in M/M/c queues. *Management Science*, v.23, No. 9, 1019-1022, 1977

/6/ Koenigsberg, E.; Lam, R.C.: Cyclic Queue Models of Fleet Operations. *Oper. Res.*, 24, 516-529, 1976

/7/ Köchel,P.: An Approximate Solution for the Fleet Sizing and Allocation Problem. *Promet-Traffic-Traffico*, 9, 195-200, 1997

/8/ Köchel,P.; Kunze, S., Nieländer, U.: Optimal control of a distributed service system with moving resources: Application to the fleet sizing and allocation problem. *International Journal of Production Economics*, vol. 81-82, 443-459, 2003

/9/ Powell, W.B.; Carvalho, T.A.: Dynamic Control of Logistics Queueing Networks for Large-Scale Fleet Management. *Transportation Sci.*, 32, 90-109, 1998

/10/ Trivedi, K.S.: Probability & Statistics with Reliability, Queueing and Computer Science Applications. Prentice Hall, 1982

/12/ Weber, R.R.: On the marginal benefit of adding servers to G/GI/m queues. *Management Science*, v.26, No. 9, 946-951, 1980

/13/ Wu, P.; Hartman, J.C.; Wilson, G.R.: Fleet Sizing in the Truck Rental Industry. *Working Report No. 98W-009*, Lehigh University, 1999

TRAVELING SALESMAN PROBLEM AT THE POST OF SLOVENIA

Andrej Lisec
Post of Slovenia, Ltd., Business Unit Postal Logistics Center Ljubljana
Cesta v Mestni log 81, SI - 1002 Ljubljana, Slovenia
E-mail: andrej.lisec@posta.si

Marija Bogataj
University of Ljubljana, Faculty of Economics
Kardeljeva ploščad 17, SI – 1000 Ljubljana, Slovenia
E-mail: marija.bogataj@guest.arnes.si

ABSTRACT

In this paper the theory and methodology of Roberto Wolfler Calvo (2000), described in the article *A New Heuristic for the Traveling Salesman Problem with Time Windows*, are presented and their application to the Post of Slovenia is made. The reduction of logistic costs using this approach is very high. Suggestions for further improvements are given.

KEY WORDS: Postal system, Traveling Salesman Problem, Postal Logistics Center, Regional Parcel Center, Parcel Post

1. INTRODUCTION

Central European regions are improving their postal services. The Post of Slovenia, Ltd. would also like to follow these efforts. This paper presents an approach to the spatial optimization of postal services, particularly as applicable to the Post of Slovenia in respect to the Traveling Salesman Problem.

The Traveling Salesman Problem with Time Windows - TSPTW developed by Wolfler Calvo [10] is described and applied to the Slovenian postal network by a set of fully interconnected posts (nodes), each one characterized by a time interval or a time window. One of these nodes represents a Parcel Post (depot), where the parcels from all local post services are collected and distributed. Each arc has an associated travel time and cost. The problem consists of finding a minimum - cost Hamiltonian tour visiting each node during its time window. The tour starts and ends on the lower level of the Parcel Post. The TSPTW is a useful generalization of the classic Traveling Salesman Problem - TSP and it has applications in many vital sequencing and distribution systems.

This problem is only a part of logistic problems of Slovenian postal network when we wish to set up an optimal allocation of modern Postal Logistics Centers in given 4-level hierarchical structure of regional central places in Slovenia. A verified model of spatial optimization by Bruns [1], developed for the Post of Switzerland, has been previously studied and has been partly applied and extended to evaluate Slovenia's second Postal Logistics Center Maribor. The previous analysis is based on the paper of Bruns [1]: Restructuring of Swiss Parcel Delivery Services, where 0 - 1 programming has been used.

So far two Postal Logistics Centers - PLCs have been opened up in Slovenia in the last nine years. We have made the evaluation of logistic costs considering two instead of one PLC and evaluated whether the two existing Postal Logistics Centers in Slovenia already have optimal macro - locations of required capacity by using the model of Bruns [1]. Simulations demonstrated that the PLC Maribor in addition to the PLC Ljubljana is acceptable if the variable costs of service from the PLC Maribor are lower or at least the same as the variable cost of the PLC Ljubljana and if the costs of services of both do not exceed a certain critical value. On the lower level Parcel Posts (with short time depots) have to be established. They do not exist so far.

Not only in Switzerland but also in other Central European Countries, the problem of postal hub location has been presented in some papers as being vital for efficient postal logistics. Wasner and Zapfel [9] have described the hub transportation network for parcel delivery service in Austria in the paper *An integrated multi-depot hub-location vehicle routing model for network planning of parcel service*. According to them the problem of several parcel posts, their location and their coverage of area by the post connected in cycles is the basic problem on the lowest level.

2. TRAVELING SALESMAN METHODOLOGY EMBEDDED IN HUB LOCATION PROBLEMS

Postal hub is a special type of central facility, which can act as transshipment point in postal transportation systems with many origins and destinations.

E-commerce influences the increase of parcel services and the number of parcels all over the world. How can this demand be addressed properly? New technologies and modern organization in postal services could improve the design of postal systems and especially the logistics of these systems.

By using the basic approaches of graph theory we improve the approach to the hub location problem with the model named Multi hierarchical hub location problem and also combine in with the vehicle routing problem on a special way.

Given a collection of Posts and the cost of travel between each pair of them, the Traveling Salesman Problem is to find the cheapest way of visiting all of the activity nodes and returning to the starting point. In the standard version, the travel costs are symmetric in the sense that traveling from Post i to Post j. The basic problems of companies that are dealing with parcels are:

1. To determine the optimal number of Posts as hubs patronizing separate areas (Regional Parcel Centers), and to determine the location and size of hubs;
2. To determine optimal fixed locations of Postal Logistics Centers, patronizing Regional Parcel Centers, location of Regional Parcel Centers, patronizing Parcel Posts and Parcel Posts patronizing area of inhabitants. In these posts we need to achieve:an effective collection and sorting of parcels;
   - a reliable delivery;
   - an effective transport.

The criterion here is that the total sum of logistic costs in this service of parcels should be minimal, often under certain capacity constraints.

The Posts on the level of local communities, patronizing a certain area, have to be assigned to the proper Parcel Post.

The transport network has to be built, which connect Posts, Parcel Posts, Regional Parcel Centers and Postal Logistics Centers, where the costs of daily transshipment of parcels would be minimal.

These problems are well discussed and overviewed in the theory and practical applications of Ebery [3], O'Kelly [7], Campbell [2], Wasner [9], Bruns [1], Marin [6], Ernst [4], Skorin – Kapov [8] and Gendreau [5].

Successful parcel service application of postal services model was developed by the Swiss post, described in Bruns [2], which deals with restructuring of the parcel service

network, with choosing transshipment points among the nodes in the network. They used a discrete facility location model.

Austrians Wasner and Zapfel [9] describe why optimal design of depot and hub transportation networks for parcel service providers makes it necessary to develop a generalized hub location and vehicle routing model.

In this article the focus is on the Traveling Salesman Problem only on the lower level. The results by Wolfler Calvo [10] are being used. Let $G=(N, A)$ be a graph, where $N=\{0,1,...,n\}$ is the set of n nodes plus the depot $\{0\}$ and the arc set is $A=\{(i,j): i,j\in N, i \neq j\}$. The time window is represented by $[e_i, l_i]$ and, to each arc $(i, j)\in A$, there is an associated cost (c_{ij}) and travel time (t_{ij}). The travel time matrix is symmetric ($t_{ij}=t_{ji}$) and with strictly positive entries ($t_{ij}>0$). Here e_i is the earliest time for delivery (or collection) and l_i the latest time for delivery (or collection).

A possible formulation of the TSPTW is as follows:

$$\min \sum_{(i,j)\in A} c_{ij} x_{ij} \tag{1}$$

subject to:

$$\sum_{J\in N} x_{ij} = 1 \qquad \forall i \in N \tag{2}$$

$$\sum_{J\in N} x_{ji} = 1 \qquad \forall i \in N \tag{3}$$

$$\text{if } x_{ij} = 1 \text{ then } p_i + t_{ij} \leq p_j \quad \forall (i,j)\in A \tag{4}$$

$$e_i \leq p_i \leq l_i \qquad \forall i \in N \tag{5}$$

$$x_{ij} \in \{0,1\} \qquad \forall (i,j)\in A \tag{6}$$

The variable $x_{ij}=1$ means that arc (i, j) is used, otherwise $x_{ij}=0$. The variable p_i represents the departure time from node v_i.

Equations 2 and 3 define an assignment problem of dimension $(n+1)$. Equations 4 and 5 represent the time window constraints.

3. APPLICATION TO THE POST OF SLOVENIA

Our application is based on an analysis of Slovenia's postal service. The country is divided according to NUTS system on NUTS 2 and NUTS 3 regions, patronizing 192 local areas.

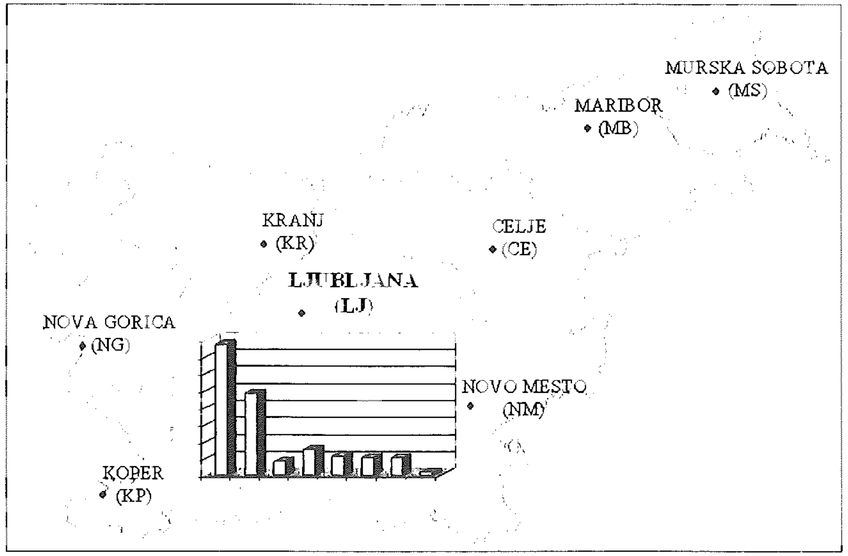
Spatial hierarchy of postal services is more or less embedded in geographically and politically determined regionalization.

We used the data from the Post of Slovenia and optimized the hierarchical structure for picking processing and delivery of parcels from Post to Post.

The Posts are allocated to Parcel Post on the criteria of minimum number of kilometers and especially on the experience of daily transport from Post to Post to Postal Logistics Center. We took the experience of postal workers in Business Unit Postal Logistics Center Ljubljana into consideration and used the method of the Traveling Salesman Problem.

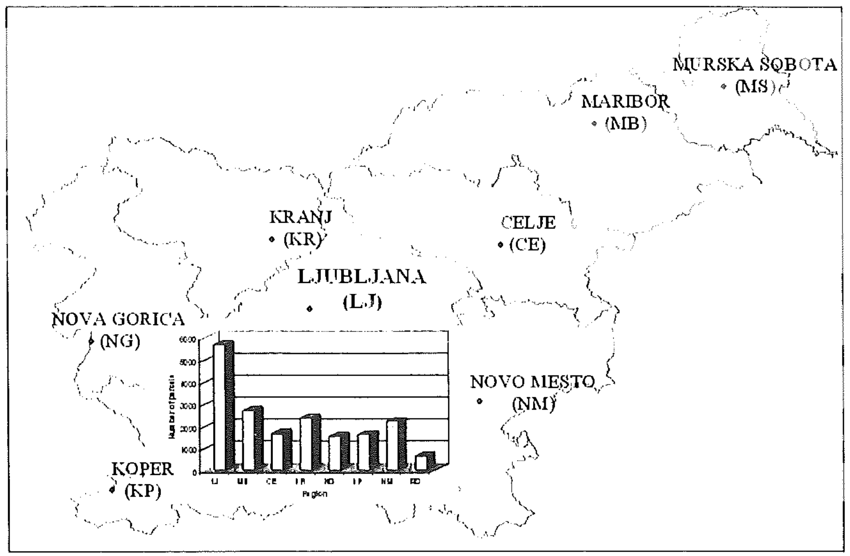
In Slovenia there are 555 posts in year 2005. The flows of parcels are directed from Post to Post until the truck is fully loaded and then sent to PLC Ljubljana or PLC Maribor and back. Potential Regional Parcel Centers are not open yet, but could be opened in Regional Business Units. They will patronize Parcel Post.

According to the results of our research, proper capacity and allocation of Regional Parcel Centers and Parcel Posts should be carried out. In this case the hierarchy for Parcel Posts would be as follows: the Parcel Posts should patronize twenty or less posts.



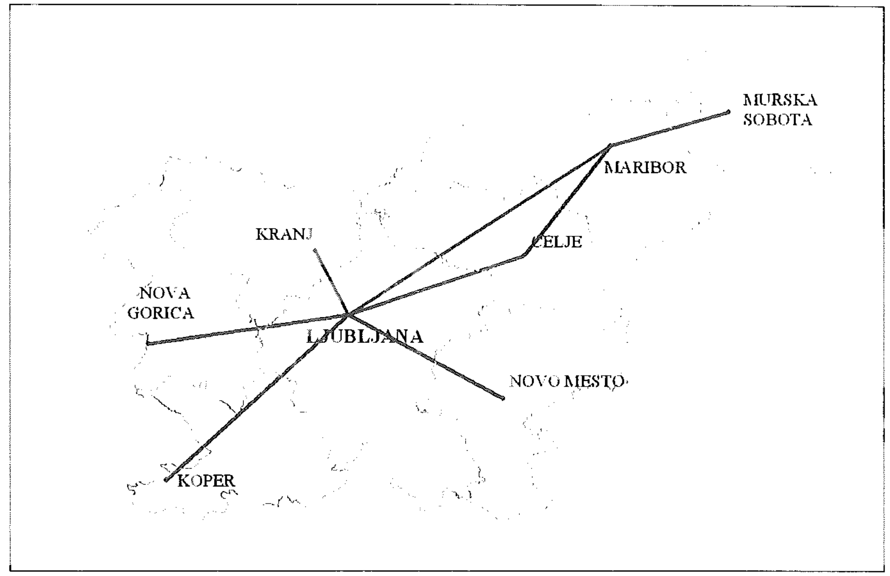
Picture 1: The average daily number of collected parcels from regions to PLC Ljubljana

Picture 1 presents the average daily number of collected parcels in PLC Ljubljana for all Business Units and Picture 2 the average daily number of delivered parcels from PLC Ljubljana.



Picture 2: The average daily number of delivered parcels from PLC Ljubljana to regions

Picture 3 shows the structure of parcel flows in the Post of Slovenia.



Picture 3: The structure of parcel flows

Posts start to operate approximately two hours before the post offices open and close two hours after the post offices do so.

Our application is made for the covering service area of the Postal Logistics Center Ljubljana.

We determined the optimal cycles for areas covered by the Postal Logistics Center Ljubljana: 4 Parcel Posts patronized by the post of Kranj, 3 Parcel Posts patronized by the post of Nova Gorica, 4 Parcel Posts in the Koper and 5 Parcel Posts patronized by the regional post of Novo mesto (Picture 4).

Our optimization is applied only to the area patronized by the Postal Logistics Center Ljubljana because we have experience here and all the necessary data also only from the PLC Ljubljana. We propose the use of the same optimization methods for the Postal Logistics Center Maribor as well.

By using the Traveling Salesman methodology, the data on distances between Posts, the statistics of parcels from the years of 2003 and 2004 and intensity of flows between posts from the year 2005 we got an optimal new transportation network connecting the post and Parcel Posts patronizing them with the criteria of 24 hour time window delivery.

Optimization is focused on the minimal length of the sum of cycles by using the shortest paths between Posts and Parcel Post.

The results are presented in the sums of lengths of cycles (in kilometers) for all Posts, which are patronized by the Parcel Post areas. This is the only possible way to compare the difference of daily length of cycles between the new and the old transport structure.

Here each Post is visited more than two times per day from patronized Parcel Post.

Our results, presented in Table 1 show sums of the lengths of cycles in the previous regime, but Table 2 presents new optimal sum of the lengths of daily cycles.

Table 1: Previous sums of length of all cycles in kilometers

old	Ljubljana	Kranj	Nova Gorica	Koper	Novo mesto	**sum**
truck	3129	1171	2047	1688	2005	***10040***
van	1776	424	367	233	597	***3397***
car	435	444	374	324	672	***2249***
sum	***5340***	***2039***	***2788***	***2245***	***3274***	**15686**

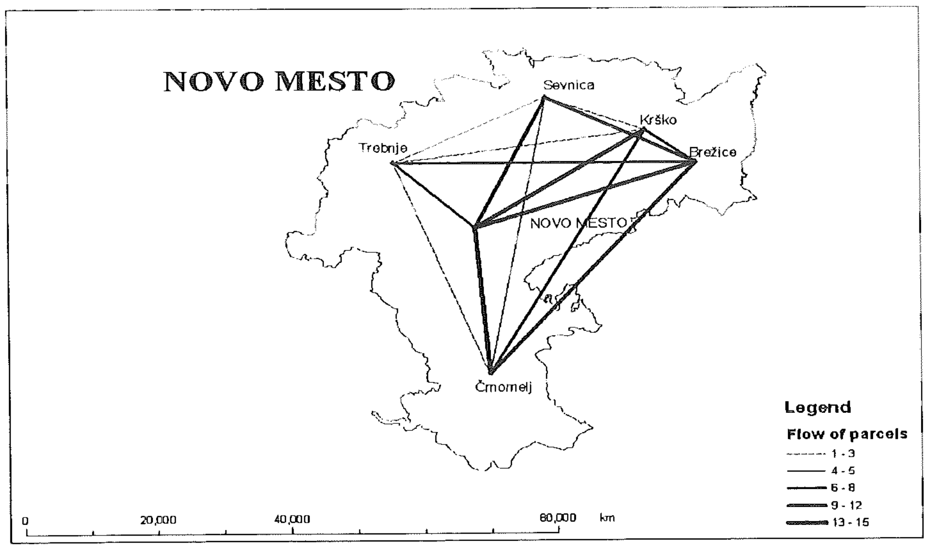
Table 2: New sums of length of all cycles – improved solutions in kilometers

new	Ljubljana	Kranj	Nova Gorica	Koper	Novo mesto	**sum**
truck	807	602	878	769	626	***3682***
van	658	538	534	135	672	***2537***
car	1406	262	665	888	665	***3886***
sum	***2871***	***1402***	***2077***	***1792***	***1963***	**10105**

The transport structure of the Business Unit Novo mesto is presented a bit closer here. It encompasses the Parcel Post Sevnica, which has the following associated distance matrix:

	Sevnica	Bostanj	Studenec	Radece	Zidani Mo.	Loka pri Z.
Sevnica	/	2	15	/	/	12
Bostanj	2	/	13	12	/	/
Studenec	15	13	/	/	/	/
Radece	/	12	/	/	2	2
Zidani Mo.	/	/	/	2	/	4
Loka pri Z.	12	/	/	2	4	/

Five iterations have been made. The value of the objective function is 58. The sequence of arcs on this cycle is the following (constrains in time windows are based on the average speed of 70 km/hour): Sevnica – Bostanj (2) – Studenec (13) – Radece (25) - Zidani Most (2) - Loka pri Zidanem Mostu (4) – Sevnica (12). The total distance is 58 km.



Picture 4: Regional Parcel Center Novo mesto and their Parcel Posts

The transportation costs are reduced (because of reducing the total number of kilometers per day) from 15686 km (which is the situation at present) to 10105 km. This will be achieved when Parcel Posts will patronize the sequence of Posts on the local area. The difference is 5581 km. If we also consider fixed and variable cost, then the costs are reduced by 30 percent in total.

5. CONCLUSIONS AND FURTHER RESEARCH

In this paper the use of the Traveling Salesman Problem method has been described and applied to the Post of Slovenia, to the area patronized by the Postal Logistics Center Ljubljana. As presented, by solving the real problem of the Postal Logistics Center Ljubljana, the number of kilometers inside prescribed time window can be drastically reduced.

Our further research will be based on the assumption that the areas, which are patronized by the PLC Ljubljana, and the complementary areas, patronized by the PLC Maribor, are not given in advance. The same relaxation will be given for Regional Parcel Centers. By using this approach, the Traveling Salesman Problem in the area of the Postal Logistics Center Maribor also has to be solved for all belonging parcel posts.

References

[1] Bruns, A., et al. (2000). Restructuring of Swiss Parcel Delivery Services, Operations Research – Spektrum, p. 285 - 302.

[2] Campbel, J. F. (1994). Integer programming formulations of discrete hub location problems, European Journal of Operation Research, 72, p. 387 - 405.

[3] Ebery, J. et al. (2000). The capacitated multiple allocation hub location problem: Formulations and algorithms, European Journal of Operational Research, 120, p. 614 - 631.

[4] Ernst, A. T., Krishnamoorthy, M. (1996). Efficient algorithms for the incapacitated single allocation p – hub median problem, Location Science, 4/3, p. 139 – 154.

[5] Genreau, M., Laporte G., Seguin, R. (1996): Stochastic vehicle routing. European Journal of Operational Research, 88, p. 3 – 12.

[6] Marin, A. (2005): Formulating and solving splittable capacitated multiple allocation hub location problems. Computers and Operations Research, 32/12, p. 3093 - 3109.

[7] O'Kelly, M. E. (1987): A quadratic integer program for the location of interacting hub facilities, European Journal of Operational Research, 32, p. 393 – 404.

[8] Skorin - Kapov D., Skorin - Kapov J., O'Kelly M. (1996): Tight linear programming relaxations of incapacitated p - hub median problems, European Journal of Operational Research, 94, 1996, p. 582 – 593.

[9] Wasner, M., Zapfel, G. (2004). An integrated multi - depot hub - location vehicle routing model for network planning of parcel service, Production Economics, 90/3, p. 403 - 419.

[10] Wolfler Calvo, R. (2000). A New Heuristic for the Traveling Salesman Problem with Time windows, Transportation Science, 34/1, p. 113 – 124.

Dealing with the fluctuations of the travel times in real-world distribution problems

Ana Osvald[a], Raffaele Pesenti[b], Walter Ukovich[c]

[a] E-mail: ana.osvald@bayern-mail.de

[b] DINFO – University of Palermo
Viale delle Scienze
I-90128 Palermo, Italy
Phone: (+39) 091 6598043 Fax: (+39) 091 6598043 E-mail: pesenti@unipa.it

[c] DEEI and
University of Trieste
Via A. Valerio, 10
I-34127 Trieste, Italy
Phone: (+39) 040 558 7848 Fax: (+39) 040 558 3460 E-mail: ukovich@units.it

Abstract

In real-life routing problems with time constraints, it is very important to consider the fluctuations of the travel times, in order to find good solutions which estimate the correct customers' delivery time. In this paper, we present a method to solve the Vehicle Routing Problem with Time Windows (VRPTW) where the travel times between two locations depends not only on the distance, but also on the time of the day. The paper proposes a solution approach to this problem by first providing a simplified estimation of the time-dependent link travel times and then using an innovative heuristic which solves the related routing problem.

Keywords: Routing, Distribution, Logistics, VRP, Time-dependent travel times.

1. Introduction

Nowadays, intelligent planning offers potential savings in costs related to stock and distribution of goods. In particular, the Vehicle Routing Problem with Time Windows (VRPTW) [11] considers the classical distribution problem for the management of a fleet of vehicles where a route for each vehicle of the fleet must be found in order to minimize the total distribution cost, to satisfy each customer's demand and to consider all operational constraints (including delivery time windows, capacity of the vehicle).

A lot of research effort has been devoted to various aspects of the VRPTW and many techniques to solve this problem have been presented in the literature. Unfortunately, in many cases these techniques do not provide useful and acceptable solutions to logistics companies, because they do not take into account the fluctuations of the travel times due to the variable traffic congestion during the day. To achieve always useful solutions, it is necessary to include the additional information on the travel time variations into road network.

Problems where a time-dependent modelisation of the travel times is used are called Vehicle Routing Problems with Time Windows and Time Dependent travel times (VRPTWTD). So far, they have not been much considered in the literature.

In this paper, we present a solution approach for the VRPTWTD suitable for logistic companies; hence the solution shall be always feasible and shall be obtained in a short time of about few minutes. We achieve it by first identifying a suitable model for representing the

time-dependent fluctuations of travel times and then solving the VRPTWTD with soft time windows [11].
In Section 2 we give a brief literature review of the VRPs with Time-Dependent travel times and show different approaches used to represent the fluctuations of the travel times. In Section 3 we present the developed method to estimate the Time-Dependent Origin-Destination (TD-OD) matrix and give an overview of the developed heuristic for the solution of the VRPTWTD. Section 4 reports the results of the developed approach on real data. The advantage of the presented approach and possible future research are presented in Section 5.

2. Literature review

Malandraki and Dial [6] introduced first in 1992 the time dependent Traveling Salesman Problem (TSP). Bentner *et al.* [1] considered for the TSP a zone in the city center with traffic jams in the afternoon and showed how the simulated annealing and threshold accepting algorithms are able to handle such time-dependent problems.
Park *et al.* [8] presented the Time-dependent Vehicle Scheduling Problem (TDVSP) in which the travel speed between two locations depends on the passing areas and the time of the day. They proposed a model for estimating the time varying travel speed.
Ichoua *at al.* [3] proposeed a time-dependent model for the VRPTW. Their model is based on time-dependent travel speeds and satisfies the First-In-First-Out (FIFO) property. They extend the Tabu Search heuristic developed by Taillard *et al.* [10] to solve the problem. They tested the algorithm in static and dynamic (number of services are not known completely ahead of time) environments and showed that the time-dependent model provides substantial improvements over a model based on fixed travel times. In that work, there is no mention of any experiment on real world problems.
All of the mentioned works use the Euclidean distances to estimate the shortest, time-dependent travel times and so also the TD-OD matrix. In the real-world problems, this approximation is usually not acceptable, and real distances and real travel times must be used.
In the literature, several representations of the link travel times and related methods to estimate the shortest path are used to cope with fluctuations. The three most common representations of the link travel times fluctuations, usually used in pre-trip planning, where all data are supposed to be known before the routing process begins, are as follows:

- Time-dependent travel times; the travel time $d_{ij}(t)$ from vertex i to vertex j is a function of the departing time t from vertex i. An example is presented in Figure 1.
- Stochastic times; travel times d_{ij} are random variables.
- Stochastic and time-dependent times; the link travel times are considered a continuous time stochastic processes, i.e., random variables and with probability distributions which are functions of the time of the day.

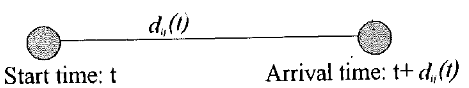


Figure 1: An example of Time-dependent representation of link travel times

The time-dependent representation of the travel time fluctuations are generally used to model time-of-day congestion effects and are indicated for short term planning [4], to which usually also the VRPTW belongs. Since we have to deal with this type of problems, we focus our attention on the time-dependent representation of the travel times.

Using this representation, it is possible in a second time to improve the calculated routes in real-time by acquiring in real-time the actual travel times and partially recalculate the routes.

3. Estimation of the time-dependent link travel time and the heuristic approach for the VRPTWTD

We split the solution of the real world VRPTWTD into two steps. Initially we estimate the TD-OD matrix and then we solve the VRPTWTD. An outline of the whole solution approach is presented in Figure 2.

First step: Estimation of TD-OD matrix

Extraction from the digital map of the road network on which the customers. are defined

Estimate TD-OD matrix, represented by 8 constant matrixes:
- Classify the arcs of the street network in 24 classes, according to their legal speed limit and traffic congestion.
- Divide a day into four time periods and to each period and for each class assign a coefficient of variation of the speed (four different time periods are used, since during the day there are usually two periods of rush hours). For each period, consider a constant average speed (so the speed is a step function of time).
- Calculate the shortest paths between all customers defined on network for 8 given start times (one at the beginning and one in the middle of each of the four time periods), using an algorithm based on the FSM algorithm and presented in [9].

⇓

Second step: Solution of VRPTWTD

Solve the Vehicle Routing Problem with Time Windows and with Time Dependent travel times with a heuristic approach:
Repeat for different Start Customers and different insertion costs
Find a feasible tour:
Solve classical VRPTW with an average travel time and travel distance
Correction Tour
end Repeat
Select the three best tours
For each tour
Tour Reduction (reduce the number of tours)
Improvement of the solution with TabuSearch
end For

Figure 2: An overview of the solution approach

The estimation of the TD-OD matrix usually requires a lot of effort, since nowadays there are very few specialized centers that manage a detailed traffic information available to the transportation companies. To represent the TD-OD matrix we use 8 constant matrixes. The travel time is in this way a step function of time. To assure that the final solution of the VRPTWTD satisfies the FIFO property it is necessary to have continuous travel times. To this end, we use modified (time continuous) travel times in the Second step of the solution approach.
To solve the VRPTWTD we have developed a heuristic that initially generates an initial solution using a constructive heuristic and then uses an improvement heuristic based on the

Tabu Search to furthermore improve the solution. The outline of this Second step is represented in Figure 2.
Initially a sequential constructive heuristic that solve a classical VRPTW is used. In this part, the average travel times and the average travel distances calculated on the TD-OD matrix are used. Then, a correction procedure is used in order to take into account the actual time dependencies of the travel times. In this procedure, the customers' delivery times are recalculated using time-dependent travel times. When necessary, some customers are relocated in the tour, or removed from it in order to make the tour feasible. The removed customers are, when possible, reinserted in other tours. Otherwise, new tour or tours are initialized for the removed and not reinserted customers.
To develop the improvement heuristic based on the Tabu Search we have used EasyLocal++, an object-oriented framework for the design and the analysis of local-search algorithms [2]. The abstract classes that compose the framework specify and implement the invariant part of the algorithm. The user can so focalize on the problem-dependent part.
In the improvement heuristic we have used the Cross Exchange technique for the neighborhood generation.
A detailed description of the whole solution approach is given in [7].

4. Performance of the proposed approach and computational results

In this section we present the performance of the proposed solution approach for the customers defined on a real road network. In the tests presented in following, three sets of 30, 50 and 100 customers, have been defined on a restricted map of the Friuli Venezia Giulia region with 50.703 geo-referenced nodes. We calculate the TD-OD matrix for each instance and solve the problem with two different approaches:

- constant travel times and travel distances, equal to the average travel times calculated on TD-OD matrixes of travel time and travel distance,
- time-dependent travel times.

	30 customers	50 customers	100 customers
Time independent approach			
Distance	245 km	376 km	859 km
Calculated time	438 min	726 min	1.326 min
Calculated lateness	42 min	75 min	351 min
Effective time	508 min	867 min	1.734 min
Effective lateness	239 min	806 min	894 min
Infeasibility	20%	25%	12,5%
Time dependent approach			
Distance	249 km	383 km	887 km
Time	493 min	842 min	1.693 min
Lateness	44 min	89 min	347 min

Table 1: Results of the VRPTWTD

Table 1 shows the results achieved with both approaches. In the firs part, we show the total distance of all routes (Distance), the calculated total travel time (Calculated time) and the calculated total lateness (Calculated lateness). The results obtained in the time independent environment do not always give a useful and feasible solution for the real-world problems. We have recalculated the routes, i.e. we have considered the sequences of customers and the

start times achieved in the time independent approach and have recalculated the delivery time at each customer, using the times given by the TD-OD matrix.
We observe that the recalculated time (Effective time) and the recalculated lateness (Effective lateness) are higher than the Calculated time and the Calculated lateness. Above all, we notice also that some of the calculated tours are not feasible anymore, since some depot's time windows turn out to be violated (in the problems with soft time windows, the time window of the depot remains hard).
The second part of the table reports the results achieved with the method presented in Section 3, that takes into account the time-dependency of the travel times.

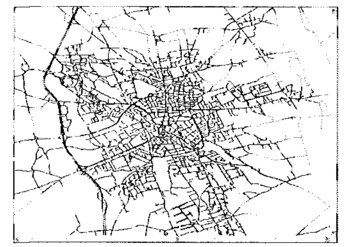


Figure 3: Time independent shortest path

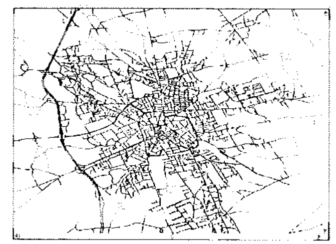


Figure 4: Time-dependent shortest path (Scenario I)

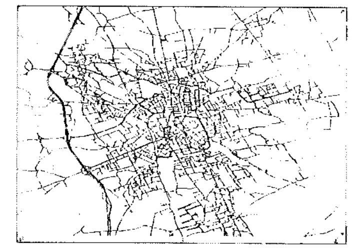


Figure 5: Time-dependent shortest path (Scenario II)

We observe that the time-dependent model provides significant improvements over the time-independent model: we have a small improvement of the total traveled time and a significant improvement of the total lateness. At last, using the time-dependent model, we achieve always useful and feasible solutions.
The cumulative computational time for the instances of 100 customers is about 105 seconds (45 for the estimation of the TD-OD matrix and 60 for the solution of the VRPTWTD) using a Pentium IV, 1.6GHz, 256MB PC.
To show how the time-dependency influences the travel times and the travel distances between the customers, we present in Figures 3-5 three shortest paths (obtained minimizing the travel time) between the same pair of customers, for three different start times. To minimize the total travel time of the path on the time dependent network, it is sometimes more useful to chose longer, but less congested route. Using the path achieved in the time-independent environment with the start time in Figure 5, we have found that the travel time lasts 23 min 26 s, about 3 min more than path in Figure 5.

	Travel time	Travel distance
Figure 4	9 min 5 s	9.979 m
Figure 5	15 min 13 s	10.148 m
Figure 6	20 min 43 s	12.474 m

Table 2: Travel time and travel distance between two points in different scenarios

5. Conclusions and further research

The results of experiments defined on real road networks show that the chosen time-dependent model provides significant improvements over the model with time-independent travel times. Above all, the developed solution approach for the VRPTWTD always produces a feasible solution, a crucial condition in the solution of the real-world distribution problems.

The computational time for the real-world distribution problems of up to 100 customers, defined on a restricted map as the one of the Friuli Venezia Giulia region, used in the tests in Section 4, is less than two minutes. Some weakness has been noticed for some problems defined on extended networks where the computational time, necessary to estimate the TD-OD matrix for problems of about 100 customers, is about 15 minutes. To speed up the calculation of the final solution, it is necessary to use alternative methods to calculate the TD-OD matrix for these problems. A useful method that allows in the standard shortest path problem to reduce the computational time of about 40-60%, is to use straight-line distance as a lower bound on the distance between any two nodes in the network [5]. This method would be useful to implement also for the time-dependent network, in order to reduce the computational time for the estimation of the TD-OD matrix on large networks.

Literature

1. Bentner J., Bauer G., Obermair G. M., Morgenstern I., "Optimization of the time-dependent traveling salesman problem with Monte Carlo methods", Physical review E, Volume 64, 036701 ,2001
2. Di Gaspero L., Schaerf A. "Writing local search algorithms using EASYLOCAL++". In S. Voß and D.L. Woodruff, editors, Optimization Software Class Libraries, OR/CS. Kluwer Academic Publishers, 2001.
3. Ichoua S., Gendreau M., Potvin J. Y., "Vehicle Dispatching With Time-Dependent Travel Times", European Journal of Operational Research, pp. 379-396, 2003
4. Kalidas A., "Estimation and Prediction of Time-Dependent Origin-Destination Flow", Ph.D. Thesis, Massachusetts Institute of Technology, 1996
5. Lysgaard J., "A two-phase shortest path algorithm for networks with node coordinates", European Journal of Operational Research, pp. 368-374, 1995
6. Malandraki C., Daskin M.S., "Time Dependent Vehicle Routing Prob-lems: Formulations, Properties and Heuristic Algorithms", Transportation Science 26 (3), pp. 185-200, 1992.
7. Osvald A., "The Vehicle Routing Problem with Time Windows and Time Dependent travel times: a heuristic approach", Ph.D. thesis, Università degli studi di Trieste, 2005
8. Park , Song, "Vehicle Scheduling Problems with Time-Varying Speed", Computers & Industrial Engineering, pp. 853-856, 1997
9. Powell W.B., Jaillet P., Odoni A., "Stochastic and Dynamic Networks and Routing", In Ball M., Maganti T., Monma C. and Nemhauser G., editors, Network Routing, North-Holland, 1995
10. Sung K., Bell M. G.H., Seong M., Park S., "Shortest paths in a network with time-dependent Flow speeds", European Journal of Operational Research, vol. 121, pp. 32-39, 2000
11. Taillard E., Badeau P., Gendreau M., Guertin F,. Potvin J. Y., "A Tabu Search Heuristic for the Vehicle Routing Problem with Soft Time Windows", Transportation Science 31, pp. 170-186, 1997.
12. Toth P., Vigo D., "The Vehicle Routing Problem", SIAM, 2002.

ALLOCATION OF SPECIALLY EQUIPPED FIRE BRIGADE UNITS: A CASE STUDY FOR INTERVENTION IN TUNNELS ON SLOVENIAN ROAD NETWORK

Dejan Paliska, Daša Fabjan
Faculty of Maritime Studies and Transportation
Pot pomorščakov 4, 6320 Portorož
Samo Drobne
Faculty of Geodesy and Civil Engineering
Jamova 2, 1000 Ljubljana

Abstract

This paper presents the application of the MCLP (Maximal Covering Location Problem) location-allocation model in a GIS (Geographic Information System) environment used to optimize the number and locations of fire brigade units with trained personnel and special equipment for intervention in case of road tunnel accidents. For simulation purposes a potential demand for intervention was calculated considering the existing locations and classification of tunnels on the Slovenian road network and the potential spatial accessibility was investigated. Stress was laid upon the tunnels that require specialized emergency vehicle intervention in cases of incidents. The paper describes the model and presents the results of the simulations.

Keywords: GIS, location-allocation problem, fire brigades, tunnel accidents, emergency systems

1. Introduction

There are almost 33 kilometers of road tunnels in Slovenia, around 30 percent of which are longer than 500 meters. Compared to other European countries, even relative to the countries' area, this is not a lot; however the possibilities of tunnel fire or similar incidents exist. With regard to the better driving conditions due to newer infrastructure and taking into account the freight traffic growth on the Slovenian road network, especially after joining the EU, the threat is even greater as most of the incidents happen due to drivers who do not respect the speed limits in tunnels. Thus it is necessary to seriously consider the risk and to foresee and take all due precautions, as well as prepare all types and means of intervention.

In case of tunnel road accidents, specially equipped fire brigade units can intervene, thus based on a demand and supply approach our work focuses on investigating the potential spatial accessibility from fire brigade locations to tunnels, which would involve as few fire brigades as possible to keep the costs lower, but at the same time to be much more efficient in time and space. For this purpose a Maximal Covering Location Problem (MCLP) location-allocation model was used in a GIS environment.

2. A GIS approach and the model used

GIS technology is increasingly used and recognized as an important planning tool for the acquisition, organization, manipulation, analysis and display of large volumes of spatially referenced data. It enables spatial data creation, task automation and enhanced map production [3]. The integration of model-based methods and GIS technology have substantial benefits for managing and analyzing data to produce information relevant to decision making and in simulating the effects of different planning decisions [4].

For the simulations, the commercial GIS package ARC/INFO 8.3 was used. There are six location-allocation built-in models available in ARC/INFO, each designed to solve a different type of problem. To effectively use these models, it is important to understand each model thoroughly - the types of problems it is designed to solve and the types of solutions it tends to produce. Location-allocation models determine the locations for centers and the allocation of demand to centers according to a specific objective. Some models also impose constraints on the solution (such as make sure no demand is further than 10 kilometers or minutes from its closest center). Among these there are emergency service location models, where the objective is to serve as many calls as possible within a given time or distance.

In our case we used MAXCOVER (Maximal Covering Location Problem) to locate a fixed number of facilities in order to maximize the demand that is covered within a specified time or distance [2], and MAXCOVER - constrained (Maximal Covering Location Problem with a Mandatory Closeness Constraint) to maximize the demand that is covered within a specified time or distance, while ensuring that no demand is further than a second time or distance threshold. Both models assign the demand to its closest facility and maximize the coverage of fixed number of located facilities linearly [8], however the first has only one coverage constraint, the second has two time or distance constraints.

A demand point is considered covered if it is within the distance or time threshold of its assigned facility. The facilities are located so that the demand covered is maximized. The actual distances traveled from demand points to centers, whether inside the distance threshold or not, are not relevant. For an adequate emergency facility location a required response time to demand points must be as short as possible. A time or distance threshold defines the surrounding area that a facility can serve (demand within this area is so covered).

When it is not possible to cover all demand with the existing facilities due to e.g. limited funds, both models try to maximize the amount of demand that can be covered with a fixed number of facilities within a given time or distance threshold. Running the MCLP model, each time adding a facility, it is possible to analyze the amount of demand covered and find out the trade off between the number of facilities and the coverage. There may be many different location configurations of facilities that provide the same degree of coverage. The Constrained MCLP model can then be used to analyze a variety of configurations.

The MCLP with Mandatory Closeness Constraints locates the facilities so that all demand is covered within a reasonable response time within a second distance threshold when a certain demand is not covered within a first time or distance threshold. This means that there is a second larger distance or time threshold beyond which there is no other demand.

One crucial limitation of the simulation is a limitation of the implicated model (MCLP) that does not consider the possible situation in which no road accident intervention squad is available at the closest fire stations to respond to a call within the time or distance threshold. The second important limitation is that potential spatial accessibility is calculated, which does not consider other geographic factors, social–economic characteristic of target populations, and organizational characteristics of emergency service. These factors and characteristics may either increase or decrease accessibility.

3. Data used in the simulations and demand and supply formulation

Data analysis and graphic presentation of the study results were effective, fast and simple due to the GIS approach, which enabled simulation and forecasting. Geographic and attribute data from different databases were merged to make the analysis simpler and clearer, which increased its efficiency in changing and testing the parameters and conditions. The data used were mostly provided by the Administration for Civil Protection and Disaster Relief of the Republic of Slovenia. Slovenia has a non-uniform population distribution, which also affects the accident location distribution.

The simulation was run on the road network in the Republic of Slovenia. The road network consists of local roads of relative importance. As the database did not include the

data on intervention access trails to the highway tunnel portals, these had to be captured on-site. Assigning the intervention trails a value 0 for drive time, the model algorithm had to include in simulation these instead of other roads. The driving time on intervention trails was added to the total time of the optimal access path later on.

3.1 Potential demand formulation

Allowing for different scenarios, three types of demand were calculated for each tunnel. The three types of demand were needed to describe the facts of a case as realistically as possible. The first type of demand for specialized emergency vehicles intervention DEMAND_1 required a 5 rank classification of tunnels. In this case, the tunnels were classified according to tunnel length (less or more than 500 m), tunnel type (single- or double-barreled), road category and threat degree. The threat degree level was determined considering traffic volume on each road.

According to the category, very short tunnels do not require specially equipped fire brigade intervention as a rule, however it is important for all categories that the location is accessible in a foreseeable time in case of an accident in the tunnel. Tunnels that fall into categories with a relatively high demand for specially equipped vehicles intervention are usually accessible by road or by special intervention trails for faster accessibility.

The Karavanke tunnel (3336 m) falls into a category with the highest demand, which is a result of its characteristics and traffic volume. Owing to the building of new highways most tunnels on these roads are more sophisticated and better equipped. There are 28 tunnels that fall into this category, and they measure approximately 29,800 m all together.

Besides the differentiated demand based on the aforementioned factors, we used two more types of demand. All the tunnels have the same value in DEMAND_2. On the other hand, DEMAND_3 consisted of simulation conditions where two emergency units had to intervene in case of an accident. In this case, very short tunnels with relatively small traffic volume were omitted (a value 0 was assigned).

3.2 Supply formulation

At present, 44 fire brigades are considerably well equipped and trained for technical rescue on roads. These are selected within a Decree on the Organization, Equipment and Training of Protection, Rescue and Relief Forces [13]. The fire brigades are contained in the GESP layer and are well located in space. The optimal special equipment distribution was investigated among them.

Taking into consideration the different scenarios, three types of supply were estimated. SUPPLY_1 took into account the fire brigade units' qualification and equipment. All units with at least two qualified teams who can respond in less than 3 minutes were given a value of 20. This means that such a unit could intervene in cases of fire or tunnel accident autonomously.

If the fire brigade unit has only one team that can respond in less than 3 minutes, a value 10 was assigned; if one team can respond in more than 3 minutes but less than 5 minutes, then a value 5 was assigned. Such a formulation assured the intervention of at least two fire brigade units with their teams in case of tunnels with a higher degree of threat.

SUPPLY_2 assumed that all the fire brigade units are equally qualified and equipped and so in such case each unit can intervene in every tunnel autonomously. While on the other hand, SUPPLY_3 presupposed that no fire brigade unit is qualified enough to intervene in any tunnel autonomously.

With respective constraints that no demand is outside 15 minutes drive time and that the demand covered in 10 minutes drive time is maximized, more joint simulations on the Slovenian road network were carried out to search for the optimal number and locations of fire brigade units:

- among the existing units,
- among the existing units and subsequently added units.

4. Simulations results

With all types of demand (DEMAND_1, DEMAND_2 and DEMAND_3) the results of the simulations showed that considering the tunnels' locations the existing fire brigade units are not optimally located. In fact, the existing fire brigade units could not cover the demand within a maximal drive time of 15 minutes. The average response time (7.71 min) remained the same whether all 44 units or just 14 chosen units were taken into account. There was also no difference in the demand area covered, which was in both cases 71 percent.

All together 12 tunnels were not covered, which implies that the potential demand with a very high threat degree due to a large traffic volume cannot be satisfied by the supply given. Further reduction to 8 fire brigade units slightly increased the average response time while the demand area covered changed almost imperceptibly.

The simulations results suggested that the maximum response time should be increased to 22 minutes to cover all demand with just 14 fire brigade units; the average time would then be 12 minutes.

In the next simulation we considered also less qualified and equipped fire brigade units in Slovenia. This simulation included only the more important tunnels with a higher threat degree owing to the traffic volume. The maximal response time was estimated to be 11.92 minutes within which all demand was covered by 7 units. Five of these were extant specialized fire brigade units, while two were extant but less qualified units. The average response time in this case was 5.88 minutes, which is shown in Table 1 and Figure 1.

Table 1: Descriptive statistics of driving times for 7 fire brigade units

Descriptive statistics for the nearest units	
Average	5.88
Standard deviation	0.42
Median	5.80
Standard deviation	0.27
Range	10.20
Min	1.79
Max	11.92
Sum	270.82
N	46

By all criteria, this solution is rather satisfactory and practical, as a maximal coverage can be achieved at lower costs, and it is relatively easy to put into effect. The two new units ought to be turned into qualified and properly equipped units to follow EU directives, and at the same time such an action would probably be the most economically justified in the long term.

Figure 2 shows that tunnels with a higher traffic volume and thus representing a higher threat could be covered by 7 fire brigade units within a response time of 12 minutes' drive, involving the five existing and two new qualified and specially equipped fire brigade units.

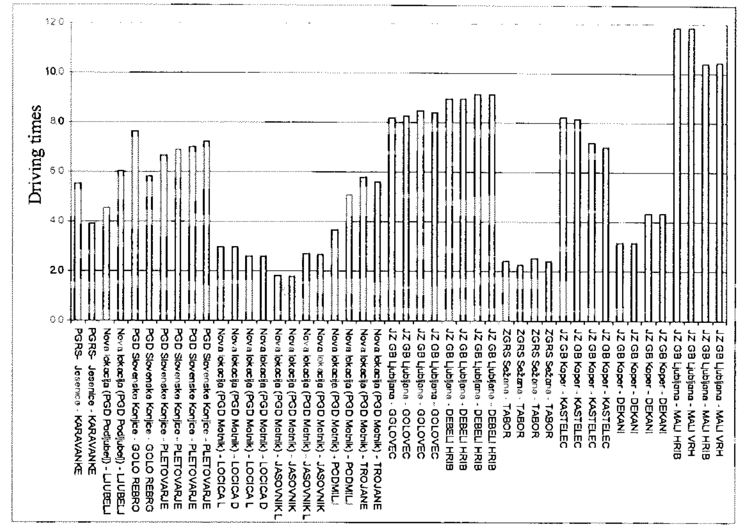


Figure 1: Driving times to tunnels with a higher threat degree for 7 fire brigade units

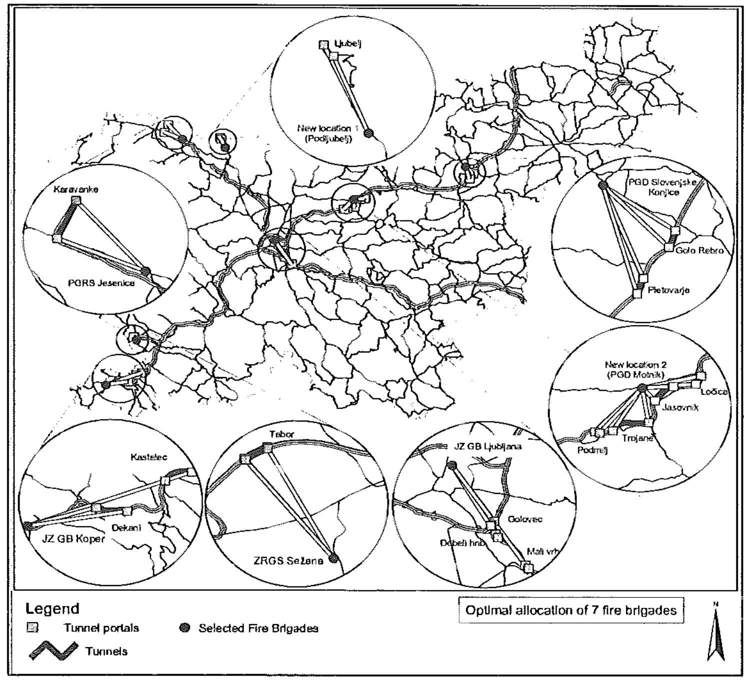


Figure 2: Optimal coverage of all potential intervention demand for more important tunnels (5 existing qualified and equipped units and 2 existing, less qualified and equipped units)

5. Conclusion

The simulations ran in the GIS environment implementing the location-allocation model led to interesting findings and suggestions for the optimal allocation of qualified fire brigade units that could intervene in case of road tunnel accidents in the Republic of Slovenia.

Testing a variety of scenarios the simulation generated more alternatives according to the constraints and conditions given.

The first simulations did not provide satisfactory results. In fact, the 44 qualified and properly equipped existing fire brigade units could not cover all demand within a maximal time of 15 minutes. Even reducing the number of units to 14 could not decrease the average response time and increase the demand area covered.

After several scenarios were run we had to consider less qualified and equipped fire brigade units. This simulation estimated the maximal response time of 11.92 minutes within which all more important tunnels were covered by 7 units. Five of these were the existing qualified fire brigade units, the others were two extant, less qualified, units. Since the less important tunnels with relatively low traffic volume do not require special equipment it is possible to state that this proved to be the optimal solution, satisfactory in the practical sense and economically justified.

The simulation results are a good basis for deciding on the number of the qualified and specially equipped fire brigade units as well as their allocation in Slovenia. However, even the fact that the most qualified squad reaction time can be as short as 1 minute, the road condition most of the times is not ideal, which in turn increases the calculated response times for 5 minutes or more. Additionally, taking into account that tunnels represent a very specific environment for fire to evolve rapidly, the tunnel road accidents with fire involved are practically impossible to dominate by experience. The fire brigade intervention in such cases can be too late. Thus, considering all these facts the education and reminders to the drivers to observe the rules in tunnels, and the control of the speed limits play an important role in preventing tunnel road accidents.

References

[1] ReVelle, C. and Snyder, S.: *Integrated Fire and Ambulance Siting: A Deterministic Model*, Socio-Economic Planning Science, 29(4), pp. 261-271, 1995.

[2] ReVelle, C.: *Siting Ambulances and Fire Companies: New Tools for Planners*, Journal of American planning Association, 57, pp. 471-484, 1991.

[3] Hall, G. Brent, Boweiman R.L. and Feick, R. D.: *GIS-based Decision Support Architecture and Applications for Developing Countries*, Waterloo: University of Waterloo, Faculty of Environmental Studies, 1997.

[4] Paliska, D.: *Road Incidents Emergency Service: a Case Study for Locating and Equipping Fire Brigades*, Promet-Traffic-Traffico, Vol. 15, No. 5, pp. 315-320, 2003.

[5] Fabjan, D., Paliska, D. and Drobne, S.: *GIS and GPS as Useful Tools to Determine Transportation Noise Levels*, 4th International Conference on Management Information Systems, Boston: WIT Press, pp. 33, 2004.

[6] Birkin, M., Clarke, G., Clarke, M. and Wilson, A.: *Intelligent GIS: Location Decisions and Strategic Planning*, New York: John Wiley & Sons Inc., 1996.

[7] Bridgehouse, B.: *Emergency Health Services: A GIS Application*, Proceedings of the 5th Canadian Conference on GIS, Ottawa, March 23-25, pp. 611- 622, 1993.

[8] Chruch, R., and ReVelle, C.: *The Maximal Covering Location Problem*, Papers of the Regional Science Association, 32, pp. 101- 119, 1974.

[9] GIS Newslink: *Emergency Mapping Response Service Implemented (Dauphin County, PA)*, GIS World, 6(7), pp. 11, 1993.

[10] Hancock, S.: *GIS for the Emergency Service*, Mapping Awareness, Volume 6, Number 5, pages 26-28, 1992.

[11] Henry, L. and Daniaud, C.: *Fighting Fire: The Montpellier Fire Brigade's Use of GIS Technology*. GIS Europe, Volume 4, Number 6, pp. xx-xxii, 1995.

[12] Official Gazette of the Republic of Slovenia, 22/99: *Decree on the Organization, Equipment and Training of Protection, Rescue and Relief Forces*, 1999.

[13] Parentela, M.E., Sathisan, S.K.: *GIS-based Allocation of Emergency Response Unit along a Transportation Route*, ESRI User Conference, 1995.

[14] Petelin, S. in drugi: *Uspešnost intervencije v predorih*. Raziskovalna razvojna naloga s področja požarne varnosti, Ministrstvo za obrambo, Uprava za zaščito in reševanje, 2005.

UNIVERSAL ROUTING ALGORITHM FOR CITIES AND OTHER BUILT-UP AREAS

Pavels Patlins

Riga Technical University, Faculty of Engineering Economy, Logistic and Transport Economics Department.

Abstract

50 years ago the main business problem was to develop a production; after that – to organize different marketing processes, but the key topic today is to organize the logistic process by optimal way. In other case it is impossible to achieve success in business. Auto transport is the corner-stone if we plan goods delivery processes for cities, towns and other built-up areas. Actually, we may use only this mode of transport to satisfy customers and organize from-door-to-door delivery technique.

Keywords: Planning, routing, logistic methods, algorithm.

1. INTRODUCTION

It is not so easy to create the optimal route within cities, towns and other built-up areas, because there are many different ways how to complete it. Sometimes logistic specialists make a typical mistake during the routing process, they try to minimize only vehicle's way, serving customers within the particular route.

The problem is really topical especially for multi-drop and multi-pick circular routes, when each vehicle should serve many customers (for instance, 50) and return to the main depot. It is necessary to optimize this process to achieve the best solution.

There are many special methods and also computer programs that may be used to solve the particular problem, connected with routing. But every method or program has different cons, talking about route planning within cities, towns and other built-up areas. In this case it is possible to achieve the optimal result, only using specials methods, programs and specialists' experience combination.

2. ROUTE CLASSIFICATION

Routing is the top problem within the modern logistic environment.

There are two main types of routes:

- circular route – connects more than 2 points. There are three types of the pendulum route: multi-pick route, multi-drop route and combined route – (figure 1);
- pendulum route – is a route, connecting only two points (for instance, it may connect the particular depot with the customer) – (figure 2).

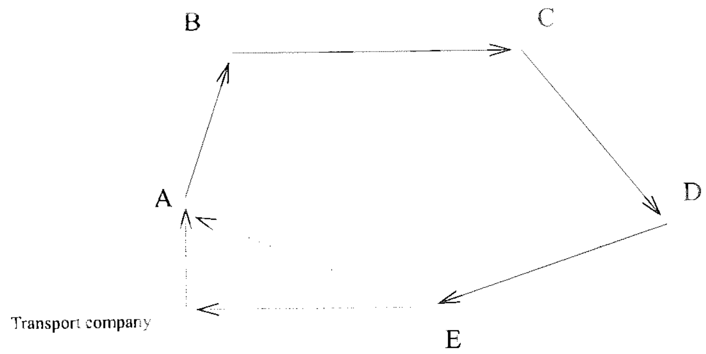


Figure 1. Example of a circular route [2,p.121]

Figure 1 illustrates a very popular mode of circular route, which is multi-drop route. Production is delivered from the one point (depot) to many different customers.

Multi-pick route means, that production is collected from many different places; after that it is necessary to delivery this production to the particular place. For instance, this route is very significant for dustman companies.

Combined circular route associate both of the above-mentioned types.

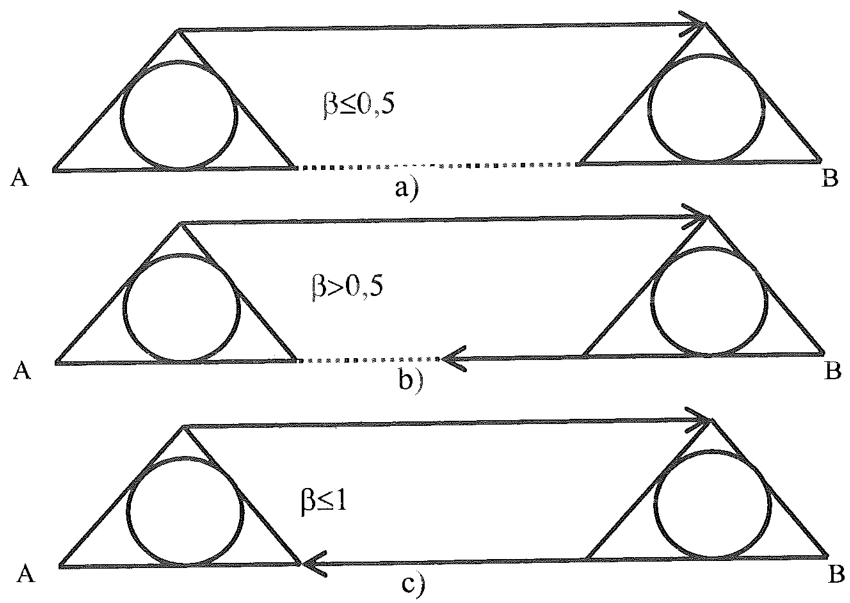


Figure 2. Pendulum route's types

Figure 2 illustrates types of the pendulum route.

a) Pendulum route with the idle running back.

b) Pendulum route with when the running back is partly-loaded.

c) Pendulum route with the full-loaded running back [2,p.121].

Actually, c) is the best type of the pendulum route; the running utilization rate (β) is equal by 1 or 100%. Unfortunately, it is quite difficult to create this route in the real life.

3. UNIVERSAL ROUTING ALGORITHM FOR CITIES AND OTHER BUILT-UP AREAS.

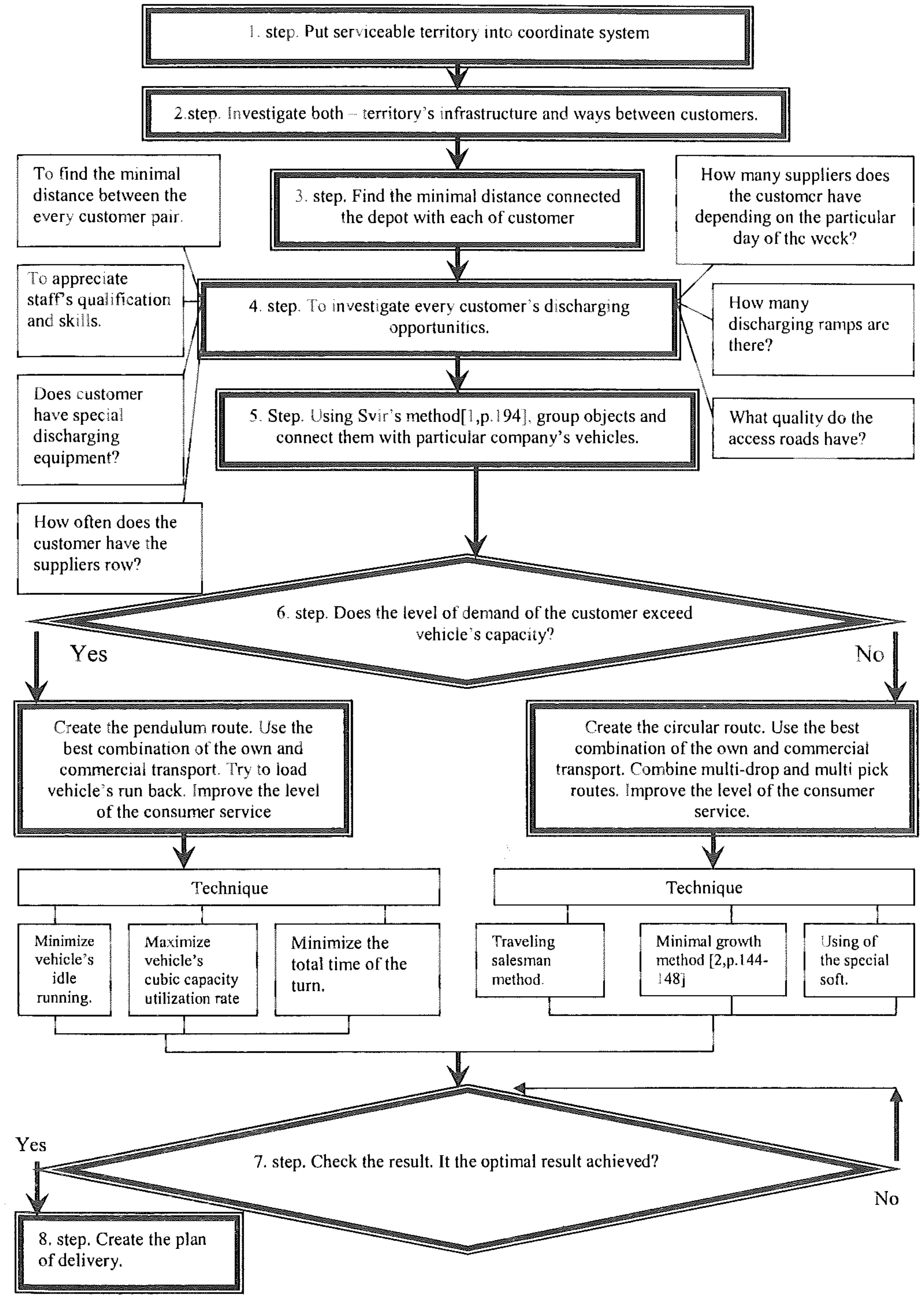


Figure 3. Universal routing algorithm for cities and other built-up areas.

According to the universal routing algorithm for cities and other built-up areas, it is expediently to use the following steps, solving routing problem:

Step 1. First of all it is necessary to put serviceable territory into coordinate system. This step provides visuality and clearness of the task.

Step 2. After that it is expediently to investigate ways between customers and territory's infrastructure in general. It is necessary to know everything about roads, its quality, distance and other aspects.

Step 3. Then to find the minimal distance between the start-point (depot, where production is located) and every customer. Minimal distance means way, which provides auto run time, going to the customer.

Step 4. Fourth step – investigate customers' particularities as well as their discharging opportunities and volume. There are 7 key-areas within this step (please, see figure 3, fourth step).

Step 5. After that routing is started. Using special Svir's method [1,p.194], it is expediently to divide the serviceable territory in some sectors as well as group objects and connect them with particular company's vehicles.

Every vehicle serves the particular sector and particular customers are connected with appropriate vehicle.

Step 6. When logistic specialists start routing, they have to answer to the following question: does the level of demand of the customer exceed vehicle's capacity? In order of the answer as well as depending on some other factors, it is necessary to model a circular route or a pendulum one. There some similar principles in the planning of both types of the route. For instance, either circular, or pendulum route requires to use the best combination of the own and commercial transport. Of course, the main task of every effective route is to improve the level of the consumer service. This step is really very significant, because it may increase or decrease company's transportation costs.

There are some additional tasks as well. For example, to reduce company's transportation costs, decreasing fuel costs, and other cost types. But this task is not the most significant in the modern business. Depending on the route type, it is expediently to use a particular technique, to achieve the best result. It is very important to use special methods, creating circular routes or use special principles, creating pendulum routes (increase the level of the cubic capacity utilization rate and decrease the idle running of the vehicle).

Step 7. The last step is: to check the result. If the optimal result is achieved, logistic specialists create the plan of delivery. If the result is not optimal, it is necessary to improve it and check one more time.

4. CONCLUSION

Route planning process is a very significant logistic chain's operation. It is necessary to draw routes very carefully, in other case delivery, transportation and logistic costs in general will be increased.

Routing process helps to satisfy all the customers in the best way, to supply them production in time. It is not so easy to achieve it, especially within city's route when one vehicle serves many clients.

It is expediently to use specials methods and techniques to design the best route for cities and other built-up areas. The most efficient way is to use these methods, techniques and professional experience combination. Universal routing algorithm for cities and other built-up areas (chapter 3) makes easy and more efficient route planning process.

5. REFERENCES

1. Гаджинский А. Логистика: Учебник для высших и средних специальных учебных заведений. - 3-е издание, перераб. и доп.- М.:Информационно- внедренческий центр "Маркетинг", 2000.- 375с.
2. Неруш Ю. Логистика: Учебник для вузов. - 2-е изд., перераб. и доп. - М.: "ЮНИТИ - ДАНА", 2000. - 389 с
3. John J. Coyle, Edward J. Bardi, C. John Langgley, Jr. The Management of Business Logistics, (8-th Edition), West Publishing Company, 2001 - pp. 580.
4. S.Engewall. Cost allocation in some routing problems – a game theoretic approach. Division of optimization, Department of mathematics, Lonkoping institute of Technology, SE-582 83, Linkoping, Sweden -pp 134.

POLICE PATROL ROUTE OPTIMIZATION

Marko Urh, Anton Čižman
University of Maribor, Faculty of Organization Sciences, Kidriceva 55a, 4000 Kranj, Slovenia
marko.urh@fov.uni-mb.si, anton.cizman@fov.uni-mb.si

Abstract
This paper treats the problem of police patrol route optimisation according to the locations with higher incident frequency. In the course of our research we have analysed 9537 traffic accidents that occurred between the years 1994 and 2002. In the case of these incidents the police patrol reaction time is extremely important. To perform the simulation of police patrol movement we have constructed a model of streets and roads in the area covered by the police station in Kranj, using a geographical information system. The model contains 1012 links and 691 nodes with a total length of 439,68 kilometers. The developed and described model was prepared in the cooperation with the Operations and Communications Centre of the General Police Directorate (OKC GPD) and can be use for solving other route problems.

Key Words: optimization, mathematical programming, geographical information systems, police

1. Introduction

Continual changes in the world as well as in Slovenia have become constant and they dictate different kind of work and life. With this change also the values of individuals and society are being modified. The fact is, that security represents one of the key points which influences on quality of life. Police has the power and legal authorization for ensuring safety and represents public service for citizens. Patrolling as the part of a police work is a diverse and dynamic process. For arrival on the place of a certain event, police needs some time, which is called police reaction time and represents time, which elapses from a citizen call on intervention number to the arrival of police patrol, criminal-police group or some other competence service to the place of intervention event. Solving complex problems has become more efficient with improvement of information technology with association with procedures and technique of optimization. Geographical information systems (GIS) represent an important tool for making analyses. Geographical information systems are computer supported information systems for capturing, saving, searching, analyzing, displaying and distribution of spatial data and information [1]. The main advantage of this kind of analyses is a chance of seeing new connections between available data. This gives us information, which we haven't had so far and can be used for better decision-making.

2. Some of the aspect of police work

Within every day work, police collects enormous number of data. Unfortunately, this data are rarely used for strategic and tactical purposes. Police should be looking for problem area using information systems, which are available according to the trends of events. Because of enormous amount of data this kind of information systems can be one of the best systems for solving and analyzing problems which are connected with some location. Results of this kind of analyses can be used for making appropriate police strategies and tactics, which are used for solving certain problems. It's a known fact that the calls on intervention number 113 (before 92) increases every year [2]. This confirms a demand for more rational and optimal work of Slovene police.

According to characteristics of individual events handled by the police and comparing to statistical reports of the Slovenia police, every year happens about approximately 40.000 traffic accidents, 35.000 violations of public order and 90.000 criminal offences.
The most obvious example of negative events are spatial. From the year 1930, when the first research in this field has been done, it has been discovered, that some kind of negative actions, methods used by perpetrators and degree of crime, differ according to the places, where they took place [3]. This is also true for traffic accidents. Reaction police time is critical when dealing with traffic accidents, violation of public order and in some criminal cases. However all these events are highly predictable considering the place where they occur.
Some facts, which confirm legitimacy of making this kind of research, are:

- number of events increases every year (positive index growth),
- number of police patrols can't be increased, so we can only optimize their routing and methods of their work,
- shorter reaction time of the police time can bring very positive consequences within the meaning of preventing violation of public order, arresting the perpetrators, rescuing lives, increasing satisfaction of citizens with services offered by police and other positive things,
- combination of information technology with procedures and technologies of optimization offers possibilities, which haven't been available before,
- methods and techniques of operation research can be used for optimizing bigger systems such as police GIS, regard to the nature of the problem and
- adequate data, skilled people and suitable information systems are the basic for conquering new knowledge, which can help us fight crime or some other unpleasant and unpredictable events.

In this research task we have tried to answer the question, how to reduce police response time. We have combined three different areas such as operation research, geographic information systems and police work.

3. Optimization model under GIS (ArcView-ESRI)

3.1. Construction of network model of streets and roads

First step, necessary for doing the computer simulation was constructing the model of streets and roads on the area covered by the police station in Kranj, which is one of the top five police stations in Slovenia according to the number of people who live there. Vectorization of streets and roads was made by software ArcView-ESRI. Model was containing characteristic of real world, such as one-way street, closed street, restraint turning left or to the right, speed limit and length of individual streets or roads. On the network model made by streets and roads on the area covered by police station Kranj, there was necessary to build a proper topology, which will represents a basic for execute network analyses. In this phase we used software ARC/INFO. The purpose of constructing topology is to define logical relations between streets and roads.

3.2. Designating the points with higher density of events

Every traffic accident treated by Slovenian police is entered into the geographical information system of the police, where it represents point on a map with all-important characteristics such as

date, time, cause, etc. In this research we analyzed 9537 traffic accidents, which happened between years 1994-2002 on area covered by police station in Kranj. Slovenian police considers all traffic accident where participant had died (4th category), was badly physically injured (3rd category), was lightly physically injured (2nd category) and traffic accident where there was only material damage (1st category), participants can't agree about responsibility or to fulfill European report and wish the police to arrive [4]. As already mentioned, we analyzed traffic accidents on area covered by police station Kranj. To accomplish this task we used computer software called ArcView Spatial Analyst. With this software we could divide researched area on squares with dimensions of 100 x 100 meters, which is appropriate resolution when analysing traffic accidents in a city. Based on such distribution we counted the events and determine the points with higher frequency of events (hot spots). We recognized 11 critical points with higher frequency of traffic accidents and most of them were taken place on crossroads.

3.3. Establishing optimal route

Points with higher frequency of traffic accidents were a base ground for determine optimal route of police patrol. Combination of suitable network model, points with higher frequency of events and with software called ArcView Network Analyst we could determine optimal route for police patrol. Assumption in this model was that police patrol must visit all 11 points in optimal order considering shortest possible time needed. Figure 1 represent optimal route of police patrol on the area covered by police station Kranj, according to the points with higher frequency of traffic accidents. At simulation we assumed that this area is covered by one police patrol responsible for traffic accidents.

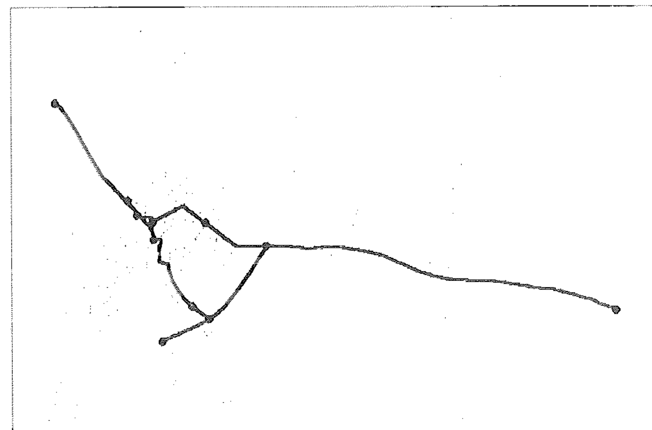


Figure 1: Optimal route for visiting all 11 points with higher frequency of traffic accidents on area covered by police station Kranj.

3.4. Realization of simulation

Realization of simulation required a suitable network model of streets and roads, established optimal route of police patrol based on suitable procedure of optimization. In first case we have assumed, that police patrol was located at a random spot somewhere on the area covered by police station Kranj. In second case we put police patrol random somewhere on a optimal route, which we have determined and described earlier. We have randomly chosen 50 traffic accidents in this area, which took place between years 1994-2002.

In first simulation we assumed that police patrol was somewhere random on the area covered by police station Kranj. Procedure of simulation was as followed:

- first it was necessary to choose one among 50 random points (traffic accidents) and it is presented on figure 2 as point A,
- then we have randomly chosen one among 1012 links (street or road), which composed network model,
- programme (written in Avenue) has random chosen length on chosen link,
- after that programme puts a point on chosen length of chosen link, which is presented on figure 2 as point B and represents starting point of police patrol,
- then the programme establishes two points, which represents starting (B) and ending point (A) or the beginning and the end of the route, which has to be traveled,
- time needed for travel the route by the police patrol is written in special file as time needed by this patrol. After 50 simulations we have calculated average time needed by patrol to arrive on one of 50 randomly chosen points,
- simulation was made for 50 randomly chosen points and for every one of them the simulation was done 50 times, which gave us 2500 simulations.

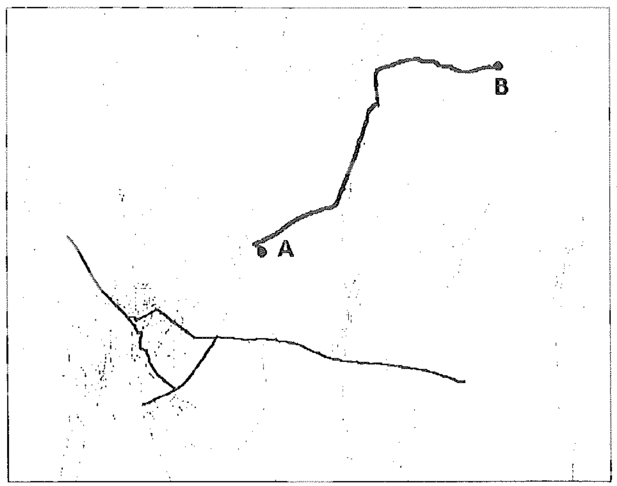


Figure 2: Route of the police patrol from random location (B) to random chosen traffic accident (A).

In second simulation we assumed that police patrol was random somewhere on the optimal route within the area covered by police station Kranj. Procedure of simulation was as followed:

- first it was necessary to choose one among 50 random points (traffic accidents), and it is presented on figure 3 as point A.

- then we have randomly chosen one among links, which represents optimal route (streets and roads on figure 1).
- programme has random chosen length on chosen link,
- after that programme puts a point on chosen length of a chosen link, which is on figure 3 presented as point B and represents starting point of the police patrol,
- then the programme establishes two points, which represent starting (B) and ending point (A) or the beginning and the end of the route, which has to be traveled,
- time needed for travel the route by the police patrol is written in special file as time needed by this patrol. After 20 simulations we have calculated average time needed by patrol to arrive on one of 50 randomly chosen points.
- simulation was made for 50 randomly chosen points and for every one of them the simulation was done 20 times, which gave us 1000 simulations.

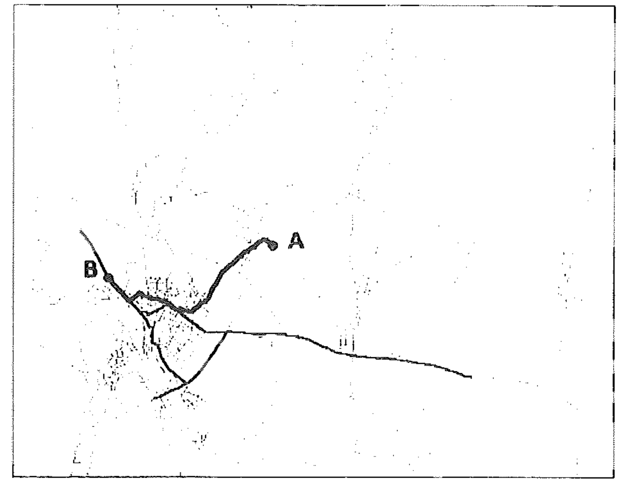


Figure 3: Route of the police patrol from random location on optimal route (B) to random chosen traffic accident (A).

4. Results

For the purpose of simulation we have chosen 50 random traffic accidents, which happened between years 1994-2002 on area covered by police station Kranj. In first scenario the police patrol was placed somewhere random on network model (on the street or road) and had to come to the place where traffic accident had happened. Second scenario assumed that the point of an accident was the same as in first case but police patrol was located somewhere random on optimal police route. In first case average time needed for arrival was 7'04" and in the second case the time was 5'26''. Comparing those times, we can clearly conclude that the second alternative routing is 23,1 % shorter. In research we additionally established, that majority of traffic accidents happen on Fridays and between 14:00 and 16:00 hour. This is the most probably a consequence of tiredness, which is result according to finished working shift and other factors connected with tiredness. According to the locations with higher frequency of traffic accidents we established, that this locations in majority represents bigger crossroads in area.

5. Conclusions

For the successfully realization of the simulation it was necessary to build a suitable network model of streets and roads on area covered by police station Kranj. Model contains 1012 links and 691 nodes, which measure 439,68 kilometers. Critical points with higher frequency of traffic accidents were established based on 9537 traffic accidents, which happened in the researched area between years 1994-2002. Reaction time of the police patrol is extremely important in these cases. Based on these points we determine optimal route, which have to be traveled by the police patrol under condition that every point has to be visit precisely once. Properly build network model in geographical information system was representing a base for realization of simulation. From the research and results we can conclude that the police patrol reaction time is shorter if we put patrol on optimal route, which was defined earlier. Researches from the past confirm the theory that there is a strong connection between the locations where certain events are taking place. Our research has shown that by planning optimal route we can reduce the time that is needed to patrol to arrive on the place where events took places. Improved reaction time of the police patrol can have positive effects. Former arrival on the place where the accident happened can save lives, prevents further accidents and has other positive effects. We believe that police offers a good service to citizens. From this point of view we can conclude that by the reducing of reaction time the satisfaction of citizens would be increased. Other positive thing about optimal police routing is that we can reduce variable transportation costs, which are nowadays more important than ever.

From logistical aspect we can say, that:

- we can reduce transport costs if we consider optimal routing for police patrol and
- proper time and spatial supply represents added value, which means, that the subjects or services are available, when and where is necessary [5].

The negative side of using an optimal route for police patrol is a chance that this kind of patrolling could become boring and cause monotony. However in spite of bad things we can say that advantage of this kind of work is essentially bigger. This kind of model can be used in other areas in Slovenia with small modification. With this kind of work we can upgrade the police planning and enable advantages of the police work and in the society.

References:

1. Korte G.B.: The GIS Book, On World Press, tretja izdaja, Santa Fe USA, 1994.
2. Internet: http://www.policija.si/si/preventiva/113/index.html
3. Trojanowicz, R.C., Kappeler, V.E., Gaines, L.K., Bucqueroux, B.: Community Policing. A Contemporary Perspective. Anderson Publishing, Cincinnati, OH, 1998.
4. Zakon o varnosti cestnega prometa (ZVCP-1) Ur.l. RS, št. 83/2004, 35/2005, 67/2005 Odl.US: U-I-32/05-13, 69/2005.
5. Čižman A.: Logistični management v organizaciji, Moderna organizacija, Kranj, 2002.

Section 6
Finance and Investment

MULTI-CRITERION APPROACH VERSUS MARKOWITZ IN SELECTION OF THE OPTIMAL PORTFOLIO

Zdravka Aljinović, Branka Marasović, Neli Tomić-Plazibat
Faculty of Economics, University of Split
Matice hrvatske 31, 21 000 Split, Croatia
E-mail: zdravka.aljinovic@efst.hr; branka.marasovic@efst.hr; neli@efst.hr

Abstract: In the paper we select an optimal portfolio on the Croatian capital market. First, selection is carried out using classic Markowitz model, which allows selection of the optimal portfolio based upon the mean-variance (M-V) criterion. Second, we illustrate how the multi-criterion approach makes it possible to integrate, within the portfolio selection process, the conventional M-V criteria with other market criteria.

Key words: portfolio selection, Croatian capital market, efficient portfolio, multi-criteria methods

1. INTRODUCTION

The Markowitz's theory of portfolio selection describes how to calculate a portfolio that gives the lowest variance of return of all portfolios having the same expected return, or the one that has the highest expected return of all portfolios having the same variance, i.e. how to calculate an efficient portfolio.

In accordance with the modern portfolio theory maximisation of returns at minimal risk should be the investment goal of any successful investor. Nevertheless, contrary to the expectations of the modern portfolio theory, the tests carried out on a number of financial markets reveal the existence of other indicators important in portfolio selection. Considering the importance of variables other than return and risk, selection of the optimal portfolio becomes a multi-criterion problem.

In the paper we select an optimal portfolio on the Croatian capital market using conventional Markowitz model and then, we illustrate how the multi-criterion approach makes it possible to integrate the conventional M-V criteria with other market criteria.

2. THE MARKOWITZ MODEL OF THE PORTFOLIO OPTIMIZATION

Now, we show how to calculate efficient portfolio and efficient frontier (the set of all efficient portfolios).

Throughout this chapter we use the following notation: There are N risky assets, each of which has expected return $E(r_i)$. The variable R is the column vector of expected returns of these assets, and S is the $N \times N$ variance-covariance matrix.

A *portfolio of risky assets* (when our intention is clear, we shall just use the word *portfolio*) is a column vector x whose coordinates sum to 1. Each coordinate x_i represents the proportion of the portfolio invested in risky asset i. The *expected portfolio return* $E(r_x)$ of a portfolio x is given by the product of x and R:

$$E(r_x) = x^T R = \sum_{i=1}^{N} x_i E(r_i).$$

The *variance of portfolio x's return*, $\sigma_x^2 \equiv \sigma_{xx}$ is given by the product $x^T S x = \sum_{i=1}^{N}\sum_{j=1}^{N} x_i x_j \sigma_{ij}$.

Mathematically, we may define an efficient portfolio as follows: For a given portfolio variance σ_x^2 (or standard deviation), an efficient portfolio x is one that solves

$$\max E(r_x) = x^T \cdot R \equiv \sum_{i=1}^{N} x_i \cdot E(r_i)$$

subject to $x^T Sx = \sigma_x^2, \sum_{i=1}^{N} x_i = 1, \quad 0 \le x_i \le x_{Mi},$

where x_{Mi} is the maximum proportion to invest in share i.

3. THE MULTI-CRITERION APPROACH

Considering the importance of variables other than return and risk, selection of the optimal portfolio becomes a multi-criterion problem.
Generally, the distinction is made in literature between two categories of criteria: the accounting criteria and those based on market value. The accounting criteria are indicators of the financial situation of a company. There is a number of such indicators and their choice depends on the manager's approach and objective.
Market criteria are indicators obtained by analysing the situation on the stock exchange. They include: expected return, total risk (variance of expected returns), system risk (beta), PER (price earnings ratio), share liquidity, etc.
As there is a strong interdependence between some of the accounting criteria and market criteria, in order to avoid their redundancy we limit this study to the following set of criteria:
1. Expected return $E(r_x)$.
2. Beta coefficient: Beta coefficient is the measure of the systematic risk. For the whole market, beta is by definition equal to 1.0; for a security with lower value of systematic risk, i.e. with lower variability of expected rate of return, beta is lower than 1.0, and higher than 1.0 for those with higher rate of risk. It is defined as:

$$\beta_j = \frac{COV(r_m, r_j)}{VAR(r_m)},$$

where r_m is the expected rate of return of the market portfolio and r_j is the expected rate of return of the share. It is the criterion to be minimized.
3. Turnover (liquidity criterion): $\Omega = \frac{\text{total number of shares treated}}{\text{total number of shares outstanding}}$.
This criterion is to be maximized.
4. Price earnings ratio (PER): In many papers it is shown that shares with lower PER have, on average, higher absolute and risk-adjusted returns, than randomly selected shares, so this criteria is to be minimized.
5. Market-to-book value ratio: Shares of the companies with higher market-to-book value ratio represent investments with higher risk, and so this criterion is to be minimized.
The method of multi-criterion programming, used in this paper, is based on PROMETHEE II approach. With PROMETHEE II, each pair of actions is compared according to each criterion. Since the overall portfolios that can be made up from a set of pre-selected shares are infinite, it is impossible to compare all of the pairs of portfolios. In the paper [5] it is proposed the method that allows an absolute evaluation of each possible portfolio by comparing it with two fiction portfolios: the ideal $(\overline{P})$ and the anti-ideal $(\underline{P})$. For criterion C_j to maximize we'll have $C_j(\overline{P}) = \max_i C_j(s_i)$, where $S = \{s_1, s_2, ..., s_N\}$ is the set of N pre-selected shares. In the same way $C_j(\underline{P}) = \min_i C_j(s_i)$. The set of possible solutions is the set of portfolios which can be formed from pre-selected shares. For any possible portfolio P has to be $\Phi(\underline{P}) \le \Phi(P) \le \Phi(\overline{P})$ (where $P = X'_p S = \sum_{i=1}^{N} x_i s_i$, x_i is proportion of the share s_i in portfolio P).
For criterion C_j to maximize we will have:

$$P_j(P, \underline{P}) = \begin{cases} 0 & if\ C_j(P) = C_j(\underline{P}) \\ H_j^-(d_j) & if\ C_j(P) > C_j(\underline{P}) \end{cases}, \quad d_j = C_j(P) - C_j(\underline{P})$$

and

$$P_j(\overline{P}, P) = \begin{cases} 0 & if\ C_j(P) = C_j(\overline{P}) \\ H_j^+(d_j) & if\ C_j(\overline{P}) > C_j(P) \end{cases}, \quad d_j = C_j(\overline{P}) - C_j(P)$$

where

$$C_j(P) = \sum_{i=1}^{N} x_i C_j(s_i),$$

$$\sum_{i=1}^{N} x_i = 1,$$

H_j^+ and H_j^- are preference functions.
According to PROMETHEE II, net flow of a portfolio P is:

$$\Phi(P) = \sum_{j=1}^{N} w_j \left(P_j(P, \underline{P}) - P_j(\overline{P}, P)\right)$$

The preference functions requires thresholds q_j^+, p_j^+, σ_j^+, i.e. q_j^-, p_j^-, σ_j^-. Finally, the optimal portfolio is one that solves:

$Max\, \Phi(P)$

subject to;

$$\sum_{i=1}^{N} x_i = 1 \text{ i } 0 \le x_i \le x_{Mi},$$

where

$$\Phi(P) = \sum_{i=1}^{N} w_j \Phi_j(P),$$

$P = X'_p S$,

$X_P = (x_1, x_2, ..., x_N)$,

$S_P = (s_1, s_2, ..., s_N)$,

x_i: proportion invested in share i $(i = 1,2,..., N)$ in portfolio P,
x_{Mi} maximum proportion to invest in share i in portfolio P,
N is the number of pre-selected shares which can be included in portfolio P.

4. AN APPLICATION

Using the above models we subsequently calculate the optimal portfolios on the Croatian capital markets. From the total number of securities quoted on the Zagreb stock exchange in 2004 and 2005 a sample of fourteen shares has been separated. The shares sample contains the most traded shares in the second part of 2004 and first quarter of 2005: CROS-R-A, PLIVA-R-A, PODR-R-A, ATLS-R-A, ADRS-P--A, DLKV-R-A, ERNT-R-A, IGH-R-A, KOEI-R-A, KRAS-R-A, PBZ-R-A, PLAG-R-A, ZABA--R-A, ZAPI-R-A. For each security from the sample we take the closing price at the end of each month. First we calculate the monthly return for each security. This is the percentage return that would be earned by an investor who bought the security at the end of a particular month $t-1$ and sold it at the end of the following month. For month t and security A, monthly return r_{At} is defined as:

$$r_{At} = ln\left(\frac{P_{A,t}}{P_{A,t-1}}\right).$$

Now, we calculate monthly mean, monthly variance and monthly standard deviation. By programming in MATLAB, we calculated ten efficient portfolios (Table 1) and efficient frontier for Croatian shares market u observed period (Figure 1). The range of solutions is given from the lowest to the highest possible monthly return, with a share of shares in the portfolio which lies on the efficient frontier, and their monthly risk and return.

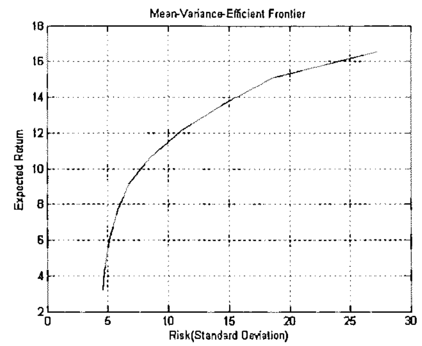


Figure 1. Efficient frontier for Croatian share market

Table 1. Efficient portfolios for Croatian shares market

Port	1	2	3	4	5	6	7	8	9	10
CROS-R-A	0,4777	0,5832	0,6281	0,6188	0,4477	0,1175	0	0	0	0
PLIVA-R-A	0	0	0	0	0	0	0	0	0	0
PODR-R-A	0	0	0	0	0	0	0	0	0	0
ATLS-R-A	0	0,0356	0,0624	0,0891	0,1210	0,1497	0,2115	0,2774	0,3899	1
ADRS-P--A	0	0	0	0	0	0	0	0	0	0
DLKV-R-A	0	0	0	0	0	0	0	0	0	0
ERNT-R-A	0	0	0	0,0641	0,1937	0,3048	0,4095	0,4763	0,2058	0
IGH-R-A	0	0	0	0	0	0	0,0665	0,1993	0,4043	0
KOEI-R-A	0	0	0	0	0	0	0	0	0	0
KRAS-R-A	0,1746	0,1790	0,1791	0,09620	0	0	0	0	0	0
PBZ-R-A	0	0	0	0	0	0	0	0	0	0
PLAG-R-A	0	0	0	0	0	0	0	0	0	0
ZABA--R-A	0	0	0,0610	0,1318	0,2377	0,4279	0,3126	0,0470	0	0
ZAPI-R-A	0,3477	0,2021	0,0695	0	0	0	0	0	0	0
Exp. Ret.%	3,2114	4,6946	6,1779	7,6611	9,1444	10,6277	12,1109	13,5942	15,0774	16,5607
Port. S:D:	4,6235	4,7931	5,2359	5,8280	6,7705	8,5713	11,1281	14,5149	18,5610	27,3281

To apply the proposed multi-criterion model, we had to calculate: values of beta, turnover rate, market-to-book value ratio and PER for pre-selected set of shares. The values of all criteria are shown in table 2.

Table 2. Stock's values for the constructed criteria

Stocks	Mean return (%)	β	Turnover (%)	PER	M/B
CROS-R-A	6,104554647	0,3617	4,35275912	13,51974	1,512619
PLIVA-R-A	-3,101549283	0,6437	7,91014276	8,521907	0,943449
PODR-R-A	4,299254057	1,0165	32,7235612	14,71509	0,670162
ATLS-R-A	16,56070642	1,0833	8,84472735	86,52308	1,131153
ADRS-P--A	3,586385321	1,0892	37,1017526	7,496381	0,597637
DLKV-R-A	9,52658376	1,7629	9,07602715	8,642446	1,595817
ERNT-R-A	11,5267951	1,0633	4,21995269	8,784485	1,543534
IGH-R-A	15,45432582	2,0765	40,8714844	13,08793	1,043264
KOEI-R-A	10,03302109	1,7023	39,1477255	52,84407	0,37464
KRAS-R-A	4,809726606	0,7125	15,743644	15,08071	0,705089
PBZ-R-A	7,884573604	1,4761	0,82862545	12,31563	2,132389
PLAG-R-A	5,861165236	1,2675	3,91054294	15,4925	0,150894
ZABA--R-A	9,154158959	0,8201	0,28580844	8,772863	1,469018
ZAPI-R-A	-1,565430255	0,547	1,61479571	15,50568	7,532247

In this case the linear preference criterion is used, which includes one threshold. For each criterion j, and for both of functions H_j^+ and H_j^- we need thresholds p_j^+ and p_j^-, which define strict preference area. Then, function Φ_j, for criterion which is to be maximized, is of the form:

$$\Phi_j(P) = P_j(P,\underline{P}) - P_j(\overline{P},P) = \begin{cases} \dfrac{C_j(P) - C_j(\underline{P}) - p_j^-}{p_j^-} & when\ C_j(\underline{P}) \le C_j(P) \le C_j(\underline{P}) + p_j^- \\ 1 - 1 = 0 & when\ C_j(\underline{P}) + p_j^- \le C_j(P) \le C_j(\overline{P}) - p_j^+ \\ \dfrac{p_j^+ - C_j(\overline{P}) + C_j(P)}{p_j^+} & when\ C_j(\overline{P}) - p_j^+ \le C_j(P) \le C_j(\overline{P}) \end{cases}$$

We propose these values for our application:

p_j^+	11,5607	1,0383	32,8715	7,5036	4,8491
p_j^-	8,1015	0,6765	7,7142	71,5231	2,5322

This set of values for the thresholds p_j^+ and p_j^- respect the relation $p_j^+ + p_j^- = \left|C_j(\overline{P}) - C_j(\underline{P})\right|$. These values must also reflect the preferences of a decision-maker, exactly as relative importance of criteria. In order to take into consideration the behaviour of investors, we proceed to the change of weights. Table 3 shows weights of three possible scenarios.

Table 3. Weight of each criterion

Criterion	Mean return	β	Liquidity	PER	M/B
Scenario 1	0.30	0.30	0.15	0.15	0.10
Scenario 2	0.50	0.10	0.20	0.10	0.10
Scenario 3	0.20	0.50	0.10	0.10	0.10

We note that this application is a simple illustration of the proposed approach. No real decision-maker is implied. By programming in MATLAB we get solutions given in table 4. Firstly, an optimal portfolio is calculated without maximum proportion constraint, secondly in considering maximum proportion constraint $x_M = 0.3$, and finally in considering $x_M = 0.2$.

Table 4. Optimal proportions to invest in shares

Stocks	Scenario 1			Scenario 2			Scenario 3		
	x_i	x_i $x_M = 0.3$	x_i $x_M = 0.2$	x_i	x_i $x_M = 0.3$	x_i $x_M = 0.2$	x_i	x_i $x_M = 0.3$	x_i $x_M = 0.2$
CROS-R-A	0	0.1	0.2000	0	0	0	0.8886	0.3000	0.2000
PLIVA-R-A	0	0	0	0	0	0	0	0	0
PODR-R-A	0	0	0	0	0	0	0	0	0
ATLS-R-A	0	0	0	0	0.3000	0.2000	0	0	0
ADRS-P--A	0.1469	0.3000	0.2000	0	0	0	0.1114	0.0445	0.2000
DLKV-R-A	0	0	0	0	0	0	0	0	0
ERNT-R-A	0.8531	0.3000	0.2000	0	0.1000	0.2000	0	0.0555	0.2000
IGH-R-A	0	0	0.2000	1	0.3000	0.2000	0	0	0
KOEI-R-A	0	0	0	0	0.3000	0.2000	0	0	0
KRAS-R-A	0	0	0	0	0	0	0	0.3000	0.2000
PBZ-R-A	0	0	0	0	0	0	0	0	0
PLAG-R-A	0	0	0	0	0	0	0	0	0
ZABA--R-A	0	0.3000	0.2000	0	0	0.2000	0	0.3000	0.2000
ZAPI-R-A	0	0	0	0	0	0	0	0	0

5. CONCLUSION

The two different approaches have provided significantly different results. Undoubtedly, the optimal relation of return rate and risk is what should be taken into account in selection of the optimal portfolio, but the question is how to consider other, very important market and accounting indicators for companies whose shares are traded. In this work we have shown how that it can be done by a new multi-criterion programming approach.That approach allows a stock market expert to build, in accordance with the decision-makers' preferences, all the available information into the model.

BIBLIOGRAPHY

[1] Aljinović Z., Marasović B., Tomić-Plazibat N., (2005): **The Selection Of The Optimal Portfolio On The Croatian Capital Market,** *Proceedings of Sixth International Conference on Enterprise in Transition (extended abstract, 299-301, full text on CD ROM)*

[2] Anderson, P. L., (2004): **Business Economics and Finance with MATLAB, GIS, and Simulation Models**, MATLAB & Simulink Based Books, Chapman&Hall/CRC

[3] Benninga, S., (2000): **Financial Modeling,** The MIT Press, Cambridge

[4] Black, F., (1972): **Capital Market Equilibrium with Restricted Borrowing,** *Journal of Business*, 45: 444 – 455

[5] Bouri A., Martel J. M., Chabchoub H, (2002).: **A Multi-criterion Approach For Selecting Attractive Portfolio,** *Journal Of Multi-Criteria Decision Analysis* 11:269-277

[6] Prohaska, Z., (1996): **Analiza vrijednosnih papira,** Infoinvest, Zagreb

[7] The Zagreb Stock Exchange (2005): **History Trading 2004-2005,** http://www.zse.hr/history.php?sessionId=62b230da7398c66c9575bdb8625d416e

THE CHOICE OF PRICE INDEX FORMULAE FOR ELEMENTARY INDICES

Draženka Čizmić
Faculty of Economics, University of Zagreb
Kennedyev trg 6, 10 000 Zagreb, Croatia
dcizmic@efzg.hr

Abstract

Recently the interest into the calculation methods that are appropriate at the lowest level of aggregation of a Consumer Price Index (CPI) has increased. It has long been known that different calculation methods for elementary indices give different results.

The geometric mean is being introduced by more and more countries. The approach is to calculate a geometric mean of prices in both periods and then derive the price relative or calculate a geometric average of the price relatives. Both calculations will yield the same results.

Keywords: Consumer Price Index, price index formula, elementary aggregate, elementary index, elementary index bias

1. Introduction

Measuring prices and their rate of change accurately is central to almost every economic issue. Consumer Price Indices (CPIs) published by national statistical agencies are mostly described as measures of price change obtained by comparing the cost of a fixed set of commodities through time.

The CPI is the best measure currently available, but it is not a true cost-of-living index. It suffers from a variety of conceptual and practical problems. There are a number of issues in the design of a CPI where there is no unique best way of doing things.

Recently the interest into the calculation methods that are appropriate at the lowest level of aggregation of a CPI has increased. It has long been known that different calculation methods for elementary indices give different results.

2. The nature of these indices

Elementary aggregates[1] play a strategic role in the index structure. They serve as strata in price sampling and provide primary blocks to build up indices at higher aggregation levels. The overall CPI is a weighted sum of all the sub-indices at the lowest level of aggregation.

The choice of elementary index formulae is a very important decision, likely more important than choosing an index formula at higher aggregation levels. Different countries use a variety of techniques to calculate elementary indices. The Boskin Commission (1996)

[1] Item, composite item and elementary aggregate are related terms. The term "item" is used loosely to mean any good or service included in the expenditure covered by the index. The term "composite item" is used more precisely to signify the most detailed level of aggregation of consumption for which weights can be estimated. If a breakdown of the consumption of a composite item by region or type of outlet is not available and cannot be estimated, then that composite item constitutes an elementary aggregate. But if such further breakdown can be made, then the composite item is decomposed into a number of elementary aggregates.

maintained that the use of particular index number formula for elementary indices imparted an upward bias of approximately 0.25 percentage points annually in the overall U.S. CPI.[2]

Research on elementary indices has mostly been restricted to unweighted formulae (arithmetic and geometric means) because detailed quantity data are not available within elementary indices. Thus the problem is to compute elementary aggregate indices using only sample price data and without any weights.

The choice of formula for elementary aggregate indices needs to take account of three features of such price data, which are of major practical importance:

1) There are sometimes missing observations.
2) Forced replacements occur – the disappearance of one product from the outlet where its price has been observed or the permanent closure of the outlet requires the selection of another product to replace it.
3) Optional many-to-many replacements occur when the sample of products and/or outlets is revised.

In response to an increase in the price, consumers could respond by adjusting their spending to favour relatively lower-priced goods or services. Substitution can take several forms (substitution among brands of products, substitution among product sizes, substitution among outlets, substitution across time, substitution among types of items within the category, substitution among specific items in different index categories).[3]

An elementary aggregate index can be computed:

- as a ratio of mean prices or a mean of price ratios
- using arithmetic or geometric or some other mean
- as a chained index or as a fixed base index

There are certain tests or axioms, which an index should meet:

- The identity test: whether or not the index has a value of unity if all prices return to their price reference-period level.
- The permutation test: if price "bounce", in the sense that the price of each of two products becomes the same as the price of the other was in the preceding month, while other prices remain unchanged, the index should show no change.
- The unit of measurement test: for example, if a price per kilo is used in all periods for any observed product, instead of the price per 500 grammes the elementary aggregate index will not be affected.

3. Recent developments in elementary indices

3.1 Australia

It was recognised that a more effective index processing system might be developed if the index compilers could select an elementary aggregate calculation formula from a suite of such formulae. The specific formulae being considered are:

I geometric mean of price relatives

[2] M.J. Boskin, E.R. Dulberger, R.J. Gordon, Z. Griliches, D.W. Jorgenson; The CPI Commission: Findings and Recommendations; American Economic Review, May 1997 (Papers and Proceedings), 87(2), pp.78-83.

[3] In response to an increase in the price charged by a store for a certain brand of ice cream, a consumer could respond by redistributing purchases along any of several dimensions represented by other priced items in the category: to another brand of ice cream whose price had not risen, to a larger package of ice cream with a smaller price per ounce, to ice cream at a different store where ice cream is on sale, or to a brand of frozen yoghurt. The consumer also could respond by postponing purchase until a later date. Finally, the consumer could substitute from the ice cream brand to a specific alternative dessert item that is in another CPI category.

II arithmetic mean of price relatives

III relative of arithmetic mean prices

The focus was placed on the use of the unweighted (or equal weighted) variants of the above formula.

The logical starting point are the known properties of the alternative approaches from the perspective of their effect on index outcomes:

- the three approaches deliver an identical outcome when all price movements are equal
- approaches II and III deliver identical outcomes when all base period prices are equal
- approaches I and III deliver identical outcomes under conditions of price "bouncing"
- approaches I delivers the same result as an approach based on the relative of geometric mean prices under all conditions

The approach based on the relative of arithmetic mean prices is attractive due to its immediately demonstrable equivalence to the standard Laspeyres formula, its ability to deal equitably with price bouncing and the fact that it produces an average price as an automatic by-product of index construction. Arguments against this approach are: 1) in practice, outlet weights are at best based on knowledge of relative value shares rather than quantities, and 2) the use of average prices alone denies the index compiler ready access to individual measures of long term price behaviour (price relatives).

The strengths of the approach based on the arithmetic mean of price relatives are: 1) the better alignment of outlet weights with the underlying source data, 2) its ability to cope more equitably with store standards, and 3) the additional information embodied in price relatives. The major weakness is seen as the counter intuitive way in which it deals with price bouncing.

The strengths of the geometric mean of price relatives approach are: 1) it assigns base period outlet weights in accordance with the source data, 2) it handles the problem of price bouncing equitably, 3) it provides the index compiler with price relatives, and 4) it delivers outcomes identical to those obtained by the relative of geometric mean prices approach. Its major drawbacks are: 1) the difficulty statisticians would encounter explaining its use to the larger community and, 2) the fact that it does not preserve fixed underlying quantity weights.

In Australia (unweighted) geometric mean formula is the preferred formula for calculating the elementary indices primarily because it models, better then alternative formulae, substitutions that consumers make in their consumption of goods and services in response to changes in relative prices.

The geometric mean cannot be used to calculate the average price in all elementary aggregates. It cannot be used in cases where the price could become zero. It is also not appropriate to use the geometric mean in elementary aggregates covering items between which consumers are unable to substitute. For these elementary aggregates the relative of arithmetic mean prices formula is used.

3.2 Canada

It was at beginning of the eighties when Prices Division encountered some anomalies in the calculation of elementary indices. The anomalies were diagnosed as a result of linking arithmetic means of price relatives. The purpose of the study, carried out in the Central Research Section of Prices Division, was to assess the long-term effect of using alternative elementary index formulae.

Six elementary index formulae have been taken into consideration, of which three are ratios of mean prices:

- a ratio of equiweighted arithmetic mean prices

- a ratio of equiweighted geometric mean prices
- a ratio of eguiweighted harmonic mean prices

Three other formulae are means of price relatives:

- an equiweighted arithmetic mean of price relatives
- an equiweighted geometric mean of price relatives
- an equiweighted harmonic mean of price relatives

There are relatively small differences between index numbers derived from elementary indices calculated as ratios of arithmetic, geometric and harmonic mean prices. For some commodities the highest index numbers were obtained with ratios of arithmetic mean prices, for some with harmonic means prices, while index numbers obtained with ratios of geometric mean prices fell between these two.

Differences between index numbers derived from elementary indices calculated as arithmetic, geometric and harmonic means of price relatives are much larger. Signs of these differences are certain: an arithmetic mean of price relatives is never lower than their harmonic mean, while a geometric mean lies between the two. The amplitude of the differences grows with the frequency of linking.

A geometric mean of price relatives is always equal to the ratio of geometric mean prices derived from the same sample. The index numbers obtained using these two elementary index formulae are also always equal to each other. This is not case with arithmetic or harmonic means.

Transitive formulae, such as ratios of mean prices or geometric means of price relatives, will give the same index number whatever the frequency of linking.

The ratio of arithmetic mean prices was used in the Canadian CPI until 1994 as the main elementary index formula to estimate month-to-month price change for any given commodity within a given spatial stratum.[4] Geometric means started being used from January 1995, to avoid problems with those categories of products that exhibit a broad spectrum of prices.

3.3 United States

The possibility of using the geometric mean formula to calculate elementary indices in the CPI was first raised by BLS[5] researchers in 1993. In December 1996, the Advisory Commission to Study the CPI (the Boskin Commission), recommended the use of the geometric mean formula for the aggregation of prices within all item categories in the CPI. In April 1997, BLS began issuing an experimental CPI that used the geometric mean estimator in the calculation of all elementary components of the index. This index was issued to provide users with a quantitative estimate of the impact that use of a geometric mean formula at the lowest level of index construction would have on the performance of the CPI.

The CPI-U-XG indices (the experimental CPI using geometric means) were a supplement to the price information available from BLS (CPI-U and CPI-W indices[6]). Historical differences between the CPI-U and CPI-U-XG arose not only because of the different formulae used, but also because of methodological changes made in the CPI-U since 1990. For historical comparison, BLS also issued an experimental Test Laspeyres series called the

[4] Each ratio was calculated taking into account only prices from "bilaterally matched samples", i.e. only prices that have been collected during two consecutive months in the same outlet and that relate to the same item. Long-term estimates of price change are obtained indirectly, through linking of these direct month-to-month elementary indices.

[5] Bureau of Labour Statistics

[6] The Consumer Price Index for all Urban Consumers (CPI-U) and the Consumer Price Index for Urban Wage Earners and Clerical Workers (CPI-W)

CPI-U-XL, which differs from the CPI-U-XG only in the use of the Laspeyres formula for aggregation of price quotations.

From December 1990 to February 1997, the CPI-U-XG rose 16.2 percent, compared with 18.6 percent for the CPI-U-XL. The average annual rate of growth in the CPI-U-XG over this period was 2.46 percent, 0.34 percentage points lower than the annual growth rate of the CPI-U-XL.

The geometric mean estimator has been introduced in both the CPI-U and the CPI-W in January 1999.[7] Components retaining the arithmetic mean (Laspeyres) formula are: 1) selected shelter services, 2) selected utilities and government charges, and 3) selected medical care services.

3.4 European Union

Within European Union, Harmonised Indices of Consumer Prices (HICPs) are consumer price indices compiled on the basis of a harmonised coverage and methodology.

Commission Regulation (EC) No 1749/96 defines elementary aggregates referring to the expenditure or consumption covered by the most detailed level of stratification of the HICP and within which reliable expenditure information is not available for weighting purposes. An "elementary aggregate index" is a price index for an elementary aggregate comprising only price data.

The ratios of geometric or arithmetic means of prices are the two formulae, which should be used in elementary aggregations. The arithmetic mean of price relatives should not normally be used, as it will in many circumstances result in failure to meet the comparability requirement. It may be used exceptionally where it can be shown not to fail comparability requirement.

The price index for an elementary aggregate may be calculated as a chain index using one of the two preferred formulae. The arithmetic mean of price relatives must not be used where chaining is more frequent than annual.

4. Elementary indices in Croatia

The CPI has become the official measure of inflation in Croatia in January 2004. It replaced Retail Price Index and Cost of Living Index. The index covers more than 540 representative goods and services divided in 12 major consumption groups according to COICOP.[8]

Most prices are collected locally. Central price collection is applied in cases of state monopolies and nationally homogeneous prices. For products for which prices are collected locally, the Central Bureau of Statistics supplies loose specifications to price collectors. Collectors are free to initially select particular product for pricing in a particular outlet. In order to make sure that the same products are then re-priced each month, the price collectors also record a pre-defined set of additional product characteristics.

Elementary aggregate indices are calculated as a ratio of geometric means of the current and reference period prices of the products within an elementary aggregate. The data required for the estimation of weights are not available within an elementary aggregate.

[7] The CPI, which is calculated using a fixed-weight Laspeyres formula, does not reflect the fact that consumers can and do change spending patterns as relative prices change. A measure of change in consumer prices that uses geometric mean formula successfully accounts for this consumer behaviour. The index using geometric means provides a closer approximation to a cost-of-living index.

[8] Classification of Individual Consumption by Purpose

5. Conclusion

The choice of elementary index formulae is a very serious business, with strong potential consequences on the resulting index numbers. In most practical cases, choosing an elementary index formula is more important than choosing a macro index formula for higher aggregation levels.

In most discussion, the elementary index number problem has been viewed as "lower level substitution" or "elementary index bias". An emerging body of research suggests that the geometric mean estimator for elementary indices adjusts for this bias fairly well.

The geometric mean is being introduced by more and more countries. The approach is to calculate a geometric mean of prices in both periods and then derive the price relative or calculate a geometric average of the price relatives. Both calculations will yield the same results.

References

1. Australian Bureau of Statistics; Australian Consumer Price Index: Concepts, Sources and Methods; http://www.abs.gov.au
2. Boskin M.J., Dulberger E.R., Gordon R.J., Griliches Z., Jorgenson D.W.; The CPI Commission: Findings and Recommendations; American Economic Review, May 1997 (Papers and Proceedings), 87(2), pp. 78-83.
3. Bureau of Labour Statistics, Planned Change In The Consumer Price Index Formula, 1998., http://www.bls.gov
4. Bureau of Labour Statistics, The Experimental CPI using Geometric Means (CPI-U-XG), 1998., http://www.bls.gov
5. Commission Regulation (EC) No 1749/96 on initial implementing measures for Council Regulation (EC) No 2494/95 concerning harmonised indices of consumer prices, OJ No L 229/3
6. European Commission, Report from the Commission to the Council on harmonization of consumer price indices in the European Union, COM/2000/0742 final
7. IMF; Croatia, Summary Methodology, Price Index: Consumer prices; http://www.imf.org
8. OECD, Main Economic Indicators, Comparative methodological analysis: Consumer and producer price indices, Volume 2002, Supplement 2, http://www.oecd.org
9. Schultz B., Choice of price index formulae at the micro-aggregation level: The Canadian empirical evidence, First Meeting of the Ottawa Group on Price Indices, Ottawa, 1994., http://www.ottawagroup.org/meet.shtml
10. Triplett J.E., Elementary indexes for a Consumer Price Index, Fourth Meeting of the Ottawa Group on Price Indices, Washington, 1998., http://www.ottawagroup.org/meet.shtml
11. Turvey R., Consumer price indices, An ILO manual, International Labour Office, Geneva, 1989.
12. Turvey R., Elementary aggregate indexes, 1999., http://turvey.demon.co.uk/index.htm
13. Woolford K., A Pragmatic Approach to the Selection of Appropriate Index Formulae, First Meeting of the Ottawa Group on Price Indices, Ottawa, 1994., http://www.ottawagroup.org/meet.shtml

SAMPLE SURVEY RESEARCH ON FINANCIAL RISK MANAGEMENT: A CASE OF CROATIAN COMPANIES

Ksenija Dumičić
University of Zagreb, Graduate School of Economics and Business, Department of Statistics
Kennedyjev trg 6, HR-10000 Zagreb, Croatia, kdumicic@efzg.hr

Mirjana Čižmešija
University of Zagreb, Graduate School of Economics and Business, Department of Statistics
Kennedyjev trg 6, HR-10000 Zagreb, Croatia, mcizmesija@efzg.hr

Anita Pavković
University of Zagreb, Graduate School of Economics and Business, Department of Finance
Kennedyjev trg 6, HR-10000 Zagreb, Croatia, apavkovic@efzg.hr

Abstract: The paper presents the authors' sample survey research findings on statistically significant dependence of usage of financial risks management instruments on characteristics of Croatian firms, such as ownership type and application of controlling function. The main activity and managers' need for further education about risk protection did not appear to be in significant dependence on managing risks activity. The liquidity risk appeared to be more serious than currency risk and interest rate risk. Low practical efficiency of instruments was the main reason for non-usage of more protection.

Keywords: financial risk management instruments, survey research, random sample, χ^2-test of independence, Levene's test for equality of population variances, t-test of means difference

1. Introduction

According to Jorion (2001), it is about a decade since financial industry emerged from pre-history of financial risk management, and now the circumstances are met for introducing Integrated Risk Management or Enterprise-wide Risk Management, which covers company – wide risks: business and non-business risks, where financial risks are considered to be a set of non-business risks including market risks, credit risks, liquidity risks and operational risks. In the praxis the role of CFO (Chief Financial Officer) becomes more and more important and more under stress, so they wish to share the weight of risk pressure using modified decisions and modified managing strategies. Financial risks come out the financial transactions, but, since, sooner or later, all the risks create financial consequences, as Peterlin (2003) said, indirectly all kinds of risks could be considered as to be financial. While economic risk is connected with uncertainty of producing real goods and services in an economy, financial risk is uncertainty connected with the price of financial contracts that perform the affirmation of these goods and services (shares, bonds, and currency itself), and this risk can not be eliminated, but could be only dispersed. After recognizing the type of risk, financial managers may use hedging, insurance or a kind of diversification strategy, prescribed instruments or financial limits.

Companies in transition countries such as Croatia are particularly exposed to the problem of financial risks, especially because they do not have enough experiences with protection against them. So, the authors found financial management instruments usage to be quite challenging topic to investigate.

The aim of this paper is to present the main results of an original empirical survey research conducted using random sample of Croatian companies in October 2004 about the extent to which large and medium-sized companies were using financial risks management instruments to protect themselves against three kinds of risks: liquidity risk, currency risk and interest rate risks. The authors have investigated the impact of companies' characteristics on attitude towards financial management means application.

The research hypothesis that financial managing in Croatian companies was independent on their selected characteristics were tested as follows: the hypothesis H1 was that the usage or non-usage of risk protection instruments was independent on Croatian companies' main activity; the hypothesis H2 stated that the usage or non-usage of protection instruments was independent on the type of ownership; the hypothesis H3 included the statement that the usage or non-usage of financial risk protection means was independent on the application of controlling function; and the last hypothesis H4 stated that the usage/non-usage of financial risk protection was independent on the managers' need for some additional education about financial protection services.

Further purpose of the paper was to discover if managers knew or used certain protection instruments for each kind of risks, why they did not use them more, or why they did not use these instruments at all.

In the analysis the appropriate statistical tests: Levene's test for equality of population variances, t-test for means difference and χ^2-test of independence were applied (for tests McClave et al., 2005, could be seen).

2. Survey Sample

Survey research based on a telephone interview as data collection mode with financial (or accounting) managers from a random sample of n=101 Croatian companies was carried out in October 2004. The methodology of stratified random sampling according to the number of employees as a stratification criterion was adopted. An approximately equal allocation of sampling units was applied: n_1=50 large (with more than 250 employees), with sampling fraction f_1=0.12; and n_2=51 medium-sized companies (51-250 employees), sampling fraction equals f_2=0.03, from each of respective strata was applied. Sampling frame used was a FINA's list of Companies from 2002. Later on, after the survey was finished, it was recognised that distinguishing the companies by number of employees was not that important for the main variable under study. But, because this was an originally conducted pilot survey for Croatia, as it has been described in more details in Dumičić et al. (2005a), it seemed to be important to test if the size of a company would influence or not their usage or non-usage of financial risks protection instruments.

Methods of estimation took into account the procedure of random sampling of units in each of defined strata, so, margin of errors could be given, as well. Considering unbiased point estimator of the proportion and with normal approximation of its sampling distribution, in each strata an interval estimate for the proportion with n=50 and 95% confidence level (z = 1.96), was within maximum margin of error of ∓ 14.2%.

In the sample the majority of 56% (23 out of 41) of financial risk protection instruments users were from large companies and 44% from those medium-sized. The users under study had got an average number of employees of 565 with coefficient of variation of 107%. Among non-users there was a majority of 55% of medium (33 out of 60) and 45% of large firms. An average number of employees for non-users was 371 with coefficient of variation of 83%. Also, 56% (28 out of 50) of large companies; and 43% (22 out of 51) those medium-sized were users of financial risk protection instruments. But, the dependence of protection instrument usage and company size was not shown to be statistically significant, because in testing the hypothesis of independence for categories in Table 1 the empirical χ^2 did not give

enough evidence for rejection of the null-hypothesis (χ^2=1.200, p-value=0.273), compare to the detailed research project description in Dumičić et al. (2005a).

Table 1. Usage of financial risks protection instruments by company size

Usage	Medium-sized companies	Large companies
Use	18	23
Don't use	33	27
Total	51	50

For companies under study that were using financial risks protection instruments an average annual revenue in the previous year (2003) was higher (35029725 €), than for the rest of companies that were not using them (16923357 €). With the assumption of different variances (Levene's empirical ratio F=15.355 and p-value=0.000), the t-test of the means difference for annual revenue for these two types of companies was applied, and, it showed that this difference was significant only with significance level α higher than calculated p-value= 0.066 (t=1.888). Also, in firms that apply risk protection, the revenue was growing more dynamically than in the rest of firms. The appropriate χ^2-test of independence gave the result which was highly significant, so it is possible that the revenue growth was statistically dependent on usage of financial risk protection, and vice versa (empirical χ^2=9.711, p-value=0.008).

3. Research Results

Considering the main economic activity of companies, for the purpose of testing the research hypothesis H1, the structure of the sample was as follows: 64% of companies were from industry sector (manufacturing; electricity, gas and water supply; and construction); and 36% from services sector (retail trade; wholesale trade; hotels and restaurants; transport, storage and communications; real estate, renting and business activities). Even though the majority of instruments users were from industry sector, the dependence of risk protection instruments usage/non-usage on the two roughly defined sectors of activity (industry sector and services sector) did not appear to be statistically significant (χ^2=0.466; p-value= 0.495), so, the research hypothesis H_1 could not be accepted.

In testing the hypothesis H2, the survey results showed that the companies that were using these instruments were more often (69%) registered as share holding companies than as limited ones, and, dependence of usage/non-usage of protection instruments on the type of ownership appeared to be significant only with significance level α higher than calculated p-value=0.071 (χ^2=3.261).

Further, the hypothesis H3 came out from the idea that companies that apply a controlling function in their business tend to apply some kind pf risk protection means, too. Among users of risks protection instruments there were 87% of companies with developed financial controlling and among non-users there was a majority of 65% of such companies. The χ^2-tests of independence allowed the rejection of the null-hypothesis (χ^2=5.097, p-value =0.024), so, upon this test, with α=0.05, the categories of risk protection usage and application of controlling were significantly dependant, so, the research hypothesis H_1 could be accepted.

The hypothesis H4 stated that active financial risk management of a company was influenced by the managers need for some additional education about financial protection services, or vice versa. It was also found out that 40% of managers interviewed wanted to get additional education about financial risk protection instruments services. When testing the independence of two categories mentioned in Table 2, the empirical χ^2 don't appear to be significant, so, the null-hypothesis was accepted as it would be possible (χ^2=0.684, p-value=0.408), and the alternative H_1 rejected.

Table 2. Demand for additional education about financial risk protection by company size

Education	Medium-sized companies	Large companies
Need	24	20
Don't need	27	30
Total	51	50

The reasons, mentioned by interviewed managers, for not using the protection instruments more often and more intensively are listed in Table 3, and this list indicates that the reasons seem to be quite different across the surveyed companies. "Protection instruments are not enough efficient" was the most often mentioned reason for that.

Table 3. The main reasons why companies did not use financial risk protection instruments more often (one answer was possible)

Reasons	Medium-sized companies	Large companies
Protection instruments are not well known	2	3
Protection instruments are not enough efficient	6	2
Protection instruments are too expensive	5	2
Financial market is not enough confident	1	2
Financial services have too many formal requirements	1	2
Something else	2	10

Also, the reasons for non-usage of the risk protection that were mentioned by interviewed managers are listed in Table 4. It could be noticed that certain reasons for no-usage of protection by surveyed companies appeared very differently. "Protection instruments are not well known" was the most often mentioned one.

Table 4. The main reasons why companies did not use financial risk protection instruments (one answer was possible)

Reasons	Medium-sized companies	Large companies
Protection instruments are not well known	14	6
Protection instruments are not enough efficient	12	2
Protection instruments are too expensive	0	3
Financial market is not enough confident	1	1
Financial services have too many formal requirements	3	2
Bad quality of financial services	1	0
Something else	2	12

Table 5 shows which instruments were known to financial managers interviewed and which of them were actually used, compare to Dumičić et al. (2005b). It is evident from the highest frequencies for liquidity risk that most of users were concentrated on protection

against this kind of risk. Currency risks, as well as interest rate risks, were not considered as to be so dangerous for Croatian managers. The frequencies for financial risk protection instruments "used" were in many cases smaller than for "known" category. This difference is the most evident for liquidity risk protection method called "credit capability analysis".

Table 5. No. of managers that were acquainted by or were using certain types of financial risk protection instruments (more than one answer was possible)

Against Liquidity Risk	Known	Used
Cash flow investments analysis	34	29
Analysis of assets, liabilities and sources	30	29
Credit capability analysis	38	20
Against Currency Risk	**Known**	**Used**
Netting	9	5
Leading and lagging	11	10
Selling prices policy	14	12
Assets and liabilities management	10	10
Currency forward	12	11
Currency futures	15	14
Currency swap	9	8
Currency options	8	6
Against Interest Rate Risk	**Known**	**Used**
Interest rate management at the money market	4	4
Forward rate agreements	4	3
Interest rates futures	6	5
Interest rate swap	5	3
Interest rate option	4	3
Caps, floors, and collars	3	3

4. Conclusion

After the survey research for the purpose of this paper was carried out the enterprises in Croatian transition economy seemed to be not enough aware of dangers that have arisen from financial risks. The results from 2004 Croatian companies sample survey have shown that two fifths of medium and large companies in Croatia did use instruments to protect themselves from them.

There was a variety of reasons for non-usage mentioned by interviewed managers, but, those that prevailed were: "Protection instruments are not well known" and "Protection instruments are not enough efficient".

The main reasons for not using protection more often seemed to be quite different across the surveyed companies, but the most significant reason was: "Protection instruments are not enough efficient".

The research results indicated that Croatian companies were not sufficiently acquainted with adequate protection instruments. Managers interviewed considered liquidity risk and currency risk to be the most serious types of risks, and interest rate risk did not seem to be that dangerous. These larger firms included in the sample were better informed and they used risk protection approach, instruments or services, more often than medium-sized ones. Numbers of companies that "used" these instruments were in many cases smaller than for the companies that only "knew" about them.

Considering research hypothesis, sample survey research results led to some new findings.

After testing the hypothesis H1, the Croatian company's usage/non-usage of financial risks protection instruments appeared not to be in statistically significant dependence on the main activity of these companies (H_0 accepted).

The test of independence of usage/non-usage of financial risks protection instruments on the type of ownership (the hypothesis H2), showed that the two categories could be considered as dependent (H_0 rejected) with significance level α higher than p-value=0.071, which is not highly significant.

The test of independence of usage/non-usage of financial risks protection instruments on applying the controlling function (the hypothesis H3), has shown that the empirical χ^2-value was significant with α=0.05, so the independence of categories under study could not be accepted as possible (H_0 rejected).

After testing the hypothesis H4, there was a surprise because nevertheless 40% of interviewed managers had got a desire for additional education about risk protection instruments, this characteristic was not in statistically significant dependence (H_0 accepted) on the fact that this financial managers used (or did not use) risk protection means.

Following the research findings, it should be recommended to banks and other financial institutions in Croatia, such as insurance companies, and consultants, to make stronger efforts to inform corporate clients about financial risks protection services, methods and instruments, and to suggest them what is available for application.

The limitation of this survey is that relatively a small sample size was applied, so, in perspective a bigger sample should be recommended. In a future sample survey the authors plan to include not only large and medium-sized, but also small Croatian companies to investigate the financial risks influence to this category of firms and their attitude towards this real life phenomenon, as well.

REFERENCES

[1] Dumičić, K., Čižmešija, M., Pavković, A., Andabaka, A. (2005a). Istraživanje primjene metoda upravljanja financijskim rizicima u hrvatskim poduzećima – Anketa na uzorku poduzeća. Projekt Ekonomskog fakulteta – Zagreb. (Unpublished).

[2] Dumičić, K., Dumičić, M., Cukrov, R. (2005b). Financial Risk Management Instruments Usage in Large and Medium-Sized Enterprises – Business Survey in a Transition Country. Bulletin of the 55nd Session of the International Statistical Institute, Sydney, Australia, April 2005, CD-rom.

[3] Jorion, Ph. (2001). Value, risk and control: a dynamic process in need of integration. In Financial Times Mastering Risk. Volume 1: Concepts. (Edt.: Pickford, J.). FT, Prentice Hall, Pearson, Edinburgh, pp. 119-124.

[4] McClave, J. T., Benson, P. G., Sincich, T. (2004). Statistics for Business and Economics. 9th Edt., Prentice-Hall International, London.

[5] Peterlin, J. (2003). Upravljanje financijskim rizicima – evidentiranje i izvješćivanje. Računovodstvo, revizija i financije, RRIF, 10, pp. 46-54.

MODELLING EXCHANGE RATE VOLATILITY IN CROATIA

Nataša Erjavec and Boris Cota
University of Zagreb, Faculty of Economics, Trg J.F.Kennedya 6, 10 000 Zagreb, Croatia
nerjavec@efzg.hr, bcota@efzg.hr

Abstract
There is no accepted model of firm behaviour subject to risk arising from fluctuations in exchange rates and other variables. The consensus on the appropriate method for measuring exchange rate volatility does not exist. A wide variety of methods that could be used to generate predicted values of exchange rate uncertainty have been proposed. We employed GARCH type models and used data on the percentage change of daily exchange rate between Croatian kuna and US dollar.

Keywords: volatility models, exchange rate volatility, GARCH

1. Introduction

The aim of this paper is to define a good volatility model that would capture the basic characteristics of the percentage changes of daily exchange rate between Croatian kuna and US dollar. Forecasting volatility of financial time series is the major task in many financial applications. A model should capture the commonly held facts about financial time series; such as volatility clustering, pronounced persistence, mean-reversion or asymmetry. Many financial series display volatility clustering or autoregressive conditional heteroscedascity, *i.e.* it is commonly seen characteristic in financial application that large changes are followed by large changes and small changes are often followed by small changes. Volatility clustering implies a strong autocorrelation in squared returns. The result of such a volatility clustering is that external shocks influence the expectation of volatility many periods in future, *i.e.* the volatility exhibits persistence.

Another characteristic of financial series volatility is that it evolves over time in a continuous manner, *i.e.* the volatility jumps are rare. Volatility does not diverge to infinity, which means that volatility varies over some fixed range. Mean-reversion of volatility is a further implication of volatility clustering. It means that there is a normal level of volatility to which volatility returns. As a consequence, there is a convergence in the term structure forecasts to the long-term average volatility level. Furthermore, innovations may have an asymmetric impact on volatility. The asymmetric structure of volatility generates skewed distribution of forecasts.

In order to define an adequate volatility model we analysed all mentioned characteristics of financial series. We employed data on percentage change of daily exchange rate between Croatian kuna and US dollar. The liberalization of capital flows and large increase in the scale of cross-border financial transactions have increased exchange rate movements. The magnitude of exchange rate movements increased in countries with underdeveloped markets and without stable economic policies (Prasad *et al.*, 2003). High exchange rate volatility is consequence of currency crises in emerging market economies. Some other changes in world economy may have reduced the impact of exchange rate volatility. For example, extensive use of financial hedging instruments could reduce firm's vulnerability to risks arising from volatile currency movements. Moreover, fluctuations in different exchange rates may have offsetting effects on profitability of multinational firms.

There is no consensus on the appropriate method for measuring exchange rate volatility. Commonly accepted model of firm behaviour subject to risk arising from fluctuations in exchange rates and other variables does not exist. Therefore, theory cannot provide definitive guidance as to which measure is most suitable. The scope of the analysis will to some extent dictate the type of measure used (IMF, 2004). There is a wide variety of methods that could be used to generate predicted values of exchange rate. We employed the univariate GARCH type of model proposed by Bollerslev (1986) in which the mean and the variance of the series are simultaneously modelled allowing the variance term to depend upon the lagged variances as well as lagged squared residuals[1]. On the basis of such a model part of the volatility can be forecasted (based on the past values of the exchange rate) and firms engaged in trade would make an effort to develop such forecasts[2]. Such a model can be further used to forecast the absolute magnitude of returns and to predict the quantities.

2. Model

The dependent variable, the input to the GARCH volatility model, is always a return series. The returns are assumed to be generated by a stochastic process with time-varying volatility. GARCH model consists of two equations. The first is the conditional mean equation and the second is conditional variance equation. In this study, we defined the mean equation in its simplest possible form, *i.e.*

$$r_t = \mu + u_t, \tag{1}$$

where μ is a constant and denotes the average of returns over the data period and $u_t \approx N(0, \sigma_t^2)$.

In GARCH model variance term depends upon the lagged variances as well as the lagged squared residuals. This allows for persistence in volatility with a relatively small number of parameters. The variance model for the standard GARCH(p,q) model is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{2}$$

The simplest case of GARCH model is the generic or "vanilla" GARCH(1,1) model:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 . \tag{3}$$

σ_t^2 is known as the conditional variance since it is a one-period estimate for the variance calculated based on any past information thought relevant. The conditional variance is changing, but the unconditional variance of u_t under a GARCH(1,1) specification is constant and given by:

$$Var(u_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)} \tag{4}$$

as long as $\alpha_1 + \beta_1 < 1$. The model with such a property is a stationary GARCH model. For the model, the conditional variance forecasts converge upon the long-term average value of the variance as the prediction horizon increases. For $\alpha_1 + \beta_1 \geq 1$ the unconditional variance is not defined.

The case of $\alpha_1 + \beta_1 = 1$ is known as a "unit root in variance" also termed Integrated GARCH or IGARCH. This phenomenon is commonly observed in practice and it leads to imposing the constraint $\alpha_1 + \beta_1 = 1$ in a GARCH(1,1) model. Thus, IGARCH models are unit-root GARCH models, with the key feature that the impact of the past squared shocks $\eta_t = u_{t-i}^2 - \sigma_{t-i}^2$ for *i*>0, on u_t^2 is persistent. Consequently, IGARCH volatility forecasts do not mean-revert at all. For such models, the conditional variance forecasts will tend to infinity as the forecast horizon increases. The IGARCH(1,1) model can be written as:

[1] In the GARCH model the conditional variance is modelled as an ARMA process

[2] Estimation and forecasting volatility is important feature of financial risk modelling, *i.e.* in risk management volatility modelling provides a simple approach to calculating value at risk of a financial position.

$$\sigma_t^2 = \alpha_0 + (1 - \beta_1) u_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \quad \text{where } 0 < \beta_1 < 1. \tag{5}$$

The GARCH(1,1) model is the most common specification for GARCH volatility models, being relatively easy to estimate and generally having robust coefficients that are interpreted naturally in terms of long-term volatilities and short-run dynamics, (Alexander, 2001). The sizes of the parameters α_1 and β_1 determine the short-run dynamics of the resulting volatility time series. Large β_1 coefficient indicates that shocks to conditional variance take a long time to dye out (so volatility is persistent) while large α_1 coefficient specifies that volatility reacts quite intensely to market movements.

A feature of many financial series that is not captured by GARCH model is the "asymmetry effect" also known as the "leverage effect" or "risk premium effect". In the context of financial time series analysis the asymmetry effect refers to the characteristic of the time series on asset prices that an unexpected drop tends to increase volatility more than an unexpected increase of the same magnitude (or that bad "news" tends to increase volatility more than the "good news". GARCH models do not capture this effect since the lagged error terms are squared in the equations for the conditional variance, and therefore a positive error has the same impact on the conditional variance as a negative error. There are a number of ways of parameterising this idea. We employed the so-called GJR model of Glosten, Jagannathan and Runkle (1993) which is a GARCH(1,1) model where the variance equation is defined:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + d_1 u_{t-1}^2 I_{u>0}(u_{t-1}) \tag{6}$$

with I being an indicator function, in this case, for $u > 0$. Negative value of d_l means that negative residuals tend to increase the variance more than the positive ones.

GARCH models are popular forecasting models when dealing with financial time series. When evaluating the accuracy of a particular model for computing conditional variance forecasts we calculated the 1-step ahead forecasts at the forecast origin h. We used the recursive formula where the forecasts of the early periods are used in the later periods.

When forecasting the conditional variance the observed values of the series are not available for comparison. As a proxy, it is traditional to use the squared values of the data. The forecasts can then be compared by with this proxy and the forecasts can be evaluated in the same way as forecasts of the series itself. We employed standard measures of forecasting accuracy derived from the forecast errors namely; mean absolute error MAE, mean absolute percentage error MAPE, mean squared error MSE and root mean squared error RMSE.

3. Empirical results

We used data on the percentage changes of the log of daily exchange rate between Croatian kuna and US dollar (K/$US) from 1 January 1997 to 15 January 2005. Excluding days for which no prices are quoted (New Year's Day, etc.), this resulted in a total of N = 2013 observations. Percentage changes of the log of daily exchange rate are calculated from indices of daily nominal exchange rate HRK vs. USD (CNB midpoint exchange rate, 2001=100). Data are taken from the Croatian National Bank. Summary statistics of the data show that the mean of the series is 0,000478. It is almost zero indicating that the average change is zero. The skewness coefficient of -0.14237 indicates that the distribution is negatively skewed which is a common feature of many financial series. The kurtosis coefficient (excess) which is a measure of the thickness of the tails of the distribution equals 0.55969 and is significantly higher than zero. A Normal distribution has excess of zero implying that the assumption of normality for the distribution of exchange rate changes is dubious for the series.

As (log) exchange rates are to a rough approximation usually described as a random walk (Tsay, 2002) we consider a model (1) where a conditional mean equation only includes an intercept. The result of the OLS estimation with standard error is:

$$r_t = \underset{(0,0034)}{0,0005} + u_t, \qquad Var(u_t) = 0,0234 \tag{7}$$

The analysis of the correlogram of the data indicates no serial correlation. However, the sample autocorrelation function (SACF) and sample partial autocorrelation function (SPACF) of the squared series indicate the possibility that the changes in exchange rate are not independent. Combining the two plots, it seems possible that changes in exchange rate are serially uncorrelated but dependent. However, the dependence is not so pronounced on the graphs and it is important to perform the formal testing for ARCH effects. The results give Ljung-Box Q-Statistics for the residual series of Q(24)= 27.5951 (p-value=0.2775) and for the squared series Q(24)= 63.1765 (p-value 0.0000) which is highly significant resulting in a clear rejection of the homoscedascity assumption.

In order to model a variance equation (3) the GARCH(1,1) showed to be the most parsimonious. The parameters of the model are estimated by maximum likelihood[3] under the commonly employed assumption that the conditional distribution of the errors is normal[4]. Thus, a joint estimation of the mean and variance equations is:

$$r_t = \underset{(0,0032)}{0,0002} + u_t, \qquad \sigma_t^2 = \underset{(0,0002)}{0,0004} + \underset{(0,0063)}{0,0270} u_{t-1}^2 + \underset{(0,0110)}{0,9561} \sigma_{t-1}^2 \tag{8}$$

From the volatility equation the implied unconditional variance of u_t is 0,0237 which is very close to that of equation (7). However, t-ratio of the parameter in mean equation suggests that the parameter is insignificant. In the variance equation, all coefficients are significant at 5% level. SACF and PACF of the standardised shocks $\hat{v}_t = \frac{\hat{u}_t}{\hat{\sigma}_t}$ and squared process $\hat{v}_t^2$ fail to suggest any significant serial correlation in the two processes. More specifically, Ljung-Box Q-Statistics for $\hat{v}_t$ series are Q(12)= 16.6429, p-value= 0.1635 and Q(24)= 27.6810, p-value = 0.2737. For the squared series Q(12)= 13.2571, p-value = 0.3506 and Q(24)= 22.4309, p-value = 0.5536. It is evident that there is no autocorrelation left in the squared standardised residuals and the model seems to be well specified. When evaluating the accuracy of a model for computing conditional variance forecasts we calculated the 1-step ahead forecasts, at the forecast origin h=1895. The measures of forecast accuracy derived from the forecast errors are presented in Table 1.

The characteristic of financial time series is that the distribution is leptokurtic. For allowing for kurtosis it is possible to explicitly allow for non-normality in the conditional distribution as well as the unconditional distribution when estimating GARCH models. Assuming that u_t follows a standardized Student t-distribution we re-estimated the GARCH(1,1) model jointly estimating the degrees of freedom of the Student t-distribution and the parameters of the model. A joint estimation of the mean and variance equations for the Student t-distribution GARCH(1,1) model is:

$$r_t = \underset{(0,0033)}{0,0013} + u_t, \qquad \sigma_t^2 = \underset{(0,0002)}{0,0004} + \underset{(0,0080)}{0,0315} u_{t-1}^2 + \underset{(0,0125)}{0,9526} \sigma_{t-1}^2. \tag{9}$$

[3] We used quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimisation method

[4] Bollerslev and Wooldrige (1992) showed that the maximum likelihood estimates of the parameters of the GARCH model assuming Gaussian errors are consistent even if the true distribution of the innovations is not Gaussian.

u_t follows Student t-distribution with estimated degrees of freedom of 13.84 ≈14 and standard error of 3.5502. Ljung-Box Q-Statistics for the standardised residual series of the model are Q(12)= 16.5531, p-value 0.1672 and Q(24)= 27.5504, p-value = 0.2795. For the squared series Q(12)=13.4008, p-value =0.34060 and Q(24)= 22.1384, p-value = 0.5710. Since there is no evidence of autocorrelation in the squared standardised residuals we can conclude that the model is well specified. Consequently, we cannot reject the hypothesis of using a standardized Student t-distribution with 14 degrees of freedom at 5% significance level.

From the volatility equation, the implied unconditional variance or long-term variance of u_t is 0,0239. Parameters estimates and their standard errors are close to those of Gaussian model. Large lag coefficient β_1 of 0,9526 indicates that shocks to long-run (conditional) variance take a long time to dye out. The value of error coefficient α_1 of 0,0315 specifies that the volatility does not react quite intensely to market movements. For the model we plotted estimated conditional variance, Figure 1. The standard measures of forecast accuracy are given in Table 1.

In both models, non-negatively constrains of the parameters are reached. However, the fitted models show that the sum of the parameters is close to 1. In Gaussian model: $\alpha_1 + \beta_1 = 0,9831$ and in the Student GARCH(1,1) model: $\alpha_1 + \beta_1 = 0,9841$, signifying that the estimated model is close to being nonstationary. The results obtained from the estimation of the IGARCH(1,1) model (5) are:

$$r_t = \underset{(0,0034)}{0,0029} + u_t, \qquad \sigma_t^2 = \underset{(0,0001)}{0,0002} + 0,0660 u_{t-1}^2 + \underset{(0,0088)}{0,9340} \sigma_{t-1}^2 \tag{10}$$

Ljung-Box Q-Statistics for the residual series is Q(24)= 22.8375, p-value=0.5294 and for the squared series Q(24)= 18.2368, p-value 0.7914. The parameter estimates are close to those of the GARCH(1,1) models estimated above but there is a major difference between the models. The unconditional variance of u_t, hence of the r_t, is not defined under above IGARCH(1,1) model. This seems hard to justify for the percentage change series. From the theoretical point of view, the IGARCH phenomenon might be caused by occasional level shifts in volatility. The actual cause of persistence in volatility deserves a careful investigation. For the model the measures of the forecast accuracy are given in Table1.

Table 1: Forecast evaluation measures for forecasts from selected models

Measures of forecast accuracy	Model		
	Gausian GARCH(1,1)	Student t-distribution GARCH(1,1)	IGARCH(1,1)
MAE	0.02110	0.02101	0.02033
MAPE	32580.58110	32122.32491	37890.70633
MSE	8.84332e-04	8.83355e-04	9.06070e-04
RMSE	0.02974	0.02972	0.03010

To test for the asymmetry in the Student t- distribution GARCH(1,1) model we employed a GJR model (6). The joint estimation of mean and variance equations gives:

$$r_t = \underset{(0,0034)}{0,0011} + u_t, \qquad \sigma_t^2 = \underset{(0,0002)}{0,0004} + \underset{(0,0089)}{0,0349} u_{t-1}^2 + \underset{(0,0119)}{0,9555} \sigma_{t-1}^2 - \underset{(0,0115)}{0,0109} u_{t-1}^2 I_{u>0}(u_{t-1}), \tag{11}$$

with estimated degrees of freedom for the Student t-distribution of 13,80 (3,533).

The asymmetry coefficient equals -0,0109 and has an expected negative sign, although its t-ratio is only -0,9477. Thus, the asymmetry is not statistically significant in the model which is not unusual finding for exchange rate, (Verbeek, 2000). The asymmetry is also rejected by the sign bias test (SBT) the negative sign bias test (NSBT) and the positive sign bias test (PSBT) proposed Engle and Ng (1993).

To find the best model for forecasting volatility of exchange rate we compared the standard measures of forecast accuracy derived from the forecast errors for the estimated models, Table 1. On the basis of these measures, Student t-distribution GARCH(1,1) model (10) produces the most accurate forecasts of the conditional variance.

The graph of the conditional variance forecasts from the Student t- GARCH(1,1) model, Figure 2, shows that model approaches a constant value within a few steps.

Figure 1: Estimated conditional variance from Student t-dist. GARCH(1,1)

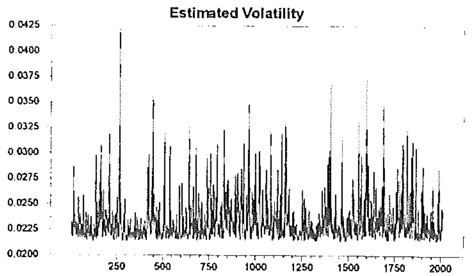
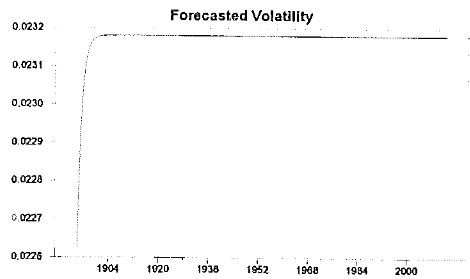


Figure 2: Forecasts of conditional variance from Student t-dist. GARCH(1,1)



4. Conclusion

The aim of the paper is to find a good model for forecasting the conditional variance or volatility of the percentage change of exchange rate. The model is chosen by its ability to forecast volatility. There is no accepted model of firm behaviour subject to risk arising from fluctuations in exchange rates and other variables. Therefore, theory cannot provide definitive guidance as to which measure is most suitable. We employed GARCH(1,1) model. Estimation of several models shoved that Student t-distribution GARCH(1,1) produces the most accurate forecasts of the conditional variance.

References:

Alexander, C. (2001). *Market models: A guide to financial data analysis*, John Wiley & Sons; Chichester, UK.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroscedasity, *Journal of Econometrics* 31, 307-327.

Bollerslev, T. and J.M. Wooldrige (1992). Quasy-maximum likelihood estimation and inference in dynamic models with time varying covaraince*s*, *Econometric Reviews,* 11(2), 143-172.

Engle, R. and V.Ng. (1993). Measuring and Testing the Impact of News on Volatility, *Journal of Finance,* 48, 1749-1778.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the relation between expected value and the volatility of excess returns on stocks, *Journal of Finance*, 48,1779-1801.

IMF (2004). Exchange Rate Volatility and World Trade, *Occasional Paper* No. 28.

Prasad, E., K.Rogoff, S. Wei and M.A. Kose (2003). *"Effects of Financial Globalization on Developing Countries: Some Empirical Evidence", IMF*

Tsay, R. S. (2001). *Analysis of Financial Time Series*, Wiley and Sons, Chichester, UK.

Verbeek, M. (2000). A Guide to Modern Econometrics, John Wiley & Sons, New York

MACROECONOMIC MODELLING RELAXING THEORETIC ASSUMPTIONS IN THE CROATIAN FINANCIAL SPHERE

Elza Jurun, Snježana Pivac
University of Split, Faculty of Economics
Matice hrvatske 31
21000 Split, Croatia
elza@efst.hr, spivac@efst.hr

Abstract
This paper establishes an approach to macroeconomic modelling under conditions of relaxing the assumptions of the classical model. Macroeconomic variables of the Croatian financial market are integrated in the same multiple linear regression model. After parameter estimation by the classic method the problem of multicolinearity appears. Based on the same database the same variables were modelled by mathematical programming. The comparative analysis of results confirms the advantages of linear programming especially when the problem of multicollinearity appears.

Keywords: parameter estimation, relaxing theoretic assumptions, linear programming, multicollinearity, Croatian financial market

1. INTRODUCTION

A faster development and wider application of the statistic modelling in the financial sphere started in the 1970s. Those were the years when various financial innovations on the capital markets of the developed countries emerged and their modelling required stochastic methods. Contrary to these trends, the Croatian financial market is still underdeveloped and poor in financial instruments. Development of the capital market in Croatia started at the end of the 1990s along with the process of transition and privatisation, which still remain its important determinant. As a consequence the stock market, although itself underdeveloped, imposed itself as one of the more important segments of the overall capital market. The situation on the developed world markets is opposite. This characteristic of the Croatian capital market, along the processes of transition and privatisation is significantly influenced by traditional financing through loans, lack of knowledge of functioning and potentials of the capital market, monopolistic banking system, inadequate legislation as well as the overall government attitude to the capital market.

2. NERVOUSNESS ON CROATIAN STOCK MARKET

After the period of dramatic war and post-war changes in the socio - economic system, Croatia has plunged in an equally unstable period of economy in transition. Within a relatively short period of a decade the transition period has encouraged fundamental changes in all aspects of life (sociological, cultural, moral, legal ...). Those changes caused feedbacks, which in the economic sphere are manifested through higher qualitative and quantitative instability of movement in almost all segments of economy.

For the purpose of this paper, this thesis will be illustrated remaining within the framework of the characteristics of movement of the basic indicators of the Croatian financial market. . The estimates show that the first effects of globalisation started determining the situation on the Croatian market after its "stabilisation" in the late 1990s. It follows the period of transition economy of the complex of financial sector, and therefore we can claim that it was then that Croatia became an integral part of the world financial market.

By choosing the variables the presence of both characteristic spheres of the financial market is wanted to be achieved: the one connected with the cash flow as well as the one connected with the securities flow. Apart from that, after the statistical-mathematical analysis only the variables with confirmed linear connection in the period from 1999 to 2005 have been introduced in the model.

In order to illustrate the movement on Croatian financial market more vividly and clearly, fundamental indicators of changes in its particular segments were chosen. In the developed economies movements in interest rates for short-term government securities are widely accepted as an indicator of stability. Due to the underdevelopment of the Croatian economy in relation to the developed countries, this indicator will be considered as a qualitative one, rather than emphasising its quantitative dimension. Figure 2.1. besides the average daily interest rates for each month on the Zagreb Money Market in the period from 1999 to 2004 also indicates weighed interest rates for treasury notes which are due in 35 days. Therefore, let the graphic illustrations replace further comments.

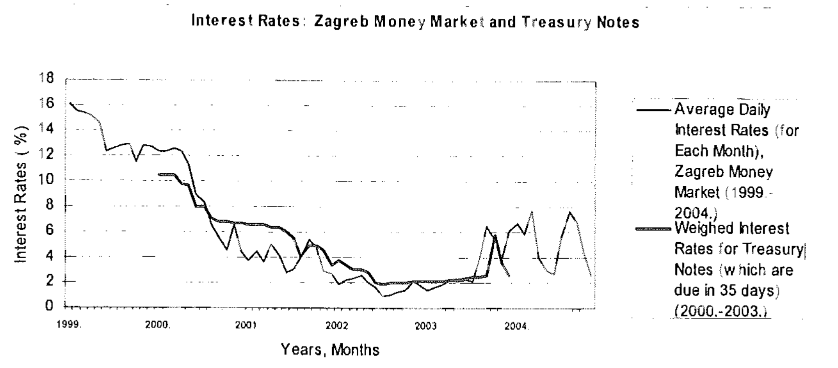


Figure 2.1: Interest Rates: Zagreb Money Market and Treasury Notes

As illustrated at Figure 2.1. the average daily interest rates on the Zagreb Money market in the given period range from 1% to 16.13 %. The lowest weighted interest rate for treasury notes due in 35 days amounted to 1.88% whereas the highest was 10.5%.

Observing the movement of average daily interest rates for each month it can be noticed not only their significant quantitative fluctuation, but also significantly different global levels of movement in particular years of the observed period, as it is shown at Figure 2.2.

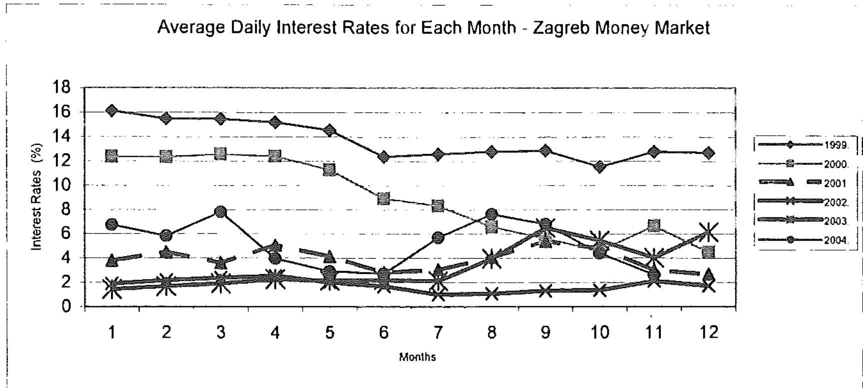


Figure 2.2: Interest Rates - Zagreb Money Market.

Instability of interest rates is combined with considerable fluctuations in other segments of financial activities, especially those under powerful influence of the exchange rate of the Croatian currency. The kuna exchange rate depends significantly on the money supply and interventions of the Croatian National Bank which aim at maintaining it within the desired limits.

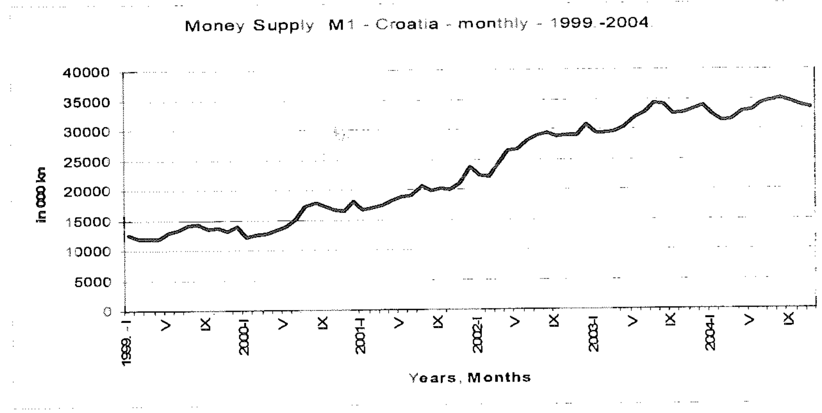


Figure 2.3: Money Supply M1 - Croatia

This paper models the dependence of CROBEX on the above mentioned variables of the Croatian financial market. CROBEX is the official index of the Zagreb Stock Exchange plc. Its issuing started on September 1st 1997. The basic date is July 1st 1997, and the basic value 1,000. It is weighted according to market capitalisation while the weight of a certain share is limited to 35%. If one of the CROBEX shares was not traded on the particular day, its last closing price is used. It is calculated continuously during the trade by using the last price in the equation:

$$I^j_t = \frac{M^j(t)}{K_T \cdot M(0)} \cdot 1000, \qquad (1)$$

where:

$M^j(t)$ - stands for the market capitalisation of the CROBEX shares on day t and time j,

$M(0)$ - stands for the market capitalisation on the basic day,

$M(0)$ - 12,942,773,686,

K_T - the factor of adaptation of the index basis on the auditing day T,

K_T - 1.1101848.

In order to be included in CROBEX, shares should meet the following conditions:

- liquidity – shares were traded on more than 75% of the total number of trading days in the observed period or are among the 25% of shares classified according to the number of trading days,
- quotation on the Stock Exchange for more than six months,
- market capitalisation higher than the median of market capitalisation of ordinary shares on the last day of the observed period.

3. PARAMETERS ESTIMATION IN MULTIPLE LINEAR REGRESSION MODELS

The database consists of time series of 72 monthly data on CROBEX movement, (CROB), average daily interest rates on the Zagreb Money Market (IR), money supply (M1) and interventions of the Croatian National Bank (INT). The data refer to the period from January 1999 to December 2004.

3.1. Parameters estimation by classic method

$$CR\hat{O}B = \hat{\beta}_0 + \hat{\beta}_1 \cdot IR + \hat{\beta}_2 \cdot M1 + \hat{\beta}_3 \cdot INT \quad (2)$$

Table 3.1. Parameters estimation of model (2) by classic method

Independent Variable	$\hat{\beta}_i$	t-ratio
Intercept	0,64329	9,6055
IR	-0,01032*	-2,7118
M1	0,18267*	8,9107
INT	0,00476**	1,7061

R^2=0,7863,
*significant at α =1%,
**significant at α =10%.

As it was expected, this model is statistically significant according to all the statistic criteria. It explains the dominant and important influence of the quantity of money supply, movement of the short-term interest rates and interventions of the Croatian National Bank on the CROBEX movement, as well as of the linearity of their interdependence. Testing the model, not only according to statistic but also according to econometric criteria, the multicolinearity was detected.

3.1.1. Testing the multicolinearity problem

Table 3.2. The correlation matrix of independent variables

	IR	M1	INT
IR	1,00	-0,69*	-0,24*
M1	-0,69*	1,00	0,17
INT	-0,24*	0,17	1,00

*significant at α =5%,

As illustrated in the above table there is, indeed, a statistically significant linear correlation between the interest rate and the money supply, and the correlation between the interest rate and the interventions is statistically more significant than the one between money supply and interventions on the money market. It leads us to the conclusion that the interest rate is a variable which in this model represents a significant factor of multicolinearity. Therefore Farrar - Glauber test was carried out. The value of the determinant of the matrix of the correlation coefficient of independent variables is: $\det R = 0.4937$. The calculated value of the chi-square test:

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2k+5)\right] \ln \det R = 48{,}114 . \quad (3)$$

When it is compared with the table value for $df = \frac{1}{2}k(k+1) = 6$:

$$\chi^2_{tab} = 12{,}59 , \quad (4)$$

it can undoubtedly be concluded that the multicolinearity exists.
Therefore the multiple linear regression model was estimated without including interest rates as independent variables.

3.1.2. Parameters estimation without including interest rates as dependent variable

$$CR\hat{O}B = \hat{\beta}_0 + \hat{\beta}_1 \cdot M1 + \hat{\beta}_2 \cdot INT \quad (5)$$

Table 3.3. Parameters estimation without including interest rates as dependent variable

Independent Variable	$\hat{\beta}_i$	t-ratio
Intercept	0,49163*	12,7618
M1	0,22056*	14,0631
INT	0,00612**	2,1290

R^2=0,76287,
*significant at α =1%,
**significant at α =5%.

For this model too statistic and common econometric tests have been carried out. It has been established that neither the problem of heteroskedasticy nor autocorrelation of errors of relation exist, and that the level of multicolinearity is below the one that would according to common levels of significance present a problem. Namely, for this model, the calculated value according to Farrar-Glauber test is $\chi^2 = 1.393$, and it does not exceed $\chi^2_{tab} = 7.815$.

In order to estimate the preciseness of the results from one more methodologically different aspect the model has been additionally estimated by linear programming using the norm l_∞.

3.2. Parameters estimation by linear programming

The linear regression model provides the possibility of parameters estimation by use of the linear program with l_∞ norm, i.e. criterion of maximal absolute deviation maximisation.
The function is introduced:

$$f(\hat{\beta}) = \left\| X\hat{\beta} - Y \right\|_\infty = \max_{i=1,\dots,n} \left| (X\hat{\beta} - Y)_i \right| \quad (6)$$

To minimise the maximal absolute deviations the problem becomes:

$$Min_{(\beta)} \max_{i=1,\dots,n} \left| (X\hat{\beta} - Y)_i \right| \quad (7)$$

The substitution is introduced:

$$z = \max_{i=1,\dots,n} \left| (X\hat{\beta} - Y)_i \right| . \quad (8)$$

Accordingly, the regression solution in the form of linear programme, where the CROBEX is the dependent variable, becomes:

$$\begin{aligned} &Min\ z \\ &X\hat{\beta} - CROB \le z\,e \\ &-X\hat{\beta} + CROB \le z\,e, \\ &z \ge 0 \ , \end{aligned} \quad (9)$$

where e is the unit vector.

Table 3.4 Parameters estimation by linear programming

$\hat{\beta}_0$	$\hat{\beta}_1^{\,+}$	$\hat{\beta}_1^{\,-}$	$\hat{\beta}_2^{\,+}$	$\hat{\beta}_2^{\,-}$	$\hat{\beta}_3^{\,+}$	$\hat{\beta}_3^{\,+}$	z
0,84220	$3{,}57\times10^9$	$3{,}57\times10^9$	0,15756	0	0,4977	0	0,23876

As it is obvious in table 3.4: $\hat{\beta}_1^{+} + \hat{\beta}_1^{-} = 0$; which means that practically variable IR has not statistically significant influence on the dependent variable.

After parameters estimation by classic method and by linear programming using norm l_∞ their comparative analysis has been carried out.

Valuating the results of the model from the point of view of their application with purpose of control and planning the significance of both approaches has been established. Namely, the estimated values in both cases adapt to actual data in a satisfying way and therefore justify their implementation not only for the purpose of analysis.

4. FINAL CONSIDERATIONS

Using the same multiple regression model this paper links the macroeconomic relations which reflect the cash flow with the movement on the stock market, as the dominant segment in the securities flow on the Croatian financial market. Modelling is carried out under conditions of relaxing the assumptions of the classical model. Along with that, after the statistical-mathematical analysis only those variables whose indisputable linear connection for the period between 1999 and 2005 was established were introduced in the model. On basis of the same database the same variables were first modelled by the classic method, and then the parameters were estimated by mathematical programming.

From the point of view of econometric criteria model estimation by linear programming has a number of advantages.

After specification and the overall procedure of parameter estimation by the classic method, statistic econometric criteria require each particular parameter as well as the model as a whole to be tested. On the other hand, model estimation by linear programming offers completely valid results from the point of view of all statistic and econometric criteria already on the first step. Such mathematical processes have been built in the procedure which were able, in our case, to offer parameter estimation but were also able to identify the variable which is the dominant factor of multicolinearity in the model. Namely, apart from all parameter estimation fulfilling both statistic and econometric criteria, the model was estimated without the interest rate variable which is highly correlated with other regression variables of the multiple model.

Therefore, wherever it is possible it is suggested parameter estimation using mathematical programming which is undoubtedly, according to statistic and econometric quality of estimation, superior to the classic estimation methods, especially under conditions of relaxing the assumptions of the classical model.

REFERENCES

1. Benninga, S., *Financial Modelling*, The MIT Press, Cambridge, 2000.
2. Jurun E., Pivac S., Zlobec S., *Influence of Fiscal Policy in Transitional Environment*, Proceedings of the 5th International Coference on Enterprise in transition, Split-Tučepi, 2003.
3. Prohaska, Z., *Analiza vrijednosnih papira*, Infoinvest, Zagreb, 1996.
4. Zlobec, S., *Stable Parametric Programming*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
5. Zlobec S., *Authorized lessons: Solving Regression Problems by Linear Programming,* Department of Mathematics and Statistics, McGill University, Montreal, 2003.
6. The Croatian National Bank: http:// www.hnb.hr
7. The Zagreb Money Market: http:// www.trzistenovca.hr
8. The Zagreb Stock Exchange: http:// www.zse.hr

A NONLINEAR MONETARY MODEL OF INFLATION

Alenka Kavkler, Bernhard Böhm

Faculty of Economics and Business, Razlagova 14, 2000 Maribor, Slovenia

Institute for Mathematical Methods in Economics, Vienna University of Technology, 1040 Vienna, Austria

e-mail: alenka.kavkler@uni-mb.si, bernhard.boehm@tuwien.ac.at

Abstract

This paper investigates a well known monetary model of the inflation process. The basic tool for identification and estimation of the model equations is the smooth transition regression approach. From the simulation of the estimated nonlinear system asymmetric policy reactions can be derived.

Keywords: Smooth Transition Regression (STR), Quantity Equation, Phillips Curve, Okun's Law.

1. Introduction

This paper is devoted to an investigation of asymmetric policy reactions. It focuses on the monetary approach to inflation control and uses a simple three equation model that is specified with the use of the smooth transition regression approach. The next section presents the theoretical specification, followed by a linear estimation attempt used as a starting point. The nonlinear estimation approach is given in section 4 and the policy simulations are presented in the final part.

2. Theoretical specification

We start with a model of monetary inflation theory which can be shortly characterized by an equation describing the monetary system augmented by a Phillips curve and the equation of Okun's law. The simple elementary system is given by (see [3]):

$$m_t = x_t + \pi_t,$$
$$\pi_t = \pi_t^* - b(u_t - u^*),$$
$$u_t - u_{t-1} = -a(x_t - x^*).$$

These three equations determine three unknown variables: real growth rate (x_t), inflation rate (π_t) and unemployment rate (u_t) in terms of given monetary growth. The first equation representing the quantity equation in growth rates (assuming constant velocity) will be substituted by a demand for real money equation, which may not be homogenous of degree 0 in its arguments. We assume the nominal money stock to be given by the monetary authority. The equilibrium in the monetary sector can be described by a general equation expressing the relationship between money stock, output and prices, respectively interest rates. The Phillips curve relates inflation to the deviation of the unemployment rate from its natural rate (u^*) augmented by backward and forward inflationary expectations. Okun's law provides a relationship between the change in unemployment rate and the deviation of the actual from the trend rate of real output growth (x^*). By considering an Okun type relationship and additional supply effects in the Phillips curve the resulting model exhibits essential features of the "triangle model" as defined by Gordon ([4]). Supply effects may include tax changes, rates of change in import or oil prices, or the change in the unemployment rate which may give rise to hysteresis. The (excess) demand side is typically represented by the unemployment gap and may also incorporate lagged effects of the growth of the stock of money.

This monetary approach to an explanation of inflation will be applied to quarterly data from West Germany between 1970:1 and 1998:4.

3. Linear econometric model

The econometric model employs transformations of variables into growth rates. In a preliminary specification, all equations are modelled as linear relationships. This simplifies the search for an appropriate nonlinear specification. On the basis of these empirical results further investigations will have to reveal any remaining nonlinearity in these relations.

The first equation specifies real money demand growth (m_t) as dependent variable explained by adjustment terms representing the adjustment of real money growth to output growth (x_t) and prices. We obtained the following results by the least squares estimation.

Dependent Variable: m_t

Sample (adjusted): 1970:1 1998:4 (Included observations: 116)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
const.	0.023727	0.005669	4.185161	0.0001
x_t	0.183880	0.127073	1.447037	0.1508
x_{t-4}	0.211157	0.098361	2.146751	0.0341
$\Delta\pi_{t-1}$	-0.582542	0.264546	-2.202045	0.0298
m_{t-1}	0.883156	0.047842	18.45990	0.0000
m_{t-4}	-0.267305	0.050737	-5.268417	0.0000
π^f_t	-0.005449	0.001437	-3.792490	0.0002
dummy1	0.112028	0.016734	6.694747	0.0000

$R^2 = 0.855$, S.E. = 0.022, SSR = 0.054, AIC = -4.690

Table 1: OLS results for the surrogate quantity equation

The Phillips curve is modelled according to the considerations in Böhm ([1]). The inflation rate (π_t) depends on the unemployment gap (u_t-u^*) (with u^* the natural unemployment rate), energy price inflation (π^0_t) and expected inflation modelled by backward and forward looking components. Forward looking price expectation (π^f_t) equals the difference between the nominal and the real rate of interest according to Fisher's formula $r^n_t = r^r_t + \pi^f_t$. As nominal rate, we use the long term government bond yield while the real interest rate is represented by a proxy variable, the ratio of the sum of real GDP over four quarters to the corresponding sum of real investment expenditures. In view of the modification of the Phillips curve by Samuelson and Solow ([7]), labour productivity growth (λ_t) is also considered in the equation. The results of the linear Phillips curve are given in the table 2.

Dependent Variable: $\Delta\pi_t$

Sample (adjusted): 1969:3 1998:4 (Included observations: 118)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
const.	0.007590	0.002537	2.991486	0.0035
$\Delta\pi_{t-1}$	0.418053	0.087216	4.793325	0.0000
π_{t-1}	-0.247808	0.055656	-4.452521	0.0000
π_{t-4}	-0.166698	0.067031	-2.486884	0.0145
$\Delta\pi_{t-8}$	-0.089575	0.056247	-1.592533	0.1144
$\Delta\pi_{t-12}$	-0.110270	0.055852	-1.974331	0.0511
π_{t-5}	0.165211	0.053915	3.064292	0.0028
π^f_t	0.000746	0.000404	1.846239	0.0678
π^0_t	0.176137	0.014223	12.38415	0.0000
π^0_{t-1}	-0.203229	0.024364	-8.341309	0.0000
π^0_{t-2}	0.098185	0.019734	4.975404	0.0000
λ_{t-2}	-0.073331	0.026445	-2.772931	0.0066
u_t	-0.000693	0.000218	-3.174645	0.0020
m_{t-2}	0.029236	0.015522	1.883461	0.0625
m_{t-3}	-0.074374	0.021147	-3.516986	0.0007
m_{t-4}	0.064679	0.016840	3.840904	0.0002
dummy2	0.010957	0.002014	5.440388	0.0000

$R^2 = 0.809$, S.E. = 0.004, SSR = 0.002, AIC = -8.027

Table 2: OLS results for the Phillips curve

Okun's law describes the short-run relationship between the GDP gap and the unemployment rate. This empirical relationship, developed by A.M. Okun in the 1970s, links the output gap (x_t-x^*) to the change in the unemployment rate. x^* denotes the expected rate of real growth following the long-run trend. For an application, see Kavkler and Böhm ([6]). We found the following reasonable linear estimates.

Dependent Variable: Δu_t

Sample (adjusted): 1970:2 2000:3 (Included observations: 122)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
$x_t - x^*$	-4.953883	0.805964	-6.146531	0.0000
const.	0.131311	0.031709	4.141159	0.0001
u_{t-1}	-0.017967	0.004797	-3.745722	0.0003
Δu_{t-1}	0.428198	0.068666	6.235993	0.0000
dummy3	0.636079	0.102657	6.196183	0.0000
dummy4	-0.433570	0.143116	-3.029498	0.0030

$R^2 = 0.695$, S.E. = 0.141, SSR = 2.298, AIC = -1.036

Table 3: OLS results for the Okun's law

The dummy variables were introduced to reduce the ARCH effects caused by the outliers in the years 1991 and 1992, when the German reunion took place. Inspection of the properties of the estimated linear equations indicates in all cases problems of parameter stability or misspecification. CUSUM an CUSUM square tests reveal structural breaks in the surrogate quantity equation in the period from 1991:3 till 1993:1, whereas the Ramsey RESET test with 3 fitted terms rejects the null hypothesis of NID errors for the Phillips curve and the Okun's law.

In order to improve specification we need to investigate the influence of nonlinearities, which we shall assume to be of the smooth transition kind.

4. Smooth transition regressions

A detailed description of this technique can be found in Teräsvirta ([8]), to which we refer the reader for further information.

Each of the three equations is now subjected to linearity tests. All independent variables may be considered as suitable candidates for a transition variable. From the table 4 with test results we obtain a set of suggested transition variables from which we first choose those with the smallest p-value. Due to occasional problems with convergence of the nonlinear optimization procedure some experimentation to find appropriate starting values was required.

Quantity equation

	df	trend F-stat. (p-val.)	x_t F-stat. (p-val.)	x_{t-4} F-stat. (p-val.)	$\Delta\pi_{t-1}$ F-stat. (p-val.)	m_{t-1} F-stat. (p-val.)	m_{t-4} F-stat. (p-val.)	π^f_t F-stat. (p-val.)
F	22, 87	2.3383 (0.0028)	2.1617 (0.0062)	1.4783 (0.1034)	2.3838 (0.0023)	2.5023 (0.0014)	2.5366 (0.0012)	1.8201 (0.0266)
F4	7, 87	2.6015 (0.0174)	1.5166 (0.1722)	0.7533 (0.6277)	1.7722 (0.1031)	1.1519 (0.3389)	3.0211 (0.0069)	2.4554 (0.0240)
F3	7, 94	2.6158 (0.0164)	1.7743 (0.1015)	1.3475 (0.2370)	2.9608 (0.0075)	3.0606 (0.0060)	0.8138 (0.5780)	1.2178 (0.3008)
F2	7, 101	1.4620 (0.1894)	3.1389 (0.0048)	2.5564 (0.0182)	2.1552 (0.0445)	3.1299 (0.0049)	3.5351 (0.0019)	1.7021 (0.1169)

Phillips curve

	df	trend F-stat. (p-val.)	$\Delta\pi_{t-1}$ F-stat. (p-val.)	π_{t-1} F-stat. (p-val.)	π_{t-4} F-stat. (p-val.)	$\Delta\pi_{t-8}$ F-stat. (p-val.)	$\Delta\pi_{t-12}$ F-stat. (p-val.)	π_{t-5} F-stat. (p-val.)	π^f_t F-stat. (p-val.)
F	49, 51	1.8274 (0.0174)	2.0394 (0.0064)	1.6366 (0.0419)	1.2605 (0.2073)	0.8453 (0.7219)	1.3384 (0.1524)	1.3137 (0.1683)	1.8682 (0.0144)
F4	16, 51	2.1217 (0.0217)	1.4443 (0.1592)	1.5258 (0.1270)	0.7590 (0.7214)	0.4812 (0.9453)	1.2311 (0.2781)	0.9873 (0.4842)	1.3288 (0.2168)
F3	16, 67	1.1073 (0.3667)	2.4285 (0.0060)	0.9073 (0.5643)	1.4331 (0.1535)	1.1130 (0.3617)	1.3955 (0.1712)	1.9154 (0.0340)	1.7354 (0.0609)
F2	16, 83	1.6001 (0.0869)	1.4994 (0.1196)	2.2298 (0.0098)	1.7142 (0.0597)	1.2651 (0.2393)	1.2288 (0.2644)	0.9602 (0.5065)	2.0472 (0.0189)

	df	π^0_t F-stat. (p-val.)	π^0_{t-1} F-stat. (p-val.)	π^0_{t-2} F-stat. (p-val.)	λ_{t-2} F-stat. (p-val.)	u_t F-stat. (p-val.)	m_{t-2} F-stat. (p-val.)	m_{t-3} F-stat. (p-val.)	m_{t-4} F-stat. (p-val.)
F	49, 51	1.6030 (0.0488)	1.3350 (0.1545)	1.3113 (0.1699)	1.0655 (0.4110)	1.8317 (0.0170)	2.4265 (0.0010)	1.9963 (0.0079)	1.9241 (0.0110)
F4	16, 51	1.2918 (0.2386)	1.2738 (0.2498)	1.1995 (0.3006)	0.9075 (0.5654)	1.8731 (0.0461)	3.0185 (0.0014)	1.6067 (0.1009)	1.6857 (0.0803)
F3	16, 67	1.2029 (0.2893)	0.6912 (0.7926)	0.7013 (0.7827)	1.2097 (0.2842)	1.1664 (0.3173)	0.9951 (0.4727)	1.9518 (0.0301)	2.2049 (0.0129)
F2	16, 83	2.0969 (0.0158)	2.0742 (0.0172)	2.1086 (0.0152)	1.1526 (0.3233)	1.8654 (0.0357)	1.9842 (0.0236)	1.6771 (0.0675)	1.1442 (0.3302)

Okun's law

	df	trend F-stat. (p-val.)	$x_t - x^*$ F-stat. (p-val.)	u_{t-1} F-stat. (p-val.)	Δu_{t-1} F-stat. (p-val.)
F	13, 103	1.3097 (0.2193)	1.5387 (0.1161)	0.8491 (0.6079)	1.2270 (0.2712)
F4	4, 103	2.4543 (0.0504)	2.0633 (0.0910)	0.2426 (0.9135)	0.6275 (0.6440)
F3	4, 107	0.9118 (0.4599)	1.9709 (0.1042)	1.4665 (0.2175)	1.8459 (0.1254)
F2	4, 111	0.8000 (0.5277)	0.8253 (0.5118)	1.1052 (0.3578)	1.5155 (0.2025)

Table 4: Linearity tests

The final set of estimates obtained for the parameters of the surrogate quantity equation and the Phillips curve is found below.

$$m_t = \underset{(0.0039)}{0.0112} + \underset{(0.1194)}{0.2213}\, x_t + \underset{(0.0878)}{0.1766}\, x_{t-4} \underset{(0.2259)}{-0.6599}\, \Delta\pi_{t-1} + \underset{(0.0500)}{0.9274}\, m_{t-1} + \underset{(0.0166)}{0.0920}\, dummy1 +$$

$$+ (\underset{(0.0793)}{-0.2096}\, m_{t-1} \underset{(0.0022)}{-0.0104}\, \pi^f_t)\, [1 + \exp\{\underset{(1.6579)}{-4.1174}\, (m_{t-4} - \underset{(0.0101)}{0.0572}) / 0.0571\}]^{-1}$$

$$T = 116,\ R^2 = 0.870,\ S.E. = 0.021,\ \hat{\sigma}_{nl} / \hat{\sigma}_{lin} = 0.723,\ JB = 5.053(0.080)$$
$$F_{AR(1)} = 1.351(0.248),\ F_{AR(4)} = 1.072(0.374),\ F_{AR(8)} = 0.915(0.507)$$
$$LM_{ARCH(1)} = 1.725(0.189),\ LM_{ARCH(4)} = 4.754(0.314),\ LM_{ARCH(8)} = 15.168(0.056)$$

Equation 1: STR estimates of the surrogate quantity equation

$$\Delta\pi_t = \underset{(0.0122)}{0.1874}\, \pi^0_t + \underset{(0.0240)}{0.0548}\, \pi_{t-5} \underset{(0.0233)}{-0.0545}\, \lambda_{t-2} \underset{(0.0154)}{-0.0582}\, m_{t-3} + \underset{(0.0133)}{0.0894}\, \pi^0_{t-2} + \underset{(0.0150)}{0.0607}\, m_{t-4} + \underset{(0.0023)}{0.0118} +$$

$$+ \underset{(0.0018)}{0.0129}\, dummy2 + \underset{(0.0291)}{0.4132}\, \Delta\pi_{t-1} \underset{(0.0478)}{-0.3286}\, \pi_{t-1} \underset{(0.0591)}{-0.1478}\, \Delta\pi_{t-12} \underset{(0.0188)}{-0.1867}\, \pi^0_{t-1} \underset{(0.0002)}{-0.0012}\, u_t +$$

$$+ (\underset{(0.0863)}{0.1768}\, \Delta\pi_{t-12} \underset{(0.0111)}{-0.0228}\, \pi^0_{t-1} + \underset{(0.0002)}{0.0004}\, u_t)\, [1 + \exp\{\underset{(0.0057)}{-42.8009}\, (m_{t-2} - \underset{(0.0027)}{0.0599}) / 0.0569\}]^{-1}$$

$$T = 116,\ R^2 = 0.817,\ S.E. = 0.004,\ \hat{\sigma}_{nl} / \hat{\sigma}_{lin} = 0.944,\ JB = 1.084(0.582)$$
$$F_{AR(1)} = 0.0004(0.995),\ F_{AR(4)} = 1.587(0.184),\ F_{AR(8)} = 1.561(0.148)$$
$$LM_{ARCH(1)} = 0.382(0.537),\ LM_{ARCH(4)} = 4.100(0.393),\ LM_{ARCH(8)} = 6.032(0.644)$$

Equation 2: STR estimates of the Phillips curve

The models passed the diagnostic checking satisfactorily. The p-values of the Jarque-Bera test and the test of no remaining error autocorrelation show that the null hypotheses of the normal distribution of the error term and of no error autocorrelation, respectively, cannot be rejected. The test of parameter constancy detects no problems and there are also no ARCH effects present in our models.

5. System of equations

The estimated nonlinear system is inspected for its dynamic properties. Contrary to estimation, the money stock is considered a policy instrument to influence the inflationary process as well as the unemployment rate. Additionally, the reaction of a change in oil prices, in the forward looking expected inflation rate and to a productivity shock is also evaluated. In the latter cases, reactions turn out to be rather symmetric, while shocks in the money stock exhibit significant asymmetries. The graphs in the figure 1 show the results of a unit shock of one standard deviation of the money stock on inflation and unemployment.

Maintained unit shock of one standard deviation of the money stock creates a permanent increase in inflation by 2.4 % points and a decrease in unemployment rate by 8.2 % points. Also in this case the asymmetric reaction to monetary policy is obvious (figure 2). This again shows that working with a nonlinear model generates a result that is known in practice, but not achieved by linear models.

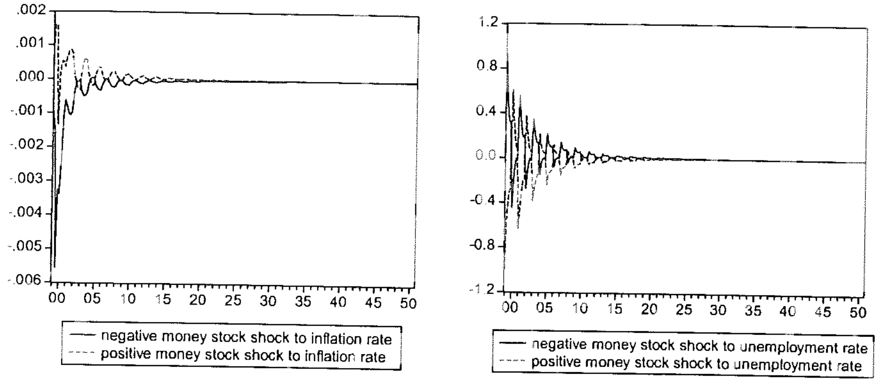


Figure 1: effect of an initial unit shock of the money stock to inflation and unemployment

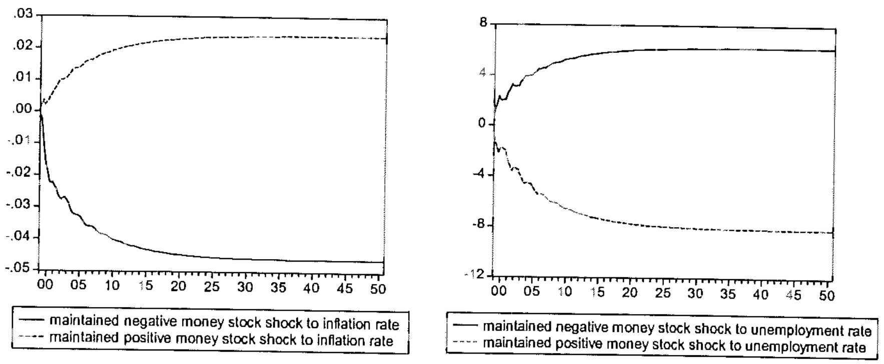


Figure 2: effect of a maintained unit shock of the money stock to inflation and unemployment

References

[1] B. Böhm: *Structural Change in the Austrian Inflation Process,* in G. Chaloupek et. al (eds.) Ökonomie in Theorie und Praxis, p. 31 – 46, Springer Verlag, 2001.

[2] B. Böhm: *Smooth Transition Models of Structural Change,* in J. Leskow, M. Puchet, L. F. Punzo (eds.) New Tools of Economic Dynamics, Lecture Notes in Economics and Mathematical Systems, Vol. **551**, Springer Verlag, 2005.

[3] H. Frisch: *Theories of Inflation*, Cambridge University Press, 1983.

[4] R. Gordon: *The Time – varying NAIRU and its Implications for the Economic Policy*, Journal of Economic Perspectives 11 **(1)**, p. 11 – 32, 1997.

[5] C.W. Granger, T. Teräsvirta: *Modelling Nonlinear Economic Relationships*, Oxford University Press, 1993.

[6] A. Kavkler, B. Böhm: *Modelling Okun's law for Germany with Smooth Transition Regression*, Symposium in Operations Research '03 Proceedings, p. 289 – 294, 2003.

[7] P.A. Samuelson, R.M. Solow: *Analytical Aspects of Anti-Inflation Policy*, American Economic Review **(50)**, p. 177–194, 1960.

[8] T. Teräsvirta: *Modelling Economic Relationships with Smooth Transition Regression*, Handbook of Applied Economic Statistics, p. 507-572, New York, Marcel Dekker, 1998.

FIVE CROATIAN BANKS FOREIGN CURRENCY POSITION ANALYSIS USING VALUE AT RISK

Dubravko Mojsinović

Privredna banka Zagreb, d.d, Račkoga 6, 10 000 Zagreb, Croatia

dubravko.mojsinovic@pbz.hr

Abstract: Foreign currency open position is the difference between assets and liabilities denominated in a certain foreign currency. The Croatian Kuna counter value of that difference changes as the exchange rate changes. The banks differ regarding the amounts of open positions in different currencies. It can be said that riskier is the bank which has relatively high long open positions in currencies whose exchange rates are volatile and positively correlated. Value at risk recognizes these conditions and therefore is used to compare five Croatian banks in 2002 and 2003.

Keywords: Croatian Banks, Foreign currency open position, Value at Risk

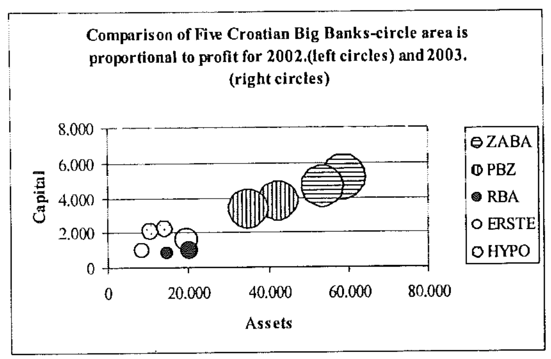
Introduction

Assets and liabilities are frequently denominated in different currencies. Croatian Kuna (HRK) counter value is subject to foreign exchange rate fluctuations. Open position is the difference between assets and liabilities denominated in a certain foreign currency.[1] Big open position suggests the existence of bigger foreign exchange rate risk. Long or positive open position means that the bank has more assets denominated in a certain foreign currency than the liabilities. Short or negative open position means that the bank has more liabilities than the assets denominated in a certain currency. More concretely, if the bank has placed 100 EUR as loans and received 50 EUR as savings, it means that the bank is 50 EUR long or that the bank has positive open position. Increase of foreign exchange rate HRK/EUR will increase the net value of the bank denominated in HRK and the decrease will lower it. Analysis of open position is conducted in a way to include open positions which are under the influence of the management and foreign exchange rates which are external to banks. This is done using VaR or Value at Risk indicator which takes into account the magnitude of the open positions, volatility and correlations of exchange rates. This is done for five big Croatian banks at the end of 2002. and 2003.

Data and preliminary analysis

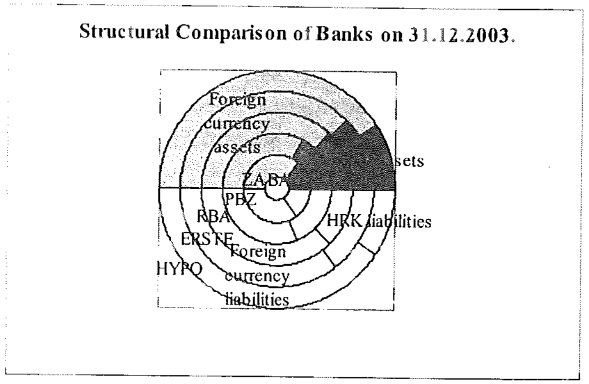
Five big Croatian banks were selected.[2] For comparison total Croatian banking system assets on 31.12.2003. amounted to 201,8 billion HRK, while profit for 2003. was 2,5 billion HRK.[3] It can be seen from the figure that all banks moved to the right from 2002. to 2003. This movement suggests bank growth. Assets, capital and profit of the selected banks increased.

Figure 1 Data for five Croatian banks in million HRK[4]



As can be seen from the figure bellow all banks have a significant amount of assets and liabilities denominated in foreign currency or linked to the foreign currency. This suggests that there is a need for managing foreign currency risk. Managing foreign currency risk is in the first place expressed through the absolute amount of open position. Foreign currency open position is the difference between assets and liabilities denominated in a certain currency. It can be assumed that the lower the amount, the better is the safety against the exchange rate change risk. Banks are longer in EUR. This means that they have more assets in EUR than the liabilities. Therefore there exists the risk resulting in unfavorable movements of EUR exchange rate. However, present Croatian National Bank policy is such that it keeps this exchange rate fairly stable.

Figure 2 Assets and liabilities in foreign currency and HRK for five banks



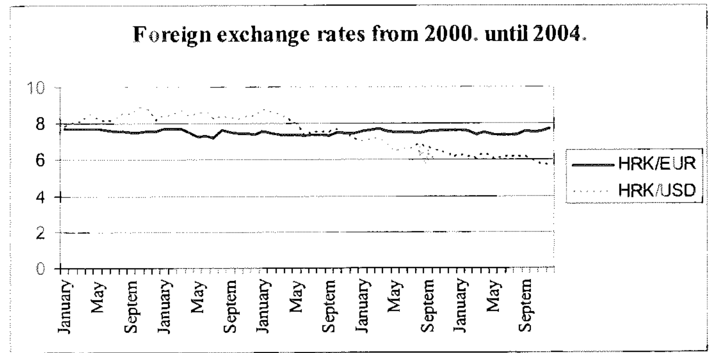
Open positions existence does not mean that there is no foreign currency risk management in place. Because of this VaR will be calculated to show how much could the bank lose with such a position with respect to the exchange rate volatility. Moreover banks regularly have to reprice their assets and liabilities according to the middle exchange rate of the Croatian National bank to meet the regulatory standards. The impact of the open position and foreign currency exchange rate movement on profit is shown in the table bellow.

Table 1 Impact of open position and foreign currency exchange rate movement on profit for the selected currency

Open position in the selected currency	Change of the exchange rate in respect to the previous value in time	
	Increase (more HRK/EUR)	Decrease (less HRK/EUR)
Assets-Liabilities>0	Profit	Loss
Assets-Liabilities =0	0	0
Assets-Liabilities <0	Loss	Profit

The figure bellow shows the movement of end of month middle Croatian National Bank exchange rates for HRK/EUR and HRK/USD[5] from 2000. until 2004.

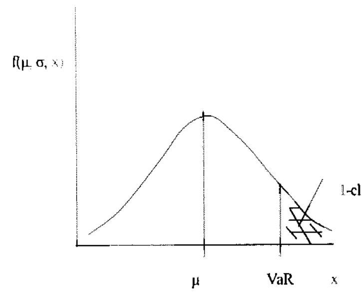
Figure 3 Middle end of month HRK/EUR and HRK/USD exchange rates



The figure suggests the stability of HRK/EUR. Since most of the open position is in EUR, the open position in EUR is also stable. HRK/USD is not stable, but since the USD position is not so significant, so is the influence on open position much lower.

VaR or Value at Risk is the maximum amount of expected loss in a certain time interval with a certain probability.[6] VaR indicates the expected loss in case when the unfavorable event does not happen. There are different ways to calculate VaR.[7] In this work VaR is calculated using assumption that the loss is normally distributed.[8] Profit in this sense is negative loss. Normal distribution is described by two parameters μ and σ. μ is the expected loss and σ is the standard deviation of loss.

Figure 4 VaR and normal distribution



$$VaR = \sigma \cdot \alpha_{cl} + \mu$$

where α_{cl} is the value of x for the standard normally distributed variable (μ=0, σ=1)[9] which divides the area bellow the function in two parts. On the left side the area is cl (or confidence level) and on the right side is 1-cl.

If expected loss is zero and standard deviation depends on the exchange rates and open positions then[10]

$$VaR = \alpha_{cl} \cdot \sqrt{\sum_{i=1}^{I}(A_i - P_i)^2 \cdot Var[p_i] + 2 \cdot \sum_{s=1}^{I-1}\sum_{r=s+1}^{I}(A_r - P_r)\cdot(A_s - P_s)\cdot \operatorname{cov}(p_r, p_s)}$$

where p_i is the exchange rate of currency i. Var is the variance and cov the covariance. A_i is the amount of assets in currency i, P_i is the amount of liabilities in currency i. I is the number of currencies.

The expression under the root is the total variance of sum of dependent normally distributed open positions in HRK or portfolio of open positions. Since the series of exchange rates are monthly, the VaR refers to the monthly loss. For the purpose of having lower VaR it is favorable to hold positions of the same sign in negatively correlated currencies and of the opposite sign in positevely correlated currencies.

Since amounts denominated in other currencies where insignificant they were added to EUR resulting in table bellow.

Table 2 Assets, liabilities and open position in millions EUR and USD for five banks

Bank		ZABA	ZABA	PBZ	PBZ	RBA	RBA	ERSTE	ERSTE	HYPO	HYPO
Date		31.12.2003.	31.12.2002.	31.12.2003.	31.12.2002.	31.12.2003.	31.12.2002.	31.12.2003.	31.12.2002.	31.12.2003.	31.12.2002.
Assets	EUR	5.032	4.743	3.034	2.626	1.730	1.295	1.752	679	1.342	994
	USD	587	625	712	555	263	218	196	172	204	167
	HRK	16.547	13.926	14.885	11.413	5.249	3.354	4.925	2.125	2.545	1.797
Liabilities	EUR	4.960	4.543	2.915	2.456	1.753	1.283	1.659	608	1.338	959
	USD	578	626	697	540	271	224	198	166	207	168
	HRK	17.156	15.412	15.887	12.784	5.028	3.396	5.619	2.706	2.562	2.052
Open position	EUR	72	200	119	170	-23	12	92	71	4	35
	USD	9	0	15	15	-7	-7	-2	7	-2	-1
	HRK	-609	-1.486	-1.002	-1.371	221	-43	-694	-580	-16	-255

Standard deviation and covariance of the exchange rate distribution is shown bellow.

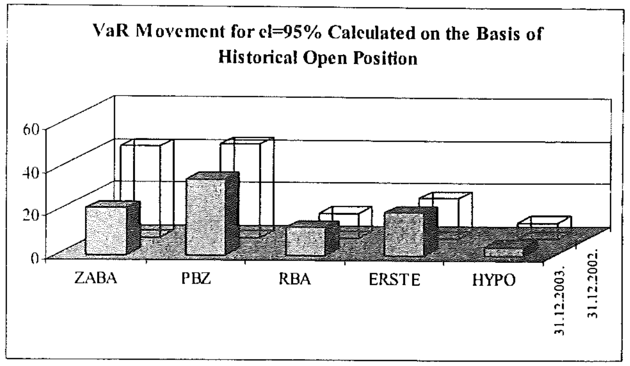
Table 3 Exchange rate distribution parameters

Currency	Expectation	Variance	Covariance		Correlation coefficient	
			HRK/EUR	HRK/USD	HRK/EUR	HRK/USD
HRK/EUR	7,5225	0,0167	0,0167	-0,0016	1,0000	-0,0122
HRK/USD	7,4885	0,9968	-0,0016	0,9968	-0,0122	1,0000

Results

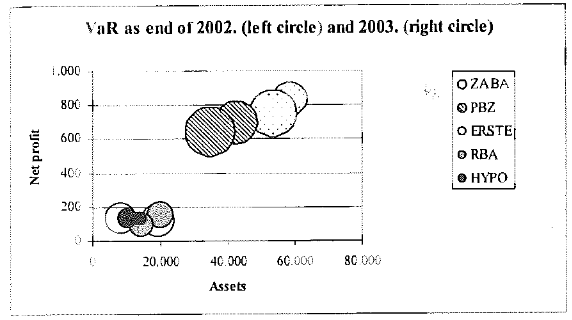
If all banks would have had the same combination of open positions, they would have had the same VaR. So the differences in VaR are explained by the different amounts of open positions. It is obvious that banks have significant differences in their VaR values.

Figure 5 Five banks VaR movements in millions HRK



Banks decreased VaR by decreasing their open positions which is a favorable direction and indicates the existence of the risk management function. PBZ has the greatest VaR and did much less on lowering it than ZABA although in 2002 they were much alike. ZABA by the end of 2003. decreased its VaR on the level of smaller banks.

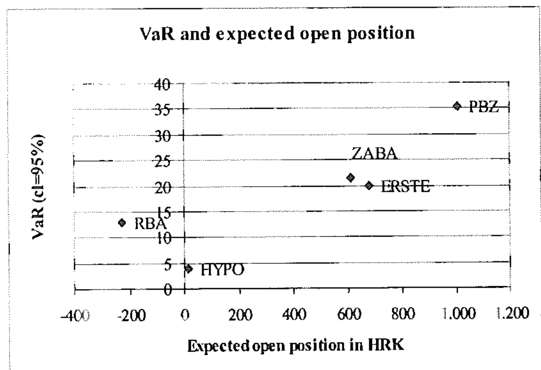
Figure 6 Comparison of VaR, profit and assets movements in millions HRK



Although RBA and ERSTE are of the similar size, RBA has smaller VaR. RBA increased its profit and assets, but did not decrease VaR. ERSTE grew, but did not decrease VaR. ZABA and PBZ grew proportionately increasing their profit and assets and decreasing VaR.

It seems that at the end of 2003. ZABA, PBZ and HYPO are in the better position comparing them to ERSTE and RBA who were not able to decrease their VaR values.

Figure 7 VaR and the expected value of open position at the end of 2003. in millions HRK



The figure shows that the greater open position means greater VaR. It shouldn't be forgotten that VaR grows with the value of open position, but that its value also depends on the volatilities and covariance of exchange rates. It can be seen that ERSTE (680, 20) has greater expected value of the open position and at this point it could be said that it has greater VaR. However this is not so because ZABA (613, 22) has lower open position in EUR, but significantly higher in USD whose exchange rate is much more volatile and thus more contributes to the increase of ZABA's VaR.

References

[1] Saunders, Anthony: Financial Institutions Management: a modern perspective, Irwin, Illinois, 1994, page 241.

[2] Hrvatska bankarska industrija, Banka, specijalno izdanje, July 2004, pages 6-7.

[3] Hrvatska bankarska industrija, Banka, specijalno izdanje, July 2004, pages 6-7.

[4] Godišnje izvješće 2003. Zagrebačka banka d.d, pages 49, 51, Godišnje izvješće 2003. Privredna banka Zagreb d.d, page 46, Godišnje izvješće 2003. Raiffeisenbank Austria d.d. Zagreb, page 33, Godišnje izvješće 2003. Erste & Steiermärkische bank d.d, pages 18, 19, Godišnje izvješće 2003. Hypo Alpe-Adria-Bank d.d, pages 50, 51.

[5] Tablica H10: Srednji devizni tečajevi Hrvatske narodne banke na kraju razdoblja, Bilten HNB number 100, January 2004.

[6] Dowd, Kevin: An Introduction to Market Risk Measurement, John Wiley & Sons, Ltd, 2002, page 19.

[7] Dowd, Kevin: An Introduction to Market Risk Measurement, John Wiley & Sons, Ltd, 2002, pages 55-87.

[8] Dowd, Kevin: An Introduction to Market Risk Measurement, John Wiley & Sons, Ltd, 2002, pages 38-39.

[9] Šošić, Ivan: Vježbe iz statistike I. dio, mikrorad d.o.o. i Ekonomski fakultet Zagreb, 1994, page 132.

[10] Saunders, Anthony: Financial Institutions Management: a modern perspective, Irwin, Illinois, 1994, page 248.

ANALISYS AND RISK MANAGEMENT OF INVESTMENT PROJECTS USING METHOD MONTE CARLO SIMULATION

Ivan Šmid
Centar za poduzetništvo Krapinsko-zagorske županije, Krapina
e-mail: i.smid@vip.hr
Valerija Bublić
Lorana d.o.o., Zagrebačka 26, Krapina
e-mail: lorana@kr.htnet.hr

Abstract: In this paper we investigate the possible influence of different input variables in the investment project Building and Equipping of Chicken Farm on the output distribution of the internal rate of return (IRR) and the net present value (NPV) of the project. After identifying which input variables have the strongest influence on the investment project, we consider the feasibility of the additional investment that influences the distribution of the identified critical inputs. In this paper we applied the method Monte Carlo Simulation.

Keywords: Investment project, internal rate of return, net present value, risk analysis, risk management, Monte Carlo simulation.

1. INTRODUCTION

In the theory and practice of management various activities that help to influence risk events or risk phases of the business process are well known. Entrepreneurs quite often do not have enough financial resources to act on all potential risks simultaneously. Therefore, in the risk management practice we try to recognize potential risk causes, which have the strongest influences on the final business result. We undertake activities to prevent risk in accordance with recognized priorities. Recognition and estimation of priorities among potential risk causes are one of the most important managerial tasks [1].

Calculations of the net present value and the internal rate of return are very important facts when estimating the profitability of the investment. Based on this information, which is an obligatory part of each investment project, banks take decisions about loans.

In accordance with the instructions of the European Commission, sensitivity analysis and risk analysis, which include probability analysis, are a recommended content of investment projects and studies [2]. It is necessary to define the critical variables of the project and on the basis of series of historical data define their probability distribution, from which follows the probability distribution of the net present value and the internal rate of return. However, sometimes it is very difficult to collect statistical data about probability progression of the critical project variables. In such cases information can be collected from literature or estimated by a group of experts.

In this paper will present how the risk management strategy can be based on the simulation results.

2. RISK ANALYSIS

Data necessary for the calculation of the parameters for estimating the profitability of the investment refer to the future events. Due to that, those inputs are uncertain and each calculation involving them contains a particular level of uncertainty. When taking a decision based on unreliable inputs, it has to be considered whether the ranges within which criteria values can be expected are acceptable or not. If there is a possibility that the investment could be unprofitable, we can talk about the investment risk. In order to minimize the investment risk it is first of all necessary to recognize the possible causes of the risk and then, to act on recognized potential causes of undesired events.

In this paper, risk analysis of the deterministic financial model is based on the application of method Monte Carlo simulation. We will try to state how the combined uncertainty of the input variables influences the risk.

3. PROBLEM DESCRIPTION

The Investor plans to build and to equip a chicken farm, whose capacity is 50.000 chickens in a cycle. Basic data are as follows:

- Total investment 4.315.295 kn: 3.976.000 kn in fixed assets and 339.295 kn in working capital.
- Financial sources: 2.315.295 kn of own capital and 2.000.000 of bank loan, payback period 10 years and interest rate 5%.
- Total income: 4.026.713 kn
- Average weight of chickens' growth: 2.05 kg/pc
- Average mortality rate: 3% per cycle
- Planned number of cycles/year: 5
- Planned sales price of chickens: 8.10 kn/kg
- Planned costs of one-day chickens: 3,00 kn/pc
- Planned feeding costs are standardized as follows:
  a) Mixture PPT1, price 2.70 kn/kg, average need is 1 kg/chicken
  b) Mixture PPT2, price 2.60 kn/kg, average need is 1 kg/chicken
  c) Mixture PPT3, price 2.50 kn/kg, average need is 1,66 kg/chicken
- Planned labour and other costs are calculated on the basis of real costs.

Prices of input and output are based on market prices in February 2005.

Henceforward the economic flow is shown (Figure 1), which derives from the projected income and cash flow statement. The internal rate of return 8.42% is calculated on the basis of the economic flow of the project, as well as the net present value in the amount of 1.086.416 kn by using 5% discount rate.

ELEMENTI	2005	2006.	2007.	2008	2009.	2010.	2011.	2012.	2013.	2014.	2015.	2016.	2017.
	EKONOMSKI TOK												
I PRIMICI	0	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	4.026.713	6.765.257
II IZDACI	4.315.295	3.570.902	3.571.652	3.573.652	3.575.652	3.577.652	3.586.692	3.588.692	3.590.692	3.592.692	3.594.692	3.596.692	3.597.942
[illegible]	3.976.000												
III NETO PRIMICI	-4.315.295	455.811	455.061	453.061	451.061	449.061	440.021	438.021	436.021	434.021	432.021	430.021	3.167.316
KUMULATIV NETO PRIMITAKA	-4.315.295	-3.859.484	-3.404.423	-2.951.362	-2.500.301	-2.051.240	-1.611.219	-1.173.198	-737.178	-303.157	128.864	558.885	3.726.201

Figure 1: Economic flow of the project

As the critical input we defined the average weight of chickens growth and average mortality rate. The duration of one cycle is 42 days. Statistical data of time series of critical input variables were not available. We know from experience that the average weight of chicken growth varies from 1,90 to 2,30 kg/piece and the most common value is 2,05 kg/piece, while the average mortality rate varies from 1% to 6% and the most common value is 3%. The reasons for such oscillations are humidity, external air temperature and other factors, which cannot be precisely defined. Furthermore, as critical variables we defined the sales price of chickens, the input costs of one-day chickens and the input costs of mixtures PPT1, PPT2 and PPT3. Sales price of chickens is now 8,10 kn/kg. Due to the hyper production on the local market we expect the price to fall to 7,90 kn/kg and the maximal price could be only 7,50 kn/kg. We also predicted probabilities of prices of several other inputs:

- Mixture PPT1: min. 2.50 kn/kg, expected 2,70 kn/kg and max 2.90 kn/kg
- Mixture PPT2: min. 2.40 kn/kg, expected 2.60 kn/kg and max 2.70 kn/kg
- Mixture PPT3: min. 2,40 kn/kg, expected 2.50 kn/kg and max 2.60 kn/kg
- One-day chickens: min. 2.90 kn/pc, expected 3.00 kn/pc and max. 3.10 kn/pc.

4. APPLICATION OF THE METHOD MONTE CARLO SIMULATION

We will apply the method Monte Carlo simulation on the above-described critical inputs. First we will assign one of the classical probability distribution to the critical inputs and then the software will generate a series of random numbers, which correspond to the chosen distribution. As a result there is an attained histogram and cumulative graph of the probability of the distribution of chosen critical inputs and their influence on NPV and IRR.

The objective of this paper is to define the probability distribution of the IRR and NPV on the real example of the investment project of a chicken farm, using the Monte Carlo simulation method. Furthermore, objectives are: to identify which of the above mentioned inputs have the strongest influence on the IRR and NPV of the project and to define the risk minimizing strategy by means of investing in the equipment which will improve the probability distribution parameters of the critical inputs. For the simulation purpose the software @RISK - Palisade Corporation [3] will be applied. First, we made the model of the investment project in excel. The following values are stipulated as uncertain input variables: average weight of chickens growth (Picture 2), average mortality rate, the sales price of chickens, costs of one-day chickens and costs of the mixtures PPT1, PPT2 and PPT3.

Software @RISK enables us the choice of a large number of different probability distributions for input variables. Because of the insufficient historical data we have applied the simple triangular distribution for each relevant input variable. Using the triangular distribution, it is necessary to define three values: the assumed minimal value of the random variable, its assumed maximal value, and the most probable assumed value of the random variable [4]. In Figure 2 we quote the distributions of inputs which we used in this case applying @RISK software.

Nb.	Input variable	Distribution
1.	Average weight of chickens growth (Picture 2)	RiskTriang (1.90; 2.05; 2.30)
2.	Average mortality rate	RiskTriang (0.01; 0.03; 0.06)
3.	Sales price of chickens	RiskTriang (7,50;7,90;8,10)
4.	Cost of one-day chickens	RiskTriang (2,90; 3,00 3.10)
5.	Cost of mixture PPT1	RiskTriang (2,50; 2,70; 2.90)
6.	Cost of mixture PPT2	RiskTriang (2,40; 2,60; 2.70)
7.	Cost of mixture PPT3	RiskTriang (2,40; 2,50; 2.60)

Figure 2. Distribution of input variables for software @RISK

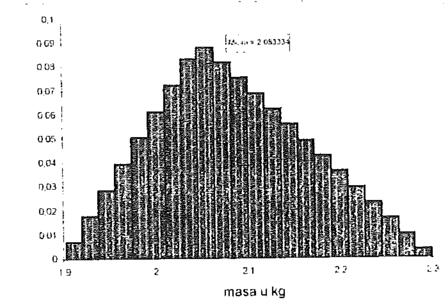


Figure 3: Average weight of chicken growth

Histograms (Figure 3, and analogically histograms for other critical variables in accordance with Figure 1) show relative frequencies of input variables based on 10.000 generated random numbers.

5. RESULTS OF MONTE CARLO SIMULATION

Since the oscillations of the critical inputs in Excel model are connected with IRR and NPV of the project (Figures 4 and 5) the results of the simulations are given in succession where all critical inputs are varied simultaneously.

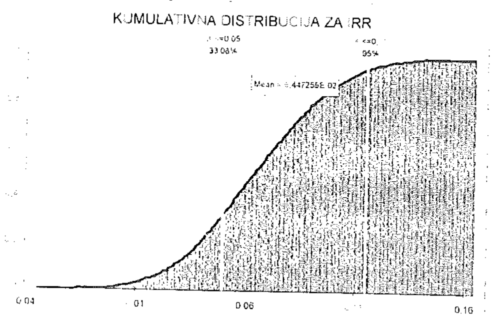


Figure 4. Cumulative probability distribution of the IRR

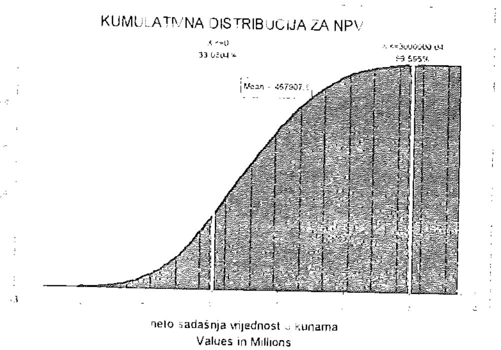


Figure 5. Cumulative probability distribution of the NPV

6. RISK MANAGEMENT

Since the probability of the project failure is 33%, which is considered to be a relatively high rate, it is necessary to define the priorities of acting on particular input variables. Thereby the software @RISK helps us to show the information about the correlation of input and output variables of the project. Information is given in the form of a so-called Tornado graph (Figure 6).

The higher the absolute value of the correlation coefficient is, the bigger is the influence of the input variable on the result. Figure 6 depicts the Tornado graph with correlation coefficients, which show how the NPV of the project depends on particular input variables. Furthermore, we will show only the Tornado graph for the NPV, as the graph for the IRR looks quite similar. Input variables which influence the project based on Figure 5, following the degree of influence from high to low are: the average weight of chicken growth, the sales

price of chickens, the average mortality rate, costs of mixtures PPT1, PPT2 and PPT3 and costs of one-day chickens.

Obtaining that valuable information, we can define the strategy of risk management for the chicken farm.

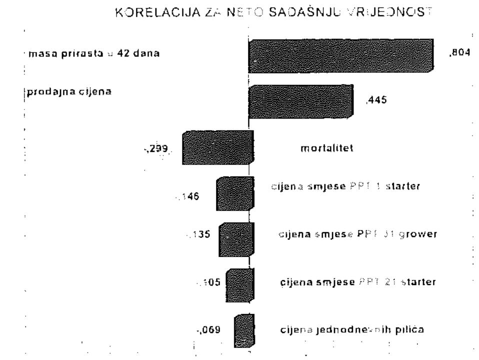


Figure 6. Tornado graph with correlations of inputs with NPV

Generally, it could be stated that the project is acceptable if the NPV is higher than zero and if the IRR is higher than the interest rate on the market (in our case 5%). As the support for decision-making in this case, the most useful are graphs of the cumulative distribution IRR and NPV (Figures 4 and 5) and the Tornado graph (Figure 6). From these graphs we can draw several conclusions. Cumulative graphs (Figures 4 and 5) depict that the probability of the NPV of the project being below zero and the IRR being below 5%, is 33.08%. Figure 4 depicts that the probability of IRR of the project being lower than 11%, is 95%.

From the Tornado graph (Figure 6) we see that there are two variables on which the Investor could act. The average weight of chicken growth and the average mortality rate are highly correlated with NPV. The correlation for the average weight of chicken growth is 0,804 and for the mortality rate is -0, 299. Acting on these variables by optimal temperature, air humidity regulation and cooling of the farm by fog dispersion, the risk of the economic failure of the project can be reduced. It is possible to act on these variables by investing in the computer air conditioning system, (producer Big Dutchman [5]), which in the case of the chicken farm is 500.000 kn more expensive than the one which the Investor originally planned to buy.

In case of utilizing the precise computer controlled air-conditioning adjustment in the farm, using the special software for the system MC 236, INFOMATIC, it is realistic to expect changes of the input probability distributions of the average mortality rate and the average weight of chickens growth (Figure 7). With these changes we will make the new model using @RISK software in order to see how probable it is that NPV of the project is bellow zero i.e. that the project is unprofitable.

Nb.	Input variable	Distribution
1.	Average mortality rate	RiskTriang (0,01; 0,03; 0,04)
2.	Average weight of chickens growth	RiskTriang (2,00; 2,20; 2,30)

Figure 7. Redefined distributions of input variables for @RISK software

As the result on the redefined model, by means of Monte Carlo simulation method, analogically to the previous procedure, a new cumulative probability distribution of the NPV is created (Figure 8).

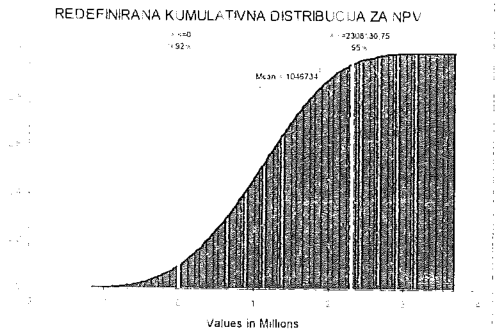


Figure 8. Redefined cumulative probability distribution of the NPV

7. CONCLUSION

Redefined cumulative probability distribution of the NPV (Figure 8) shows that the probability that NPV could be negative and the project unsuccessful is 9.92%, which is lower than 33.08%.

The price of reducing this risk is 500.000 kn which should be invested additionally in computerized air conditioning system, and it is an acceptable price for reducing the risk of the project failure below 10%. The mean of the NPV without the additional investment (Figure 5) is 467.908 kn and the mean of the NPV with additional investment (Figure 6) is 1.046.734 kn.

The difference between means of the net present value with and without the additional investment is 596.826 and it is higher than the estimated costs of additional investment. Thus this investment is reasonable if the critical inputs will be put in the frame of Figure 2.

In this paper it is shown how we can use the method Monte Carlo simulation on the deterministic financial model like a business plan by risk management when we calculate the IRR and the NPV of the project. The manager has to identify and to focus on the most important possible risk sources. Based on the example of the investment project, where we illustrated the procedure of the risk management, it is shown how the combination of the common managerial knowledge and methods of quantitative analysis form excellent foundations for defining the risk management strategy.

LITERATURE

[1] Hunjak, T., Redep, M.: Upravljanje rizicima u planiranju trajnih obrtnih sredstava pomoću Monte Carlo simulacije, časopis RRiF, br. 7/2003, str. 74

[2] ***Guide to Cost Benefit Analysis of Investment Project, Prepared for Evaluation unit DG Regional Policy European Commission – Working Team Coordinated by Professor Massimo Florio

[3] www.palisade.com

[4] Hunjak, T., Redep, M.: Upravljanje rizicima u planiranju trajnih obrtnih sredstava pomoću Monte Carlo simulacije, časopis RRiF, br. 7/2003., str. 76

[5] ***Katalog MC 236 Der Computer fuer ein optimales Stallklima in jedem Gefluegelstall

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 7
Multicriteria Decision Making

MULTIPLE PRICES AND DE NOVO PROGRAMMING

Zoran Babić[*], *Tihomir Hunjak*[**]
[*]Faculty of Economics, 21000 Split, Matice hrvatske 31, Croatia, e-mail:babic@efst.hr
[*]Faculty of Organization and Informatics, 42000 Varaždin, Pavlinska 2, Croatia, e-mail:thunjak@foi.hr

Abstract: Optimization of production program is one of the crucial problems of optimization. For that purpose different methods and procedures were developed in order to present the decision maker with solutions which can provide the maximal realization of his aims. These various methods and procedures were mainly looking for solutions in accordance with the previously defined resources and constraints.

In real production planning situations linear programming models may be difficult to apply because of its assumption of proportionality. One common phenomenon is multiple pricing of raw materials because companies may have more than one source of raw material at different costs, or quantity discounts being offered for purchases.
Key words: Multiple prices, raw materials, De Novo Programming

1. Introduction

De Novo presents a special approach of optimization. Instead of "optimizing a given system" De Novo suggests a way of "designing an optimal system". De Novo approach does not limit the resources as most of the necessary resource quantities can be obtained at certain prices. Resources are actually limited because their maximum quantity is governed by the budget, which is an important element of De Novo. Using De Novo most various cases can be handled more effectively than by the standard programming model. Changes in prices, technological coefficients, increasing costs of raw materials, quantity discounts and other similar and real production situations can be easily incorporated in De Novo model and can give very satisfactory solutions.

In this paper we will show the application of De Novo in the real production system which produces various ferroalloys and in its production uses a set of raw materials with different prices.

2. Problem Setting

This paper analyses the production planning problem in one ferroalloys factory. It produces four types of ferroalloys: FeCr, FeMn, FeSiMn and FeSi. These alloys are produced in three electric furnaces by special carbon-electrochemical treatments.

The ferroalloys furnaces work continuously through the whole year except during the repairing and cleaning break in the second part of the year.

The Table 1 contains data relating to annual capacities of the furnaces expressed in hours, time needed for one ton production of ferroalloys and net-income per ton of the product.

Table 1. Furnace capacities and net income for ferroalloys production

	Fe Cr	Fe Si	Fe Mn	Fe Si Mn	Capacity (in hours)
Furnace 1	0.2084				7920
Furnace 2			0.2604	0.48	7160
Furnace 3		0.495			7920
Net-income	42.284	54.104	18.409	35.663	

For the ferroalloys production various raw materials are used. Available quantities of the five main raw materials as well as the use of these raw materials per ton of ferroalloys are presented in the Table 2. In the last column there are unit prices of these raw materials.

Table 2. Main raw materials

Raw materials	Fe Cr	Fe Si	Fe Mn	Fe Si Mn	Available quantities (b_i)	Unit price of raw materials (p_i)
Electric energy (000 kwh)	4	9.6	2.63	4.75	357056	8
Manganese ore (t)	0	0	2.04	0.5	42795.7	19.5
Coke (t)	0.5	0.29	0.465	0.48	33361.6	35.5
Electrode mass (t)	0.035	0.07	0.022	0.06	2953.8	86
Quartz (t)	0.28	2	0.02	0.58	42407.3	7.1

According to these data the production planning problem can be posted as the linear programming model with one or more objective functions. If x_i is the production quantity of i-th ferroalloy, objective function (total net-income) which has to be maximized is:

$Max\ (42.284x_1 + 54.104x_2 + 18.409x_3 + 35.633x_4)$.

Capacity constraints for the first and the third furnace can be modified by bounding first two variables, i.e. $x_1 \le 38004$, and $x_2 \le 16000$.

Furthermore, because of connected production of FeMn (x_3) and FeSiMn (x_4) our problem has imposed a need for introducing the additional constraint: $x_3 \ge 3.4x_4$.

Market constraints (purchased quantities) also require limitation onto production of FeSiMn ($x_4 \ge 4000$), so that our complete L.P. model acquires the standard L.P. form, and its optimal solution is:

$x_1^* = 38\ 004$ t (Fe Cr), $x_2^* = 14\ 128$ t (Fe Si),
$x_3^* = 17\ 940$ (Fe Mn), $x_4^* = 4\ 000$ t (Fe Si Mn).

with maximum value of objective function: $z^* = 2\ 844\ 269$ currency units.

Let us note that this model does not consider prices of raw materials, as it assumes that the available quantities have been already purchased. From the Table 2 the costs of purchased raw materials can be calculated and they amount to $\sum b_i p_i = 5430419.58$ of currency units

3. De Novo Approach

De Novo formulation of the problem assumes that raw materials quantities are not limited, i.e. that any quantity can be purchased depending on the means available for that purpose. In other words, instead of the constraint on the available quantities of raw materials we obtain only one constraint, which is the constraint of the available budget (B). If raw material unit price is p_i (i = 1, ...,m) and the required quantity is b_i, then the production model can have the following form:

$$\text{Max } z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \quad (1)$$

$$\text{s. t.} \quad \begin{aligned} a_{11} x_1 + a_{12} x_2 + \ldots + a_{1n} x_n &= b_1 \\ a_{21} x_1 + a_{22} x_2 + \ldots + a_{2n} x_n &= b_2 \\ &\ldots \\ a_{m1} x_1 + a_{m2} x_2 + \ldots + a_{mn} x_n &= b_m \end{aligned} \quad (2)$$

$$p_1 b_1 + p_2 b_2 + \ldots + p_m b_m \le B \quad (3)$$

$$x_j, b_i \ge 0,\ j = 1, 2, \ldots, n;\ i = 1, 2, \ldots, m \quad (4)$$

Solving that model can be made simpler by the substitution of b_i equations into the budget equation, where

$$p_1 a_{1j} + p_2 a_{2j} + \ldots + p_m a_{mj} = v_j \quad (5)$$

represents the unit variable cost of producing product j.

In that way we obtain a very simple linear programming model with only one budget constraint, and possibly with additional constraints for market and technological reasons.

$$\text{Max } z = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \quad (6)$$

$$\text{s.t.} \quad v_1 x_1 + v_2 x_2 + \ldots + v_n x_n \le B \quad (7)$$

$$x_j \ge 0,\ j = 1, 2, \ldots, n \quad (8)$$

Let us assume that the available budget in our problem is exactly the amount spent on the purchase of raw materials from the Table 2, i.e. 5430419.58 of currency units.

Using that we can reformulate our ferroalloys production model and obtain the optimal solution:

$x_1^* = 38\ 004$ t (Fe Cr), $x_2^* = 16\ 000$ t (Fe Si),
$x_3^* = 16\ 142$ t (Fe Mn), $x_4^* = 4\ 747$ t (Fe Si Mn)

with maximum value of objective function: $z^* = 2\ 939\ 105$ currency units.

It can easily be seen that, in that case, maximum value of the objective function is greater than in the previous model, while the same amount of money is spent for raw materials purchasing as in first model. In this model raw materials are to be purchased in exactly determined quantities (there is no surplus and unused quantities of raw materials), and the necessary quantities b_i (i = 1,...,5) can easily be determined by inserting variables x_j in raw materials constraints, i.e. in relations (2).

4. Increasing cost of raw materials

The linear programming model is sometimes difficult to apply in real business situations due to its assumption of proportionality. A frequent phenomenon arising in practice is the varying price of the same resource. Namely, if a company needs additional quantities of raw materials it is possible to buy then from another supplier but at a different (usually higher) price. Let us assume that i-th raw material can be purchased at the price p_i, but only for the quantity lower than Q. To purchase i-th raw material above that quantity it is necessary to take another supplier whose price is $p_i' > p_i$. Then the relation for the i-th raw material is transformed into:

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{in} x_n = b_i + d_i, \quad (9)$$

with additional constraint $b_i \le Q$, where d_i is the additional quantity of the i-th raw material with unit price p_i'.

Since the same raw material has different price variable, income from end product unit is not constant anymore. Therefore, maximizing the sum of $c_j x_j$, would not be an accurate measure of net income. Net income equation (6) should be recalculated as the difference between sales and total cost of materials, where the objective function will include materials at both prices. Consequently, if s_j is the sales price of j-th product, the objective function has the following form:

$$\text{Max } z = \sum_{j=1}^{n} s_j x_j - \sum_{i=1}^{m} p_i b_i - \sum_{k=1}^{l} p_k' d_k \quad (10)$$

where d_k ($k = 1, \ldots, l$) are those materials which in additional quantities can be bought only at a higher price (p_k').

In the budget equation (3) it is also necessary to introduce costs for additional quantities of raw materials, so that it now takes the following form:

$$\sum_{i=1}^{m} p_i b_i + \sum_{k=1}^{l} p_k' d_k \leq B \tag{11}$$

There is no need to specify that b_i should reach the maximum value of Q first, before allowing d_i greater than zero. The optimization model ensures b_i reaching the maximum value of Q because of the lower penalty, i.e. lower price p_i.

4. Quantity discounts

Let us now consider such production situation when there are quantity discounts granted for bulk orders of raw materials. Therefore, in addition to the increasing cost effect we have to introduce this possibility into the model. Let us assume that for the k-th resource (b_k) the valid price is p_k as long as the purchased quantity is below Q, and the discounted price p_k' is valid for the entire quantity if the purchased quantity is higher than Q. Consequently the assumption is opposite to the one in the previous model, i.e. $p_k' < p_k$.

The previous formulation is not applicable since the optimization model will prefer using the less expensive material without satisfying the quota (Q). A different model has to be formulated with a slightly more complicated procedure.

Let

- b_k, p_k - the amount and price of k-th raw material if it is purchased at less than the quantity discount volume;
- d_k, p_k' - the amount and price of k-th raw material if it is purchased with the quantity discount.

The new model, in that case, instead of one equation for k-th raw material has two relations, and those are:

$$a_{k1} x_1 + a_{k2} x_2 + \ldots + a_{kn} x_n \leq b_k + M y_1 \tag{12}$$
$$a_{k1} x_1 + a_{k2} x_2 + \ldots + a_{kn} x_n \leq d_k + M y_2 \tag{13}$$

and, according to this, two relations as budget constraints:

$$p_1 b_1 + p_2 b_2 + \ldots + p_k b_k + \ldots + p_m b_m \leq B + M y_1 \tag{14}$$
$$p_1 b_1 + p_2 b_2 + \ldots + p_k' d_k + \ldots + p_m b_m \leq B + M y_2 \tag{15}$$

where M is a very great positive number ($M >> 0$).

Besides that new variables y_1 and y_2 are integer 0 - 1 variables, for which is:

$$y_1 + y_2 = 1 \tag{16}$$
$$y_1, y_2 = 0 \text{ or } 1 \tag{17}$$

In the above model there are two 0-1 variables y_1 and y_2, where due to the relation (16) only one of them always equals 1, and the other equals zero. Naturally, if the model comprises a number of resources that can be purchased at discounted price then there are more 0-1 variables.

The problem of mutual exclusiveness of the variables b_k and d_k can be introduced in the following way:

If $d_k = 0$ (there is no quantity discount) then the relation $b_k < Q$ is valid i.e. the needed quantity of raw materials is below the quantity required to obtain the discount.

Similarly, if $b_k = 0$ then the relation should be $d_k \geq Q$ (the quantity required for the discount is reached). Due to that, let us introduce two additional constraints:

$$b_k + M d_k \leq Q + N y_1 \tag{18}$$
$$d_k + N y_2 \geq Q + M b_k \tag{19}$$

where $N >> M >> 0$, i.e. N and M are very great positive numbers, but N is also much greater than M.

The model objective function remains the same, while p_k' stands for raw material prices, which are either discounted or rising.

Let us now consider such situation for our ferroalloy production model. The third raw material (coke) can be purchased at a discounted price if the bought quantity is $Q > 25\,000$, and this reduced price is valid for the entire quantity supplied, i.e. $p_3' = 30$. In addition to this, let us assume increasing costs for electrical energy and quartz in this way:

Permanent lack of electrical energy (b_1) and limited possibilities of quartz (b_5) that has to be imported result in different (higher) purchasing prices. The limit of electrical energy purchased at a lower price is 250 000 (000 kWh), while this limit in quartz is 25 000 t.

The purchasing price of the additional quantity of electrical energy is $p_1' = 12$, and of quartz $p_5' = 10$ currency units. Assuming the same budget level $B = 5\,430\,000$, and product sale level $s_1 = 257$, $s_2 = 242$, $s_3 = 135$ and $s_4 = 148$, the production model now is:

$$Max\ (257x_1 + 242x_2 + 135x_3 + 148x_4 - 8p_1 - 12d_1 - 19.5b_2 - 35.5b_3 - 30d_3 - 86b_4 - 7.1b_5 - 10d_5)$$

$$\begin{aligned}
0.2604\, x_3 + 0.48\, x_4 &\leq 7160 \\
x_3 - 3.4\, x_4 &\geq 0 \\
4x_1 + 9.6\, x_2 + 2.63\, x_3 + 4.75\, x_4 - b_1 - d_1 &= 0 \\
2.04\, x_3 + 0.5\, x_4 - b_2 &= 0 \\
0.5\, x_1 + 0.29\, x_2 + 0.456\, x_3 + 0.48\, x_4 - b_3 - M y_1 &\leq 0 \\
0.5\, x_1 + 0.29\, x_2 + 0.456\, x_3 + 0.48\, x_4 - d_3 - M y_2 &\leq 0 \\
0.035\, x_1 + 0.07\, x_2 + 0.022\, x_3 + 0.064\, x_4 - b_4 &= 0 \\
0.28\, x_1 + 2\, x_2 + 0.02\, x_3 + 0.58\, x_4 - b_5 - d_5 &= 0 \\
8\, b_1 + 12\, d_1 + 19.5\, b_2 + 35.5\, b_3 + 86\, b_4 + 7.1\, b_5 + 10\, d_5 - M y_1 &\leq 5\,430\,000 \\
8\, b_1 + 12\, d_1 + 19.5\, b_2 + 30\, d_3 + 86\, b_4 + 7.1\, b_5 + 10\, d_5 - M y_2 &\leq 5\,430\,000 \\
b_3 + M d_3 &\leq Q + N y_1 \\
d_3 + N y_2 &\geq Q + M b_3 \\
y_1 + y_2 &= 1 \\
y_1, y_2 &= 0 \text{ ili } 1 \\
0 \leq x_1 \leq 38004,\ 0 \leq x_2 \leq 16000,\ x_3 &\geq 0,\ x_4 \geq 4000 \\
b_1, b_2, b_3, b_4, b_5, d_1, d_3, d_5 &\geq 0
\end{aligned}$$

This is the problem of mixed integer programming which will be solved by branch and bound algorithm by Winqsb program. With the budget amounting to $B = 5\,430\,000$ of currency units, the optimal solution obtained is:

$x_1^* = 38\,004$ t (Fe Cr), $x_2^* = 15\,545.69$ t (Fe Si)
$x_3^* = 13\,600$ t (Fe Mn) $x_4^* = 4\,000$ t (Fe Si Mn), $y_1^* = 1$, $y_2^* = 0$

with maximum value of objective function:

$z^* = 10\,527\,090$ currency units,

while the optimal quantity of raw materials is:

$b_1^* = 250\,000$, $d_1^* = 106\,022.6$
$b_2^* = 29\,744$
$b_3^* = 0$ $d_3^* = 31.754.25$
$b_4^* = 2\,957.54$
$b_5^* = 25\,000$, $d_5^* = 19\,324.5$.

It is obvious that this solution provides a higher value of objective function than in the previous models for the same budget, which is only logical, as the entire quantity of the required third raw material (d_3) is purchased at a discounted price.

5. Conclusion

De Novo presents a special approach to optimization. Instead of "optimizing a given system" it suggests a way of "designing an optimal system". The basis of the standard approach is a production model with resources defined in advance, so that the constraints and the feasible set are fixed.

However, De Novo approach does not limit the resources as most of the necessary resource quantities can be obtained at certain prices. Resources are actually limited because their maximum quantity is governed by the budget which is an important element of De Novo.

Using De Novo approach most various cases can be handled more effectively than by the standard programming models. Changes in prices, technological coefficients, increasing costs of raw materials, quantity discounts and other similar and real production situations can be easily incorporated in De Novo model and provide very satisfactory solutions.

6. References

1. Z.Babić, N.Plazibat (1985): *Application of the Goal Programming to the Production Planning Problem,* Seventh European Congress on Operations Research - EURO VII, Bologna, Italy, June 16-19, 1985.
2. M Zeleny (1986): *Optimal System Design with Multiple Criteria: De Novo Programming Approach*, Engineering Costs and Production Economics, No. 10, p. 89-94.
3. M.T. Tabucanon (1988): *Multiple Criteria Decision Making in Industry*, Elsevier, Amsterdam.
4. R.J. Li and E.S. Lee (1990): *Multicriteria De Novo Programming with Fuzzy parameters*, Computers Math. Applic. Vol. 19, No. 5, pp. 13-20
5. Z.Babić, I.Pavić (1991): *De Novo Programming in MCDM,* 16th Symposium on Operations Research, September 9-11. 1991, Physica-Verlag, Heidelberg, 1992, p- 579-582.
6. Y. Shi (1995): *Studies on Optimum Path Ratios in Multicriteria De Novo Programming Problems,* Computers Math. Applic. Vol 29, No. 5, pp.43-50.
7. Z.Babić, I.Pavić (1996): *Multicriterial Production Programming by De Novo Programming Approach*, International Journal of Production Economics, Vol. 43, No.1, 1996. p-59-66.

COMPARISON OF THE APPLICABILITY OF COMPUTER SUPPORTED MULTI-CRITERIA DECISION-MAKING METHODS

Vesna Čančer
Faculty of Economics and Business Maribor, Razlagova 14, 2000 Maribor, Slovenia
E-mail: vesna.cancer@uni-mb.si

Abstract
This paper introduces and discusses the applicability of several multi-criteria decision-making (MCDM) methods. In addition to theoretical aspects of different types of scales and the assumptions about decision makers' abilities and preferences, it summarizes some of the conclusions, made by the MCDM methods' users in business practice. By utilizing the author's experience, it offers a critical overview of MCDM methods, together with an evaluation of adequate computer programs.

Keywords: decision support, methods, multi-criteria decision-making (MCDM), scales

1. INTRODUCTION

Methods for supporting business decisions are inevitable tools in problem solving, especially in the phase of problem definition (to identify, analyze and define problem) and in decision-making (to choose and verify possible solutions to the problem) [1, 2]. By utilizing the author's experience in using multi-criteria decision-making (MCDM) methods, the paper offers a critical overview of these methods (emphasizing MAVT, SMART, SWING, SMARTER, AHP, and discussing ANP, outranking approaches, interactive methods and preference programming), together with an evaluation of adequate computer programs (HIPRE 3+ and Web-HIPRE, Expert Choice, Logical Decisions® for Windows). Put into groups, they are introduced and discussed regarding their basic advantages and disadvantages, convenience in problem solving, applicability for different types of problems, the types of the obtained results, and even the participants' knowledge.

MCDM methods have already turned out to be applicable in business practice (see [1, 6, 11]). The author presents some of the real-life applications in Slovenia that result from her involvement in several research projects:

- creditworthiness assessment [4], benchmarking of business processes, environmentally oriented business decision-making, investments in new production technology [3], selection of organizational structures, selection of information systems (to support business decisions – on the micro level), and
- measuring a country's position and potential in a period of contemporary globalization [5] (to support national economic policymaking – on the macro level).

Some practical solutions that were found to solve the problems arising in these research projects are introduced as well: regarding, for example, phases of the process for the goal fulfillment, model structuring, hierarchy (re)structuring and weights (re)calculation, and consistency improvement.

2. SOME ASPECTS OF THE SELECTION OF APPROPRIATE MCDM METHODS

Multi-criteria decision-making describes the set of approaches that can help individuals or groups in researching important complex decision-making problems (see [1, 2]). They should be used when intuitive decision-making is not enough for several reasons: because of the conflicts between criteria or because of disagreement between decision makers about relevant criteria or their importance and about acceptable alternatives and preferences. To their applicability in solving complex problems contribute the following facts:

- They do not replace intuitive judgement or experience and they do not oppress creative thinking; their role is to complement intuition, to verify ideas and support problem solving.
- In MCDM we take into account multiple, more or less conflicting criteria.
- In this type of decision-making process we structure the problem.
- Users can compare different methods and assess their convenience. The most useful approaches are conceptually simple, transparent and computer supported.
- The aim of MCDM is to help decision makers learn about the problem, express their judgements about the criteria importance and preferences to alternatives, confront the judgements of other participants, understand the final alternatives' values and use them in the problem solving activities.

When applying these methods, the assumptions about decision makers' abilities and preferences are as follow:

- A decision maker is able to decide between two possibilities and to express his/her preferences.
- Relation 'is preferred to' is transitive:
  $$A\,P\,B \wedge B\,P\,C \Rightarrow A\,P\,C. \qquad (1)$$
  Transitivity is a basis for measuring decision makers' inconsistency.
- Considering (1), the relation 'A is preferred to C' is stronger than the relation 'B is preferred to C'.
- Solvability. When measuring the alternatives' values with respect to an attribute by using value functions, it is possible to assign scale values, e.g. 50, 25, 75.
- Finite upper and lower bounds of the alternatives' values.

In MCDM, *ranging* has a principal role. An analyst has to find a relative importance of each criterion or a relative preference to each alternative. Scales can be defined as nominal, ordinal, interval and ratio [2]. Nominal scales are the least restrictive and - consequently - the least informative: numbers, symbols or descriptions are used for identification. Ordinal scales enable ordinal ranging, where specific differences between the criteria importance or preferences to alternatives are not known. On the contrary, interval and ratio scales enable adjudging exact numerical level of importance or preference. With interval scales it can be expressed for how much the criteria importance or preferences to alternatives differ from each other. With ratio scales it can be expressed how many times the criteria importance or preferences to alternatives differ from each other.

One of the most widely applied sets of multi-criteria methods is *multi-attribute value (or utility) theory* (MAVT or MAUT) (see [1]). From the late 1960's, this set of methods has been developed not only by management scientists, mathematicians, psychologists, but also by practitioners in management, economic, environmental and public fields. The need to include different scientific, professional fields in the development of these methods results from the need to manage complexity. It has been improved to SMART (a simplified multi-attribute rating approach) and other approaches (for example SWING, SMARTER). They are supported by several computer programs, e.g. HIPRE 3+, Web-HIPRE [8] and Logical Decisions® for Windows [9]. One decade later the *Analytic Hierarchy Process* (AHP) method was developed [11], together with computer program Expert Choice [7]. It was completed to the Analytic Network Process (ANP), which is supported by Super Decisions. Thus it overcomes the traditional OR/MS approaches.

The use of the discussed methods would lead to over-complications when decision makers do not need so detailed results as they are obtained with these methods. Namely, often it is

good enough to find out which of the alternatives is the most preferred. Therefore the so-called *"outranking" approaches* have been developed since 1970's. The most widely applied are ELECTRE in more variants and PROMETHEE (see [12]). Further, *interactive methods* are especially applicable when a complete preference model is not constructed a priori and when alternatives need improvements (see [12]).

Since traditional methods can not satisfactorily support many complex processes, methods for the approximate specification of preferences are gaining force in enterprises. Decision makers can express approximate preference statements through interval judgements. *Preference programming* describes approaches (like PAIRS, PRIME and RICH) that can be helpful in group decision-making, too [10]. Easy-to-use software has been developed to support the interval techniques: WINPRE for supporting the preference programming, PAIRS and Interval SMART/SWING; PRIME Decisions as a software implementation of the PRIME method; RICH Decisions for supporting the RICH method [10].

When applying MCDM methods to several decision-making problems, we concluded that they should be approached step-by-step. We followed the phases of decision-making processes that are commonly acknowledged in the literature [1]: from identification of a problem, through problem structuring - model building, its use to inform and challenge thinking, to the creation and analysis of a plan of activities that can solve a problem, but we adapted and completed them for the problem's type [3, 4, 5].

3. COMPARISON OF THE MCDM SOFTWARE APPLICABILITY

Table 1 introduces software products for MCDM that have been paid much attention among experts in different business fields (because of user capabilities, availability of graphical elicitation techniques, and the possibility to transform subjective judgements into objective measures). It delineates our findings about their applicability according to our experiences.

Table 1: Some Findings About the Most Preferred Software Support for MCDM.

Software	Applicability (according to our experience)
HIPRE 3+, Web-HIPRE [8]	Especially applicable for the methods based on ordinal and interval scales: SMARTER, SMART, SWING, and for the measurement of alternatives' values with respect to each attribute by value functions, although it supports also the AHP method in the sense of pair-wise comparisons, and direct measurement of alternatives' values
Expert Choice [7]	Especially applicable for the AHP method that is based on a ratio scale (pair-wise comparisons), although it supports the measurement of alternatives' values with respect to each attribute by value functions and direct method
Logical Decisions® for Windows [9]	Especially applicable for problems where describing the alternatives is of special value (utility functions, AHP, adjusted AHP), and for weights' assessment with tradeoffs, as well as by direct entry, the SMARTER and the SMART method, weight ratios and the AHP

In the last year we are faced with the group decision-making upgrades of the computer programs that have been most preferred for individual MCDM in the last two decades. They are, for example, HIPRE 3+ Group Link, Expert Choice 11 with EC Decision Portal, and Logical Decision for Groups.

Representations of available decision analysis packages contain the information provided by the vendors and surveyed by the OR/MS researchers. The appropriate information can be easily found on the world's web pages. Before buying such a decision support package, experts in enterprises can use trial-free versions of computer programs to find out whether a product offers enough possibilities for a convenient preparation of their decisions. However, when using the results in answer and sensitivity reports, decision makers in enterprises must be provided with appropriate knowledge about the basics of the applied methods.

4. REAL-LIFE APPLICATIONS: FINDINGS

Among MCDM methods that we applied in several research projects in Slovenian enterprises, we found the AHP method specially applicable in connection with other decision-making tools, e.g. in:

- *Creditworthiness evaluation.* We delineated the process of the assessment of an enterprise's business partners' creditworthiness that is made by a firm itself (so-called internal ratings) where the problem is approached step-by-step [4]. When verifying its applicability for the creditworthiness assessment, and therefore for the selection of business partners, we concluded that the technique should involve the following steps: problem definition, elimination of unacceptable alternatives, problem structuring (building a model), establishing priorities (on importance and preferences), synthesis and sensitivity analysis with verification. Since qualitative factors come into play in the creditworthiness evaluation, the AHP technique was applied. Special attention was paid to the assessment of the criteria importance: we presented the procedure for the improvement of the decision makers' consistency. Studying the corrected intensity levels, they improved their understanding of the relationships among the criteria, and of the criteria meaning and importance as well.
- *Environmentally oriented business decision-making.* We developed the method for selecting among environmentally oriented business processes with respect to multiple criteria. Since the particularities of the business processes in the sample enterprise, decision makers' preferences, their judgements on importance, practical data about the business processes in this enterprise, as well as the research results on environmental management in the sample enterprises of the processing industry and those found in eco-balances were taken into consideration, the AHP technique was found appropriate in connection with other decision-making tools [3]. Pair-wise comparisons were successfully applied in the criteria determination, the assessment of the criteria importance and the data calculation for different business processes.
- *Process benchmarking.* The multi-criteria method for benchmarking of environmentally oriented business processes was developed and presented with a real-life case from the Slovene enterprise in the processing industry. We considered not only ecologically most acceptable manufacturing, but also other, different and conflicting criteria that are relevant to this complex goal. For this reason, the AHP technique was used; together with other decision-making tools, it is suitable for benchmarking in order to help in decision-making about business process re-engineering and selection of new production processes. The activity of benchmarking can be decomposed in several steps [6]. The approach used in this study involves the following steps: define the business process's critical success factors, identify the business processes to be included in the analyses, make analyses to determine strengths and weaknesses, and set performance goals for improvements.

Furthermore, we compared some of the popular MCDM methods in other real-life business applications, e.g.:

- *Investments in new production technology.* The methods for determination of criteria weights which base on an ordinal (SMARTER), interval (SWING, SMART) and a ratio

scale (AHP), as well as the ways of measurement of local alternatives' values (direct input, use of value functions and pair-wise comparisons) were used in a real-life problem of estimation of alternatives for acquisition of pattern sets in a Slovenian foundry. We were involved as outside experts (tutors) in this application. One expert, employed in this foundry, has appropriate knowledge about the used quantitative methods. By using interviews and questionnaires, the experts in this foundry determined the criteria, their importance and the alternatives for the model's building. The obtained final results showed the most convenient alternative for the acquisition of pattern sets and were used in decision-making. These experts evaluated SMART, SWING and SMARTER as more demanding in comparison with the AHP in this application. Similarly, pair-wise comparisons and the distributive and ideal modes of synthesis were recognized as the most important advantages of Expert Choice.
- *Selection of organizational structures.* Decision-making about the reorganization in a Slovenian service company was based on the evaluation of the internal and external factors that influence the selection of an organizational structure. The enterprise's top managers used the questionnaire (prepared by an employed expert with the appropriate knowledge about MCDM) to evaluate the factors' importance. The decision model was entered by computer programs Web-HIPRE and Expert Choice. Decision makers can select the alternative with the highest final value as the most appropriate organizational structure. Completed with sensitivity analysis, this practical business research can be used for the verification of the current organizational structure and as a guidance for further organizational aims.
- *Selection of information systems.* The discussed methods were used also in the selection of suitable approaches to further information systems' development in a Slovenian building company. The method based on an ordinal scale SMARTER is evaluated convenient in decision situations, in which it is possible to evaluate only the rank of the criteria importance. The enterprise's experts evaluated SMART, SWING and appropriate software as excellent tools in solving complex problems. However, they emphasized that the quality of the decisions made on the basis of their results depends on the responsibility in establishing priorities about the criteria importance and preferences to alternatives.

On the macro level, we developed a multi-criteria method for *measuring and analyzing the globalization* of national economies with the intention to establish and study thoroughly a country's position and potential for international integration [5]. The method includes some advantages of the AHP, emphasizing model structuring and establishing priorities on indicators' importance. When applying this methodology to measuring the globalization of national economies, we concluded that it should involve the following steps: problem definition, model structuring, data collecting and measuring, establishing priorities, model restructuring and weights (re)calculation (in some cases), synthesis, and sensitivity analysis.

5. CONCLUSIONS

Experts and managers in co-operative Slovenian enterprises found MCDM methods applicable in business practice for solving several complex management problems. Since common-practice methods cannot satisfactorily support many complex processes, including national policymaking, the multi-criteria method for measuring and analyzing the globalization of national economies can define their key success and failure spheres.

Let us briefly emphasize the most evident common conclusions that arise from the use of the above-mentioned methods in the presented applications. Experts and managers in the co-operative enterprises accepted MCDM methods as a new approach in solving complex problems since they considerably contribute to favorable business decisions. Because of the real-life problems' complexity, a chosen method (e.g. the AHP) should be used in connection with other decision-making methods. The most important advantages of the AHP that were mentioned by the analysts, experts and managers are: pair-wise comparisons, measuring inconsistency, the distributive and ideal modes of synthesis and sensitivity analysis. The methods based on an interval scale SMART and SWING were chosen as more convenient for solving problems with sufficient information basis. With some adaptations, the models structured and the methods developed in the described research projects can be used in similar cases on the micro and macro level.

When using the results in answer and sensitivity reports, decision makers must be provided with appropriate knowledge about the basics of the used methods in order to read, interpret and use these results for problem solving. The ability to learn and to co-operate with experts inside and outside an enterprise is, therefore, of high importance.

References

1. Belton, V., Stewart, T. J. (2002): *Multiple Criteria Decision Analysis: An Integrated Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers.
2. Čančer, V. (2003): *Analiza odločanja (Decision-making analysis. In Slovenian)*. Maribor: University of Maribor, Faculty of Economics and Business.
3. Čančer, V. (2004): *The multicriteria method for environmentally oriented business decision-making*. Yugoslav Journal of Operations Research, 14, 1: 65-82.
4. Čančer, V., Knez-Riedl, J. (2005): *Why and How to Evaluate the Creditworthiness of SMEs' Business Partners*. International Small Business Journal, 23, 2: 141-158.
5. Čančer, V., Bobek, V., Korez-Vide, R. (forthcoming in 2005): *The multi-criteria method for measuring economic globalization of national economies*. In: R. Scitovski (ed.), Proceedings of the 10th Conference of Operational Research KOI '04. Osijek, Zagreb: University of Osijek, Croatian Operational Research Society.
6. Dey, P. K. (2002): *Benchmarking project management practices of Caribbean organizations using analytic hierarchy process*. Benchmarking, An International Journal, 9, 4: 326-356.
7. Forman, E. H., Saaty, T. L., Shvartsman, A., Forman, M. R., Korpics, M., Zottola, J., Selly, M. A. (2000): *Expert Choice 2000*. Pittsburgh: Expert Choice; Inc.
8. Helsinki University of Technology: *Web-hipre help*. http://www.hipre.hut.fi, consulted April 2005.
9. Logical Decisions: *Logical Decisions® for Windows*. www.logicaldecisions.com, consulted April 2005.
10. Mustajoki, J., Hämäläinen, R. P., Salo, A. (2005): *Decision Support by Interval SMART/SWING – Incorporating Imprecision in the SMART and SWING Methods*. Decision Sciences, 36, 2: 317-339.
11. Saaty, T. L. (1999): *Decision Making for Leaders*. Pittsburgh: RWS Publications.
12. Vincke, Ph. (1992): *Multicriteria Decision-Aid*. Chichester: John Wiley & Sons.

AN INTERVAL AHP MODEL FOR EFFICIENCY EVALUATION OF PRODUCTION UNITS

Josef Jablonsky
Department of Econometrics, University of Economics
Praha, 130 67 Czech Republic
jablon@vse.cz, URL: http://nb.vse.cz/~jablon/

Abstract
The paper deals with models and methods for evaluation of efficiency of production units. The standard modeling approach for evaluation of efficiency is data envelopment analysis (DEA) based on the definition of efficiency as the ratio of outputs produced by the unit and inputs spent in the production process. Standard data envelopment analysis models divide the units into inefficient and efficient ones. The efficient units receive the efficiency score 100% by standard DEA models and can be further classified by super-efficiency DEA models. The paper discusses the possibility of using the AHP model with interval pairwise comparisons for evaluation and classification of efficient units and compares given results with super-efficiency DEA scores. The proposed approach is applied in assessing the efficiency of pension funds in the Czech Republic – the results given by super-efficiency DEA models and by the interval AHP model are compared and discussed.

Keywords: AHP, interval AHP, data envelopment analysis, efficiency

1. Introduction

There are several standard tools for evaluation of efficiency of production units. One of the most popular modelling approaches in this field is data envelopment analysis (DEA). DEA models measure relative efficiency of decision making units (DMU) that are usually described by several inputs spent for production of several outputs. Let us consider the set E of n decision making units E = {DMU_1, DMU_2, ..., DMU_n}. Each of the units produces r outputs and for their production m inputs are spent. Let us denote $\mathbf{x}^j = \{x_{ij}, i=1,2,\dots,m\}$ the vector of inputs and $\mathbf{y}^j = \{y_{ij}, i=1,2,\dots,r\}$ the vector of outputs of the DMU_j. Then **X** is the (m, n) matrix of inputs and **Y** the (r, n) matrix of outputs. A DEA model evaluates efficiency of the DMU_q, $q\in\{1,2,\dots,n\}$ by looking for a virtual unit with inputs and outputs defined as the weighted sum of inputs and outputs of the other units in the decision set - $\mathbf{X}\lambda$ a $\mathbf{Y}\lambda$, where $\lambda=(\lambda_1, \lambda_2,\dots, \lambda_n)$, $\lambda>\mathbf{0}$ is the vector of weights of the DMUs. The virtual unit should be better (at least not worse) than the analysed unit DMU_q. The problem of looking for a virtual unit can be generally formulated as a standard linear programming problem:

$$\begin{aligned} &\text{minimize} && z = \theta - \varepsilon(\mathbf{e}^T\mathbf{s}^+ + \mathbf{e}^T\mathbf{s}^-), \\ &\text{subject to} && \mathbf{Y}\lambda - \mathbf{s}^+ = \mathbf{y}^q, \\ & && \mathbf{X}\lambda + \mathbf{s}^- = \theta\mathbf{x}^q, \\ & && \lambda, \mathbf{s}^+, \mathbf{s}^- \geq \mathbf{0}, \end{aligned} \qquad (1)$$

where $\mathbf{e}^T = (1,1,\dots,1)$ and ε is a infinitesimal constant. The variables s^+, s^- are just slack variables expressing the difference between virtual inputs/outputs and appropriate inputs/outputs of the DMU_q. Obviously, the virtual inputs/outputs can be computed using the optimal values of variables of the model (1) as follows:

$$\mathbf{x}^{q'} = \mathbf{x}^q\theta^* - \mathbf{s}^-,$$
$$\mathbf{y}^{q'} = \mathbf{y}^q + \mathbf{s}^+.$$

The DMU_q is to be considered as efficient if the virtual unit is identical with evaluated unit (does not exist a virtual unit with better inputs and outputs). In this case $\mathbf{Y}\lambda = \mathbf{y}^q$, $\mathbf{X}\lambda = \mathbf{x}^q$ and the minimum value of the objective function z = 1. Otherwise the DMU_q is not efficient and minimum value of $\theta<1$ can be interpreted as the need of proportional reduction of inputs in order to reach the efficient frontier.

The number of efficient units identified by DEA models and reaching the maximum efficiency score 100% can be relatively high and especially in problems with a small number of decision units the efficient set can contain almost all the units. In such cases it is very important to have a tool for a diversification and classification of efficient units. That is why several DEA models for classification of efficient units were formulated. In these models the efficient scores of inefficient units remain lower than 100% but the efficient score for efficient units can be higher than 100%. Thus the efficient score can be taken as a basis for a complete ranking of efficient units. The DEA models that relax the condition for unit efficiency are called super-efficiency models. The super-efficiency models are always based on removing the evaluated efficient unit from the set of units. This removal leads to the modification of the efficient frontier and the super-efficiency is measured as a distance between evaluated unit and a unit on the new efficient frontier. Of course several distance measures can be used - this leads to different super-efficiency definitions. The first super-efficiency DEA model was formulated in (Andersen and Petersen, 1993).

The paper discusses possibility to evaluate the efficiency of production units by an alternative approach based on the analytic hierarchy process (AHP) model with interval judgements (IAHP model). It is organized as follows. The next section contains a brief introduction to AHP models with interval judgements. Section 3 formulates an IAHP model for evaluation of efficiency of decision making units. The numerical experiments are realised on the real-world example that consists in comparison of 12 pension funds operating in the Czech Republic. The results of the analysis by both the techniques are presented and discussed. The last section contains summarization of the paper and discussion about the future research.

2. The interval AHP model

The AHP is a powerful tool for analysis of complex decision problems. The AHP organizes the decision problem as a hierarchical structure containing always several levels. The first (topmost) level defines a main goal of the decision problem and the last (lowest) level describes usually the decision alternatives or scenarios. The levels between the first and the last level can contain secondary goals, criteria and subcriteria of the decision problem. The number of the levels is not limited, but in the typical case it does not exceed four or five. Let us consider a simple three-level hierarchy that can represent a standard decision problem with finite set of alternatives - evaluation of n-alternatives $X_1, X_2, \ldots, X_n$, by k-criteria $Y_1, Y_2, \ldots, Y_k$, (Figure 1). The decision maker expresses his preferences or compares importance of the elements on the given level with respect to the element of the preceding level. The information resulting from decision maker's judgements in the given level of the hierarchy is synthesised onto the local priorities. They can express, e.g. relative importance of criteria (weight coefficients - in Figure 1 denoted by v_j, i=1,2,...,k) or preference indices of the units with respect to the given criterion (w_{ij}, i=1,2...,n, j=1,2,...,k). In the standard AHP model the decision maker judgements are organised into paired comparison matrices at each level of the hierarchy. The judgements are point estimates of the preference between two elements of the level. Let us denote the paired comparison matrix $\mathbf{A} = \{a_{ij} | a_{ji} = 1/a_{ij}, a_{ij}>0, i,j=1,2,\ldots,k\}$, where k is the number of elements of the particular level. Saaty (1990) proposes to use for preference expression a_{ij} integers in the range 1 through 9, where 1 means that the i-th and the j-th element are equally important and 9 that the i-th element is absolutely more important than the j-th element. The local priorities are derived by solving the eigenvector problem (2)

$$\mathbf{A}.\mathbf{v} = \lambda_{max}\mathbf{v}, \qquad \sum_{i=1}^{k} v_i = 1, \tag{2}$$

where λ_{max} is the largest eigenvalue of **A** and **v** is the normalised right eigenvector belonging to λ_{max}.

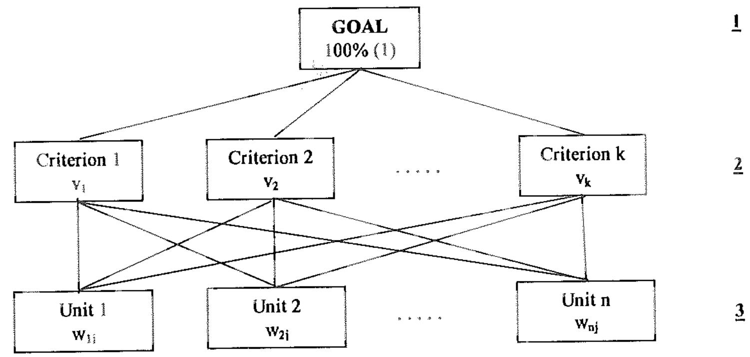


Figure 1: Three-level hierarchy.

In the standard deterministic AHP approach the decision maker always specifies point estimates that express his preference relations between two elements in the given hierarchical level. It can often be very difficult to fulfil this condition for decision makers. They feel much better and closer to have the possibility to express their preferences as interval estimates. For instance, instead of giving that the i-th element is four times as preferable as the j-th element, he can assert that the i-th element is at least two but no more than five times as preferable as the j-th element. The AHP model with interval decision maker judgements is usually called interval AHP (IAHP) model. It is characterised by interval pairwise comparison matrices given as follows:

$$A = \begin{bmatrix} 1 & <p_{12},q_{12}> & \cdots & <p_{1k},q_{1k}> \\ <p_{21},q_{21}> & 1 & \cdots & <p_{2k},q_{2k}> \\ \vdots & \vdots & \cdots & \vdots \\ <p_{k1},q_{k1}> & <p_{k2},q_{k2}> & \cdots & 1 \end{bmatrix} \tag{3}$$

where p_{ij} is lower bound and q_{ij} upper bound for preference relation (a_{ij}) between the i-th and j-th element. Due to the reciprocal nature of the pairwise comparison matrices the relation $p_{ij}.q_{ji} = 1$ holds for all i,j=1,2,...,k.

The judgements in the IAHP can be considered as random variables defined over the given interval. In this way the IAHP changes from the deterministic model to the model with some stochastic features. That is why it cannot be analysed in the traditional way – by solving the eigenvector problem (2). It is necessary to look for new approaches that will respect stochastic features. The random variables for description of interval judgements can be selected from the available probabilistic distributions. We will use the uniform distribution defined over the interval $<p_{ij}, q_{ij}>$ and in our numerical experiments below. The preferences of elements derived form from matrix A are random variables. Their characteristics can be computed by several approaches – we used a Monte Carlo simulation that is very simple and offers lower and upper bounds for the preferences in a very short time.

3. The IAHP model: evaluation of the Czech pension funds

The discussion concerning the DEA and IAHP models in evaluation of efficiency and especially super-efficiency will be demonstrated on a small numerical example with a real economic background. It is the problem of evaluation of efficiency of available pension funds in the Czech Republic. We worked with the data set for 12 pension funds, each of them was characterized by the following seven criteria (the data are from year 2003):

- INP 1 - the number of customers [thousands],
- INP 2 - total assets [mil. CZK],
- INP 3 - equity capital [mil. CZK],
- INP 4 - total costs [mil. CZK],
- OUT 1 - appreciation of the customer deposits for the last year (2003) [%],
- OUT 2 - average appreciation of the customer deposits for the last three years (2001 – 2003) [%],
- OUT 3 - net profit [mil. CZK].

For DEA analysis, first four criteria were taken as inputs and the remaining ones as outputs of the model. The criterion matrix is given in Table 1.

	#of cust.	assets	equity	tot. costs	appr. 1	appr. 3	profit
Allianz	106	4095	77.0	49.5	3.00	3.69	1.29
Credit Suisse	611	22592	549.0	454.1	3.36	3.67	5.22
CSOB Progres	18	452	56.0	15.1	4.30	4.15	1.13
CSOB Stabilita	304	8508	298.6	203.3	2.30	2.83	10.87
Generali	23	789	74.0	15.5	3.00	3.90	0.45
ING PF	346	9767	289.1	221.7	4.00	4.27	0.26
CP PF I	225	6348	290.7	184.7	3.34	3.65	6.83
CP PF II	518	12441	522.5	297.3	3.10	3.37	6.90
CS PF	401	10954	223.5	238.8	2.64	3.31	1.10
KB PF	285	11776	441.6	166.0	3.40	4.14	6.40
PF Ostrava	19	935	71.0	18.2	2.44	2.68	0.04
PF Zemsky	14	468	87.9	23.2	4.01	4.24	2.03

Table 1: Pension funds – criterion matrix.

It is clear that the funds listed in the previous table are of different nature. Four of them are very small (CSOB Progres, Generali, PF Ostrava and PF Zemsky) and the remaining ones are significantly bigger. That is why we decided to analyse them separately. In Table 2, there are results of DEA analysis of both groups of funds (big and small). We used the basic envelopment DEA model with variable returns to scale with output orientation and the super-efficiency model under the same assumptions. First column of Table 2 contains efficiency scores of the evaluated units - higher score corresponds to more efficient unit. The same holds for super-efficiency scores presented in the last column of Table 2. Of course the super-efficiency scores are available for units indicated as efficient by the standard model only. The word "infeasible" for super-efficiency score of Allianz fund means that the corresponding variable returns to scale (VRS) super-efficiency model has no feasible solution. This situation can occur in VRS super-efficiency models very often and in this case it disables the possibility to classify the efficient units.

DEA/VRS model	efficiency score	super-eff score
Big funds		
Allianz	100.00	infeasible
Credit Suisse	96.12	
CSOB Stabilita	100.00	159.24
ING PF	100.00	119.87
CP PF I	100.00	122.40
CP PF II	95.64	
CS PF	82.14	
KB PF	100.00	112.40
Small funds		
CSOB Progres	100.00	infeasible
Generali	93.72	
PF Ostrava	64.06	
PF Zemsky	100.00	infeasible

Table 2: Efficiency measures given by DEA models.

The IAHP model for evaluation of efficiency is very simple and it is presented on Figure 2. In this model, q_1 is the total weight of inputs and q_2 is the total weight of outputs, $q_1 + q_2 = 1$. qi_j, j=1,2,...,m, are the weights of single inputs and qo_j, j=1,2,...r, are the weights of single outputs. Preference indices u_{ij}, i=1,2,...,n, j=1,2,...,m+r, express the preference of the i-th alternative (pension fund) with respect to the j-th input/output.

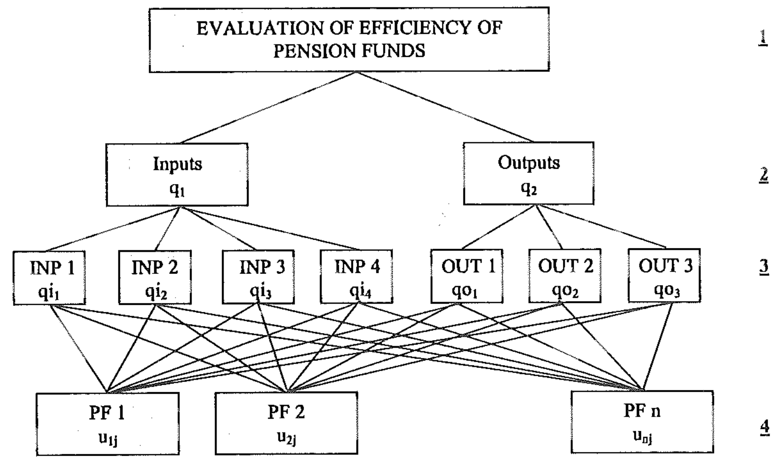


Figure 2: The IAHP model for evaluation of efficiency.

The global preference indices of alternatives p_i, i=1,2,...,n, are synthesized from previous hierarchical levels as follows:

$$p_i = \sum_{j=1}^{m+r} u_{ij}, \; i = 1,2,\ldots,n.$$

In the analysis, the AHP model does not derive the weights of the inputs and outputs but they are either to set up directly as constants or optimised as variables of the proposed simulation model. In the model the pairwise comparisons of alternatives (pension funds) with respect to all inputs and outputs are given as random variables with uniform distribution defined over the interval <a,b>. The comparisons reflect given criterion values but by this way it is possible to use different returns to scale for different inputs and outputs. It is one of the advantages of this approach. Our numerical experiments were realized on the set of five big pension funds identified as efficient by DEA model with variable returns to scale. The pairwise comparisons of all alternatives with respect to first input (number of employees) and output (appreciation of customer deposits) are presented in Table 3 (free cells are the reciprocal values). Similarly the comparison matrices for the remaining inputs and outputs are given.

	Number of customers				
	Allianz	CSOB	ING	CP PF	KBPF
Allianz	1	<6,8>	<7,9>	<3,5>	<5,7>
CSOB		1	<1,2>		
ING			1		
CP PF		<1,3>	<3,5>	1	<2,3>
KB PF		<1,2>	<2,4>		1

Appreciation of customer deposits				
Allianz	CSOB	ING	CP PF	KBPF
1	<3,5>			
	1			
<4,6>	<5,7>	1	<3,5>	<3,5>
<1,3>	<3,5>		1	1
<1,3>	<3,5>			1

Table 3: Pairwise comparisons of the IAHP model.

The DEA models maximize the individual efficiency of evaluated units by looking for optimum weights of inputs and outputs. That is why we did not work with weights derived by the IAHP model but we tried to optimise the weights in order to reach the best efficiency score of the evaluated alternative. We used the following requirements to the set of the weights:

1. The sum of weights equals to 1, all the weights have to be greater than 0.05 (AHP I).
2. The sum of weights of inputs equals to 0.5, the same holds for the weights of outputs. All the weights have to be greater than 0.05 (AHP II).
3. All the weights are fixed to value $1/(m+r)$, ie. 1/7 in our example (AHP III).

Pension funds	DEA/VRS super-eff.	AHP I	AHP II	AHP III
Allianz	infeasible	0.4574	0.3262	0.3334
CSOB Stabilita	159.24	0.3608	0.2583	0.1449
ING PF	119.87	0.3652	0.2764	0.1894
CP PF	122.40	0.2123	0.2056	0.1699
KB PF	112.40	0.2397	0.2138	0.1625

Table 4: Comparison of super-efficiency measures.

The optimisation run was realized by means of Crystal Ball software that is an MS Excel add-in application for Monte Carlo experiments. Crystal Ball contains a special tool for optimisation under stochastic conditions called OptQuest. This tool can find optimum values of variables (weights of inputs and outputs in our case) in stochastic environment that can be modelled within MS Excel. The optimisation criterion is the efficiency score p_i, $i=1,2,\dots n$, of the evaluated alternative that is to be maximized. Because the efficiency score under our stochastic conditions is a random variable we tried to maximize its mean value. We always used a five minutes optimisation run for all the alternatives with 100 trials per one simulation.

The results are presented in Table 4. The first column of Table 4 contains super-efficiency scores computed by the Andersen and Petersen DEA model with variable returns to scale, the remaining three columns contain maximized efficiency score of pension funds given by the presented IAHP model with different weight constraints (weight sets I, II and III).

The DEA/VRS model is not able to evaluate the Allianz fund, the best among the others is the CSOB Stabilita fund and the worse is the KB pension fund. The results given by the IAHP models are quite different. If we consider the same weights for all the inputs and outputs (AHP III), the Allianz fund is classified on the top and the remaining funds are very close each other. The results for the set of weights AHP I and AHP II are almost identical. The best is the Allianz fund, the mean efficiency score of the CSOB and ING pension funds is more or less the same and on the bottom of the ranking are the last two funds.

4. Conclusions

The aim of the paper was to verify how the AHP models can be used for efficiency evaluation of production units and compare the results given by proposed interval AHP model with efficiency scores computed by DEA models. In contrary to super-efficiency DEA models the advantage of the IAHP approach consists in several points:

- the IAHP model can use different scales for different inputs and outputs according to the decision maker preferences,
- in super-efficiency DEA models with VRS not all the units receive their super-efficiency score (the problems need not be always feasible),
- the IAHP model can deal with categorical inputs and outputs without any transformation,
- the IAHP model offers a possibility of sensitivity analysis of results with respect to the inputs and/or outputs.

The main disadvantage of the AHP (IAHP) model comparing to the DEA models consists in preparing the data set for analysis (pairwise comparison matrices) and in time consuming length of the optimisation analysis. Nevertheless, by using the AHP model the decision maker can receive new information useful for the global analysis of the efficiency of the evaluated set of units.

Future research in this field will be focused on comparison other super-efficiency DEA models (slack based models) with IAHP models and other multiple criteria decision making techniques. Several real-life economic applications will deal as background for numerical experiments.

Acknowledgements

The research is partially supported by the GA CR - project no. 402/03/1360.

References

[1] Andersen, P. and Petersen, N.C. (1993) "A procedure for ranking efficient units in data envelopment analysis", Management Science, 39, 1261-1264.
[2] Cooper, W.W., Seiford L.M. and Tone, K. (2000) Data Envelopment Analysis, Boston, Kluwer Publ.
[3] Jablonsky, J. (2004) "Models for Evaluation of Efficiency of Production Units (in Czech)" Politicka ekonomie, 2, pp.206-220.
[4] Saaty, T.L. (1990) The Analytic Hierarchy Process, Pittsburgh, RWS Publications.
[5] Zhu, J. (2003) Quantitative Models for Performance Evaluation and Benchmarking. Boston, Kluwer Publ.

DEVELOPMENT OF A MULTI-CRITERIA EVALUATION MODEL FOR CLASSIFYING WOOD PRODUCTS

Igor Lipušček, Lidija Zadnik Stirn, Leon Oblak
University of Ljubljana, Biotechnical Faculty
1000 Ljubljana, Slovenia
igor.lipuscek@bf.uni-lj.si
lidija.zadnik@bf.uni-lj.si
leon.oblak@bf.uni-lj.si

Abstract: This paper presents the process of developing a multi-criteria decision model for classifying wood products according to their environmental burden during their life cycle. The aim of the model development was to include the most important criteria of burdening the environment appearing in life cycle of wood products and to calculate the uniform estimate of environmental burden. For the generation of the model we used the Expert Choice methodology. The model is based on expert opinions gathered with the Delphi method and mathematically and statistically processed.

Keywords: wood industry, classification of wood products, multi-criteria decision support model, Delphi method, analytic hierarchy process

1. INTRODUCTION

Environment preservation and protection belong to the greatest challenges the world has set for the next decade. The consequence is an increasing number of environmental protection organisations, environmental standards, legislative regulations and limitations forcing companies in production which impacts the environment as little as possible. Furthermore, the consumers are becoming more ecologically conscious and are deciding on purchasing products whose raw materials, production, usage and removal do not represent too much impact on the environment. They are realising that certain products and their manufacturing have a different effect on burdening the environment they live in. If in the past, environment protection with a specific product was considered as merely something of no consequence and an additional unnecessary expense, it is an important marketable argument today, and a fundamental reason for the products to be sold at all in the future (Oblak 1999).

Dealing with these problems are also wood manufacturing companies that otherwise manufacture natural material, but the process of production is environmentally controversial. Therefore, the manufacturers of these products are facing a dilemma - which products to produce in order to impact the environment as little as possible.

From the viewpoint of the wood industry the environment is impacted by a number of factors: releasing of solid wastes (wood wastes, wastes of wood based materials, artificial substances wastes, wastes from lacquering, etc.); burdening of air with smoke gases from fire places and with waste air appearing in production, where wood dust, dissolvent and diluter vapours in gluing and lacquering are arising; appearance of waste waters with organic and inorganic parameters, etc. These parameters have various effects on the environmental burden. We namely distinguish several categories of impact, such as global warming, stratospheric ozone depletion, photochemical smog formation, eutrophication, human carcinogenicity, atmospheric acidification, aquatic toxicity, terrestrial toxicity, habitat destruction, depletion of non-renewable resources, waste heat, malodorous air and water, noise, etc. (Guinee et al. 2002; Rosselot and Allen 1999). Establishment and estimation of the influences on the environment, respectively a comparison of all these influences among themselves due to every enumeration, therefore, presents an enormous problem that needs dealing with. With the methods known up to the present day (ISO 14000) it is possible to estimate environmental burden for every category of impact separately (Guinee et al. 2002). However, it is impossible to give an overall estimate from the results of each category because they are incompatible. In presenting an overall estimate of environmental burden we use different models (Oblak and Zadnik Stirn 2000) based on methods of operations research, especially methods of multi-criteria decision-making (Bohanec and Rajkovič 1995, Zadnik Stirn 2001; Bell et al. 2003). Therefore, this crucial problem of calculating the overall estimation of environmental burden is separated into several sub-problems (Lai et al. 2002). These are then estimated separately according to all impact criteria they have an affect on. For the overall estimation of environmental burden we use the multi-criteria methods of the analytic hierarchy process (Saaty 1994, Winston 1994).

2. MODEL FORMULATION

For this reason we generated a multi-criteria mathematical model for classifying products according to their environmental burden in the entire life cycle (LC) of product which enables us a critical judgement of products inside a certain organisation, ranking of product concepts from the point of view of environmental burden and a comparison of environmental burden caused by wood and substitution products.

In developing of the decision-making model, we considered all most important parameters of environment burdening appearing in wood industry. In structuring the decision-making tree we considered the mutual dependencies, textual connections and the principles of perfection, operation, dismantlement, unredundancy and minimalism. The decision-making problem was divided into a larger number of sub-problems according to the principles of multi-criteria decision-making, which were then later divided into more levels and sub-levels. A decision tree is presented in Figure 1.

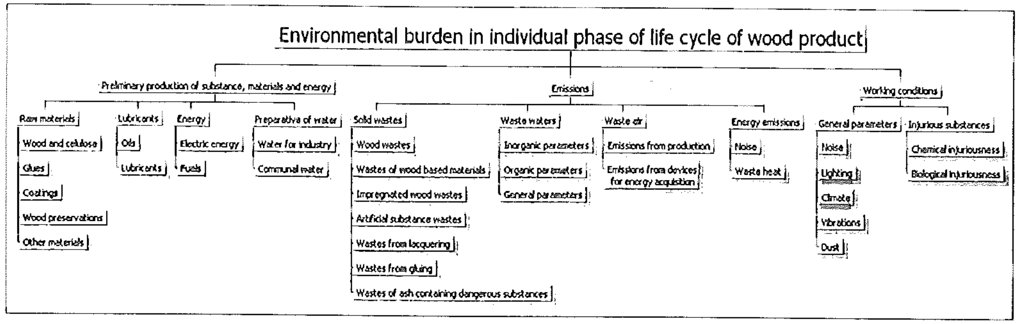


Figure 1. Decision tree for wood product classification regarding their impacts on environment

On the basis of the structured decision-making tree we developed a questionnaire for each of the decision-making levels, in which we pair-wise compared all the criteria at an individual level. It was determined which of the compared criteria presents a bigger impact on the environment and what the difference between them was.

The questionnaire was distributed to experts licensed to perform opinions on environment impacts, experts from the field of wood science and technology, ecology, environment protection and safety at work. The expert findings were handed over from 48 experts from university, Ministry of environment, research institutes, institutions for ecological research and other organisations for studying and protection of the environment. In pair-wise comparisons we were establishing which of the compared criteria represented a bigger burden on the environment and what the difference between the compared criteria of the burden is. In defining the differences between the compared criteria of the burden we derived from the assumption that there was no absolute environment suitability and from the fact that the impacts on the environment are so complicated that it is impossible for us to evaluate them correctly with the present knowledge of the consequences and the present level of information. The differences between the compared criteria were therefore presented by the experts on the basis of their expert knowledge of the criteria and the most important and well-known characteristics of the studied criteria. A sample of the questionnaire for the first level of the decision tree is shown in Figure 2.

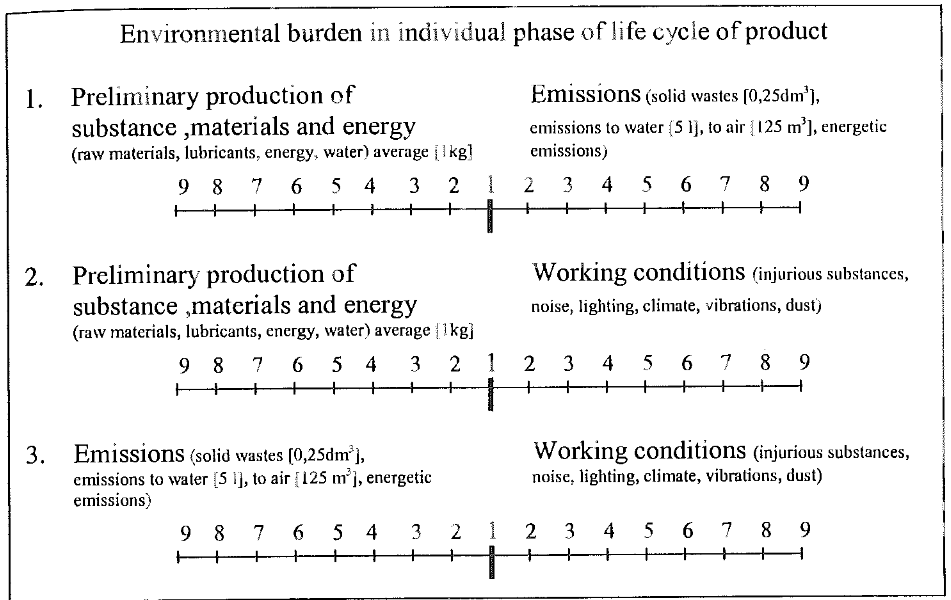


Figure 2. Sample of questionnaire for expert findings of pair-wise comparisons

The instructions for completing the questionnaire were as follows: "In continuation different criteria are being compared pair-wise. We are interested in which of the enumerated criteria presents a bigger impact on the environment and what the difference in intensity of the impact of a specific criterion is (according to all categories of environmental burden which the criterion appears in) in comparison with the compared one, if measured on a scale from 1 to 9."

In presenting the pair-wise comparison estimations three possible groups of answers were possible:

a) the two criteria have the same impact on the environment. This means there is no difference between the intensity of the environment impact, meaning the ratio between the two criteria is 1:1 or the estimation is 1, situated in the middle of the assessment scale shown in Figure 2;

b) the criterion on the left hand side of the assessment scale has a bigger impact on the environment than the criterion on the right hand side of the assessment scale. This means that the ratio of impact intensity of the left criterion against the right criterion is from 2:1 to 9:1 or the estimation is from 2 to 9 on the left side of the scale shown in Figure 2. The estimation depends on how much bigger the impact intensity of the left criterion is.

c) the criterion on the right hand side of the assessment scale has a bigger impact on the environment than the criterion on the left hand side of the assessment scale. This means that the ratio of impact intensity of the left criterion against the right criterion is from 1:2 to 1:9 or the estimation is from 2 to 9 on the right side of the scale shown in Figure 2. The estimation depends on how much bigger the impact intensity of the right criterion is.

The gathering of expert estimations was carried out with the Delphi method (Pečjak 2001; Ronde 2003; Chang et al. 2002; Handfield et al. 2002). In the first round, we collected ten expert opinions and processed them statistically-mathematically. We entered the statistical values into the questionnaire and repeated the round of gathering information on a larger group of forty-eight experts. The acquired data was processed again and we filtered disturbances.

After the expert opinion gathering was finished we calculated the mean values of the pair-wise comparisons. Individual estimation subjectivity was made unbiased by using mean estimation values of the impact intensity ratio for further calculations. For the measure of the variable mean we used the median, which is the most representative mean measure for data measured with an ordinal unit scale (Košmelj 2001). The consistency for the calculated mean values was verified.

On the basis of the expert opinion analysis, we used the AHP method and the Expert Choice software to calculate and define utility functions – in Figure 3 shown as priority factors written in brackets by parameters (L means local priority; G means global priority), which we determined the rules of movement on the decision-making tree with from the lowest level all the way to the final estimation of the variant, i.e. the product (Lai et al. 2002, Saaty 2003). Part of the decision-making tree is presented in Figure 3.

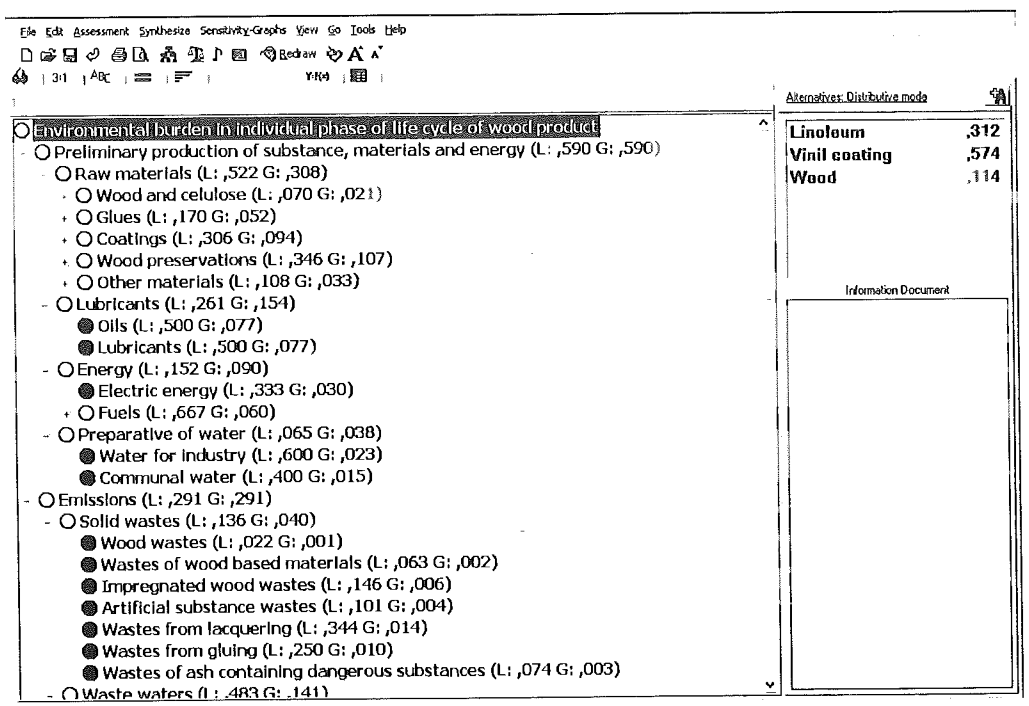


Figure 3. Decision tree for product classification with recorded priority factors

After calculating the priority factors we formed the final model for classifying wood products according to their impact on the environment during the whole life cycle of product. In the presented hierarchy structure of Figure 1 it is evident that the decision-making problem of evaluating environmental burden is first divided into three sub-problems. These are further on divided into even smaller sub-problems, etc., all the way to the lowest criteria of the decision tree. The functions of passing over to a higher level are shown in the form of priority factors (in Figure 3 shown in brackets by parameters), which are ascribed on each criterion of the decision tree.

3. CONCLUSION

The represented mathematical model for classifying products according to their environmental burden in the entire wood product life cycle contains all the most important criteria of environmental burden present in wood industry itself. It is shaped and made according to the principles of the multi-criteria decision-making, which enables easy deducting and adding of new criteria into the model if the model turns out to be set too widely or too narrowly. In the case of adding new criteria these are set at a definite level of the decision-making tree and we can estimate pair-wise comparisons between the entered and the existing criteria at this level and by using the AHP method we can re-calculate priority factors and correct utility functions of the changed decision-making tree. Utility functions at all other levels need not be changed.

The model is shaped for the entry of numerical data, which are measured for these compared products in the stage of the state of the studied variant life cycle inventory. In as far these values for a certain criteria cannot be measured directly, the AHP method enables us to enter the model in the shape of indirect comparisons and estimations as well. The AHP method can namely operate also with descriptive variables and estimations.

The mathematical model for classifying wood products according to their environmental burden in the entire life cycle, which is briefly presented here, enables not only calculating the final estimations of the differences in environmental burden among the studied products but also comparisons at all sub-levels of the decision-making tree and thus establishing distinct environmental advantages and disadvantages of a certain product compared to other products included in the research. The methodology of the multi-criteria decision-making also enables the sensitivity analysis of the final estimations to the changes of priority factors at any level of decision-making. The presented model thus enables us a critical judgement of products inside a certain organisation, ranking of product concepts and a comparison of wood products with substitute products. The model of classifying products according to their environmental burden can serve as support in choosing among the alternatives and deciding on a product, which in its entire life cycle, not only in an individual stage, represent the smallest impact on the environment. The model can also be used as help in judging the ecological quality of products.

REFERENCES

Bell, M. L., Hobbs, B. F. and Ellis, H. (2003), The use of multi-criteria decision-making methods in the integrated assessment of climate change: implications for IA practitioners, Socio-Economic Planning Sciences 37, 289–316.

Bohanec, M. and Rajkovič, V. (1995), Multi-criteria Decision-making Models (in Slovene, *translated by the authors*). Organizacija 28(7), 427-438.

Chang, P. C., Wang, C. P., Yuan, B. J. C. and Chuang, K. T. (2002), Forecast of development trends in Taiwan's machinery industry, Technological Forecasting & Social Change 69, 781–802.
Guinee, J.B., Gorree, M., Heijungs, R., Huppes, G., Kleijn, R., Koning, A., Oers, L., Wegener Sleeswijk, A., Suh, S. and Udo de Haes, H.A. (2002), Handbook on Life Cycle Assessment - Operational Guide to the ISO Standards. Dordrecht, Kluwer Academic Publishers.
Handfield, R., Walton, S. V., Sroufe, R. and Melnyk, S. A. (2002), Applying environmental criteria to supplier assessment: A study in the Application of the Analytical Hierarchy Process, European Journal of Operational Research 141, 70-87.
Košmelj, K. (2001), Applied Statistics (in Slovene, *translated by the authors*). Ljubljana, Biotechnical Faculty.
Lai, V. S., Wong, B. K. and Cheung, W. (2002), Group decision making in a multiple criteria environment: A case using the AHP in software selection, European Journal of Operational Research 137, 134-144.
Oblak, L. and Zadnik Stirn, L. (2000), The Possibility of Solving Economic and Environmental Protection Problems in Wood Industry Companies by the Application of the Method of Fuzzy Goal Programming. Ekólogia Bratislava 19(4), 409-419.
Oblak, L. (1999), Ecological Information System to Support the Decision-making Process in Manufacturing Firms, Proceedings of Slovenian Informatics Conference, DSI 1999, 794-800.
Pečjak, V. (2001), Ways to New Ideas: Techniques of Creative Thinking, Ljubljana, New moment.
Ronde, P. (2003), Delphi analysis of national specificities in selected innovative areas in Germany and France, Technological Forecasting & Social Change 70, 419–448.
Rosselot, K. and Allen, D.T. (1999), Life-Cycle Concepts, Product Stewardship and Green Engineering, http://www.epa.gov/opptintr/greenengineering/ch13_summary.html
Saaty, T. L. (2003), Decision-making with the AHP: Why is the principal eigenvector necessary, European Journal of Operational Research 145, 85-91.
Saaty, T.L. (1994), Fundamentals of Decision Making and Priority Theory, Pittsburgh, RWS Publications.
Winston, W.L. (1994), Operations Research: Applications and Algorithms, Bellmont Duxbury Press.
Zadnik Stirn, L. (2001), The Optimal Management of a Natural System Using the Multi-objective Saaty's AHP Method, Proceedings of Slovenian Informatics Conference, DSI 2001, 452-460.

MULTI-ATTRIBUTE UTILITY THEORY IN SUSTAINABLE RURAL LAND MANAGEMENT

Anka Lisec
University of Ljubljana, Faculty of Civil and Geodetic Engineering
Jamova 2, SI-1000 Ljubljana, Slovenia
e-mail: anka.lisec@fgg.uni-lj.si

Lidija Zadnik Stirn
University of Ljubljana, Biotechnical Faculty
Večna pot 83, SI-1000 Ljubljana, Slovenia
e-mail: lidija.zadnik@bf.uni-lj.si

Abstract
In this paper the multi-attribute utility theory is discussed in the context of complex realistic decision problems in the rural land planning and rural land management. The multi-attribute utility theory is presented at a general level, where the main stages of model development are described. At the end, the multi-attribute utility methodology is exposed as a sample case for supporting the decision process when implementing land consolidation in the rural landscape, with special emphasis on sustainable rural land management.

Keywords: rural land management, sustainable development, decision theory, multi-criteria decision making, multi-attribute utility theory.

1 INTRODUCTION

Decision making is of importance in many fields in social and natural sciences, including rural land management and rural land planning. The types of decision problems in the rural landscape (as for example the implementation of land consolidation) involve a large set of feasible alternatives, and multiple conflicting and incommensurate evaluation criteria with the main aim of providing the decisions which are economically, environmentally, and socially effective.

Rural land development and rural land management need to be based on balanced information about multiple objectives and require the integration of interdisciplinary knowledge from specialized fields, such as agriculture, biology, hydrology, economics, soil science and management science. Therefore, the use of multi-objective models has become a necessity to support and to guide constructive thinking about management decisions in rural areas [7]. Over the years, several models of how individuals should make multi-criteria choices have been developed. In general, the main role of multi-criteria analysis techniques is to deal with the difficulties that human decision-makers have been shown to have in handling large amounts of complex information in a consistent way. All approaches make the options and their contributions to the different criteria, and all require special techniques of judgement. They differ however in how they combine the data and usually provide an explicit weighting system for the different criteria. The methods can be used to identify the single most preferred option, to rank options, or to distinguish acceptable from unacceptable possibilities.

One of the approaches to the complex decision making process is to develop a model that evaluates alternatives giving an estimate of their utility for the decision-maker. Based on this estimate, the options are ranked and/or the best one is identified [1]. In the presented paper, the methodological framework of the multi-attribute utility theory (MAUT) is presented, and a sample problem is discussed.

2 MULTI-ATTRIBUTE UTILITY THEORY (MAUT): AN OVERVIEW

One of the major analytical tools associated with the field of decision analysis is multi-attribute utility theory, which derives from the work of von Neumann and Morgenstern (1947, In: [2]), and of Savage (1954, In: [2]). These authors aimed to derive a theory of how rational individuals ought to choose between alternatives. Their work provided powerful theoretical basis that is needed in order to make a quantification of preferences in the process of making decisions under uncertainty. Progress in the sense of developing methodology for the practical usage was made in the seventies, when Keeney and Raiffa (1976, In: [2]) developed a set of procedures, consistent with earlier normative foundations, which would allow decision-makers to evaluate multi-criteria options.

Multi-attribute utility analysis of alternatives explicitly identifies the attributes that are used to evaluate alternatives, and helps to determine those alternatives that perform well on majority of the selected attributes. The basis for the MAUT is a utility theory, which is a systematic approach to quantifying an individual's preferences. It is used to rescale a numerical value on several attributes of interest onto a 0-1 scale where 0 represents the worst preference and 1 the best one. This allows the comparison of many diverse measures, which is of significant importance in the multitude of relevant attributes. The end result is a rank ordered evaluation of alternatives that reflects the decision-makers' preferences [5].

The multi-attribute utility methodology for the evaluation of a set of alternatives typically consists of the following steps [5]:

- Identification of alternatives and attributes.
- Estimation of the performance of the alternatives with respect to the attributes.
- Development of utilities and weights for the attributes.
- Evaluation of the alternatives and sensitivity analysis.

2.1 Structuring the alternatives and attributes

Let as assume that there are n attributes $X_1,\ldots,X_n$ whose values are denoted by $x_1,\ldots,x_n$. The alternatives (d) and attributes (X) form a matrix in which each row corresponds to an attribute and each column describes the performance of the options against each criterion (see Table 1). The individual performance assessments are often numerical, but may also be expressed as a descriptive or other non-numerical valuable. Therefore, the information in the basic matrix has to be converted into consistent numerical values.

The next step of the procedure generates (I) a single attribute utility function over each attribute that is scaled from 0 to 1, a weight for each attribute, and (II) a multiple attribute utility function derived from the single attribute utility functions and the weights. The assessment of the single attribute utility function obtains possible ranges for the measures x_i of each attribute X_i, scaling the measures by assigning preference value of 0 to the worst level and 1 to the best level. Common forms of this function include concave for risk averse behaviour, convex for risk seeking behaviour, linear for risk neutral behaviour, and 'S' shaped for hybrid of the convex and concave functions [5]. According to Clemen (1991, In: [5]) one popular form for single attribute utility function u(x) is presented by the following equation (1) where the quantities A, B, and RT are parameters that must be set by the decision-maker.

$$u(x) = A - Be^{-(x/RT)} \tag{1}$$

There are several assessment techniques for eliciting utility functions from decision-makers. Several methods also exist for assigning weights (w) to the performance measures. One of those is the Trade-off method where decision maker provides the preference ordering among the attributes. This procedure in conjunction with the constraint that the weights must sum to one uniquely determines weights [6]:

$$w_1 + w_2 + \ldots + w_n = 1 \tag{2}$$

2.2 Evaluating the decisions

To quantify the preferences of the decision-maker the utility u is assigned to each of possible consequences $(x_1,\ldots,x_n)$. In the case when $u(x_1",\ldots,x_n") > u(x_1',\ldots,x_n')$, the decision-maker prefers the consequences $(x_1",\ldots,x_n")$ of the alternative d". The aggregation of the seemingly disparate measures is based on utility theory, that is presented by Keeney and Raiffa (1976, In [5]), and von Winterfeldt and Edwards (1986, In [5]). The use of utility theory ensures that any recommendation reflects [5]:

- the interactions between measures, if any,
- the relative attractiveness of a specific level on a measure,
- the relative attractiveness of performance on different measures.

When the decision-maker's preferences are consistent with the independence conditions, a multi-attribute utility model $u(x_1,\ldots,x_n)$ can be decomposed into an additive, multiplicative, or other well structured forms that simplify assessment. An additive function of multi-attribute utility model could be presented as follows [5]:

$$u(x_1,\ldots,x_n) = \sum_{i=1}^{n} w_i u_i(x_i), \tag{3}$$

where x_i represents the level of performance on measure i; $u_i(x_i)$ is a single attribute utility function over attribute x_i, scaled from 0 to 1; w_i is the weight for attribute x_i, and the weights w sum to 1.

An example of the multiplicative model, which is based on a weaker independence condition, is given as [5]:

$$1 + ku(x_1,\ldots,x_n) = \prod_{i=1}^{n}\left[1 + kk_i u_i(x_i)\right] \tag{4}$$

where k_i is a positive scaling constant satisfying $0 \le k_i \le 1$, and k is an additional scaling constant that characterizes the interaction effect of different measures on preference. When the sum of k_i amounts to one, the multiplicative model (4) reduces to the additive model (3). The advantage of using more complex and nonlinear functions is that more complex preferential behaviours can be represented, although the assessment of such nonlinear functions may require more time and possibly more cognitive burden [4].

3 SAMPLE PROBLEM

The main aim of land consolidation is on one hand the contribution of the productivity, efficiency, and competitiveness of the agricultural sector, and on the other the sustainable development of the rural areas. The objectives of land consolidation are to improve the land holdings of farmers by concentrating their farms in as few plots as possible, and to support the farms with roads and infrastructure, when needed. It leads to better land use planning and land management and if it is carried out in a comprehensive way, it should support environmental protection and natural resource management.

By the implementation of land consolidation, the multi-functionality of the rural landscape should not be overlooked. To a great degree, the rural landscape serves as agricultural production area and on the other side as habitat for different animal and plant species. Furthermore, it is part of the cultural landscape and provides different possibilities for development of tourism, recreation, relaxation, and sport activities in the countryside. Rural landscape with intensive agricultural production might cause visual and psychological side effects (Figure 1) and could hinder the additional activities. In the last decades, sustainable land management is becoming increasingly important in the process of the rural land consolidation, when a special attention to the conservation of natural and semi-natural ecosystems is being given [3].

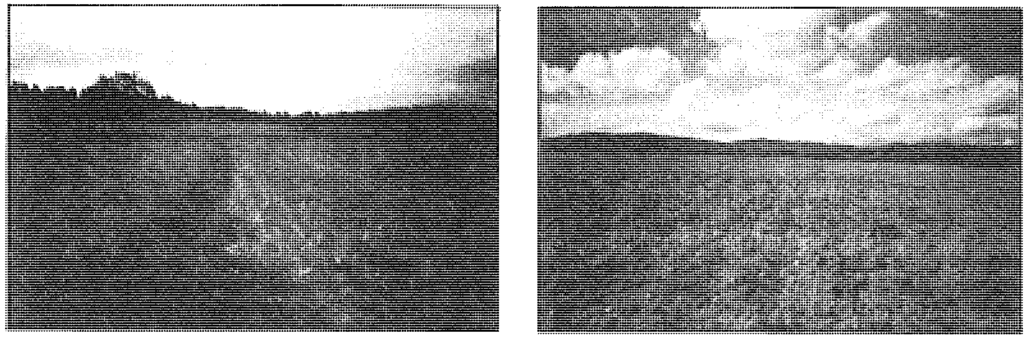


Figure 1: The rural landscape with variegated composition (left), and of intensive agriculture production (right).

3.1 Land consolidation objectives and attributes

To address the problem of estimating land consolidation alternatives in the selected rural area, we adopted the multi-attribute utility theory (MAUT) approach of Kenney and Raiffa (1976, In: [2]). Let us take a simple example of the multi-attribute utility model, where a linear, risk neutral, single attribute utility function is assumed. Further, we will assume that the decision-maker's preference structure is independent. Therefore, we may suppose a separable preference structure which is in the form of additive functions (3).

In the first stage of the multi-criteria decision analysis, the structure of multi-attribute model for the evaluation of alternatives has to be developed. The decision-maker must have information on different attributes that characterize the selected criteria, i.e. economics, environment, social etc. (Figure 2).

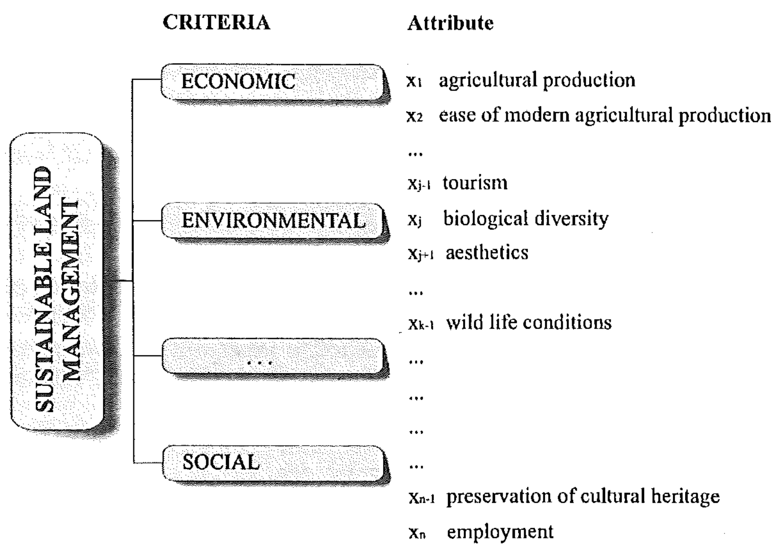


Figure 2: Criteria and attributes for policy choice model.

In the second stage, the alternatives are identified and described by the values of the chosen attributes. Each attribute is measured in different dimension and scales. All the values of the attributes have to be converted to the numerical values and could be collected in the performance matrix (Table 1). For the illustrative purposes, we assume that the decision maker has to choose between four alternatives (d) with respect to the sustainable land management, where plots' amalgamation and plots' re-allotment consider:

- d' Conservation of over 75 % of natural ecosystems in the original place;
- d'' Conservation of over 75 % of natural ecosystems, with the possibility of removing some of them to new location or replacing with new ones.
- d''' Conservation of 50 % of natural ecosystem, with the possibility of removing some of them to new location or replacing with new ones.
- d'''' No special attention to the natural ecosystem conservation – less than 50 % of such ecological elements remain in the consolidated area (along the river bank, roads etc.).

Attribute	**Decisions/ Alternatives**			
	d'	**d''**	**d'''**	**d''''**
X_1	x_1'	x_1''	x_1'''	x_1''''
X_2	x_2'	x_2''	x_2'''	x_2''''
...	...	...	...	...
...	...	...	...	...
X_n	x_n'	x_n''	x_n'''	x_n''''

Table 1: Performance matrix or consequence table.

As some of the attributes like social acceptance might be inherently inaccurate or difficult to obtain, some variations of the descriptions could be formed, anticipating either an optimistic or pessimistic development. Bohanec and Rajkovič [1] suggested the solution for taking into account such variability of the descriptive attributes with the increased number of alternatives (for example optimistic and pessimistic social acceptance) that provide a foundation for the subsequent what-if analysis.

In the next stage, a single attribute utility function has to be assigned to each of the attribute (Table 2). It is a scoring function that maps a performance measure from its range of possible values to [0,1]. We assume that the decision maker is risk neutral and the linear utility function for single attribute is given with (Anandalingam, 1987, In: [7]):

$$u(x) = \frac{x - x_{worst}}{x_{best} - x_{worst}} \quad (5)$$

Otherwise, according to (1), non-linear single attribute function would be assigned, where parameters A, B and RT have to be set by the decision-maker.

Attribute	**x_{worst}**	**x_{best}**	**u(x)**
X_1	x_{1worst}	x_{1best}	$(x-x_{1worst})/(x_{1best}-x_{1worst})$
X_2	$x_{2\ worst}$	x_{2best}	$(x-x_{2worst})/(x_{2best}-x_{2worst})$
...	...	...	...
...	...	...	...
X_{n-1}	$x_{(n-1)\ worst}$	$x_{(n-1)best}$	$(x-x_{(n-1)worst})/(x_{(n-1)best}-x_{(n-1)worst})$
X_n	$x_{n\ worst}$	x_{nbest}	$(x-x_{nworst})/(x_{nbest}-x_{nworst})$

Table 2: Utility function for single attribute.

Once the performance of each alternative on each attribute has been obtained, the next step in the analysis is the aggregation of the seemingly disparate attributes and determination of the desirability of each alternative. Assuming that decision-maker's preference structure is independent, an additive multiple attribute utility model can be used (3). In that case, the overall evaluation of attribute X is defined as a weighted sum of its evaluation with respect to its relevant value dimensions. The weight (w_i) determines the impact of the alternative policy on the measurable attributes (X_i), which are estimated by the decision-maker. The attribute weights are normalized using the fact that the weights should sum to one (2).

The evaluation of the alternatives rank the alternative policies according to the value of the multiple attribute utility function $u(x_1,\ldots,x_n)$. In the case that the decision maker prefers the alternatives like d" > d' > d''' > d'''', the values of the utility function are ranked as follows:

$$u(x_1'',\ldots,x_n'') > u(x_1',\ldots,x_n') > u(x_1''',\ldots,x_n''') > u(x_1'''',\ldots,x_n''''). \qquad (6)$$

CONCLUSION

Rural landscape management has been identified as providing important non-market benefits to society, which includes landscape preservation, environmental protection, preservation of cultural heritage etc. Decisions in the rural land management involve comparing alternatives that have strengths or weaknesses with regard to the multiple objectives of interest. It is difficult to derive a measure for non-agriculture uses of the rural landscape. Progress in this direction may be achieved also with the proposed multi-attribute utility approach. Special challenges to implementing MAUT in the rural landscape management are the valuations of the various externalities, and the determination of the interactions between different measures.

REFERENCES

[1] Bohanec, M., Rajkovič, V., 1999. Multi-Attribute Decision Modelling: Industrial Applications of DEX. In: Informatica, Number 23, Ljubljana, pp. 487-491.

[2] Dodgson, J., Spackman, M., Pearman, A., Phillips, L., 2005. Multi-criteria analysis manual. Office of Deputy Prime Minister, London, Internet source: http://www.odpm.gov.uk/stellent/groups/odpm_control/documents/contentservertemplate/odpm_index.hcst?n=7096&l=2

[3] Lisec, A., Pintar, M., 2005. Conservation of natural ecosystems by land consolidation in the rural landscape. In: Acta agriculturae Slovenica, Number 1, Volume 85, Ljubljana, pp. 73-82.

[4] Malakooti, B., Subramanian, S., 1999. Generalized polynomial decomposable multiple attribute utility functions for ranking and rating multiple criteria discrete alternatives. In: Applied Mathematics and Computation, Number 106, pp. 69-102.

[5] Morrice, D.J., Butler, J., Mullarkey, P.W., 1998. An approach to ranking and selection for multiple performance measures. In: Proceedings of the 1998 Winter Simulation Conference. http://www.informs-sim.org/wsc98papers/096.pdf

[6] Zadnik Stirn, L., 2002. Operacijske raziskave - Interno gradivo na podiplomskem študiju. Biotechnical Faculty, Ljubljana, pp. 134-146.

[7] Zadnik Stirn, L., 2001. Analysing rural development projects under uncertainty. In: Proceedings of the 6th International Symposium on Operational Research SOR'01, Slovenian Society Informatika, Ljubljana, pp. 313-318

The 8th International Symposium on Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 8
Networks

ON DISTRIBUTED SOLVING OF THE CAPACITATED VEHICLE ROUTING PROBLEM WITH BRANCH-AND-CUT ALGORITHMS

Matko Botinčan, Goranka Nogo
Department of Mathematics, University of Zagreb
mabotinc@math.hr, nogo@math.hr

Abstract. In this paper, we describe theoretical and implementation issues concerning solving the capacitated vehicle routing problem with branch-and-cut algorithms. A distributed solver for the capacitated vehicle routing problem has been implemented within the branch-and-cut framework SYMPHONY. We present results obtained during experimental evaluation of the solver on a cluster of computers and analyze its scalability.

Keywords: vehicle routing problem, branch-and-cut, distributed computing, experimental evaluation

1. Introduction

Vehicle routing problem (VRP) is the name for a class of problems in which a minimum-cost set of routes for a fleet of delivery vehicles originating and terminating at a depot must be determined to serve known customer demands. In most practical applications, the VRP is enriched with additional particularities such as time windows, backhauls, pickup and delivery, etc. However, in this paper we shall concentrate on solving its most basic variant in which no additional constraints are presented and every vehicle has a limited capacity. This variant of the VRP is commonly referred as the capacitated VRP (CVRP).

The purpose of this paper is to present basic theoretical and implementation issues concerning a branch-and-cut CVRP solver and report some of the experimental results obtained during evaluation of the distributed version of the solver on a cluster of computers. Particular contribution of the paper is that sequential and parallel running times for one concrete CVRP instance are thoroughly analyzed and related to major implementation bottlenecks of a branch-and-cut algorithm.

The rest of the paper is organized as follows. In Section 2, we give a precise mathematical definition of the CVRP and its integer programming formulation. A theoretical overview of branch-and-cut algorithms is given in Section 3. In Section 4 we describe our implementation of the CVRP solver within the branch-and-cut framework SYMPHONY. In Section 5, we present obtained experimental results and analyze scalability of the solver. The final Section 6 gives concluding remarks.

2. Problem definition and its integer programming formulation

Let $G = (V, E)$ be a complete graph with the vertex set $V = \{0, \dots, n\}$ and the set of edges E. We assume that vertices $1, \dots, n$ correspond to the customers, while the vertex 0 corresponds to the depot. A nonnegative cost c_{ij} is associated with each edge $(i, j) \in E$ representing the travel cost between vertices i and j. The cost structure is assumed to be symmetric, i.e. for all i and j holds $c_{ij} = c_{ji}$ and $c_{ii} = 0$. Each customer $i \in \{1, \dots, n\}$ is associated with a quantity $d_i \geq 0$ of a single commodity which needs to be delivered. A set of K identical vehicles, each with capacity C, is available at the depot (it is assumed that $d_i \leq C$ for all $i \in \{1, \dots, n\}$). The objective of the CVRP is to find a collection of K simple circuits in the graph G with a minimum cost (each circuit corresponding to a vehicle route) such that each customer is served by a single vehicle route, each vehicle route leaves from and returns to the depot, and the sum of the demands of the customers visited by each vehicle route does not exceed given vehicle capacity C.

From this description it is clear that the CVRP subsumes several difficult combinatorial optimization problems as its special cases. For instance, if we take $C = \infty$ we get an instance of the multiple traveling salesman problem. If we additionally take $K = 1$, a TSP instance is obtained. On the other hand, one can obtain an instance of a bin packing problem by putting $c_{ij} = 0$. It is worth noting that this BPP aspect makes the CVRP a considerably more difficult combinatorial optimization problem than the TSP.

In order to solve a given CVRP instance with a branch-and-cut algorithm, we first need to formulate the problem as an integer programming problem. Hence, we associate an integer variable x_e with each edge $e \in E$ and obtain the following integer programming formulation of the CVRP:

$$\min \sum_{e \in E} c_e x_e \tag{1}$$

$$s.t. \sum_{e=\{i,j\} \in E} x_e = 2, \quad \forall i \in V \setminus \{0\} \tag{2}$$

$$\sum_{e=\{0,j\} \in E} x_e = 2K \tag{3}$$

$$\sum_{\substack{e=\{i,j\} \in E,\\ i \in S, j \notin S}} x_e \geq 2b(S), \quad \forall S \subseteq V \setminus \{0\},\ S \neq \emptyset \tag{4}$$

$$0 \leq x_e \leq 1, \quad \forall e = \{i,j\} \in E,\ i,j \neq 0 \tag{5}$$

$$0 \leq x_e \leq 2, \quad \forall e = \{0,j\} \in E \tag{6}$$

$$x_e \in \mathbb{Z}, \quad \forall e \in E. \tag{7}$$

The quantity $b(S)$ represents a lower bound on the number of vehicles needed to serve customers in the set S. In this paper we take an obvious lower bound $b(S) = \lceil \sum_{i \in S} d_i / C \rceil$, whereas the resulting family of facet-defining inequalities (4) is then referred as capacity constraints.

3. Branch-and-cut algorithms

In this section we give a brief overview of a general class of methods for solving hard combinatorial optimization problems which are encompassed under the term branch-and-cut. Intuitively, every branch-and-cut algorithm consists of a classical branch-and-bound approach strengthened with an additional use of LP relaxations and problem specific cutting planes at every node of the enumeration tree.

Branch-and-bound algorithms employ divide-and-conquer strategy for partitioning the space of feasible solutions into smaller subproblems and solving each of the subproblems recursively. Let $\mathcal{S}$ denote the subproblem which is analyzed in a step of the branch-and algorithm. Initially, $\mathcal{S}$ represents the set of all feasible solutions of a given minimization problem. In its bounding phase, $\mathcal{S}$ is relaxed and the solution of the relaxation gives a lower bound for the value of an optimal solution to the subproblem. If its

value is larger than the value of the best upper bound previously found, the subproblem $\mathcal{S}$ is not considered further. If the solution of the relaxation is a member of $\mathcal{S}$, then it is optimal and the subproblem is solved. Otherwise, a partition $\{\mathcal{S}_1, \ldots, \mathcal{S}_p\}$ of the set $\mathcal{S}$ is identified, where each $\mathcal{S}_i$ represents a new candidate subproblem. The algorithm continues with selecting one of the candidate subproblems and processes it in the same way.

The bounding operation in the branch-and-bound algorithm is often accomplished by using linear programming techniques. Typically, the integrality constraints of an integer programming formulation of the problem are relaxed and the obtained LP relaxation is solved with a LP solver. If globally valid (usually faced-defining) inequalities are used in the LP relaxation, the resulting algorithm is called *branch-and-cut* algorithm.

For a more precise description, let us assume that the subproblem $\mathcal{S}$ is described by a set of inequalities $\mathcal{L}_\mathcal{S}$ such that $\mathcal{S} = \{x \mid x \text{ is integral and } \alpha^T x \leq \beta,\ \forall(\alpha, \beta) \in \mathcal{L}_\mathcal{S}\}$. According to the classical result of Farkas, Weyl and Minkowski [4], there exists a finite set $\mathcal{L}$ of inequalities such that $\text{conv}(\mathcal{S}) = \{x \mid \alpha^T x \leq \beta,\ \forall(\alpha, \beta) \in \mathcal{L}\}$. The inequalities in $\mathcal{L}$ represent *cutting planes* which will be used for strengthening the LP relaxations. As the problem of finding a member of $\mathcal{L}$ which is violated by a given point in space has the same computational complexity as the original optimization problem [7], it is not possible to enumerate all of the inequalities in $\mathcal{L}$ in practice. Instead, they are defined implicitly and separation algorithms have to be used to attempt to find at least some of these inequalities when they are violated.

However, separation algorithms need not always succeed to find a violated member of the set $\mathcal{L}$. If no inequalities from $\mathcal{L}$ can be found, the branch-and-cut algorithm is forced to branch. The branching operation is accomplished by specifying a set of planes which will divide the subproblem such that the current solution will not be feasible for any of the new subproblems. One simple method is to select an inequality $\alpha^T x \leq \beta$ having integer coefficients on the left side and $\beta \notin \mathbb{Z}$ on the right side, and then branch by generating two subproblems defined by $\alpha^T x \leq \lfloor \beta \rfloor$ and $\alpha^T x \geq \lceil \beta \rceil$.

4. Implementation issues

To implement a CVRP solver we have decided to use one of the existing open source branch-and-cut software frameworks. We have chosen SYMPHONY [8], developed by T. K. Ralphs and L. Ladany, which is a part of open source COmputational Infrastructure for Operations Research COIN-OR [3]. The design goal of SYMPHONY is to provide an easy-to-use framework within which distributed solvers for various types of optimization problems can be implemented independently of the underlying hardware architecture. The internal library, which contains majority of the subroutines needed for an implementation of a distributed branch-and-cut algorithm (tree management, cutting planes pool, interface to the LP solver, inter-process communication, etc.), interfaces with a user through a well-documented application programming interface. In order to implement a distributed solver, the user only needs to provide problem-specific methods and connect them to the SYMPHONY's internal library.

SYMPHONY's functions are grouped into five independent computational modules. This modular implementation allows easy and configurable parallelization. The modules can be compiled as a single sequential code as well as separate processes intended for running within the PVM distributed environment, hence, with a little additional effort, a user can develop a sequential and a distributed version of the solver at the same time. As we are interested in the analysis of scalability issues, it is important that the sequential and parallel version of the solver are developed under the same framework. Distributed version of our CVRP solver was compiled in such a way that SYMPHONY's master module, tree manager and cutting planes pool constituted one master process, while the cut generator and the linear programming module were built as a stand-alone slave process (several of which may run concurrently).

As it is not possible to give here a detailed description of the implementation of our CVRP solver within SYMPHONY's framework, we shall stress out only the most important remarks. Our implementation is based on the VRP sample application publicly available as a part of SYMPHONY. However, it differs from it in a number of ways, most notably in the use of separation procedures. Namely, we have provided own implementation of procedures for separation of the capacity constraints. Procedures are based on connected components and shrinking heuristics given in [6]. Moreover, we have employed several separation procedures for other classes of facet-defining inequalities available in the CVRPSEP package developed by J. Lysgaard [2] as well as those available in the CGL (Cut Generator Library) which is an integral part of COIN-OR [3]. Nevertheless, for ease of presentation and possibility to compare the obtained results with those in [5], experiments in this paper employed procedures for separation of capacity constraints only.

5. Experimental results

Experimental evaluation of our branch-and-cut CVRP solver has been performed on the cluster Isabella located at the University Computing Centre, University of Zagreb. In our experiments we have used 16 dual-processor nodes Pyramid GX28 (AMD Opteron 248). Unfortunately, due to limitations on the size of this paper we are able to present only a small part of the obtained experimental results.

Our main interest was comparison of the running time of the sequential version of the solver (denoted as T_0) with the running time of the distributed version of the solver. The distributed version of the solver consists of one master process and p slave processes (incorporating cut generator and LP solver), each process running on its own processor. Its running time we denote with T_p. For each $p \in \{1, \dots, 16\}$ we calculate the parallel speedup, defined as the ratio T_0/T_p, which describes the speed advantage of the parallel algorithm against the sequential one. We also calculate the parallel efficiency which is defined as the ratio $T_0/(pT_p)$ of the parallel speedup to the number od processors.

As a first example, we present results obtained for the P-n50-k7 CVRP instance. Running time of the sequential version of the solver on that CVRP instance was 551.472 seconds. Running times, parallel speedups and efficiencies of the distributed version of the solver for a number of processors ranging from 1 to 16 are presented in Table 1.

As we can see from the table, parallel efficiency is strictly less than one due to parallel overhead. Parallel overhead generally has tree basic components: idle time, redundant work and communication overhead. As argued in [5] and evidenced by a detailed analysis of our experimental results not shown here, communication overhead (time spent for packing and unpacking messages) and redundant work (time spent performing a work that would not have been performed in the sequential algorithm) is not an issue. However, idle time, which main contributors are ramp-up time (time occurring at the beginning of the algorithm when there are not enough jobs available to employ all processors) and handshaking time (time spent by the node processing modules waiting for

Number of processors	Running time [s]	Parallel speedup	Parallel efficiency
1	728.829	0.757	0.757
2	366.152	1.506	0.753
3	285.473	1.932	0.644
4	219.365	2.514	0.628
5	172.705	3.193	0.639
6	143.431	3.845	0.641
7	119.860	4.601	0.657
8	106.169	5.194	0.649
9	95.631	5.767	0.641
10	82.285	6.702	0.670
11	74.552	7.397	0.672
12	68.437	8.058	0.672
13	65.980	8.358	0.643
14	50.973	10.819	0.773
15	47.413	11.631	0.775
16	45.606	12.092	0.756

Table 1: Experimental results obtained for P-n50-k7 CVRP instance

requests to be serviced by another module — the tree manager or the cutting planes pool module) presents a problem which cannot be significantly improved and this type of overhead presents a very challenging task for implementors of branch-and-cut frameworks in general.

As a second example, we show a graphical representation of running times of the distributed version of the solver against the number of processors for three CVRP instances: A-n53-k7, B-n51-k7 and P-n50-k7. These CVRP instances are moderate in size and are chosen so that corresponding running times would be approximately of the same order of magnitude and hence could be drawn on the same graph. The graph is presented in Figure 1. Description of the analyzed CVRP instances can be found at [1].

6. Conclusions and future work

Branch-and-cut is an algorithmic approach for solving hard combinatorial optimization problems which has become very popular in the last fifteen years. In this paper, we have presented its theoretical aspects and described a distributed CVRP solver implemented within the branch-and-cut framework SYMPHONY.

Although the performance of our implementation is not comparable to that of the best state-of-the-art solvers which can solve CVRP instances of more than 100 nodes, the results we have obtained are still interesting. Our goal was not to compete with the best state-of-the-art solvers, but to analyze scalability of our CVRP solver on moderate-size CVRP instances. As there is a lack of detailed experimental evaluation and performance analysis of distributed CVRP solvers in the literature, our results though present a worthy contribution.

Moreover, we hope that with additional experiments we will be able to analyze limitations of our implementation more precisely and improve its bottlenecks. Also, a more clever utilization of additional separation procedures as well as adaptive branching criteria could improve the performance of our CVRP solver on larger CVRP instances.

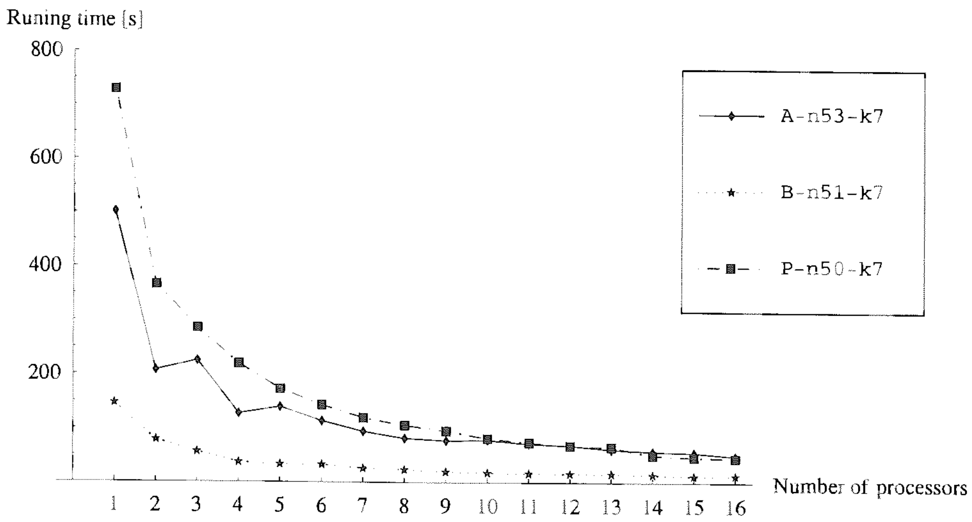


Figure 1: Running times for three CVRP instances with respect to the number of processors

References

[1] Capacitated Vehicle Routing Problem Data Sets. *http://branchandcut.org/VRP/data*

[2] J. Lysgaard. CVRPSEP package. *http://www.hha.dk/ lys/CVRPSEP.htm*

[3] COIN-OR web page. *http://www.coin-or.org*

[4] G. L. Nemhauser and L. A. Wolsey. Integer and Combinatorial Optimization. John Wiley & Sons, 1999.

[5] T. K. Ralphs. Parallel Branch and Cut for Capacitated Vehicle Routing. Parallel Computing 29(5): 607 – 629, 2003.

[6] T. K. Ralphs, L. Kopman, W. R. Pulleyblank and L. E. Trotter Jr. On the Capacitated Vehicle Routing Problem. Mathematical Programming Series B 94: 343 – 360, 2003.

[7] A. Schrijver. Theory of Linear and Integer Programming. John Wiley & Sons, 1986.

[8] SYMPHONY web page. *http://branchandcut.org/SYMPHONY*

[9] P. Toth and D. Vigo. The Vehicle Routing Problem. SIAM monographs on discrete mathematics and applications, 2002.

SENSITIVITY ANALYSIS OF THE MINIMAL COST SOLUTION BY AND/OR GRAPHS

DUŠAN HVALICA

UNIVERSITY OF LJUBLJANA, FACULTY OF ECONOMICS

ABSTRACT. In scheduling, 'system nervousness' can be addressed by applying AND/OR graphs, specifically, by performing sensitivity analysis of the optimal solution subgraph in the corresponding AND/OR graph. A suitable redefinition of the cost of solution subgraph allows for considering the cost dimension of alternative solutions; we show how sensitivity analysis can be performed in this context.

Keywords: system nervousness, alternative process scheduling, AND/OR graph, sensitivity analysis

1. INTRODUCTION

In the last years, *alternative process scheduling* has been steadily gaining importance, since, among other things, it allows for integration of planning and manufacturing [1, 2]. As opposed to common scheduling, here the choice to be made is not only *when* to perform every activity, but also *which* of available sets of activities is to be used. As shown in [6, 5], this approach can be effectively modelled by AND/OR graphs.

In [6] it is also shown that using AND/OR graphs to model scheduling provides a way to address the issue of '*system nervousness*' (the fact that a relatively small change of the environment can result in substantial change of the system [4, 7], so that a complete schedule regeneration is needed). To be more specific, it is shown that in the case when the minimal total duration of activities is of primary interest, the sensitivity analysis (the determination, for any given optimal solution, of the bounds within which it remains optimal) can be performed and thus help in resolving the question, whether to reschedule or not.

Sometimes, on the other hand, the costs of alternative solutions are the most important. Here we show that AND/OR graphs can still be used to model such situations and that sensitivity analysis can be performed in this case as well.

2. AND/OR GRAPHS

In any directed graph, for any node x, $\Gamma(x)$ will denote the set of child nodes of x and for any arc (x, y) its cost will be denoted by $c(x, y)$. It will be assumed that $c(x, y) \in \mathbb{R}^*$, where $\mathbb{R}^* = \mathbb{R} \cup \{\infty\}$.

An *AND/OR graph* is a directed graph G for which a partition of the set of its nodes $G = O \cup A \cup N_G$ is given, where N_G denotes the set of the tip nodes (the nodes without child nodes); the nodes in O and A are called *OR-nodes* and *AND-nodes*, respectively. AND- and OR-nodes differ by the way in which they are *established* — an OR-node is established when any of its child nodes is established, an AND-node is established when all of its child nodes are established, while a tip-node is established if and only if it belongs to some given set $S \subset N_G$; the nodes in S are called *solved nodes*.

AND/OR graphs were introduced to model problem reduction — the situation

B can be solved by solving B_1 or by solving both B_2 and B_3

can be represented by the AND/OR graph in Figure 1; a node is established when the problem, corresponding to that node, is solved. Nodes in N_G represent problems which either can not be reduced to subproblems or can be solved without reduction (hence the term solved nodes). Node C is an AND-node; it is established when both B_2 and B_3 are. Node B — an OR-node — can be established by establishing any of B_1, C.

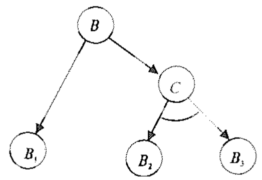


Figure 1

3. Alternative solutions

A project is traditionally described as a set of activities, equipped with the precedence relation

$$A_1 P A_2 \quad \Leftrightarrow \quad A_1 \text{ must finish before } A_2 \text{ can start} \tag{1}$$

It is tacitly understood that every activity starts as soon as possible, which implies that A_2 starts as soon as every A, such that APA_2, is finished.

More flexibility can be gained by introducing yet another binary relation

$$A_1 E A_2 \quad \Leftrightarrow \quad \text{the finish of } A_1 \text{ enables the start of } A_2 \tag{2}$$

thus allowing alternative ways of accomplishing a goal. For instance, if the conditions for the start of activity A_2 can be attained by performing any of A_1, A_3 then we have A_1EA_2 and A_3EA_2.

It will be tacitly understood that when introducing alternatives one is thorough, i.e., when A_2 starts at least one of A, for which AEA_2, is finished. Hence the following will apply:

$$\forall A\ (AEA_2 \Rightarrow BPA) \;\Rightarrow\; BPA_2 \tag{3}$$

It can be shown [6] that under these assumptions a new relation $P' \subset P$ exists for which

$$R_{P'} \cap R_E = \varnothing,$$

i.e., no activity is in the range of both P' and E, but which, by (3), nevertheless implies complete original information about precedence.

Clearly such a project can be represented by an AND/OR graph — activities in the range of P' and E are represented by AND-nodes and OR-nodes, respectively; a node is established when the corresponding activity is finished. For instance, the situation

$$B \text{ is triggered by the finish of } B_1 \text{ or by the finish of both } B_2 \text{ and } B_3 \tag{4}$$

can be represented by the AND/OR graph in Figure 1.

A weighted AND/OR graph is then obtained by setting

$$c(x, y) = d(y)$$

where $d(y)$ denotes the duration of the activity represented by y.

Thus, AND/OR graphs provide a suitable environment for modelling alternative solutions; the areas where such approach would be advantageous, are manufacture-to-order, concurrent engineering, design management, etc.

4. Solution subgraphs and their cost

A *solution subgraph* of $x \in G$ is such a subgraph $G' \subset G$ that

- $x \in G'$ and for every $t \in G'$ there is a path $x, \ldots, t$ in G',
- for every OR-node $t \in G'$ it contains exactly one arc leaving t,
- for every AND-node $t \in G'$ it contains all arcs leaving t,
- every tip node $t \in G'$ is solved,
- it contains no cycles.

Clearly every node of a solution subgraph is established; to establish x it suffices to find any of its solution subgraphs.

When an AND/OR graph represents a project with possible alternative realizations then every solution subgraph corresponds to a feasible realization of the project; the same applies to AND/OR graphs associated with job-shop problems.

In any AND/OR graph the *cost of the minimal solution subgraph* can be defined in several ways.

If it is given by the function $w : G \to \mathbb{R}^*$, satisfying

$$w(u) = \begin{cases} h(u), & u \in N_G \\ \min_{v \in \Gamma(u)} \{w(v) + c(u,v)\}, & u \in O \\ \max_{v \in \Gamma(u)} \{w(v) + c(u,v)\}, & u \in A \end{cases} \tag{5}$$

where

$$h(u) = \begin{cases} 0, & u \in S \\ \infty, & u \in N_G \setminus S \end{cases}$$

then when $w(x,y)$ is equal to the duration of activity y, the cost of the minimal solution subgraph is equal to the minimal time of the completion of the project. If AND/OR graph is associated to a job-shop problem then any minimal solution subgraph (in the sense (5)) corresponds to the solution of the problem [5].

Here, we shall define the cost of a solution subgraph P by

$$w(P) = \sum_{(x,y)\in P} c(x,y) \tag{6}$$

If $c(x,y)$ is equal to the cost of activity y, then the cost of a solution subgraph equals to the total cost of the corresponding realization of the project and any minimal solution subgraph gives the least expensive realization.

It should be noted that for the cost defined by (6) the computational complexity is much greater than for (5) — the problem of finding the minimal solution subgraph with respect to (6) is $\mathcal{NP}$-hard [3]. Of course, this is yet another reason speaking for performing sensitivity analysis, as then scenarios can be made in advance that allow for quick response to sudden changes.

5. Sensitivity analysis

In this section we shall see, how — for any given optimal solution subgraph — to determine for each arc the bounds within which its weight must lie — all other weights being constant — if the optimal solution subgraph is to remain optimal.

Our solution is algorithmic, i.e., we present algorithms that for each arc (x,y) determine the bounds of the interval $[\alpha,\beta]$ such that for $c(x,y) \in [\alpha,\beta]$ the minimal solution subgraph remains unchanged.

We shall use the following notation: $G'_{(x,y)}$ will denote the subgraph of G, spanned by the nodes that lie on any path P in G which starts at t such that

- $(x,y) \in P$ or
- P splits from P' at an AND-node and $(x,y) \in P'$.

Clearly, every solution subgraph of t which contains (x,y) must lie in $G'_{(x,y)}$.

By $G''_{(x,y)}$ we shall denote the subgraph of G obtained by removing (x,y). Furthermore, $S_{\min}$, $S'_{\min}$ and $S''_{\min}$ will denote minimal solution subgraphs of t in G, $G'_{(x,y)}$ and $G''_{(x,y)}$, respectively.

Consider the following procedures:

Procedure $L(x,y)$;
  begin
  $\alpha \leftarrow 0$;
  if $(x,y) \notin S_{\min}$ **then**
    if $S'_{\min}$ exists **then**
      $\alpha \leftarrow \min(0, c(x,y) - (w(S'_{\min}) - w(S_{\min})))$;
  end;
Procedure $U(x,y)$;
  begin
  $\beta \leftarrow \infty$;
  if $(x,y) \in S_{\min}$ **then**
    if $S''_{\min}$ exists **then**
      $\beta \leftarrow c(x,y) + (w(S''_{\min}) - w(S_{\min}))$;
  end;

It turns out that the following applies:

Proposition 1. *If $S_{\min}$ is a minimal solution subgraph of t then for any arc (x,y) procedures L and U determine the lower and upper bound, respectively, of the interval $[\alpha,\beta]$ such that for $c(x,y) \in [\alpha,\beta]$ — all other weights being constant — the minimal solution subgraph remains unchanged.*

Proof. To prove that L determines α correctly, observe that when $(x,y) \in S_{\min}$, lowering $c(x,y)$ will not spoil the minimality of $S_{\min}$, while when $(x,y) \notin S_{\min}$, the subgraph $S_{\min}$ will cease to be minimal only if lowering $c(x,y)$ results in $w(S'_{\min}) < w(S_{\min})$. The correctness of U follows from the facts that when $(x,y) \notin S_{\min}$, raising $c(x,y)$ will not spoil the minimality of $S_{\min}$, while when $(x,y) \in S_{\min}$, the subgraph $S_{\min}$ will cease to be minimal only if raising $c(x,y)$ results in $w(S''_{\min}) < w(S_{\min})$. ■

6. Conclusion

We have shown how to conduct sensitivity analysis when AND/OR graphs are used to model alternative realizations of the project and when the minimization of the cost is in focus. In the description of our algorithms we have made use of an algorithm for searching for a minimal solution subgraph. Unfortunately, we are still short of an effective algorithm for this problem (however, it is not difficult to construct an algorithm with exponential time complexity). As already mentioned, by performing sensitivity analysis in advance this drawback can be, to a certain extent, avoided.

References

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

[2] G. Erdos, T. Kis, P. Xirouchakis, Modelling and evaluating product end-of-life options, Int'l Journal of Production Research 9 (2001), 1203-1220.

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

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

[5] D. Hvalica, L. Bogataj, Sensitivity results considering rescheduling by AND/OR graphs, International Journal of Production Economics 93-94, (2005), 455-464.

[6] D. Hvalica, AND/OR graphs and project management, Journal of Computing and Information Technology (CIT) 5, No 3 (1997), 193 - 197.

[7] O. Tang, R. W. Grubbström, Rescheduling Considerations for Production Planning Using Control Theory, in M. T. Hillery and H. J. Lewis, (Eds.), Proceedings of the Fifteenth Conference of the International Foundation for Production Research, Limerick, Ireland, Vol. 1 (1999) 447-450.

CHINESE POSTMAN PROBLEM WITH PRIORITIES

Tomaž Kramberger[1], Janez Žerovnik[2,3]
[1] University of Maribor, Faculty of Logistics, Mariborska 2, 3000 Celje, Slovenija
[2] University of Maribor, Faculty of Mechanical Engineering, Smetanova 17, 2000 Maribor, Slovenija
[3] IMFM/DTCS, Jadranska 19, 1000 Ljubljana, Slovenija
Emails: tomaz.kramberger@uni-mb.si, janez.zerovnik@imfm.uni-lj.si

Abstract. A generalization of the Chinese Postman Problem is considered, in which a linear order on a set of important nodes is given and the task is to traverse all edges at least once in such a way that the higher priority nodes are visited as soon as possible.

Keywords: graph, eulerian tour, snow plowing, salt gritting

1. Introduction

Let $G(V,E,u)$ be an undirected weighted graph, where V is the set of nodes, E is the set of edges and $u: E \rightarrow R$ is a weight function assigning a positive cost of traversing edges. The well-known Chinese Postman problem is to find the shortest postman tour traversing each edge of a graph at least once. Several real-world problems, such as street sweeping, mail delivery, solid waste collection, salt gritting and snow plowing can be modelled as Chinese postman problems with some additional constraints.

This article deals with optimal organization of scattering icy roads. To be carried out properly, we should take into consideration both security and economical effects. Regarding security, the most exposed and first icy road net spots have priority. From the economical point of view, all these roads have to be scattered one after another using the cheapest route. Here we assume that the shortest route is also the cheapest.

Winter ploughing and scattering of the road is of most importance and of great expense. If the roads are not ploughed or slippery roads are not scattered, participants in traffic are exposed to great danger. Weather conditions often cause traffic jams and have negative economic effect, at least causing great dissatisfaction in people. In several papers (see, for example article [2], [8], [7] or [6]) this problem is considered using different formulations. Here we give a formal definition that is not equivalent to those in [2] and [8]. While it is natural to model the problem with a variant of the Chinese postman problem, the most appropriate definition of the cost function is not obvious. Here we define a multi objective cost function where the most important goal is to minimize the total cost (time, mileage), and second, among all Chinese postman solutions we are looking for walks that will visit the priority nodes as soon as possible.

2. Problem definition

The problem considered here is formally defined as follows:

Given an undirected weighted graph $G(V,E,u)$, and a (short) sequence of priority nodes v_1, v_2, ..., v_k, the objective is to find a walk, which is

(1) the shortest walk, which traverses each edge at least once, and

(2) among the solutions which satisfy (1), find the walk that visits v_1 as soon as possible

(3) among the solutions which satisfy (2), find the walk that visits v_2 as soon as possible

....

(k) among the solutions which satisfy (k-1), find the walk that visits v_k as soon as possible

3. The algorithm

We propose a solution which is a combination of the well-known algorithms for minimum matching, Fleury's algorithm for constructing an Eulerian walk and the Dijkstra's algorithm for computing shortest paths (see, for example [1, p.269-271] or [5, p.225-227]). In short, the algorithm proceeds as follows. First, as for the basic Chinese postman problem, the set of odd nodes is identified and shortest paths between pair of nodes are added to the original graph, introducing double edges if necessary (see, for example [9, p.132], or any textbook in combinatorial optimization [3, 4, 5]). The new graph will be denoted by G_0. Any polynomial algorithm for minimum matching of an auxiliary graph can be used. In the second phase, an eulerian walk on the new graph is constructed. First, a walk P_1 from depot to the first priority node v_1 is chosen according to the following rule: it is one of the walks from v_0 to v_1 which are shortest under condition that G_0 without P_1 is connected. Denote $G_1 = G_0 - P_1$. Then, a walk P_2 is chosen such that it is a shortest walk from v_1 to v_2 such that $G_2 = G_1 - P_2$ remains connected. If v_2 has already been visited, then P_2 is empty. In this way, all priority nodes are visited and finally, a walk back to v_0 is constructed which covers all the edges in G_k.

Using the construction, it is easy to prove

Theorem. The algorithm is optimal.

4. Example

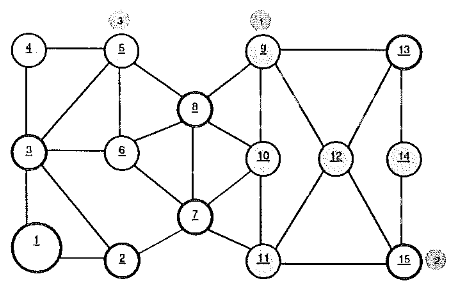


Fig. 1: Graph with six nodes of odd degree

Let us consider node v_1. Vehicle should drive all the roads and scatter them in the fastest possible way visiting priority spots/arcs within the shortest time. The priority nodes are, in this order: crossing v_9, crossing v_{15} and crossing v_5.

According to the idea of the algorithm presented in Section 2, we have to construct an Eulerian graph (i.e. a graph with all nodes of even degree). To the graph with odd degree nodes we add new connections, so that all odd degree nodes will have even degree. We choose all the odd degree nodes and form an auxiliary graph, which is a complete graph and the edge weights correspond to the distances in the original graph. In our example, we have six odd degree nodes and the edge weights (or, distance) matrix:

	v_2	v_3	v_7	v_8	v_{13}	v_{15}
v_2		1	1	2	4	3
v_3	1		2	2	4	4
v_7	1	2		1	3	2
v_8	2	2	1		2	3
v_{13}	4	4	3	2		2
v_{15}	3	4	2	3	2	

Table1: Matrix of distances between nodes of odd degree.

A minimum matching has to be found, for example using the algorithm for minimum perfect matching with time complexity $O(n^4)$ (see, for example [3, p.236] or [10]). In our case, the optimal solution is $v_2 \rightarrow v_3$, $v_7 \rightarrow v_8$, $v_{13} \rightarrow v_{15}$ which can be found quickly even by inspecting all possible cases. Thus, we get a new graph G_0, in which all nodes have even degree:

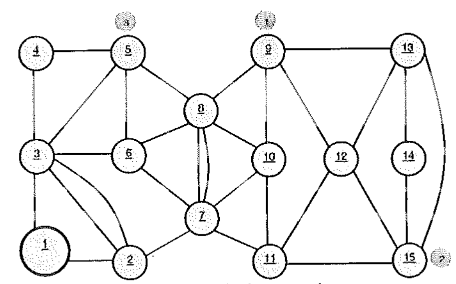


Fig. 2: Eulerian graph.

Note that the new edge $v_{13} \rightarrow v_{15}$ is in fact a shortest path $v_{13} \rightarrow v_{14} \rightarrow v_{15}$.

Once we have the eulerian graph, we can find the eulerian walk with Fleury's algorithm. Recall that in Fleury's algorithm we have a lot of freedom for the construction of the walk as long as we do not cross bridges (see, for example [9, p.126]). In other words, we must not disconnect the remaining graph. In our case we have priority nodes, therefore want to visit the priority nodes as soon as possible. However, not any shortest path can be taken. We have to find the path that is the shortest among the paths such that $G_1 = G_0 - P_1$ is a connected graph. To find the shortest paths, we can apply Dijkstra's algorithm to calculate the value of the node distances d_i from node v_1 to the node v_i. Dijkstra's algorithm runs in time $O(n^2)$. In our example, we get value of the distances d_i and other nodes, as seen in Fig. 3.

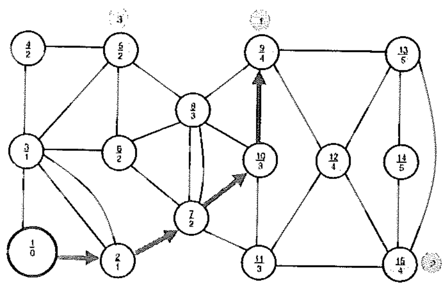


Fig. 3: Shortest path from v_1 to v_9

To find the shortest path from v_1 to v_9, we start from the node v_9 and construct the shortest path backwards. Let d_i be the value of the distance from the node to the v_1, u_{ij} and distance between node v_9 and observed neighbouring nodes. Clearly, neighbours with value $d_i = d_9 - u_{ij}$ are on the shortest paths from v_1 to v_9. When there are more nodes each is labelled and for each one the same procedure is repeated. In our case nodes v_8 and v_{10} should be considered because $d_8 = 5-2 = 3$ and $d_{10} = 5-2 = 3$. We move to node v_{10}, and repeat the whole procedure. From this node we observe only one node, v_7, for which it holds $d_7 = 3-1 = 2$. Procedure should be repeated for nodes v_2 and v_1. In this way we have found a shortest path P_1 between v_1 and v_9. As $G_0 - P_1$ is connected, we set $G_1 = G_0 - P_1$ and continue.

As the next priority node has not been visited yet, we are looking for a shortest path from v_9 to v_{15} which satisfies the connectivity condition.

After repeating above procedure we get the shortest path from v_9 to v_{15} as in Fig. 4.

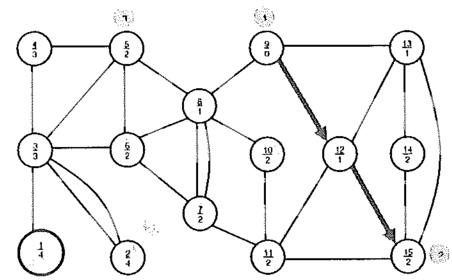


Fig. 4: Shortest path from v_9 to v_{15}.

Finally, we repeat the procedure from the last priority node v_{15}. We find the shortest path between nodes v_{15} and v_5. The resulting graph is given in Fig. 5.

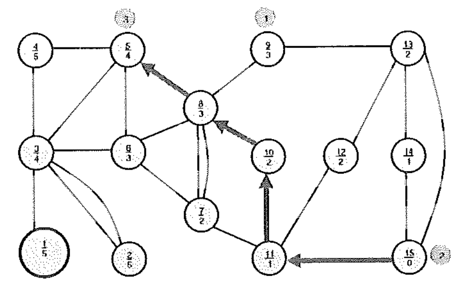


Fig. 5: Shortest path from v_{15} to v_5.

After we have visited all priority nodes, the graph G_3 is connected with exactly two nodes of odd degree, v_5 and v_1. Fleury's algorithm can be used to trace all the edges starting from v_5 and ending in v_1.
Summarizing, we have constructed the walk as showed in Fig. 6.

5. Conclusion

As illustrated by the example, the algorithm given in Section 3 provides a solution to the Chinese postman problem which has additional property that it visits the first priority node as soon as possible (similarly, it also visits the other priority nodes as soon as possible subject to the previous conditions). For the cost function defined in Section 2, the algorithm provides an optimal solution. The proof is straightforward, but is omitted here due to page limit.

However, it is perhaps more realistic not to insist on traversing the minimal length (i.e. the Chinese postman tour), but rather to assign some cost to the edges traversed. On the other hand, a cost can be assigned to any delay in visiting the priority nodes. For various ratios between the costs assigned, we would in general get different optimal solutions to the optimization problem.

For example. the solutions found here are valid for the case, where the cost of additional edges traversed is much bigger that the cost of delays and delay for the first priority node is much bigger than the cost of delay for the second node, and so on. More detailed study of various cost functions may be an interesting topic for future research.

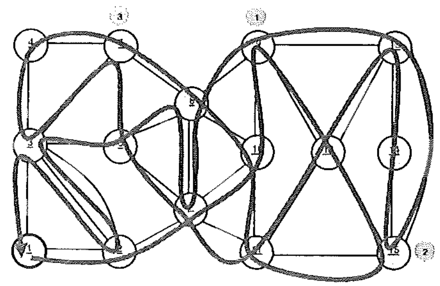


Fig. 6: Eulerian walk.

References

1) E. W. Dijkstra, A note on two problems in connection with graphs. Numerische Mathematik 1 (1959) 269-271.

2) G. Ghiani, G.Improta, An algorithm for the Hierarchical Chinese Postman Problem, Operations Research Letters 26 (2000) 27-32.

3) B. Korte, J. Vygen, Combinatorial optimization: Theory and algorithms, Springer, Berlin, Heidelberg, New York 2002.

4) C. H. Papadimitrou, K. Steiglitz, Combinatorial Optimization, Algorithms and Complexity, Dover publications inc., New York 1998.

5) S. Skiena, Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica, Addison-Wesley, Redwood City, CA, 1990, 225-227.

6) Shao, J., Lister P.J., Hart G.D., Pearson H.B., Thermal mapping: reliability and repeatability, Meteorol. Appl. 4 (1997) 131-137.

7) Thornes, J.E., Thermal Mapping and road- weather information systems for highway engineers, In Highway Meteorology, ed. By A. H. Perry and L. J. Symons, (1991) 39-67.

8) H.-F. Wang, Y.-P. Wen, Time-Constrained Chinese Postman Problems, Computers and Mathematics with Applications 44 (2002) 375-387.

9) Wilson, R.J., Watkins, J.J., Graphs, An introductionary approach, John Wiley, New York, 1990. (Slovene translation: Uvod v teorijo grafov, DMFA Ljubljana 1997.)

10) P. E. Black, "Munkres' assignment algorithm", from Dictionary of Algorithms and Data Structures, ed., NIST. http://www.nist.gov/dads/HTML/munkresAssignment.html

AN IMPROVED EVOLUTIONARY ALGORITHM FOR SOLVING THE VEHICLE ROUTING PROBLEM

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

Abstract. An improved evolutionary algorithm for solving the capacitated vehicle routing problem is presented. The algorithm employs repeated mutations in a manner similar to local search. Experiments are described, where the algorithm has been evaluated on a well known family of benchmark problem instances.

Keywords: capacitated vehicle routing problem, evolutionary algorithms, experimental evaluation, repeated mutations, local search.

1. Introduction

The *vehicle routing problem* (VRP) is an interesting combinatorial optimization task, which deals with scheduling a fleet of vehicles to distribute goods between depots and customers [11]. The basic version of the VRP is called the *capacitated* VRP (CVRP). It may be described by using a complete directed graph, whose vertices correspond to the customers, and one additional vertex corresponds to the depot. A nonnegative cost is assigned to each arc, and it represents the travel cost spent to move between the incident vertices. Each customer vertex is associated with a nonnegative demand to be delivered. A set of identical vehicles with equal capacities are available at the depot. The CVRP consists of finding a collection of elementary cycles in the graph with minimum total cost of the involved arcs, such that each cycle visits the depot vertex, each customer vertex is visited by exactly one cycle, and the sum of the demands of the vertices visited by a cycle does not extend the vehicle capacity. Obviously, the solution to a CVRP instance specifies an optimal schedule for the vehicles delivering goods from the depot to the customers, so that the demand of any customer is satisfied and no vehicle is overloaded. Each cycle in the solution corresponds to a vehicle route.

Evolutionary algorithms (EAs) are a popular metaheuristic which tries to solve optimization problems by mimicking natural phenomena, such as reproduction, mutation, survival of the fittest, etc [7]. An EA maintains a population of *chromosomes* where each of them encodes a potential solution to a particular problem instance. The population is iteratively changed, thus giving a series of population versions which are called generations. During one generation lifetime, new chromosomes are created and existing ones are altered by means of randomized "genetic" operators called *crossovers* and *mutations*, respectively. A new generation is formed by choosing more fit chromosomes from the previous generation, where the fitness measure is related to the objective function of the original optimization problem. It is expected that the best chromosome in the last generation represents a near-optimum solution.

The VRP is known to be an NP-hard problem [8]. Thus it makes sense to consider applications of metaheuristics, such as EAs, to the VRP. In recent years, there have been many attempts to solve the VRP by EAs. The obtained performance results are reported in [1,9]. General impression is that a pure evolutionary approach is not yet competitive on the VRP compared to other metaheuristics, particularly tabu search [11]. It seems that the presently used chromosomes and genetic operators are not able to capture the full essence of the problem itself. Therefore, many authors have proposed hybrid algorithms [2,3], where the performance of an EA has been improved by replacing its mutation operator by a traditional *local-search* procedure [8].

The aim of this paper is to present and evaluate yet another EA for solving the CVRP. The same algorithm has already been introduced in our previous paper [10], but now it has been improved and tested on a larger set of problem instances. Our approach is purely evolutionary in the sense that only "genetic" operators are used for producing or altering chromosomes. The novelty within our approach is something we call *repeated mutations*. Namely, genetic mutation operators are evaluated many times in order to produce similar effects as local search in hybrid algorithms.

The paper is organized as follows. Section 2 reviews some building blocks of EAs for solving the CVRP, which have been used by other authors. Section 3 explains how those building blocks have been rearranged in a novel way to form our algorithm. Section 4 reports on experiments where the algorithm has been implemented and tested on a well known library of benchmark problem instances. The final Section 5 gives a conclusion.

2. Chromosomes, crossovers and mutations

In our algorithm, a chromosome is built as proposed in [6], thus it is a list of integers representing a *permutation* of customer vertices. This permutation is interpreted as a large elementary cycle, which is obtained from a CVRP-instance solution by concatenating the vehicle routes and by omitting visits to the depot. Note that the same chromosome can in fact represent many different solutions. Still, we use a unique decoding, which is based on the greedy approach. Thus it is assumed that the first vehicle visits as many customers from the initial part of the chromosome as it is possible according to the vehicle capacity, the second vehicle serves as many customers as possible from the following part of the chromosome, etc. For instance, let the vehicle capacity be 20, and suppose that we have 9 customers whose demands are in turn:

$$2,\ 4,\ 7,\ 5,\ 3,\ 5,\ 8,\ 6,\ 1.$$

Then the chromosome

$$p = (2\,5\,8\,9\,3\,7\,6\,1\,4)$$

is decoded by the greedy approach to the following three vehicle routes:

$$(2\,5\,8\,9)\ (3\,7\,6)\ (1\,4).$$

It is assumed, but not explicitly written, that each route starts and ends in the depot.

In our algorithm, we use the crossover operator called *order-crossover* (OX) proposed in [9]. The operator creates a new chromosome (child) by combining parts from two existing chromosomes (parents). First, two cut points are randomly selected, and the chromosome part located between those cut points on the first parent is assigned to the child. The remaining positions are then filled one at a time, starting after the second cut point, by considering the customer vertices in order found on the second

chromosome (wrapping around when the end of the list is reached). For instance, let the two parents and the two cut points be as follows:

$$p_1 = (1\,2\,3\,|\,5\,4\,6\,7\,|\,8\,9),\quad p_2 = (4\,5\,2\,|\,1\,8\,7\,6\,|\,9\,3).$$

Then we can use the parents in both orders, thus obtaining the following two children:

$$c_1 = (2\,1\,8\,|\,5\,4\,6\,7\,|\,9\,3),\quad c_2 = (3\,5\,4\,|\,1\,8\,7\,6\,|\,9\,2).$$

In our algorithm, we use three different mutation operators, called *remove-and-reinsert* (RARM), *swap* (SM) and *invert* (IM), as proposed in [6], [9] and [5], respectively. Generally, a mutation makes a small random change in a single chromosome, thus creating its new version (mutant). Our operators start in the same way by randomly choosing two positions within the chromosome, and then they proceed in different ways. Namely, RARM removes the vertex (customer) from the first position and reinserts it to the second position, SM swaps the vertices at the two positions, while IM inverts the sequence of vertices between the two positions. For instance, let the starting chromosome be

$$p = (1\,2\,3\,4\,5\,6\,7\,8\,9).$$

Suppose that the randomly chosen positions are 3 and 7. Then the three operators produce in turn the following three mutants:

$$\bar{p}_{\text{RARM}} = (1\,2\,4\,5\,6\,7\,3\,8\,9),\quad \bar{p}_{\text{SM}} = (1\,2\,7\,4\,5\,6\,3\,8\,9),\quad \bar{p}_{\text{IM}} = (1\,2\,7\,6\,5\,4\,3\,8\,9).$$

Each of the three mutation operators is further on used in two variants called *global* and *local*, respectively, thus making altogether six variants denoted by RARMG, RARML, SMG, SML, IMG and IML. In the global variant, all positions within the chromosome are considered. In the local variant, both the old and the new positions of the moved customers should belong to the route of the same vehicle.

3. Algorithm description

The overall structure of our evolutionary algorithm follows the outline from Section 1. The initial population is created by producing random permutations according to the uniform probability distribution. A series of generations is produced by using the crossover operator OX and six mutation variants RARMG, RARML, SMG, SML, IMG, IML. Evaluation of chromosomes is done directly by using the objective function of the VRP, i.e. the fitness of a chromosome is equal to the total transportation cost of the corresponding solution. The algorithm stops when no further improvement is possible or when a prescribed time limit is reached.

As already mentioned, our algorithm uses mutations in a special way called *repeated mutations*. It means that mutation operators are never applied directly or separately. Instead, they are grouped within special procedures. Such a procedure considers some (or all) possible mutations of a given chromosome, and finally applies only the best one among considered, i.e. the one that maximally increases the fitness of that chromosome. A separate procedure is built for each of the six variants of mutations.

If the number of considered mutations within a repeated-mutations procedure is small, then those mutations are chosen randomly, and the whole procedure is still randomized. However, in the extreme case when all possible mutations are considered, the procedure becomes deterministic and very similar to a local search procedure.

Now follows a detailed description of the way how our algorithm transforms one generation of chromosomes into the next generation.

- Two "good" chromosomes are chosen from the current population, by using the so-called *tournament selection* [7]. The chosen chromosomes serve as parents for crossover. A prescribed number of children is generated from the same parents with different randomly generated cut points in the OX operator. Only the most fit child is retained, while the others are discarded.
- The only remaining child is improved by repeated mutations. First, the child is improved by applying for instance the RARMG procedure; then the improved child (mutant) is further improved by the same procedure; then the further improved child (mutant's mutant) is even further improved again by the same procedure, etc. When there is no further improvement, we switch to the next procedure, e.g. RARML, ... and so on until all procedures are tried.
- The final improved child is inserted into the current population so that it replaces a "similar" chromosome. Two chromosomes being similar means that relative difference of their fitness measures is below a prescribed threshold. If there is no similar chromosome, then the improved child replaces a "bad" chromosome chosen by a form of tournament selection with *elitism* [7].

As we see, two consecutive generations differ in only one chromosome.

4. Experiments

To enable experimenting, we have developed a C++ implementation of our algorithm. The implementation consists of three C++ classes, whose objects correspond to CVRP instances, chromosomes and populations, respectively. Various components of the algorithm have been realized as methods of those classes, for instance crossovers and mutations are methods of the chromosome class, while tournament selections are methods of the population class. Repeated-mutation procedures have been implemented in a flexible way, so that they can consider either all possible mutations of a given chromosome, or only a specified number of randomly chosen mutations.

By a series of preliminary tests, we have first determined acceptable values for some of the free parameters in our algorithm. At the same time, we have designed an additional procedure that dynamically adjusts the remaining parameters depending on the algorithm performance. Thus the population size has been fixed to 30. Similarly, the percent of population to be considered by tournament selection of good and bad chromosomes has been set to 70% and 15%, respectively. Next, the execution time limit has been chosen as 10 minutes, 45 minutes or 3 hours, depending on the problem instance size. On the other hand, the number of mutations considered by a repeated-mutations procedure, the number of children of the same parents and the similarity threshold have been left over for dynamic adjustment.

The implemented algorithm has been experimentally evaluated on seven benchmark CVRP instances from the so-called Christofides-Mingozzi-Toth library. The whole library is available at the on-line repository [4]. Table 1 gives some basic properties

CVRP instance	Number of customers	Number of vehicles	Cost of the best known solution	Costs of our solutions:			
				Avg	Min	Max	StdDev
CMT01	50	5	521	521.7	521	528	2.2
CMT02	75	10	830	848.0	830	859	9.5
CMT03	100	8	815	841.7	817	860	13.5
CMT04	150	12	1015	1124.6	1028	1198	49.2
CMT05	199	16	1289	1390.7	1349	1422	20.5
CMT11	120	9	1034	1140.7	1034	1181	50.0
CMT12	100	10	820	869.4	838	893	16.0

Table 1: Summary of experimental results.

of the considered test examples, including the costs of their optimal solutions or best solutions known so far. All seven instances in fact belong to a more restricted problem called the *Euclidean symmetric* CVRP, where the vertices are associated with points in the plane, and the cost of an arc is defined as the Euclidean distance between the two involved points.

The results of experiments are summarized again in Table 1. Each row presents our solutions for one particular problem instance. The presented values can easily be compared with the results obtained by other authors with other metaheuristics. Since our algorithm relies on random numbers, different runs with the same data and parameters can produce different solutions. Table 1 therefore presents statistical values computed over exactly 10 runs.

As we can see from Table 1, our algorithm solves smaller problem instances to optimality. The obtained best solutions to medium test examples are up to 2% worse than the best from literature. The algorithm is slightly less successful in solving larger instances, where the relative errors range from 2% to 5%. Our best result for CMT03, being only 0.25% above optimum, is visualized in Figure 1, where the customer coordinates and the vehicle routes are shown.

During experimenting, we have also gained some rough impressions about computing times required by our algorithm. On a PC with a 2.4 GHz Pentium processor, the solution is usually reached in 5-10 minutes. Smaller problem instances are often solved faster, i.e. in less than 1 minute. For larger instances, it can happen that the algorithm converges only after 1-2 hours.

5. Conclusion

The presented results clearly indicate that our evolutionary algorithm is competitive compared to the other metaheuristics, and that our approach based on repeated mutations can assure similar performance results as the approach based on local search. Indeed, on the chosen set of benchmark problem instances, we have obtained nearly-optimal solutions at tolerable computational costs.

Our future plan is to further improve the algorithm by developing a more sophisticated procedure for dynamic adjustment of parameters. Also, we plan to try more appropriate chromosomes or better ways to decode a solution from a chromosome.

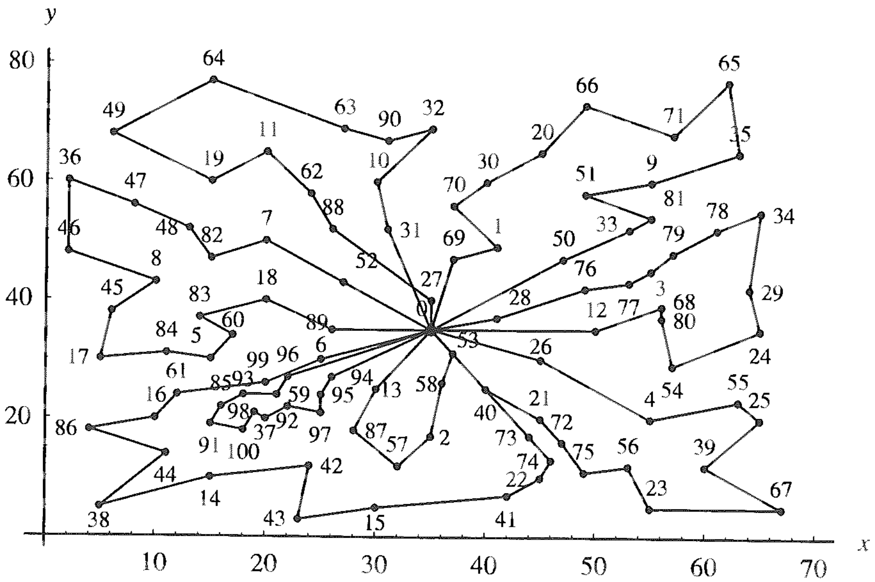


Figure 1: Our best solution to CMT03 (cost: 817).

References

1. Baker B.M., Ayechew M.A., "A genetic algorithm for the vehicle routing problem", *Computers and Operations Research*, Vol 30 (2003), 787-800.
2. Beullens P. et al., "A guided local search heuristic for the capacitated arc routing problem", *European Journal of Operational Research*, Vol 147 (2003), 629-643.
3. Braysy O. et al., "A multi-start local search algorithm for the vehicle routing problem with time windows", *European Journal of Operational Research*, Vol 159 (2004), 586-605.
4. Diaz B.D., *The VRP Web*, Languages and Computation Sciences Department, University of Malaga, 2004, `http://neo.lcc.uma.es/radi-aeb/WebVRP`
5. Hwang H-S., "An improved model for vehicle routing problem with time constraint based on genetic algorithm", *Computers and Industrial Engineering*, Vol 42 (2002), 361-369.
6. Larranga P. et al., "Genetic algorithms for the travelling salesman problem: a review of representations and operators", *Artificial Intelligence Review*, Vol 13 (1999), 129-170.
7. Michalewicz Z., *Genetic Algorithms + Data Structures = Evolution Programs*, Third Edition, Springer, New York, 1995.
8. Papadimitriou C.H., Steiglitz K., *Combinatorial Optimization - Algorithms and Complexity*, Dover, Mineola New York, 1998.
9. Prins C., "A simple and effective evolutionary algorithm for the vehicle routing problem", *Computers and Operations Research*, Vol 31 (2004), 1985-2002.
10. Puljić K., Manger R., "An evolutionary algorithm with repeated mutations for solving the vehicle routing problem", in: Lužar-Stiffler, V., Hljuz Dobrić, V. (editors), *Proceedings of the 27th International Conference on Information Technology Interfaces - ITI 2005, Cavtat, Croatia, June 20-23, 2005*, University Computing Centre, Zagreb, 2005.
11. Toth P., Vigo D. (editors), *The Vehicle Routing Problem*, SIAM Monographs on Discrete Mathematics and Applications, SIAM, Philadelphia, 2002.

The 8th International Symposium on Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 9
Production and Inventory

A PC-BASED DECISION SUPPORT SYSTEM FOR TAILORING OF LOGS IN VENEERS PRODUCTION

Anton Čižman and Marko Urh
University of Maribor, Faculty of organizational sciences
Kranj, Kidričeva cesta 55a
E-mail: anton.cizman@fov.uni-mb.si, marko.urh@fov.uni-mb.si

Abstract
We report on decision support system (DSS) that recommends solutions of specific cutting stock problems from the production of veneers. A prototype DSS is developed for use in wood-processing enterprises and is designed to aid them in creating or improving existing tailoring of logs using commercially available Operations Research software. The results of testing a typical cutting-stock problem are shown to point out how such user-friendly DSS, which utilizes linear and mixed-integer programming, can reduce inventory costs and improves the exploitation of input material.

1. Introduction

Decision Support Systems (DSS) have emerged as the computer-based system to assist decision makers address semi structured problems by allowing them to access and use data and analytic models. Such systems have the following characteristics: they are interactive computer-based systems; they are aimed at semi-structured problems. They utilize models with internal and external databases, and they emphasize flexibility, effectiveness, and adaptability. These characteristics have guided much of the research in the DSS area, but the potential benefits of DSS in the business environment have not been fully realized.

Although the definition of the DSS concept has been elusive [1, 2, 3], the field has flourished with the development of computer technology. PC technologies are becoming accepted and incorporated into organizations and our personal lives. PC-based systems have the potential to improve both individual and organizational performance. As decision makers recognize the potential benefits, many companies are investing in information technology. PC-based systems have been generally hailed as a revolution that will change the nature of professional work and transform the way people work. It is expected that almost all knowledgeable workers are likely to have their own PC to perform both stand-alone tasks and network services.

Despite the proliferation of microprocessor-based systems, the potential benefits of these systems, as aids to managerial decision making, may not be fully realized due to poor design and low acceptance by users. It is recognized that individuals are sometimes unwilling to use these systems, even if the system may increase their productivity. While some of these systems may have an impact on individuals and organizations, the adoption and acceptance of these systems among decision-makers has been limited. This may be due to the inflexibility in the systems as well as their narrow design. Therefore, it is important to understand the environment of the decision makers and the type of support they need in order to make effective decisions, and to examine the models appropriate for addressing their problems [1, 5].

In this paper we describe a PC-based DSS which addresses the optimal cutting of logs in the manufacturing of veneers. The development of DSS is proposed on pragmatic approach [2, 3] that leads to more efficient and simpler use of such systems. The effectiveness of the DSS is shown by relatively simple example which utilizes the pattern-oriented LP-based methods for solving one-dimensional cutting stock problem. The results of testing show that the application of such managerial DSS enhances problem solving capability for achieving greater competitiveness of an organization.

In the first part of the paper, we shortly present a description of major operations management problems in the production of veneers focusing on the issue of efficient log-cutting. The second part of the paper describes the formulation and development of a DSS for this situation. Finally, we illustrate how the DSS can be used to improve existing decision making process using scientific approach.

2. Analyzing the decision-making process

Primary analysis of the decision-making process applied to the management of the veneers production operations described in the paper [3] reveals a number of decision-making situations. One of the relevant problems in this regard, such as determination the optimal log-cutting strategy and the quantity of logs to be purchased, is related to optimal exploitation of input material (logs). This situation is shown in Fig. 1.

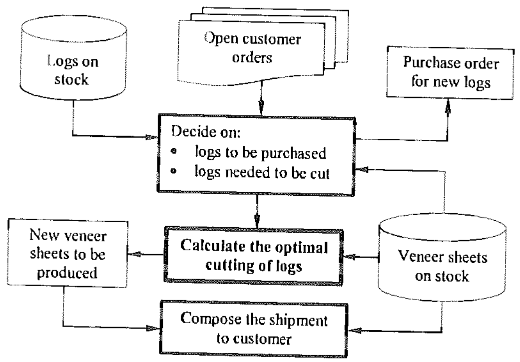


Fig. 1. The decision-making process in the production of veneers

The discussion is focused to the indicated initial decision-making problem from Fig. 1, i.e. the optimal cutting of logs. The current decision-making model in the production of veneers is based on obtainable data samples about logs and related veneer sheets that have been produced in several periods. Up to the present, the **longitudinal tailoring** was performed by an intuitive "rule of thumb". Thereby, the goal of the operations manager is to establish a "reasonable minimum" of the edge-waste remainders of logs which, unfortunately, does not take into account the end-waste remainders of logs. This simple approach to log cutting causes gradual accumulation of an undesired inventory of veneer sheets in the warehouse which significantly increases inventory costs. Thus the idea is to improve the existing decision-making process using scientific approach to minimize at the same time both, the edge waste as well as the end waste from cutting the logs.

3. The problem-solving paradigm

The problem of optimal cutting of logs in veneer production is called *one-dimensional cutting stock problem* (1d-CSP). This problem consists of determining the smallest number of logs of length L_p, ($p = 1, 2, \ldots, s$) that are available in sufficiently large quantity and have to be cut in order to satisfy the required number b_i sub-logs of lengths l_i, ($i = 1, 2, \ldots, m$). The lengths of sub-logs l_i are determined by the required length of ordered veneer sheets. On the other hand, the required number of sub-logs (b_i,) can be calculated from the required area and thickness of ordered veneer sheets taking into account the diameter and quality of logs to

be cut. A combination of required sub-logs lengths in the log of length L_p is called cutting pattern. For each log length L_p, the number n_p ($p = 1, 2, \ldots, s$) of all different cutting-patterns can be determined with regard to the particular sub-log length l_i. The element a_{ijp} of matrix A that describes each cutting pattern, represents the number of sub-logs of length l_i obtained in cutting pattern j related to log length L_p. The number of logs of length L_p to be cut according to cutting pattern j is denoted by a decision variable x_{jp}.

In this problem situation, the objective is to minimize the number of logs (x_{jp}) of length L_p, that have to be cut according to the j-th pattern. Thus, our 1d-CSP represents a pure integer programming (IP) problem that can be modeled as follows:

$$\min \sum_{p=1}^{s} \sum_{j=1}^{n_p} L_p x_{jp} \qquad (1)$$

subject to

$$\sum_{p=1}^{s} \sum_{j=1}^{n_p} a_{ijp} x_{jp} \geq b_i, \quad i = 1, \ldots, m \qquad (2)$$

where

$$x_{jp} \geq 0 \text{ and integer}, \; j = 1, \ldots n_p; \; p = 1, \ldots, s$$

For cutting pattern to be valid:

$$\sum_{i=1}^{m} a_{ijp} l_i \leq L_p \qquad (3)$$

$$a_{ijp} \geq 0 \text{ and integer}, \quad j = 1, \ldots n_p; \; p = 1, \ldots, s$$

The IP problem described above can be solved using different LP-based methods [3, 4]. In practice, most IP problems can be solved by the technique of Branch-and-Bound (BB) or by using the LP-based Gilmore and Gomory's "delayed column generation" [3, 6].

4. The system structure

Based on work done with a case study from a Slovenian veneers factory, a simple prototype DSS was developed for solving the specific cutting stock problem. The purpose of the study was to check whether a simple-to-use managerial decision support system can be developed efficiently for users lacking detailed Operations Research (OR) knowledge, starting from available commercial Mathematical Programming (MP) software. The DSS is divided into three components: the user-interface, the modeling base and the database. The integration of the three components of the DSS, i.e., the structure of our DSS, is presented in Fig. 2.

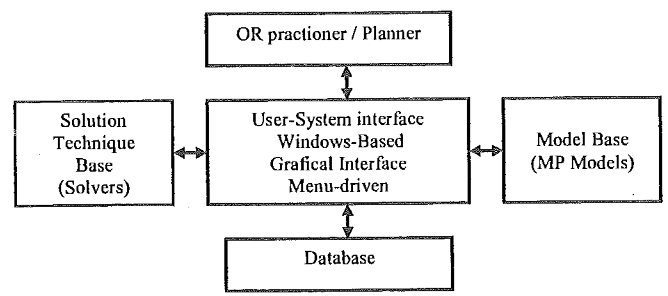


Fig. 2: The structure of a decision support system

For the purpose of selection locally appropriate commercial MP software, the process of setting up requirements for a proper SW tool was considered from the broad viewpoint of the enterprise, the viewpoint of the OR expert and the pragmatic viewpoint of the user. If compared on the basis of given requirements, a couple of MP software tools that are available in Slovenia today have favorable features as candidates to be selected for further development of a manager-friendly DSS. They appear to have an almost equal rank from the viewpoint of the OR expert. Therefore, the choice of the basic DSS development software may depend largely on the particular needs and circumstances of the enterprise where the new DSS will be used.

A general argument for SAS/OR software selection was the fact that the SAS System offers to an user a wider selection of software tools and technology, not only for MP [7, 8, 9]. Another important argument for having selected the SAS System is the possibility of effective exploitation of modern technologies for effective development and implementation of end-user oriented DSS such as: Data Warehousing and On-Line Analytical Processing (OLAP), object-oriented programming concepts (the SAS/AF software) with powerful Screen Control Language (SCL), and interfaces that provide easily and timely access to data from SAS-format and outside databases (the SAS/ACCESS software).

5. DSS design, development and testing

The goal was to build an user-friendly screen-based DSS to be used by the operations manager for cutting of logs in the production of veneers, such that it does not require any special MP knowledge for operating, can be used through simple point-and-click commands, has simple data-input and data-output procedures, can be operated from a standard personal computer running the Microsoft Windows operating system, exploits as much as possible the features and functions of the selected SAS/OR software and is flexible and easy to modify as well as to maintain.

The standard SAS/OR graphical user interface provides a specific programming language that was not considered appropriate for an operations manger. Therefore, further development was focused around building a special window-based graphical user interface (GUI) allowing easier use of typical MP functions. The graphical user interface was developed specifically for this application by means of the SAS/AF software. With this software, the developer was able quickly to prototype the basic functions and to verify the functionality of implemented functions.

5.1. Test example

Let suppose that a customer orders three different quantities of veneers of equal thickness and of three different sheets lengths: $l_1 = 182$ cm and total area A_1 m^2; $l_2 = 151$ cm and total area A_2 m^2; $l_3 = 112$ cm and total area A_3 m^2. In this situation it is calculated that the 350 sub-logs (b_1) of length l_1, 250 sub-logs (b_2) of length l_2 and 400 sub-logs (b_3) of length l_3 are needed if the original logs of tree type 3 with length of 446 cm is used. All possible cutting patterns that can be created from selected log of length 446 cm are determined. Results of this step are given in Table 1.

The operations manager has to input data into the DSS, according to the Table 1. Input data relate to the computation objective (min or max), to the number of constraints and decision variables and to the file name (existing or new). Fig. 3 and the upper-left/central part of Fig. 4 illustrate how the test data can be fed into the new DSS by means of the two data-input GUI screens.

Table 1: Possible cutting patterns for a log of length $L_l = 446$ cm, $j = 1, 2, \ldots, 12, p = 1$

Cutting patterns (j)	182-cm sublogs	151-cm sublogs	112-cm sublogs	Edge waste (cm)
Pat. 1 - x1	1	0	0	264
Pat. 2 - x2	2	0	0	82
Pat. 3 - x3	1	1	0	113
Pat. 4 - x4	1	1	1	1
Pat. 5 - x5	1	0	1	152
Pat. 6 - x6	1	0	2	40
Pat. 7 - x7	0	1	0	295
Pat. 8 - x8	0	2	0	144
Pat. 9 - x9	0	2	1	32
Pat. 10 - x10	0	0	1	334
Pat. 11 - x11	0	0	2	222
Pat. 12 - x12	0	0	3	110
Req.nb. of sublogs	350	250	400	

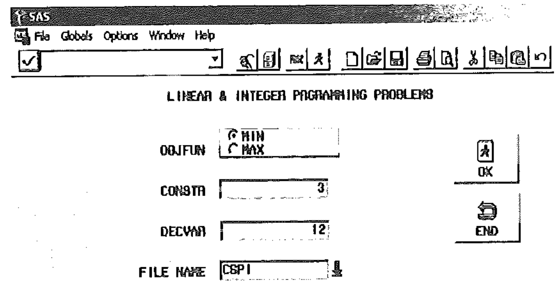


Fig. 3. Basic data-definition screen

After performing data input the presented test case brings the operations manager to the results of the SAS "PROC LP"/IP optimization procedure. To get them, he has to click on the "SOLVE" button shown in Fig. 4. The results of the optimal solution appear in the left-bottom part of the same (main data input/output) screen. In this solution, the values of decision variables (X1, X2, ..., X12) practically represents the number of 446-cm logs that have to be cut according to twelve ($n_l = 12$) cutting patterns given in Tab. 1. In this case, the minimally required number of original logs to fulfill this customer order is **338**. Optional sensitivity analysis is available by clicking either to the »RANGE/PRICE« or to the »RANGE/RHS« button. Recalling Fig. 4, the user can combine input data from other databases by clicking the ACCESS button.

The results of the optimal solution appear in the left-bottom part of the same (main data input/output) screen. In this solution, the values of decision variables (X1, X2, ..., X12) practically represent the number of 446-cm logs that have to be cut according to twelve ($n_l = 12$) cutting patterns given in Tab. 1. In this case, the minimally required number of original logs to fulfill this customer order is **338**. Optional sensitivity analysis is available by clicking either to the »RANGE/PRICE« or to the »RANGE/RHS« button. Recalling Fig. 4, the user can combine input data from other databases by clicking the ACCESS button.

Let us discuss the **benefits of using the new DSS** in comparison to the intuitive decision-making used in the veneers manufacturing firm to date. The operations manager may intuitively decide to cut one original 446 cm log directly into the required lengths of sub-logs of 182, 151 and 112 cm, leaving an edge-waste of 1 cm . This requires to cut 400 original 446-cm logs and gives a total **edge waste of 4 m** only and a total **end waste of 317,5 m.** The optimal solution result in total 338 original logs (OBJF in Fig. 4) required for cutting, where **total edge waste is 44,98 m** that is much greater then 4 m, but total **end waste is zero.**

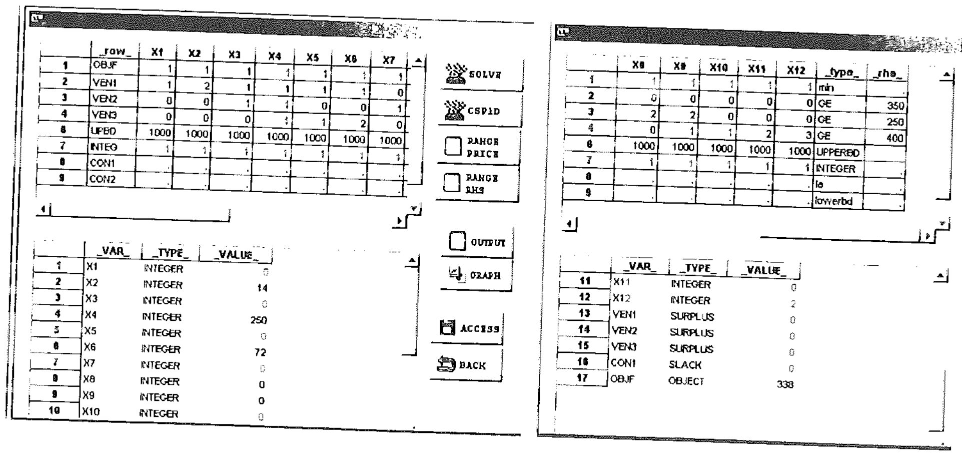


Fig. 4. Main data I/O screen: input test data and the results of optimal solution

6. Conclusion

The significance of this paper is primarily practical. The proposed approach represents an affordable alternative for a firm to get a simple-to-use DSS. In particular, this approach may represent a convenient opportunity for small and medium-size enterprises. Another significant contribution of this case study is in rising the awareness of DSS developers and OR researchers about the real needs of operations managers as a special class of DSS and OR users. These managers can contribute very much to improving the competitiveness of their firms when they have readily available IT support in the form of simple-to-use DSSs. This is in accordance with recent findings of Foulds and Thachenkary (2001) who urge the OR community to empower front-line decision makers in firms by means of end-user focused DSS.

References

1. Basnet, C., Foulds, L., Igbaria, M., 1996. Fleet Manager: a microcomputer-based decision support system for vehicle routing, Decision Support Systems 16, 195-207.
2. Chaudhry, S., Sallhenberger, L. and Beheshstian, M., 1996. A small business inventory DSS: design, development and implementation issues, Computers & Operations Research 23 (1), 63-72.
3. Čižman, A., Černetič, J., 2004. Improving competitiveness in veneers production by a simple-to-use DSS, European Journal of Operational Research 156, 241 – 260.
4. Dyckhoff, H., 1990. A typology of cutting and packing problems, European Journal of Operational Research 44, 145 - 159.
5. Foulds L. and Thachenkary C., 2001. Empower to the people – how decision support systems and IT can aid the OR analyst, OR/MS Today, June 2001, p. 3. (Downloadable from website http://www.lionhrtpub.com/orms/orms-6-01/foulds.html).
6. Gillmore, P.C. and Gomory, R.E., 1961. A linear programming approach to the cutting stock problem, Operations research, 9, pp. 849 - 859.
7. SAS, 1999. SAS/OR User's Guide: Mathematical Programming, Version 8, SAS Institute Inc., Cary, NC.
8. SAS, 1995. SAS/AF Software Frame Application Development Concepts, Version 6, First Edition, SAS Institute Inc., Cary, NC.
9. SAS, 1994. The SAS System for Information Delivery-Scientific and Technical Applications, SAS Institute Inc.

CONJOINT ANALYSIS IN WOOD PRODUCTS' DESIGN

Matjaž Feltrin, Zakraj 1a, 5283 Slap ob Idrijci, Slovenia, e-mail: matjaz_feltrin@email.si
Lidija Zadnik Stirn, Jasna Hrovatin, University of Ljubljana, Biotechnical Faculty, 1000 Ljubljana, Slovenia, e-mail: lidija.zadnik@bf.uni-lj.si; jasna.hrovatin@bf.uni-lj.si

Abstract

Operations research tools are widely used in wood products' market research which highly affects the products' design. For successful wood products' design different tools can be applied. In this paper Conjoint Analysis (CA) is used for determining the benefits of the attributes and their levels which best correspond to analyzed wood products or services. CA application is also employed in evaluating already market placed wood products where redesigning or repositioning of products should be established.

Key words: operations research, wood products' design, market research, conjoint analysis (CA)

Introduction

The problem we are dealing with is to analyze the main attributes and their levels of three different furniture elements (living room table, dining room table and a chair), and to evaluate, by presented attributes, three different chairs, one dining room table and one living room table, which participated at an exhibition (only dinning room chair is displayed in this paper to show the evaluation method). The method used for the analysis and evaluation is conjoint analysis (CA). CA is widely used as market research tool for developing and designing of new or redesigned products or services (Green and Srinivasan, 1978; Dahan and Hauser, 2000; Wang et al., 2004). Using CA, we can answer questions such as: which product's attributes are important or unimportant to the consumer, what levels of product's attributes are the most or the least desirable in the consumer's mind, how large is the market share of leading competitors' products versus our product, and many other market focused issues (Green and Srinivasan, 1978; Gustafson et al., 1999; Zadnik, 2001). Its main advantage is that it asks the respondent to make choices in the same way as the consumer is when consuming (Green and Srinivasan, 1978). Attributes and levels to evaluate wood products, used in the presented research, are given in Table 1.

Table 1: Analyzed attributes and their levels

Innovativeness	not innovative	Aesthetics	low level
	medium innovative		medium level
	high innovative		high level
Functionality	limited functionality	Safety	not appropriate
	functional		appropriate

The present research includes personal interview with four respondents of different working areas. The results demonstrate which attributes and which levels must be of highest consideration when designing the treated wood products. Further, analyzing different products, the objective is to focus on repositioning of wood products, demonstrating which attributes and which their levels were not considered sufficiently when designing.

Methods and data

The first part of the paper includes three furniture elements (living room table, dining room table and a chair) analyzed by CA, regarding respondents' preferences to attributes presented in Table 1. Further, the second part of the research is focused on five furniture elements which competed on Biennale of industrial design (BIO) in Ljubljana in the year 2004, where redesigning considering analyzed attributes was our interest. As noted, only the analysis of dinning room chair is shown in this paper.

Method used to analyze the data collected from respondents included in this research is CA, performed by SPSS program. The name "Conjoint Analysis" implies the study of the joint effects and as a multivariate technique has its applications in wood products' design where the decisions are studied (Green and Srinivasan, 1978; Gustafson et al., 2001). Benefit of using CA rather than any other existing method lies in the methods ability to incorporate all relevant aspects of the decision making process, what means that the decision maker is better able to asses the importance of attributes as well as the levels of attributes (Green and Srinivasan, 1978). Marketing managers, as well as decision makers, are faced with numerous difficult tasks when developing new wood products or modifying existing products. These tasks include prediction of:

- the profitability and/or market share for proposed new product,
- the impact of new competitors' products on profit,
- customer's acceptability of a new product (Green et al., 1986),
- competitive reaction to sellers' strategies when introducing a new product,
- the impact of situational variables on customer preference,
- response to alternative advertising strategies,
- the customers' response to alternative pricing strategies (Green and Srinivasan, 1978).

CA used for evaluation of competitive strategy of a new wood product may be implemented by modifying the marketing mix, new product identification, pricing, advertising and distribution (Green and Srinivasan, 1978; Green et al., 1986; Gustafson et al., 1999; Haaijer, 1999; Dahan and Hauser, 2000; Gustafson et al., 2001; Wang et al., 2004). Such strategies may focus on new segments or products' re-positioning, what is considered in this paper. Further, CA involves the measurement of psychological judgments, such as consumer preferences, or perceived similarities or differences between choice alternatives. The main methodologies which model CA are:

- Stimulus Construction: Two Factors at a Time; Full Factorial and Fractional Factorial Design,
- Data Collection: Two Factors at a Time, Tradeoff Analysis; Full Profile Concept
- Model Type: Compensatory and Non-Compensatory Models, Part Worth Function; Vector Model; Mixed Model; Ideal Point Model,
- Measurement Scale: Rating Scale; Paired Comparisons; Constant Sum; Rank Order,
- Estimation Procedure: Metric and Non-Metric Regression; MONANOVA; PREFMAP; LINMAP; Non-metric Tradeoff; Multiple Regression; LOGIT; PROBIT; Hybrid; TOBIT; Discrete Choice,
- Simulation Analysis: Maximum Utility; Average Utility; LOGIT; PROBIT (Green and Srinivasan, 1978; Gustafson et al., 1999; Gustafson et al., 2001).

Several different implementations of conjoint measurement are manifested in CA algorithms and computer programs. The most noticeable are: categorical conjoint measurement, monotone ANOVA models, OLS regression methods and linear programming methods (Green and Srinivasan, 1978; Green et al., 1986; Haaijer, 1999; Gustafson et al., 2001). For purposes of this paper, only OLS regression methods are discussed. The Ordinary Least Squares (OLS) regression approach to CA offers a simple method of deriving alternative forms of respondent utilities (part-worth, vector, or ideal point models). The attractiveness of the OLS model is partly a result of the ability to scale respondent choices using rating scales, rather than rankings. Based on this rated input, the CA measurement procedures are applied to identify a mathematical function of attributes. Further, the objective of OLS CA is to produce a set of additive part-worth utilities that identify each

respondent's preference for each level of a set of product's attributes (Green and Srinivasan, 1978; Gustafson et al., 2001). This model is expressed as:

$$r_j = b_0 + b_1 x_{1,j} + b_2 x_{2,j} + b_3 x_{3,j} + \ldots + b_m x_{m,j} + \varepsilon_j \quad (1)$$

where:

r_j - observed evaluation of profile j (respondent's rating), j=1,…,p; (p-th attribute)

b_0 - constant term

$b_1,\ldots,b_4$ - model parameters quantifying the effect of attribute level variation on profile evaluations

$x_{1,j},\ldots,x_{m,j}$ - variables indicating the combination of attribute level of profile j

ε_j - exogenous stochastic distributions

In CA contribution of attribute's level to the total utility is called "part worth", and the total utility of a profile is equal to the sum of all the part worth. In the part-worth model, the utility function s_i is defined as (Haaijer, 1999):

$$s_i = \sum_{p=1}^{i} f_p y_{ip} \quad (2)$$

where:

s_i - preference for the stimulus object at level i,

f_p - function representing the part worth of each of i different levels of the stimulus objects for the p-th attribute,

y_{ip} - level of the p-th attribute for the stimulus object at level i.

The estimation of utilities for each attribute permits the estimation of average attribute's importance A_p which is computed as (Green and Srinivasan 1978; Haaijer, 1999):

$$A_p = \frac{(\max_{ip} - \min_{ip})}{\sum (\max_{ip} - \min_{ip})} \quad (3)$$

where:

A_p - importance of attribute p

$\min_{ip}$ - lowest utility of analyzed attribute p

$\max_{ip}$ - highest utility of analyzed attribute p

In our research 4 respondents were identified to answer questions about three products (designed as profiles) and five selected furniture elements. The CA process started with stimulus construction as the basis for CA. It was focused on the related problems of determining which attributes to present to the respondent (Gustafson et al., 2001). In the first part of the research, each product considered, was described by profiles containing four attributes with three, respectively two levels (Table 1). The number of selected stimuli (profiles), to be presented to the respondents, depends on the type of the model and must be carefully constructed, so that the stimuli are orthogonal. Regarding the fact that, if using the full profile, it is an unmanageable for the respondent to evaluate all 36 (3x2x3x2) profiles, we have chosen the additive rule. So we were able to use fractional factorial statistical design to reduce 36 possible profiles to 9 profiles (Table 2). It is from this reduced set of profiles that we estimated the set of choice utilities associated with each of the individual attribute and their associated levels. Table 2 also presents the questionnaire which was handed to the respondents. Further, the CA experiment was performed by personal interview with respondents. They were also presented with a fold-out form that had 9 response categories, where 1 means the best and 9 the worst category (Table 2). The rated profile numbers were used for the evaluation of utility in SPSS program.

In the second part of the research five products, which competed at an exhibition, were analyzed. The objective was to evaluate the products by presented attributes and their levels (Table 1). First, respondents were asked to evaluate five furniture elements using only their pictures. These results were further used for the descriptions of furniture elements with attributes. The aim was to illustrate how well the already made products correspond to the consumers' preferences derived from the previous (first) part of the research. The questionnaire designed in the second part of the research is given in Table 3.

Table 2: Stimulus combinations of fractional factorial design

Profile	Innovativeness	Functionality	Aesthetics	Safety	(rate 1 to 9)
#1	High innovative	Functional	High level	Inappropriate	
#2	High innovative	Limited functionality	Low level	Appropriate	
#3	Medium innovative	Limited functionality	High level	Appropriate	
#4	Medium innovative	Limited functionality	Medium level	Inappropriate	
#5	Medium innovative	Functional	Low level	Inappropriate	
#6	Low innovative	Limited functionality	High level	Inappropriate	
#7	Low innovative	Limited functionality	Low level	Inappropriate	
#8	High innovative	Limited functionality	Medium level	Inappropriate	
#9	Low innovative	Functional	Medium level	Appropriate	

Table 3: Evaluation of products

Attribute	Level	Evaluation- sign (x)
Innovativeness (I)	Low innovative (1)	
	Medium innovative (2)	
	High innovative (3)	
Function (II)	Limited functionality (1)	
	Functional (2)	
Aesthetics (III)	Low level (1)	
	Medium level (2)	
	High level (3)	
Safety (IV)	Inappropriate (1)	
	Appropriate (2)	

Results

Utilities s_i (2) and the relative importance A_p (3) of each attribute were calculated for each respondent separately with SPSS CA program. The total worth for the selected respondent was calculated using the additive rule (1). The results for dining room chair when evaluating the 9 selected stimuli are shown in Table 4.

These utility values were further used to obtain a total utility for each of the profile from Table 2. For example, to find the utility of the first combination in Table 2, considering first respondent we simply add the part worth of the respective levels, using equation (2).

The derived utilities calculated for each respondent separately were then aggregated to obtain an overall result using equation (3), and are presented in Table 6.

Table 4: Respondents' evaluation and derived utility for dining room chair

	Evaluation (r_j)								
Respondent	#1	#2	#3	#4	#5	#6	#7	#8	#9
A	4	2	3	7	5	8	9	6	1
B	3	5	2	8	4	6	9	7	1
C	1	8	2	7	6	4	9	5	3
D	4	2	3	8	5	6	9	7	1
Respondent	Derived utility for profile (s_i)								
A	6,3	8,0	7,3	2,7	5,0	1,8	1,5	3,7	8,7
B	7,5	5,2	7,3	2,7	5,0	3,7	1,3	2,8	9,5
C	9,3	2,5	7,5	3,8	3,7	6,2	0,8	4,2	7,0
D	6,2	7,2	7,7	2,7	4,7	2,7	1,7	3,2	9,2

Table 5: Demonstration of evaluation of respondent A for profile #1

	Innovativeness	Functionality	Aesthetics	Safety	Constant	Total
Part worth	3	5	0,5	4,5	-6,7	6,3
max	3	5	0,5	9,0		10,8
min	1	2	0,2	4,5		1,0

Table 6: Relative importance for dinning room chair

	Relative importance (%)				Sum
Respondent	Innovativeness	Function	Aesthetics	Safety	
A	21	27	4	48	100
B	4	36	24	36	100
C	7	26	56	11	100
D	11	28	11	50	100
Total	11	29	24	36	100

From the relative importance (Table 6) it is evident that safety (36 %), functionality (29 %) and aesthetics (24%) must be of great significance when designing new products, however taking into account consumers' preferences (Table 5).

In the second part of the research products that already participated at an exhibition were analyzed. The results collected from a questionnaire (Table 3) were compared with the maximum and minimum utilities collected in the first part of the research. In Table 7 we can see that among different chairs when analyzed by respondents, different acceptance levels are displayed. This is due to different working areas of respondents. To get more reliable results research should include more respondents.

Table 7: Evaluation and derived utility of three dinning room chairs

Resp.	Evaluation (see Table 2)			Derived utility			Comparison	
	chair 1	chair 2	chair 3	ch.1	ch.2	ch.3	max	min
A	I2-II2-III2-IV1	I2-II2-III1-IV2	I1-II1-III2-IV1	4,4	9,5	1,7	10,8	1,5
B	I1-II1-III2-IV1	I1-II1-III2-IV2	I1-II2-III3-IV2	6,0	6,0	10,7	11,0	2,3
C	I2-II2-III3-IV2	I1-II2-III2-IV2	I2-II1-III3-IV2	10,0	7,0	7,5	10,3	0,8
D	I2-II2-III3-IV1	I2-II2-III3-IV2	I2-II1-III2-IV2	5,7	10,1	7,2	10,7	1,7

Conclusion

We used conjoint analysis to find out which product's attributes are of highest importance to consumer and must be of great concern when designing wood products - furniture. We also analyzed five products that participated at an exhibition, by the same attributes. The results showed that safety, which is at the highest level, must be of highest importance when designing furniture. Second most important attributes are functionality and aesthetics (at the highest level) and the last is innovativeness. Results are reasonable as it is known that everyone pays the attention first at product's safety and functionality, and then how the product looks out and how innovative it is. Regarding innovativeness, we can conclude that it is considered only after safety, functionality and aesthetics are established.

The second part of the research revealed which attributes and which their levels were, or were not considered enough when designing the analyzed products. These results could be used with real wood commodities when repositioning or redesigning the considered products. Further, from this research we can conclude that among the participated products only one nearly responded to the attributes at the highest level, which were of the highest importance to respondents. This means, that when redesigning the analyzed products, attention has to be paid on attributes by which the highest high level was revealed by the research results.

Tools of operation research such as CA can be used in designing processes in wood production companies. With CA we demonstrated how operations research tools can be and must be used for solving the problems when designing furniture. In the first phase the importance of derived attributes should be determined, in order to improve them in the re-designing phase.

References

1. Dahan, E. and Hauser, J.R., 2000. Managing a Dispersed Product Development Process. Paper 103. The Handbook of Marketing. Cambridge. Center for Innovation in Product Development at M.I.T.: 62 pages.
2. Green, P. E. and Srinivasan, V., 1978. Conjoint Analysis in Consumer Research. Issues and Outlook. Journal of Consumer Research, 5: pp 103-123.
3. Green, P. E., Carroll, J.D., Goldberg, S.M., 1986. A general approach to product design optimization via conjoint analysis. Journal of Marketing Research, 45, 3: pp 17-37.
4. Gustafson, A., Ekdahl, F., Bergman, B., 1999. Conjoint analysis: a useful tool in the design process. Total Quality Management, 10, 3: pp 327-343.
5. Gustafson, A., Herrmann, A., Huber, F., 2001. Conjoint measurement. Methods and applications. Second edition. Springer Verlag, Berlin., 542 pages.
6. Haaijer, M.E., 1999. Modeling conjoint choice experiments with the probit model. Faculty of Economic Sciences. Groningen. University Library Groningen, 20 pages http://www.rug.nl/bibliotheek/collecties/dissertaties/index, 20.07.2005.
7. Wang, Q., Shi, G., Halbrendt, C.C., 2004. Market potential for fine furniture manufactured from low-grade hardwood. Evidence from a conjoint analysis in the north-eastern United States. Forest Products Journal. 54, 5: pp 19-25
8. Zadnik Stirn, L., 2001. Operacijske raziskave (podiplomski študij biotehnike), for internal use, Ljubljana, Biotechnical Faculty, pp 149-153.

LTO Heuristics for Capacitated Lot Sizing Problem with Sequence Dependent Setups and Overtimes

Zrinka Lukač, Kristina Šorić, Višnja Vojvodić Rosenzweig
Faculty of Economics Zagreb
Kennedyev trg 6
10000 Zagreb, Croatia
e-mail: zlukac@efzg.hr, ksoric@efzg.hr, vvojvodic@efzg.hr

June 15, 2005

Abstract

The wellknown CLSP problem is generalized including sequence dependent setup times and overtimes and modeling it as a quadratic mixed 0-1 integer programming problem called Capacitated Lot Sizing Problem with Sequence Dependent Setups and Overtimes. We develop a heuristics based on Lagrangean relaxation and tabu search for solving the problem. At the end, some computational results are presented.

Key words: capacitated lot sizing problem, two machines, capacity limitations, overtimes, sequence dependent setup times, quadratic mixed 0-1 integer programming problem, heuristics, tabu search, Lagrangean relaxation

1 Introduction

In this paper we are generalizing the wellknown deterministic, single-level dynamic lot sizing - the capacitated lot sizing problem (CLSP) in the sense that the setup times are sequence dependent and the overtimes are introduced. Also, the availability of two machines is presented. The objects is to minimize the sum of the costs of production, storage, setup and overtime. A setup may imply two kinds of machine consumption. One is setup cost, expressed in monetary terms; the other is setup time, consuming a certain amount of machine-hours. Having the machine capacities expressed in time units, it is important to have the setup time as small as possible because in this way there is more time available for the production. This problem can be formulated as a quadratic mixed 0-1 integer programming problem called Capacitated Lot Sizing Problem with Sequence Dependent Setups and Overtimes.

Capacitated lot sizing problem has been shown to be NP-hard (Bitran and Yanasse, 1982). Also, when setup times are included in the model, finding a feasible solution to the capacitated lot sizing problem also becomes an NP-complete problem (Garey and Johnson, 1979). Therefore, research on developing effective heuristics has been a profitable research area for a long time.

The paper is organized as follows. In the second section we are presenting the mathematical model describing the considered problem. The third section is devoted to the relaxations. Combining these relaxations we are constructing the LTO heuristics in the fourth section and the computational results are presented. We performed three kinds of numerical simulations. First, we linearized the problem for the small dimensions in order to get the optimal solution. The same problems were solved using only Lagrangean relaxation and after that, applying the LTO heuristics. For the problems of larger dimensions we applied the Lagrangean relaxation and after that, the LTO heuristics. The comparison results are presented in the table considering the quality of the solutions and computational times.

2 Formulation of the problem

In order to formulate the above problem, let us introduce the following notations:

1. Index: i - item type, $i = 1, 2, ..., n$, j - machine type, $j = 1, 2$ and t - planning period, $t = 1, 2, ..., T$.

2. Parameters: d_{it} - the demand for product i in period t, p_{itj} - the unit production cost of product i in period t on machine j, h_{it} - the unit storage cost for product i in period t, k_{itj} - the fixed setup cost for item i in period t on machine j, u_{iltj} - the setup time from item i to item l in period t on machine j, q_{tj} - the overtime cost of machine j in period t, c_{tj} - the capacity of machine j in period t, a_{ij} - the consumption of machine j per unit of item i and O_{tj} - the maximum overtime of machine j in period t.

3. Variables: x_{itj} - the amount of item i produced in period t on machine j, s_{it} - the inventory (stock) of item i in period t, o_{tj} - the overtime of machine j in period t

and

$$y_{itj} = \begin{cases} 1, & \text{if machine } j \text{ is set up for item } i \text{ in period } t \\ 0, & \text{otherwise} \end{cases}$$

The quadratic mixed 0-1 integer programming formulation called (P) is

$$minf(x, y, s, o) = \sum_{i=1}^{n}\sum_{t=1}^{T}(\sum_{j=1}^{2} p_{itj}x_{itj} + h_{it}s_{it} + \sum_{j=1}^{2} k_{itj}y_{itj}) + \sum_{t=1}^{T}\sum_{j=1}^{2} q_{tj}o_{tj}$$

$$s_{i,t-1} + \sum_{j=1}^{2} x_{itj} = d_{it} + s_{it} \quad \forall i, t \qquad (1)$$

$$\sum_{i=1}^{n}\sum_{l=1}^{n} u_{iltj}y_{i(t-1)j}y_{ltj} + \sum_{i=1}^{n} a_{ij}x_{itj} \leq c_{tj} + o_{tj} \quad \forall t, j \qquad (2)$$

$$x_{itj} \leq M_{itj}y_{itj} \quad \forall i, t, j \qquad (3)$$

$$\sum_{i=1}^{n} y_{itj} \leq 1, \quad \forall t, j \qquad (4)$$

$$o_{tj} \leq O_{tj}, \quad \forall t, j \qquad (5)$$

$$x_{itj}, s_{it}, o_{tj} \geq 0, \quad y_{itj}, \in \{0, 1\} \qquad (6)$$

where M_{itj} is the upper bound on the production capacities, $M_{itj} = \frac{c_{tj}+O_{tj}}{a_{ij}}$. The constraints (1) represent the flow conservation constraints for each item in each period. Also, the constraints (2) describe the capacity limitations and overtime decision for each machine in each period. The constraints (4) refer to a single mode of production.

For small dimensions we linearized the model and applied the branch and bound method offered by Cplex. But in this way the number of variables and constraints increased a lot. In order to solve the problem for larger dimensions we introduce the heuristics based on tabu search and the Lagrangean relaxation.

3 Relaxations

In order to solve the NP-hard problem defined in the previous section, we propose three relaxations.

The first relaxation: decomposition $(RP1)$

The motivation for the generation of the heuristics starting points is as follows. We introduced the 0-1 variables

$$z_{ij} = \begin{cases} 1, & \text{if item } i \text{ is produced on machine } j \\ 0, & \text{otherwise} \end{cases}$$

and constraints $y_{itj} \leq z_{ij}$, $\forall i, t, j$. If $z_{ij} = 1$, $\forall i, j$ (every item can be produced on every machine), the dimension of the problem is very large. The feasibility set is as large as possible. But, by fixing some z_{ij} to 0, we decrease the number of variables and constraints. Let $\mathcal{N} = \{1, ..., n\}$. We assign the items to the machines in order to allow their productions on them. Let $\mathcal{N}_1$ be the set of all items that are going to be produced on machine 1, let $\mathcal{N}_2$ be the set of all items that are going to be produced on machine 2, and let $\mathcal{N}_0$ be auxiliary set. We allow an item to be produced on both machines as well. Because of machine deterioration, without a loss of generality we can assume that setup times u_{iltj} are proportional by some coefficient of proportionality $b > 1$, i.e. $u_{il,t+1,j} = bu_{iltj}$. Therefore we fix time to be $t = 1$. Then we fix the machine. If for the sum of the setup times on the first machine for a pair of items i and l the inequality $u_{il11} + u_{li11} \leq k$ holds for a certain value of k, then we allow both items i and l to be produced on the first machine, i.e. $i, l \in \mathcal{N}_1$. We do the same for the second machine, i.e. if $u_{il12} + u_{li12} \leq k$, then we allow items i and l to be produced on the second machine, $i, l \in \mathcal{N}_2$. If there exists an item i not assigned to any machine, we put it to set $\mathcal{N}_0 = \{i \in \mathcal{N} : i \notin \mathcal{N}_1 \bigcup \mathcal{N}_2\}$. After we've done this, the items from the set $\mathcal{N}_0$ are allowed to be produced on the machines, randomly. Now the dimension of the problem (P) is reduced significantly.

Using the notations $\mathcal{N} = \{1, ..., n\}$, $\mathcal{N}_j = \{i \in \mathcal{N} : z_{ij} = 1\}, j = 1, 2$, the indexing $i = 1, ..., n; j = 1, 2$ in every constraint is substituted by $i \in \mathcal{N}_j, j = 1, 2$ resulting in the new problem $(RP1)$.

The second relaxation: fixing F_{tj} and o_{tj} $(RP2)$

Let us consider the quadratic constraints (2). In order to eliminate the quadratic term $F_{tj}(y) = \sum_i \sum_l u_{iltj} y_{i(t-1)j} y_{ltj}$ in the capacity constraints $g_{tj}(x, y, o) = F_{tj}(y) + \sum_i a_{ij} x_{itj} - c_{tj} - o_{tj} \leq 0$, we will fix $F_{tj}(y)$ randomly (using the uniform distribution offered by Cplex) to some value F^r_{tj} from the interval $[0, maxF_{tj}]$, where $maxF_{tj} = max\{u_{iltj} : i, l \in N_j\}, j = 1, 2; t = 1, ..., T$. Also, we will fix the term o_{tj} to some value o^r_{tj} from the interval $[0, O_{tj}]$, where O_{tj} is the maximum overtime of machine j in period t. In this way we construct the problem $(RP2)$ which has the same constraints as the problem (P), but the constraints (2) are substituted with the constraints obtained in the following way:

$$\sum_i a_{ij} x_{itj} \leq c_{tj} - F^r_{tj} + O^r_{tj}, \quad j = 1, 2; \quad t = 1, ..., T \tag{7}$$

Let $c_{tj} - F^r_{tj} + o^r_{tj} = b^r_{tj}$. Now we can substitute the constraints (2) by $a_{ij} x_{itj} \leq b^r_{tj} y_{itj}$ $\forall i, t, j$ and reduce the constraints (3). Fixing F_{tj} and o_{tj} and solving the resulting capacitated lot sizing problem without setup times and overtimes, $(RP2)$, R times will give us the sequence of optimal solutions (x^r, y^r, s^r, o^r), $r = 1, ..., R$. These optimal solutions may satisfy all the constraints of (P), except maybe the constraints (2). If some of the solutions (x^r, y^r, s^r, o^r) is feasible for (P) we have an upper bound on the optimal objective function value f_{min}. Having more than one upper bound, we take the minimal one.

The third relaxation: Lagrangean relaxation (LR)

In order to obtain the Lagrangean relaxation, we do as follows. Let us dualize the constraints (2) of the problem (P) obtaining the problem (LR):

$$v(\lambda) = \{minf(x, y, s, o) + \sum_{t,j} \lambda_{tj} g_{tj}(x, y, o)\}$$

subject to (1), (3), (4), (5) and (6) of the problem (P). The problem $max\{v(\lambda) : \lambda \geq 0\}$ is the Lagrangean dual of (P) relative to the complicating constraints $g_{tj}(x, y, o) \leq 0, \forall t, j$.

How to select an appropriate value for the λ? Having the sequence of optimal solutions for $(RP2)$, (x^r, y^r, s^r, o^r), $r = 1, ..., R$, let us solve the problem $(\overline{D})$

$$max \ w$$

$$w \leq f(x^r, y^r, s^r, o^r) + \sum_{t,j} \lambda_{tj} g_{tj}(x^r, y^r, o^r), r = 1, ..., R \tag{8}$$

$$\lambda_{tj} \geq 0, \quad t = 1, ..., T, \ j = 1, 2$$

The optimal solution $\lambda = (\lambda_{tj})$, $t = 1, ..., T$, $j = 1, 2$ of the $(\overline{D})$ represents the initial Lagrangean multipliers for (LR). Let us note the optimal solution of (LR) by $(x(\lambda), y(\lambda), s(\lambda), o(\lambda))$. Now, we have the lower bound for the optimal objective function value of the problem (P) as follows:

$$f(x(\lambda), y(\lambda), s(\lambda), o(\lambda)) + \sum_{t,j} \lambda_{tj} g_{tj}(x(\lambda), y(\lambda), o(\lambda)) \leq f_{min}$$

Let $(x(\lambda), y(\lambda), s(\lambda), o(\lambda)) = (x^{R+1}, y^{R+1}, s^{R+1}, o^{R+1})$ and solve the problem $(\overline{D})$ for $r = 1, ..., R + 1$. We applied the constraint generation method also called cutting plane method (CP). One substantial advantage of (CP) over subgradient algorithms for (LR) is the existence of a true termination criterion incorporated in the algorithm.

4 LTO Heuristics and Computational Results

In order to define finaly the tabu search heuristics, we have to construct the neighborhood. For each point of the tabu search heuristics we construct exactly four neighborhood points. First we find the pair of items i and l allowed to be produced on the first machine for which the setup time u_{ilt1} is maximum. Then we construct the first neighborhood point by leaving item l to be produced on the first machine and sending item i to be produced on the second machine, and second neighborhood point by leaving item i to be produced on the first machine and sending item l to be produced on the second machine. The other two neighborhood points are obtained by finding the pair of items allowed to be produced on the second machine for which the setup time u_{ilt2} is maximum, and by sending either item i or l to be produced on the first machine.

The heuristics is implemented in the AMPL programming language and uses CPLEX. All computations were performed on a PC having Pentium IV 2.4 GHz processor and 1Gb RAM.

The next table shows the comparison of the results in terms of computational time, where Time-Imp LL denotes the average improvement (in %) of the computational time of Lagrange relaxations compared to the linearized problem, and Time-Imp TL denotes the average improvement of the computational time (in %) of the tabu search heuristics compared to the linearized problem

	Average time (in seconds)				
Problem Class (T,n)	Linearized model	Lagrange relaxation	Tabu search heuristics	Time-Imp LL (in %)	Time-Imp TL (in %)
(10, 5)	2.8	2.4	11.9	14.29	–
(15, 5)	4.8	2.8	16.8	41.67	–
(15, 8)	4301.9	265.7	354.3	93.82	–
(20, 5)	8.4	4.2	21.1	50	–
(20, 10)	110425.1	> 130000	483.5	–	99.56
(20, 15)	> 130000	> 130000	1523.4	–	> 98

Table 1. Comparison of results in terms of computational time

Comparing the results in the terms of the quality of solution, the average difference between the objective function value obtained by Lagrange relaxation and obtained by solving the linearized model is 0.42% and the average difference between the objective function value obtained by tabu search heuristics and obtained by solving the linearized model is 0.89%.

In terms of computational time, for smaller dimensions of the problem tabu search heuristics is not as effective as Lagrange relaxation method, or even linearized model. However, for higher dimensions of the problem it is by far the most effective method in terms of computational time, producing solutions within 1% gap and significantly reducing computational time.

References

[1] G.R. Bitran, H.H. Yanasse, Computational complexity of the capacitated lot size problem, *Management Science* 28 (1982) 1174-1186

[2] M. Garey, D. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, *Freeman, San Francisco, CA* (1979)

[3] B. Karimi, S.M.T. Fatemi Ghomi, J.M. Wilson, The capacitated lot sizing problem: a review of models and algorithms, *Omega, The International Journal of Management Science* 31 (2003) 365-378

[4] J. Maes, J.O. McClain, L.N. Van Wassenhove, Multilevel capacitated lot-sizing complexity and LP-based heuristics, *European Journal of Operational Research* 53 (1991) 131-148

MARKOV CHAINS IN MRP MODEL OF A MULTI-LEVEL PRODUCTION SYSTEM

Aleš Silič
Goriške opekarne d.d., Merljaki 7, 5292 Renče, Slovenia
e-mail: ales.silic@go-opekarne.si

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

Janez Kušar, Marko Starbek
University of Ljubljana, Faculty of Mechanical Engineering, Aškerčeva 6, 1000 Ljubljana, Slovenia
e-mail: janez.kusar@fs.uni-lj.si, marko.starbek@fs.uni-lj.si

Abstract: Material requirements planning (MRP) deals with relationship between the demand for the final product and the components used to make it. Thus in manufacturing systems MRP are used to schedule production. The paper presents a MRP model which is based on the concept of input-output analysis. Further, the MRP model incorporates stochastic external demand using Markov chains. In order to analyze the dynamic behavior of the multi-level production system the Laplace transform in continuos time model developed by Grubbstroem and Molinder was used. A simplified version of the presented model is illustrated on a final product – radial ventilator.

Keywords: material requirements planning, stochastic external demand, Laplace transform, Markov chains, radial ventilators

1. INTRODUCTION

Material requirements planning (MRP) is a management information system to determine schedules in multi-level manufacturing systems. Normally a multi-level system consists of raw materials, components, subassemblies, and final products, which are also known as items. MRP reduces a master schedule of final products into a time-phased schedule of requirements of intermediate items. In other words, it suggests decisions to purchase manufacture and assemble items based on the external demand for final products [4]. There are mainly three ways for representing a product and its quantity relationship with its components. These three methods are the product structure tree [6], the Gozinto graph [8] and the input matrix of input-output analysis [4].

Important concept is application of input-output analysis in the MRP models, which has been investigated by Grubbström and Ovrin [2]. Whereas input-output analysis uses matrices for the representation of the requirements of subcomponents for the production of items at later stages, the MRP approach makes use of tableaus for describing the way in which the demand for final products are broken down into requirements (internal demand) of subcomponents, taking lead time, initial inventories and order quantities into account.

In order to analyze the dynamic behavior of the multi-level production system, Grubbström and Molinder [3] used the Laplace transform theory in continuous time models. MRP models have been upgraded in several directions to come closer to real situations. In this paper we consider stochastic external demand in the MRP model including Markov chains.

2. MATERIAL REQUIREMENTS PLANNING

For the problem considered here the following basic notations are introduced:

N = number of different items as raw materials, intermediates, as well as final products

Q_i = order quantity of item i, $i = 1, 2, \dots, N$

κ_i = number of setups of item i at some point in time (production frequency)

τ_i = lead time for the production of item i from the time its production is ordered until the time the order is completed

S_i = total inventory of item i at some point in time

R_i = available inventory of item i at some point in time

d_i = external demand rate of item i at some point in time (the master production schedule)

H_{ij} = number of units of item i required for the production of one unit of item j

I = identity matrix

s = complex frequency of the Laplace transform.

For convenience, the above quantities are arranged into vectors and matrices according to Q being $(N \times N)$-dimensional diagonal matrix with Q_i in diagonal; H an $(N \times N)$-dimensional input matrix with H_{ij} in position (i,j). $H^{*}=(I - H)^{-1}$ is Leontief inverse of H. κ, d, S, R all being N-dimensional column vector with ith components κ_i, S_i, R_i, respectively.

In order to analyze the dynamic behavior of MRP system we make use of Laplace transform of a function of continuous time [1], which is defined as:

$$L\{f(t)\} = \tilde{f}(s) = \int_0^{\infty} e^{-st} f(t)dt. \tag{1}$$

The function $f(t)$ is inverse transform of $\tilde{f}(s)$. Important property of Laplace transform is translation property: $L\{f(t-\tau)\} = e^{-s\tau} L\{f(t)\}$. The lead times $\tau_1, \dots, \tau_N$ form internal demands for components of these times ahead of the completion of the products they enter. Hence, they correspond to translations backwards in time. For this purpose we introduce a diagonal matrix, the lead time matrix $\tilde{\tau}$, having $e^{-s\tau_i}$ in its diagonal position. The product $\tilde{H} = H\tilde{\tau}$ is defined as the generalized input matrix. When production takes place instantaneously in lots according to Q, the gross production of all products at any particular point in time is given by $Q\tilde{\kappa}$. This production requires components in the amounts of $HQ\tilde{\kappa}$. Taking all lead times into consideration, the internal demand (when components are reserved) is given by $H\tilde{\tau}Q\tilde{\kappa}$, which translates the requirements backwards in time compared to $HQ\tilde{\kappa}$. Due to the lead times, we make a distinction between total inventory $\tilde{S}$, available inventory $\tilde{R}$ and allocated component stock $\tilde{S} - \tilde{R}$. Making use of a time integration operator s^{-1}, total inventory accounts for all items in the system and is initial stock plus cumulate net production less external demand [5]:

$$\tilde{S} = \frac{S(0) + (I - H)Q\tilde{\kappa} - \tilde{d}}{s}. \tag{2}$$

As seen from the point of view of total inventory, the internal demand is instantaneous. Available inventory is given by [3, 5]:

$$\tilde{R} = \frac{R(0) + (I - H\tilde{\tau})Q\tilde{\kappa} - \tilde{d}}{s}. \tag{3}$$

Nonnegativity of available inventory is written $L^{-1}\{\tilde{R}(s)\} \geq 0$ and constitutes the major set of constrains in related optimization problems. Given fixed order quantities the decision variable is the cumulate number of setups $K_i(t)$ of item i in time t, which is generated as the nonnegative integer providing a nonnegative available inventory [3]:

$$K_i(t) = \left[Q_i^{-1} L^{-1} \left\{ \frac{\sum_{j=1}^{i-1} H_{ij}\tilde{\tau}_j Q_j \tilde{\kappa}_j}{s} + \left(\frac{\tilde{d}_i - R_i(0)}{s} \right) \right\} \right], \tag{4}$$

Where i means item i and j item j with lower index ($j < i$).

3. MARKOV CHAINS

According to [9], a discrete-time stochastic process is a Markov chain, independent of time t for all states i and j ($i,j=1,2,\dots,s$), if:

$$P(X_{t+1}=i_{t+1}|X_t=i_t, X_{t-1}=i_{t-1}, \dots, X_1=i_1, X_0=i_0) = P(X_{t+1}=i_{t+1}=j|\, X_t=i_t=i) = p_{ij} \tag{5}$$

Where p_{ij} is the probability that given the system is in state i at time t, it will be in a state j at time $t+1$. p_{ij} are referred as the transition probabilities and displayed as an sxs transition probability matrix P:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{bmatrix}. \tag{6}$$

Given that the state at time t is i, the process must be somewhere at time $t+1$. This means:

$$\sum_{j=1}^{s} P(X_{t+1} = j \mid P(X_t = i)) = \sum_{j=1}^{s} p_{ij} = 1, \quad i = 1, \dots, s. \tag{7}$$

A concept of steady-state probabilities is used to describe the long-run behavior of the Markov chain. There exists a vector $\pi = [\pi_1\ \pi_2 \cdots \pi_s]$ such that

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_s \\ \pi_1 & \pi_2 & \cdots & \pi_s \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_s \end{bmatrix}; \ \lim_{n\to\infty} P_{ij}(n) = \pi_j. \tag{8}$$

Where $P_{ij}(n)$ is called the n-step probability of a transition from state i to state j and is the ijth element of P^n. For large n, P^n approaches to a matrix with identical rows, and for all i, $P_{ij}(n+1) \approx P_{ij}(n) \approx \pi_j$:

$$P_{ij}(n+1) = \sum_{k=1}^{k=s} P_{ik}(n)p_{kj}, \text{ if } n \text{ is large } \pi_j = \sum_{k=1}^{k=s} \pi_k p_{kj} \tag{9}$$

Where $\pi_1 + \pi_2 + \dots + \pi_s = 1$.

4. AN ILLUSTRATION – RADIAL VENTILATORS AS FINAL PRODUCTS

Suppose a company assembles radial ventilators which are also the final products of the company. The final product denoted A is composed by an electromotor denoted B, screws denoted C, a flange denoted D, an impeller denoted E, a frame denoted F and legs of the frame denoted G. Figure 1 illustrates the product structure tree [6] of the radial ventilator.

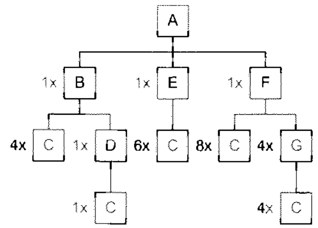


Figure 1: Product structure tree of the product A (radial ventilator)

Order quantities Q_i for components (A, B, F, E, D, G, and C) are 600, 1100, 1650, 1300, 2250, 4100 and 23000 pieces. The lead times τ_i are 1, 2, 3, 2, 2, 1 and 1 week and the initial available inventories $R_i(0)$ are 450, 1000, 1300, 1200, 2000, 5000 and 21500 pieces. The company orders components at the end of every week for the next week based on their consumption in the current week. The company has sales contract with one minor and two major purchasers. If demand for final products in separate week comes only from the minor purchaser, then the order is 100 pieces of radial ventilators. If demand of one major purchaser is included in the same week, then the order is 400 pieces. If both major purchasers are included, then the order is 700 pieces. All transition probabilities are obtained based on past sales experiences with these three purchasers and illustrate transition probability matrix P as defined by equation (6) and (7):

$$P = \begin{array}{c} \\ 100 \\ 400 \\ 700 \end{array} \begin{array}{c} \begin{matrix} 100 & 400 & 700 \end{matrix} \\ \begin{bmatrix} 0{,}35 & 0{,}5 & 0{,}15 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}7 & 0{,}15 & 0{,}15 \end{bmatrix} \end{array}$$

Company wants to know the expected external demand of radial ventilators after a long time, number of setups and component stocks per week. The demand of the final products in each week with transition probabilities shown in transition probability matrix P is described as a Markov chain with states X_1 = 100, X_2 = 400, X_3 = 700 [piece]. To find out the expected external demand E(d) after a long time we first solve the equations (9) for steady-state:

$$[\pi_1 \quad \pi_2 \quad \pi_3] = [\pi_1 \quad \pi_2 \quad \pi_3] \begin{bmatrix} 0{,}35 & 0{,}5 & 0{,}15 \\ 0{,}4 & 0{,}4 & 0{,}2 \\ 0{,}7 & 0{,}15 & 0{,}15 \end{bmatrix}$$

$$\pi_1 = 0{,}35\pi_1 + 0{,}40\pi_2 + 0{,}70\pi_3$$
$$\pi_2 = 0{,}50\pi_1 + 0{,}40\pi_2 + 0{,}15\pi_3$$
$$\pi_3 = 0{,}15\pi_1 + 0{,}20\pi_2 + 0{,}15\pi_3$$

Replacing the third equation by the condition $1 = \pi_1 + \pi_2 + \pi_3$, we obtain the steady-state probabilities $\pi_1 = 0{,}43, \pi_2 = 0{,}4$ and $\pi_3 = 0{,}17$.

Long-run behaviour of the Markov chain is given by matrix from equation (8):

$$\lim_{n \to \infty} P^n = \begin{bmatrix} 0{,}43 & 0{,}40 & 0{,}17 \\ 0{,}43 & 0{,}40 & 0{,}17 \\ 0{,}43 & 0{,}40 & 0{,}17 \end{bmatrix}$$

Computing several n-step transition of the P^n matrix, steady state is reached (to three decimal places – 0,1% mistake) after only 5 transitions (weeks).
Now we obtain the expected external demand E(d):

$$E(d) = \sum_{i=1}^{3} \pi_i X_i = 0{,}43 * 100 + 0{,}4 * 400 + 0{,}17 * 700 = 322 \text{ Pieces}$$

The input matrix H is based on the product structure tree of the radial ventilator and the lead time matrix $\widetilde{\tau}$ is based on the lead times τ_i. The product $\widetilde{H} = H\widetilde{\tau}$ is defined as the generalized input matrix. Order quantity diagonal matrix Q and initial available inventory vector $R(0)$ are based on order quantities Q_i and initial available inventories R_i.

$$H = \begin{array}{c} \\ A \\ B \\ F \\ E \\ D \\ G \\ C \end{array} \begin{array}{c} \begin{matrix} A & B & F & E & D & G & C \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 4 & 8 & 6 & 1 & 4 & 0 \end{bmatrix} \end{array}, \quad \widetilde{\tau} = \begin{bmatrix} e^{s} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{2s} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{3s} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{2s} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{2s} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{s} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & e^{s} \end{bmatrix}, \quad R(0) = \begin{bmatrix} 450 \\ 1000 \\ 1300 \\ 1200 \\ 2000 \\ 5000 \\ 21500 \end{bmatrix}$$

$$\widetilde{H} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{s} & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{s} & 0 & 0 & 0 & 0 & 0 & 0 \\ e^{s} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{2s} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4e^{3s} & 0 & 0 & 0 & 0 \\ 0 & 4e^{2s} & 8e^{3s} & 6e^{2s} & 1e^{2s} & 4e^{s} & 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 600 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1100 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1650 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1300 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2250 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4100 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 23000 \end{bmatrix}$$

Considering only request of final products which happen once a week, the external demand vector $\widetilde{d}$ is according to [7] given as:

$$\widetilde{d} = \begin{bmatrix} 322 * (e^{-s} + e^{-2s} + e^{-3s} + \cdots) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 322 * \frac{1}{e^{s} - 1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Because of limited space we give here according to (4) only a numerical example for cumulate number of setups $K_i(t)$ for component G, i.e. the sixth component - legs of frame, to fifth week $K_6(5)$. It is generated as the nonnegative integer providing a nonnegative inventory according to $L^{-1}\{\widetilde{R}(s)\} \geq 0$:

$$K_6(5) = \left[Q_6^{-1} L^{-1} \left\{ \frac{\sum_{j=1}^{6-1} H_{6j} \widetilde{\tau}_j Q_j \widetilde{\kappa}_j}{s} + \left(\frac{\widetilde{d}_6 - R_6(0)}{s} \right) \right\} \right] = \left[\frac{1}{4100} L^{-1} \left\{ 4e^{3s} 1650 \frac{K_3(8)}{s} e^{-8s} + \frac{0 - 5000}{s} \right\} \right] =$$

$$K_6(5) = \left[\frac{1}{4100} L^{-1} \left\{ \frac{13200}{s} e^{-5s} - \frac{5000}{s} \right\} \right] = \left[\frac{8200}{4100} \right] = 2$$

Cumulate number of setups and component stocks for eight weeks shows Appendix1.

5. CONCLUSIONS

In this paper we incorporated stochastic external demand in the MRP model using the Markov chains theory. Important concept is application of input-output analysis in the MRP models, which has been investigated by Grubbström and Ovrin. In order to analyze the dynamic behavior of the multi-level production system, we used the Laplace transform theory in continuous time models treated by Grubbström and Molinder.

A numerical example is shown based on structure A – a final product which represents a radial ventilator (Figure 1) illustrating the simplified model proposed above. The traditional MRP model with deterministic external demand with one known state in each week we upgraded with stochastic external demand with three states (three purchasers) with transition probabilities shown in transition probability matrix P using the Markov chains theory which is normally more close to a real situation. To find the expected external demand $E(d)$ after a long time we solved the equations (9) for steady-state distribution. The nonnegativity of available inventory $L^{-1}\{\tilde{R}(s)\} \geq 0$ constituted the major set of constrains in a related optimization problem. Given fixed order quantities the decision variable was the cumulate number of setups $K_i(t)$ of item i in time t, which was generated as the nonnegative integer providing a nonnegative available inventory. Cumulate number of setups and component stocks for eight weeks are represented in Appendix 1.

6. REFERENCES

[1] Andersson, L. E.: An introduction to Laplace transformation and applications, Input-output analysis and Laplace transforms and material requirements planning: symposium, Storlien, FPP Portorož & Linköping Institute of Technology, Portorož, 1998, pp 16-31

[2] Grubbström, R. W., Ovrin, P.: Intertemporal generalization of the relationship between material requirements planning and input-output analysis, International Journal of Production Economics, Vol 26, 1992, pp 311-318

[3] Grubbström, R. W., Molinder, A.: Further Theoretical Considerations on the Relationship between MRP, Input-Output Analysis and Multi-Echelon Inventory Systems, International Journal of Production Economics, Vol 35, 1994, pp 299-311

[4] Grubbström, R. W., Tang, O.: Early writings related to the topic of MRP and input-output analysis, and Laplace transforms and material requirements planning: symposium, Storlien, FPP Portorož & Linköping Institute of Technology, Portorož, 1998, pp 33-44

[5] Horvat, L.: MRP, input-output analiza in večnivojski sistemi proizvodnje in zalog, Slovenska ekonomska revija, Ljubljana, letnik 45, št. 4, 1994, 336-349

[6] Orliky, J. A.: Material Requirements Planning, McGraw-Hill, New York, 1975

[7] Peterle, P.: Vključitev teorije Markovskih verig v model planiranja materialnih potreb, magistrsko delo, Ekonomska fakulteta, Ljubljana, 2002

[8] Vazsonyi, A.: Scientific programming in business and industry, John Wiley, New York, 1958

[9] Winston, W. L.: Operations research: applications and algorithms, Duxbury, Belmont, 1994.

Appendix 1: Numerical example based on structure A in Figure 1

Week (t)	0	1	2	3	4	5	6	7	8
Item A									
Expected external demand $E(d)$		322	322	322	322	322	322	322	322
Available inventory R	450	128	406	84	362	40	318	596	274
Delivery Q		0	600	0	600	0	600	600	0
Net requirement		0	194	0	238	0	282	4	0
Setups κ		0	1	0	1	0	1	1	0
Cumulate setups K		0	1	1	2	2	3	4	4
Order start Q		600	0	600	0	600	600	0	600
Item B (1xItem A)									
Internal demand		600	0	600	0	600	600	0	600
Available inventory R	1000	400	400	900	900	300	800	800	200
Delivery Q		0	0	1100	0	0	1100	0	0
Net requirement		0	0	200	0	0	300	0	0
Setups κ		0	0	1	0	0	1	0	0
Cumulate setups K		0	0	1	1	1	2	2	2
Order start Q		1100	0	0	1100	0	0	0	1100
Item F (1x Item A)									
Internal demand		600	0	600	0	600	600	0	**600**
Available inventory R	1300	700	700	100	100	1150	550	550	**1600**
Delivery Q		0	0	0	0	1650	0	0	**1650**
Net requirement		0	0	0	0	500	0	0	**50**
Setups κ		0	0	0	0	1	0	0	**1**
Cumulate setups K		0	0	0	0	1	1	1	**2**
Order start Q		0	1650	0	0	1650	0	0	**0**
Item E (1x Item A)									
Internal demand		600	0	600	0	600	600	0	600
Available inventory R	1200	600	600	0	0	700	100	100	800
Delivery Q		0	0	0	0	1300	0	0	1300
Net requirement		0	0	0	0	600	0	0	500
Setups κ		0	0	0	0	1	0	0	1
Cumulate setups K		0	0	0	0	1	1	1	2
Order start Q		0	0	1300	0	0	1300	0	0
Item D (1x Item B)									
Internal demand		1100	0	0	1100	0	0	0	1100
Available inventory R	2000	900	900	900	2050	2050	2050	2050	950
Delivery Q		0	0	0	2250	0	0	0	0
Net requirement		0	0	0	200	0	0	0	0
Setups κ		0	0	0	1	0	0	0	0
Cumulate setups K		0	0	0	1	1	1	1	1
Order start Q		0	2250	0	0	0	0	0	0
Item G (4xItem F)									
Internal demand		0	6600	0	0	**6600**	0	0	0
Available inventory R	5000	5000	2500	2500	2500	**0**	0	0	0
Delivery Q		0	4100	0	0	**4100**	0	0	0
Net requirement		0	1600	0	0	**4100**	0	0	0
Setups κ		0	1	0	0	**1**	0	0	0
Cumulate setups K		0	1	1	1	**2**	2	2	2
Order start Q		4100	0	0	4100	**0**	0	0	0
Item C (4xItem B + 1xItem D + 6xItem E + 8xItem F + 4xItem G)									
Internal demand		20800	15450	7800	20800	13200	7800	0	4400
Available inventory R	21500	700	8250	450	2650	12450	4650	4650	250
Delivery Q		0	23000	0	23000	23000	0	0	0
Net requirement		0	14750	0	20350	10550	0	0	0
Setups κ		0	1	0	1	1	0	0	0
Cumulate setups K		0	1	1	2	3	3	3	3
Order start Q		23000	0	23000	23000	0	0	0	0

IMPROVEMENT DESIGN AND MANAGEMENT OF WOOD PROCESSING MANUFACTURE SYSTEMS

Lyudmyla Zahvoyska
Department of Ecological Economics, National University of Forestry and Wood Technology, Gen Chuprynky Str., 103, 79057, Lviv, Ukraine.
E-mail: zld@forest.lviv.ua

Abstract
Production and consumption reshaping according to the needs of sustainable development force us towards stepwise and rational use of all kind of production resources, including material and energy flows. In this sense production systems' operation time is a unique and non-renewable resource and its losses become an aggregated indicator of production system efficiency. Approaches, methods and models of production logistics which lead to diminished time losses in wood processing manufacturing systems are highlighted in the paper.

Key words: eco-efficiency, manufacturing systems, transfer lines management, operation time losses, stochastic simulation models, analytical models, optimal performance.

1. Introduction

Time is unique and non-renewable resource not only for some persons but also to all mankind. So that, its rational use in all fields of human activity is of great importance, especially in industrial production. To estimate efficiency of firms' industrial activity, namely manufacturing systems (MS), we use such indicators as system capacity, prime cost of production, income etc. These indicators strongly depend on level of MS' time use. Rapid development of modern facility productivity induces simultaneous price increasing. Losses of time that occur in MS involve losses of capacity, which increase together with cost of MS. Hence, the important task of production logistics optimisation, especially in case of pulsation production (De Cristofaro, 2003), we face now, is reduction or even elimination losses of time, which arise in multistage transfer line systems because of mutual equipment blocking, mutability of inputs, failures of operators and equipment. All these perturbations particularly acute emerge in wood processing MS.

Stochastic simulation models are commonly used for MS investigation but this multi-purpose instrument of operations research needs sound soft, relevant computer capacity and is time consuming. From this point of view analytical models seem to be faster and more convenient for practical application. But the way of their generation, verification and validation is not very straightforward and comprehensive.

In this contribution we highlight the results of our exploration of analytical models, which in proper way reflect dependence MS productivity on their structure, stability, reliability, composite elements' characteristics, and linking way. Models of queuing theory, which were used as a benchmark in our research, were improved and developed by means of stochastic simulation and compared with relevant results, obtained by other scientists, and real industrial MS.

2. Approaches to MS Parameters/Performance Modelling

Performance of MS strongly depends on a set of perturbing effects, such as equipment reliability, adequacy of operators' knowledge and ability, internal correspondence of successive and parallel departments' productivity. Minor divergence from prescribed level could lead to significant shutdown losses. Common circumstance, which causes them, is an equipment failure. How to run MS to minimise the time losses is question of critical importance to modern entrepreneurs.

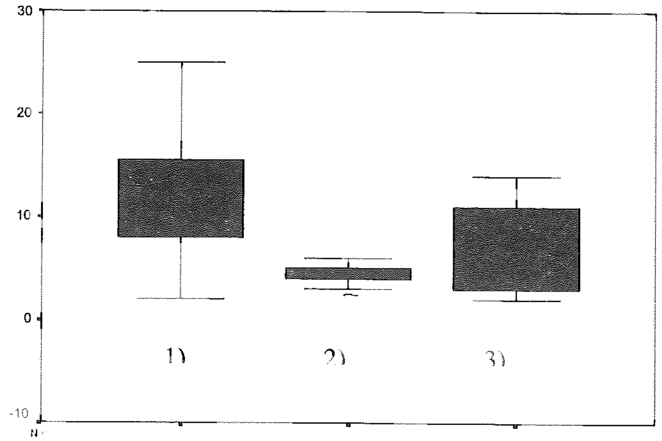


Fig.1. Box-Whisker Plot of wood processing operations
1. Veneer cutting; 2. Scarf joint;
3. Details calibration

Some types of MS, for instance wood processing manufacturing systems, are especially exposed to such stochastic perturbations due to biological origin of processed inputs, which are wooden details. Ontogenesis and phylogenies of wooden details being processed, changeability of their physical and chemical characteristics bring strong variability into operation duration (Fig. 1). This peculiarity noticeably differs production logistics of wood processing enterprises, makes it more complicated, production recourse use less efficient, and task of such systems parameters optimisation more evident.

The first significant contribution to MS problems investigation made K. Palm and P. Finch in 1947. They proposed to describe duration of technological operation by exponential probability distribution. Since then investigations of MS, especially in engineering industry, deal with Markov chain process models. R.Jackson (1954), G.Gantt (1956), E.Koenigsberg (Koenigsberg, 1956) proposed a new model of transfer line employing queuing theory. But these approaches did not find an appropriate application because of two reasons: they were not enough powerful and gave only approximate results. They fail to generate meaningful analytic solutions for non-Poisson tandem systems like $E_k/E_k/M$, which real wood processing MS are. That is why we use the models of queuing theory as a benchmark in our investigation of MS performance.

Stochastic simulation is commonly used in MS problems investigation due to its ability to incorporate in models different sources of perturbances. This kind of modelling enables investigator to make direct experiments with model of real object and let make the model as precise as investigator would like. But the process of simulation model generation, verification and validation tend to be quite sophisticated. It takes a lot of time even for experienced in simulation (Sargent, 2001).

It is known that simulation is good for didactic (teaching) purposes. Simulation not only helps acquire a good understanding of overall systems, it also gives good reason for that. It is able to provide us with missing information for different situations and case studies. It may illustrate the model development process, or allow one to better comprehend process in question, or better understand relationships within decision-making setting and to develop students' decision-making skills. Simulation tends to be relatively straightforward. Non-technical decision-makers and students prefer them especially when compare them with

sophisticated analytical methods (Budnick, McLeavey, Mojena, 1988). It is more easily to understand the modeling process and obtained results.

Hence, following the idea of lean production and high profit, modern entrepreneur needs quicker analytical models, which deals properly with types of MS, operations time, buffers capacity, and machines reliability (Gershwin, 1994). In table 1 we provide information about availability of models (both analytical (A) and empirical (E)) for transfer lines' performance optimisation.

Table 1: **Availability of Models for Transfer Lines' Performance Optimisation**

Number of machines (n) and their parameters: • Erlangian parameter (K) • Productivities ratio (μ)			**Transfer line machinery composition and rhythm**				
			Serial Composition		**Parallel Composition**		
			Without buffer	With buffer	Without buffer		With buffer
Number of machines, n	Erlangian parameter, K	Productivity ratio	Single rhythm	Individual rhythm	Single rhythm	Individual rhythm	
1	*1*		L/A			$T_l/A, E$ $Stages \le 5$	
	≥ 1		$L; T_L/E$				
2	*1*	*Equal*	$T_L; B/A$				
		No equal	$T_L; \mu; B/A$				
	≥ 1	*Equal*	$T_L; B/E$				
		No equal	$T_L; \mu; B/E$				
≥ 3	*1*	*Equal*	T_l/A	$T_L; \mu; B/E$ $for\ n \le 500$		T_L/E	
	≥ 1		T_l/A			T_L/E	
		No equal		*Virtual Pairs Method* $for\ n \le 30$			

Notes:
Shape of the records in the table: Transfer Line Performance Characteristics / Type of Models.
Transfer Line Performance Characteristics: L = expected number of units in system.
T_L = additional time losses in transfer line.
μ = machines in line productivity ratio.
B = buffer capacity.

Type of Models: A = analytical; E = Empirical.
Models in the shadowed sells are ours.

As one can see from table 1, we consider three cases of lines length, when number of machines in transfer line (***n***) equals one, two or greater than two. Ratio (μ) of machines productivity and stability of technological operations, which is described by Erlangian parameter K ($K=t^2/S^2$, where t – means duration of operation and S^2 – variance of operation duration) are analyzed in our investigations, too. For cases, when K is approaching unity ($K \to 1$) duration of operations tends to be a stochastic variable. We can see it in such wood processing operations as veneer cutting, scarf joint, details calibration, etc. If Erlangian parameter K is approaching infinity ($K \to \infty$) duration of operations tends to be a deterministic variable. This situation is typical for transfer lines, especially in engineering. As one can easily see from table 1, the most of indicated types of transfer lines there are analytical or empirical models.

Time losses unambiguously determine transfer line capacity that depends on its structure, stability of operations, machines reliability, composite elements' characteristics, facility layout (Zahvoyska, 1999, Zahvoyska, 2003). Examining transfer lines performance we marked out two types of transfer lines: lines with finite buffers between them for details storing and with direct connections between working machines (without buffers). Each of these two types of transfer lines can work in two different conditions. The first of them we call single rhythm line. In this case all machines of transfer line begin their technological operations simultaneously. Finishing their own operations all machines are waiting for the longest operation completion and only after that all of them begin the next operation. The second type of conditions is free individual rhythm of transfer line. In this case each machine can start its operation independently at any time whenever the buffer before it is not empty (Okamura and Yamashina, 1983.

Transfer lines as the main elements of MS belong to the real time systems for which time is the most valuable and non-renewable resource. Additional operation time losses induce chain of wasted resources: first of all energy, material and financial flows, overused of machinery system resource, and last but not least losses of labour. From this perspective operation time losses become an aggregated indicator of production system efficiency. Their diminishing is generally viewed as one of the most important tasks of MS management.

3. Results of Investigations

For all shadowed cells from table 1, which represent specific types of transfer lines, we developed event oriented stochastic simulation models (Zahvoyska, 2003), which provide us with missing information about time losses in MS in questions.

Further statistical elaboration of generated results gives us empirical models aimed to optimise design and management of transfer lines. In particular one can estimate value of operations time losses, the optimal productivities ratio and the optimal buffer capacity (table 2). Besides we propose to analyze transfer line performance with different machines productivity (when number of machines is less then 30) using virtual pairs method (Dudyuk, Zahvoyska, 2003). Following our investigation this method gives the most reliable information for such complex cases.

Using models, introduced in table 2, it is possible to estimate value of time losses coefficients for more than 20 types of MS, to find dependencies for determination values of the optimal machinery loading, the optimal productivities ratio, and the optimal buffer capacity for the different types of transfer lines.

Engineers, designers or MS managers can predict or assess time losses value for different types of transfer lines and make proper decision concerning their design or management. For instance, using optimising techniques one can find the optimal lines' parameters in question, like buffer availability and capacity, way of machinery composing or other relevant matters which production logistics deals with. In case of transfer lines the time losses elimination means the production of some given output level using a minimum amount

of resources (machinery, energy, and labor flows). Such models let engineers design efficient transfer lines with minimal losses of energy and material flows.

From other point of view they enable managers to use or to redesign MS in a proper way. Such improvements make systems in question leaner and nearer to eco-efficient, adequate to the needs of new concept of technological progress.

Table 2: Assessment of Additional Time Losses in Transfer Lines (Serial Composition)

Number of machines and their parameters			Transfer line machinery composition and rhythm		
			Without buffer		With buffer
n	K	μ	Single rhythm	Individual rhythm	
1	1		$L_1 = T_L^{-1} - 1, \quad L_{q1} = T_L + T_L^{-1} - 2; \; T_L = 1 - \lambda/\mu_1$		
1	≥ 1		$L_{qK} = L_{ql}(l + K_\mu^{-l})/2, \quad K_\lambda = l;$	$L_{qK} = L_{ql}(K_\lambda^{-l} + K_\mu^{-l})/2 \approx L_{ql}/K$	
2	1	=	$T_L = (B+3)^{-1}; \; B = B_2, \; B_l \to \infty$	$B = B_2 - l, \; B_l = B_2$	
2	1	≠	$T_{L1} = \mu^{B+2} T_{L2}; \quad T_{L2} = (1-\mu):(1-\mu^{B+3}); \; \mu = \mu_l / \mu_2$		
2	≥ 1	=	$T_L \approx \left(KB + \sqrt{\pi K} + 1\right)^{-1}, \quad K = 2/\left(K_l^{-l} + K_2^{-l}\right)$		
2	≥ 1	≠	$T_{L1} = \mu^{1/T_L - 1} T_{L2}, \; T_{L2} = (1-\mu)/(1-\mu^{1/T_L}),$ $K = (l+\mu)^2 / 2\left(K_l^{-l} + K_2^{-l}\mu^2\right)$		
≥ 3	1	=	$T_L = 1 - \left(\sum_{1,n} i^{-1}\right)^{-1}$	$T_L \approx (1{,}900 - 1{,}800/n)/(B+3), \; n \leq 500$	
≥ 3	≥ 1	=	$T_L \approx \left(\sqrt{K} / \sum_{2,n} i^{-1} + 1\right)^{-1}$	$T_L \approx \dfrac{1{,}900 - 1{,}800/n}{KB + \sqrt{\pi K} + 1}, \; K = \left[n \sum_{i=1}^{n} \left(t_{ri}^2 / K_i\right)\right]^{-1}$	
≥ 3	≥ 1	≠	-	Virtual Pairs Method, $n \leq 30$, $p = \sqrt[3]{n/2}\left(T_L^{-1} - 1\right)$	

Notes:

Shape of the records in the table: Transfer Line Performance Characteristics / Type of Models.

Transfer Line Performance Characteristics: L_q = expected number of units in queue.

λ = arrival rate.

μ_i = service rate.

4. Conclusions

Quoting P.Hennicke, it is worthwhile to stress, that modern environmental policy becomes aware that "the very quantities alone of energy used and substances moved as well as increased land use create a problem. Every product and every service is linked throughout its entire life cycle with energy and material throughputs" (Hennicke, 2003). Reshaping physical exchange of anthroposphere and the environment in eco-efficient way society will encourage environmentally benign technological progress. Sustainable enterprises will be one of the numerous steps in this way. And we hope the proposed models will be useful instrument to get this new strategy of development implemented.

References

1. Budnick, F., McLeavey, D., Mojena, R., 1988. Principles of Operations Research for Management. Irwin, Illinois.
2. De Cristofaro, E, 2003. The Efficiency Revolution. In: Proceedings of the 5th Int-l Conference on Ethics and Environmental Polices "Business Styles and Sustainable Development".
3. Dudyuk, D., Zahvoyska, L., 2003. Estimation and Elimination of Time Losses in Wood Processing Manufacturing Systems. IZMN, Kyiv, (Ukr.).
4. Gershwin, S.B. Manufacturing Systems Engineering, 1994. PTR Prentice Hall, New Jersey.
5. Koenigsberg, E. Production Lines and Internal Storage // Management Science. – 1956. – 5. – 410 p.
6. Okamura, K., Yamashina, H., 1983. Analysis of In-Process Buffers for Multi-Stage Transfer Line Systems // Int. J. Prod. Res. Vol. 21. No 2, p.183-195.
7. Hennicke, P., 2003. Decoupling Well-being from GDP. – Towards a New Kind of Technological Progress. In: Proc. of 5-th International Conference on Ethics and Environmental Polices "Business Styles and Sustainable Development. Kyiv, April 2nd-6th, 2003.
8. Sargent, R., 2001. Some Approaches and Paradigms for Verifying and Validating Simulation Models Intelligence // Proc. of the 2001 Winter Simulation Conference / B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer (Eds.). – P. 106-114.
9. Zahvoyska, L., 1999. Analysis of Multi-Stage Transfer Line Systems using Simulation. In: From Theory to Practice. Gaps and Solutions in Managerial Economics and Accounting in Forestry. Sisak L., Joebstl H., Merlo M. (Eds). Proc. of the Symposium in Prague. May 13-15, 1999. – P. 239-243.
10. Zahvoyska, L., 2003. Towards Eco-Efficiency: Operations research in wood-processing enterprises management. In: Proc. of the 7th International Symposium on Operational Research in Slovenia. L.Zadnik-Stirn, M.Bastic and S.Drobne (Eds.). – Podcetrtek, Slovenia, September 24-26, MIGRAF, Ljubljana. – P. 209-214.

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Section 10
Education and Statistics

TESTS OF NO ASSOCIATION BETWEEN VARIABLES - COMPONENTS OF CONSUMER CONFIDENCE INDICATOR AND STRATIFICATION VARIABLES FOR CROATIAN CONSUMER SURVEY

Vlasta Bahovec
Mirjana Čižmešija
University of Zagreb - Graduate School of Economics & Business
Trg J. F. Kennedya 6
10000 Zagreb, Croatia
bahovec@efzg.hr
mcizmesija@efzg.hr

Abstract

Consumer survey is an important tool for tracking and predicting changes in private consumption and in a whole country economy. Like a business survey, this is a qualitative economic survey. Such surveys are intended for short-term economic analysis, especially for predicting changes in the economic cycles. They are important because survey results of economic variables are available before the same or the similar data from the official statistics is published. Time series of variable balances and time series of Consumer Confidence Indicator for Croatia are too short for these predictions. However, the survey results can be used to find out whether there is no association between component variables of CCI and variables: region, settlement size, household size and children (households with or without children under 14).

Keywords: CCI – Consumer Confidence Indicator, Pearson Chi square Test, ML Chi Square Test, Phi Coefficient, Contingency Coefficient, Cramer's V

Introduction

Consumer surveys are conducted in all EU countries[1]. It can be seen that business and consumer surveys are also carried out in the future member states like Croatia. This is important for the integration of Croatian business and consumer surveys in The Joint Harmonised EU programme of these surveys, and for the integration of Croatia into the EU. The results of consumer survey are of the best quality if consumer opinions in the whole country about variables in consumer survey are independent of stratification variables. The aim of this paper is to conduct tests of no association consumer survey variables and stratification variables for Croatia.

Consumer survey and Consumer Confidence Indicator - CCI

Consumer surveys are conducted on a monthly or quarterly basis, by means of questionnaires sent by post or as a telephone interview. The questionnaire has fifteen questions[2]. The most used survey sampling method is applying stratified sample with a random choice. The survey data are presented as "balance". Balance is defined as a difference between weighted

[1] The first harmonised consumer survey was started in May 1972 (The Commission of the European Communities), on the basis of a decision by the Permanent Representatives Committee, 1970.
[2] If consumer survey is conducted monthly, the number of questions is twelve, and three questions are asked quarterly.

percentages of positive and negative answers of consumers on corresponding variables.[3] Balance is calculated for all questions (variables). Questions are grouped as: an assessment of the present situation, an assessment of the year-on-year change and an expectation.

With the aim to summarise consumers´ subjective assessments of the economic and social trends in a country, Consumer Confidence indicator – CCI can be calculated. Consumer confidence indicator (CCI) is a composite indicator. This is a simple average of seasonally adjusted balances[4] of four variables:

(1) Financial position of households
(2) The general economic situation
(3) Unemployment expectations (with inverted sign)
(4) Saving over the next 12 months [5]

$$CCI = \frac{B_{FS} + B_{GES} + (-B_{UE}) + B_S}{4} \quad (1)$$

Where are:

CCI - Consumer Confidence Indicator
B_{FS} - Seasonally adjusted balances of *financial situation of households*,
B_{GES} - Seasonally adjusted balances of *the general economic situation*,
B_{UE} - Seasonally adjusted balances of *unemployment expectations*,
B_S - Seasonally adjusted balances of *expected savings*.

Qualitative results collected through the consumer survey are closely correlated with the quantitative statistical data. "The role of this data is to provide an initial advance indicator of development in certain areas of the economy. They are not a replacement for the quantitative data, which describe the actual situation." (...The joint harmonised.., 2004, p. 387). Survey results are used as such or summarised in synthetic, confidence indicators.

Each variable in business and consumer survey, such as expectation and confidence indicator, must be a short-term forecasting indicator for the corresponding variables. It is of interest here to track the direction of changes in a variable and in a referent series. This is important because the business and consumer survey data are usually available before the corresponding official statistics is published (Bahovec, V., Čižmešija, M., 2003).

Referent series in official statistics for the series CCI can be private consumption (growth percentage compared with that in the same quarter of the preceding year). Selected questions (variables) can be compared with inflation, unemployment, and so on. The highest correlation coefficients for most EU countries are between the CCI and private consumption (private final

[3] More detailed in ...The joint harmonised EU programme of business and consumer surveys, user guide. (updated: 26/05/2004). European Economy. European Commission, Directorate-General for Economic and financial affairs., p. 8-9

[4] Because series of variable-balances in Croatian Consumer Survey are short, balances are not seasonally adjusted.

[5] (1) «How do you think the financial position of your household will change over the next 12 months? (get a lot better, get a little better, stay the same, get a little worse, get a lot worse, don´t know)», (2) «How do you think the general economic situation in this country will develop over the next 12 months? (get a lot better, get a little better, stay the same, get a little worse, get a lot worse, don´t know)», (3) «How do you think the level of unemployment in the country will change over the next 12 months? (Will it: increase sharply, increase slightly, remain the same, fall slightly, fall sharply, don´t know)», (4) «Over the next 12 months, how likely are you to be able to save any money? (very likely, fairly likely, fairly unlikely, very unlikely, don´t know)».

consumption expenditure at constant prices). It means that the series of variable balances and CCI series can be used to predict changes in private consumption. This is especially useful for countries which do not have quarterly national accounts.

Conducting consumer survey is important for the calculating *ESI – Economic Sentiment Indicator* which is a composite indicator (more detailed: Nikić, G., Šošić, I., Čižmešija, M., 2004) and which is important for predicting changes in GDP and in a country's economy as a whole.[6] In our country the series CCI is too short to make this calculation.

Consumer survey in Croatia

In addition to the business surveys in Croatia, carried out since 1995, Centre for market research *Privredni vjesnik* started in the first quarter of 2003 conducting consumer surveys too. Like business surveys, consumer surveys are carried out on a quarterly basis. The results of the tests have been regularly published in «Privredni vjesnik». Survey is based on a panel of 620 households. The panel reflects regional structure of population (5 regions), size of the settlements, size of the households, and households with and without children below the age of 14. The test was carried out by the GfK branch in Zagreb.[7] The sample size in the last Consumer test conducted for the 2004, IV quarter is 600 households and is representative of the entire Republic of Croatia.

No association test results

The aim of this paper was to test a hypothesis of no association between CCI component-variables and four stratification variables:

1) Region: (1=Zagreb, 2= Middle Croatia, 3= North Croatia, 4= Slavonia, 5=Dalmatia)
2) The settlements size: (1="less then 1000", 2="1000-4999", 3="5000-9999", 4="10000-49999", 5="50000-99999", 6="more then 99999" residents)
3) The households size: (1=one member, 2=two members, 3=three members, 4=four members, 5= five members, 6= six and more members)
4) Children (1= households with children under 14, 2=households without children under 14)[8]

One of the most widely applied statistic tools for testing the presence of assotiation between two categorical variables in the contingency table is *Pearson Chi-square test*, as well as other types of *chi-square tests (likelihood-ratio chi square (ML- chi square), Mantel-Haenszel chi square)* and *Fisher exact test* (McClave, J. T., Benson, G. P., Sincich, T. (2005)). Th e measures of association evaluate the strength of the association between the variables.

The analysis of sample data was made using SAS computer package. In the SAS output the *Pearson* and *ML chi square test* results are selected[9], as well as measures of association between two categorical variables (*Phi Coefficient, Contingency Coefficient* and *Cramer's V*.)

[6] Referent series in official statistics for the series ESI is GDP -Gross Domestic Products.

[7] There are many European Countries in which GfK is the organization for conducting consumer surveys.

[8] On the basis of consumer survey results for 2004, IV quarter.

[9] Fisher exact test is a more appropriate test of no association than chi square test when the sample size is small. This test was not selected, because the computational time can be prohibitive, depending on the size of the table and the sample size.This test is not practical (in terms of CPU time) for $\frac{n}{(r-1)(c-1)} > 5$.

In the *Pearson Chi-square test* the null hypothesis (H_0) of no association between two row and column classification variables, and alternative hypothesis (H_1) which contradicts the null hypothesis are:

$$H_0 : p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \ \forall i, j,$$
$$H_1 : \exists p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j}, \qquad (2)$$

where $p_{ij}, p_{i\cdot}, p_{\cdot j}, \ i = 1,2,\cdots,r, \ j = 1,2,\cdots,c$ are probabilities (for cells ij, i=1,...,r; j=1,...,c) of two dimensional discrete random variable, marginal column and marginal row probabilities, respectively. If the components of two dimensional discrete random variables are independent, each two-dimensional probability equals product of marginal column and row probabilities.

The expected frequencies are the frequencies that would be expected under null hypothesis that there was no relationship between the variables. The *Pearson chi-square statistic* is computed as:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(m_{ij} - e_{ij})^2}{e_{ij}} \qquad (3)$$

where $m_{ij}, e_{ij}, \ i = 1,2,\cdots,r, \ j = 1,2,\cdots,c$ are empirical and expected frequencies for cell ij respectively. The expected frequencies are the frequencies that would be expected under null hypothesis that there was no relationship between the variables.

For preceding statistics to be valid, the expected number of observations in each cell should be at least 5. If this assumption is met, sampling distribution of the test value is approximately a *chi-square distribution* with (r-1)(c-1) degrees of freedom.

The null hypothesis will be rejected if the test value is greater than critical value at given significance level α or, equivalently if the empirical significance level (p-value) is less than theoretical significance level α.

Likelihood ratio (ML ratio) chi square is formulated in the same way as *Pearson chi-square*, but the computation of the test variable is based on the maximum likelihood method. So computed test value is usually very close in magnitude to *Pearson chi square*.

Coefficient Phi is a measure of the association between two categorical variables, defined as a following function of chi square statistics:

$$\Phi = \sqrt{\frac{\chi^2}{n}}, \qquad (4)$$

and takes nonnegative values. If the value of coefficient *Phi* is zero, the variables are not associated.

The contingency coefficient is a *chi square measure of association* between two variables measured on the nominal scale. Its range is [0,1] (where 0 means complete independence).

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}. \qquad (5)$$

Cramer's V is a measure of assotiation computed as a function of chi square statistics:

$$V = \sqrt{\frac{\chi^2}{n(L-1)}} \qquad L = \min(r,c) \qquad V \in [0,1]. \qquad (6)$$

For the contingency table of «region» by «financial position» the results of tests are summarized in the SAS printout (Table 1):

Table 1

Statistics for Table of region by financial position

Statistic	Value	Prob
Chi-Square	21.7352	0.1519
Likelihood Ratio Chi-Square	22.0999	0.1400
Phi Coefficient	0.1903	
Contingency Coefficient	0.1870	
Cramer's V	0.0952	

Chi square test value is 21.7352 and the related empirical significance level *(p-value)* is 0.1519. The null hypotheses is accepted, since the *p-value* is greater than the theoretical significance level α= 0.05.The sample data lead to the conclusion that the households in different regions have similar statements of their financial position.

In this example the *ML chi square* is 22.0999 and has p-value 0.1400. This result leads to the same conclusion as *Chi square test*. All measures of association have small values, and it can be concluded that there is no association between two considered variables. The same conclusion at the 5% significance level is made for all other contingence tables under investigation. The statistics are summarized in Table 2 and in Table 3.

Table 2

Statistics	Value	Prob	Statistics	Value	Prob
Table of region by financial position			Table of settlements by financial position		
Chi-Square	21.7352	0.1519	Chi-Square	17.2012	0.3727
L R Chi-Square	22.0999	0.1400	L R Chi-Square	17.1559	0.3756
Phi Coefficient	0.1903		Phi Coefficient	0.1693	
Contingency Coeff.	0.1870		Contingency Coeff.	0.1669	
Cramer's V	0.0952		Cramer's V	0.0847	
Table of household size by financial position			Table of children by financial position		
Chi-Square	12.0228	0.7424	Chi-Square	2.3510	0.6715
L R Chi-Square	12.1409	0.7342	L R Chi-Square	2.3478	0.6721
Phi Coefficient	0.1416		Phi Coefficient	0.6721	
Contingency Coeff.	0.1402		Contingency Coeff.	0.6721	
Cramer's V	0.0708		Cramer's V	0.0626	
Table of region by economic situation			Table of settlements by economic situation		
Chi-Square	18.9353	0.2720	Chi-Square	24.2268	0.0847
L R Chi-Square	19.6319	0.2373	L R Chi-Square	24.1620	0.0860
Phi Coefficient	0.1776		Phi Coefficient	0.2009	
Contingency Coeff.	0.1749		Contingency Coeff.	0.1970	
Cramer's V	0.0888		Cramer's V	0.1005	
Table of household size by economic situation			Table of children by economic situation		
Chi-Square	19.8366	0.2277	Chi-Square	3.9787	0.4089
L R Chi-Square	19.3238	0.2522	L R Chi-Square	4.0672	0.3970
Phi Coefficient	0.1818		Phi Coefficient	0.0814	
Contingency Coeff.	0.1789		Contingency Coeff.	0.0812	
Cramer's V	0.0909		Cramer's V	0.0814	
Table of region by unemployment			Table of settlements by unemployment		
Chi-Square	14.9567	0.5278	Chi-Square	19.9108	0.2243
L R Chi-Square	15.0362	0.5220	L R Chi-Square	17.1559	0.2477
Phi Coefficient	0.1579		Phi Coefficient	0.1822	
Contingency Coeff.	0.1560		Contingency Coeff.	0.1792	
Cramer's V	0.0789		Cramer's V	0.0911	

Table 3

Statistics	Value	Prob	Statistics	Value	Prob
Table of household size by unemployment			Table of children by unemployment		
Chi-Square	0.0911	0.4840	Chi-Square	0.9190	0.9218
L R Chi-Square	0.0911	0.5141	L R Chi-Square	0.9221	0.9214
Phi Coefficient	0.1610		Phi Coefficient	0.0391	
Contingency Coeff.	0.1590		Contingency Coeff.	0.0391	
Cramer's V	0.0805		Cramer's V	0.0391	
Table of region by saving money			Table of settlements by saving money		
Chi-Square	13.0088	0.3684	Chi-Square	18.5294	0.2359
L R Chi-Square	12.7383	0.3883	L R Chi-Square	19.2771	0.2015
Phi Coefficient	0.1472		Phi Coefficient	0.1757	
Contingency Coeff.	0.1457		Contingency Coeff.	0.1731	
Cramer's V	0.0850		Cramer's V	0.1015	
Table of household size by saving money			Table of children by saving money		
Chi-Square	8.5885	0.7376	Chi-Square	7.7301	0.1020
L R Chi-Square	8.8225	0.7180	L R Chi-Square	8.3738	0.0788
Phi Coefficient	0.1196		Phi Coefficient	0.1135	
Contingency Coeff.	0.1188		Contingency Coeff.	0.1128	
Cramer's V	0.0691		Cramer's V	0.1135	

Conclusion

The sample data lead to the conclusion that at 5% significance level there is no association between component-variables of CCI and stratification variables: region, settlement size, household size and children (households with or without children under 14). The same results are conducted for the tests of no association between other variables (no components of CCI) in consumer test and stratification variables. It means that all consumers in Croatia have the similar assessments of the economic and social trends. Therefore, we believe the consumer survey has to be continued and the time series of the survey results (especially CCI) should be used for observing, following, explaining and forecasting changes in the private consumption (as a referent series) and for changes in Croatia's economy.

References

[1] Bahovec, V., Čižmešija, M. (2003). *Some experience in correspondence of factors selected by factor analysis of business survey results and the industrial production in Croatia.* SOR Proceedings (6 p.). The 7th international symposium on operational research in Slovenia. Podčetrtek, 24-26. September 2003.

[2] McClave, J. T., Benson, G. P., Sincich, T. (2005). *Statistics for business and Economics*, 9th edition. Prentice Hall. New York

[3] Nikić, G., Šošić, I., Čižmešija, M. (2004). *Estimate of ESI in Croatia – a case study of an economy in transition.* 27th CIRET Conference, Warsaw, September 2004. http://www.ciret.org/user/warsaw2004

[4] ...*The joint harmonised EU programme of business and consumer surveys*, User guide (2004). European Economy. European Commission, Directorate-General for Economic and financial affairs.

[5].................Program support SAS

MODEL OF BASIC TRAINING FOR INFORMATION SYSTEM DEVELOPERS

Mile Pavlić, Marija Marinović, Sanja Čandrlić
Faculty of Philosophy, University of Rijeka, Rijeka, Croatia
mpavlic@ffri.hr, marinm@ffri.hr, sanjac@ffri.hr

Abstract:
The paper presents the training model for information system designers, providing training on basic knowledge in business processes analyses and data modeling, reading models and participation in project teams. The training model is used with information technology professionals actively included in information system development.

Keywords: information system, design, methodic, training

1. Introduction

Learning methods for business system analysis and data modeling is an education process for training information system designers, aiming to produce a business system model, which can be used as a prototype for building an information system (IS). Teaching models have been improved during past twenty years by the authors due to the following reasons: improving of methodic and methods, emphasizing particular didactic approaches and practical work in education of information system designers.

Over time, several various training models have been developed resulting in the one-year training model, weekly training model and daily training model. The objective of one-year training is to acquire design skills, while in the daily training model the objective is to make students familiar with methods and provide guidelines for further education.

The objective of the basic training for designers (or so called weekly training) is to, during six working days, present methods and techniques of information systems design, with enough details, however, with an abridged contents, yet not omitting relevant modeling concepts, so that students, after completing their lessons, are able to read models and take part in design and IS development teams. Training is based on various didactic theories, with prominent role of didactic as a curriculum theory, i.e., objective approach [3].

Training of information system designers includes the implementation of various working models. The training was introduced in 1984 and extensive experience has been gained since. A total of 16 one days seminars, 24 weekly courses and 14 one-year studies for designers has been performed. Eleven generations of full time students were included (approximately 30 students per year), totaling to over 1000 students [9].

The paper presents experiences gained in education of information technology (IT) professionals working in IT centers in various companies (andragogy education in design school through seminars) and IT science students. The content is represented by standard process modeling methods defined in the MIRIS methodic.

2. Information system design methodology

The basic problem in modeling is to find a precise enough system (method) for recording relevant knowledge. In drawing up various types of diagrams using various methods, basic principles of organization are implemented, such as: classification, aggregation, generalization and hierarchy [7].

The methodic prescribes phases of development and activities within a particular phase up to the required extent of details, defines relations between particular activities and sequence of executing activities. The basic task of methodic is to define methods to be used by information system designers in building a system model. The methodic defines the connectivity of the models at the system design level and, on the base of these, the procedures for producing software as a part of a future IS.

Various methodic solve the problem of designing and building an IS by defining phases of the processes and implementing certain methods in these. The schedule of information system life cycle phases mostly differs for particular methodic (cascade, radical, prototype, «V» approach as a version of the cascade and radical approach and others). All of these have life cycle phases, but put forward some own features: methods, tools, design swiftness, implemented measurement, improvements and adaptations.

The basic issues included in life cycle phases, according to the cascade model, are as follows: design, realization and maintenance [2] or analysis, design and implementation [11] with overlapping activities of some phases and evaluation of the system during maintenance phase.

In more details, the breakdown of life cycle into phases is represented by: strategy – why (problem definition), analysis – what (analysis of a problem), design – how (working out a solution), building, i.e., programming – production (making an IS), testing – quality assessment (check), introducing – the first implementation with testing and maintenance [14], [12], [6].

A methodic is said to be well defined if it: decomposes life cycle phases into activities, prescribes method and output models for each activity, describes relations among particular activities, i.e., input models and information required for modeling, determines conditions for starting and finishing the modeling, prescribes quality of output documentation and it is proved efficient in real life projects.

3. The MIRIS methodic life cycle

The MIRIS methodic (acronym from Croatian **M**etod**I**ka za **R**azvoj **I**nformacijskog **S**ustava = Methodic for Development of Information System) is a set of methods and activities in time, where objectives are designing and building an information system (IS). It has been developed since 1984 and it was published in 1995 [10]. The MIRIS methodic was developed through expansion of and learning from other methodologies and testing methods in research projects [8]. It was tested in approximately hundred projects of various sizes and it was improved on the fly.

The fundamental hypothesis on which the methodic has been developed is the "system decomposition" and it is as follows:

- Dividing life cycle in three phases.
- In the first phase, producing an abstract model of the entire system, and dividing system into subsystems.
- In the second phase, producing a model of the subsystem and setting appropriate method for process modeling.
- In the third phase, modeling data for relevant processes and defining application architecture depending on the process model and data model.

Unlike this approach, it is possible to assume «decomposition of the design process».

It is quite natural to answer what to do first and what to do later, so that the hypothesis that life cycle is to be ordered according to the activities execution sequence has become the «actual» assumption in most of the methodic [5], [4].

The life cycle suggested by the MIRIS methodic is quite original. It resembles to the «V» model, so that on the left side there are phases in which a business system is decomposed, while on the right side, linear development and introduction of application is proposed regardless of respective complexity.

The basic concepts of the methodic are as follow: methods, phases, projects, forms and team members. The structure model shows their mutual relations and connectivity.

The MIRIS methodic uses three basic methods, namely: data modeling method, process modeling method and application modeling method. These methods are adapted and expanded following the requirements and goals set and have original concepts and original process modeling. The APP method (in Croatian: method for modeling "***Arhitekture Programskog Proizvoda***" = Software Product Architectures) is an original method leaning on DTP and ER methods.

Phases of the life cycle in the MIRIS methodic are divided in two groups of phases: logical modeling (IS design) shown in Table 1 and physical modeling (building IS) shown in Table 2. Each group has three phases, so there are six phases in total (in previous version there were seven phases defined). Phases are further divided in activities.

Table 1: Logical modeling - design

Phase 1: STRATEGIC PLANNING OF IS (SP)
1.1. Analysis: Defining and training a team, decomposition of a process, list of documentation and navigation through the system
1.2. Subsystems: Defining subsystems and relations
1.3. Priorities: Determining priorities
1.4. Resources: Defining complete infrastructure
1.5. Plan: Planning main project and activities
Phase 2: MAIN PROJECT (MP)
2.1. PT: Drawing up project task
2.2. DTP: Interview, analysis, process modeling (DTP)
2.3. MP Processes: Analysis of process, problem and proposal for improvements
2.4. MP Data: Data description
2.5. MP Plan: Planning project realizations
2.6. MP Resources: Defining resource models of the main project
Phase 3: PROJECT REALIZATION (PR)
3.1. DEV: Interview, abstraction, data modeling (EV)
3.2. Translation: Translation of data model into DB scheme (RM)
3.3. PR Architecture: Defining architecture of program product (APP)
3.4. PR Operations: Design of operation on DB scheme

Table 2: Physical modeling – building

Phase 4: : SOFTWARE REALIZATION (SR)
4.1. Designing physical data base
4.2. Registering DB scheme in data dictionary
4.3. Producing prototype and generating application tree
4.4. 4GL or 3GL programming
4.5. Writing instructions and explanation of program solutions
4.6. Entering test data in BP and testing
Phase 5: IMPLEMENTATION AND APPLICATION (IAA)
5.1. Training user
5.2. Entering initial data in data base
5.3. Testing suitability to users' requirements
5.4. Writing help system for users
5.5. Optimization and parallel work of the new and old system
5.6. Final testing - delivery
5.7. Application of program product
Phase 6: MAINTENANCE (MAI) Performing previous activities in order to: introduce new business processes, replace existing business processes and correct errors

4. Description of the MIRIS methodic phases

Strategic planning is the first phase of the informatization, where a series of models are defined, describing a system clearly but not in detail. The important outcome is the definition of subsystems and informatization priorities.

In strategic planning phase, there are numerous other good and detailed methodic. This phase is not a specialty of the MIRIS methodic, so some other methodic can be used here, for example SPIS [1].

Main design is a phase of process modeling. This phase is performed for each particular subsystem within a strategic plan.

In this phase, a set of documents is defined with a scope to prepare users before the interview. Following the preparation, interviewing, process modeling as well as drawing up of data flow diagrams are facilitated.

Detailed design is the phase of data modeling and operations on data using a software product. One subsystem from the main design can be divided into several components and after the main design phase, detailed design phase can be performed for several times. The result is a model describing a part of a system in detail, which represents a solid basis for building the applications. Basic methods in this phase are: method for semantic data modeling (ER from entity-relationship), method for presentation of database relational scheme and method for application modeling.

Software production is the stage where various development tools are used (languages of the III and IV generation, SUBP, RAD tools, CASE tools, etc.) and the objective is to produce complete software.

The methodic is and it aims to be independent in terms of production tool and process of the software production process. If CASE tools are used, process and data models are adapted to the tool selected, that is, it is necessary to set translation of the model from the main and detailed design into a model acceptable for the CASE tool. Model produced in design phases represents the initial version of the project documentation, which is translated into the model readable by the selected tool, while the further documentation maintenance is performed within a CASE tool. The power of concepts established in the MIRIS methodic is then lost, but the power of the generator and CASE tool are gained.

Introduction and implementation is the phase where produced software becomes tool in user's hands for carrying out business processes and it can take as much as twice longer than software production phase. The management should move an organization towards changes in behavior and implementation of completed solution. The responsibility for introducing is on management, while IT professionals should actively cooperate in training users and improvements of the software.

Maintenance is the phase consisting of executing activities from previous required phases due to introducing new or changing the existing business processes and correcting failures.

5. Model of basic information system design classes

The objective of the basic training for designers is to teach methods of business operation analysis, interviewing users, organization of databases, building applications, program design and documenting information systems.

On the basis of the development of information system phases and methods therein used for training information system designers; a training model for the MIRIS methodic is defined, as shown in Table 3. Number of seminars is not compliant with phases or design sequence; rather, it has it origins in year's long research of the contents and sequence that produced the best results in basic training about IS design.

A more complete model of one-year training differs from the basic model.

The basic training of designers consists of logical units (seminars) shown in Table 3 The sequence, duration and classification into development phases are shown in Table 3.

Table 3: Model of lessons in basic training for IS designers

Class	Duration in 60 minute periods	Phase acc. to MIRIS
Introduction to IS design	2	1
Analysis of business operation and process modeling	8	2
Conceptual modeling and database design	10	3
Hands on design and production of own model	8	2,3
Exam – production of a required model	4	-

The basic training model does not include building and software introduction phases, but these are mentioned in the introduction.

5.1. A brief description of particular seminar

Topics from a particular seminar are stated further on in this chapter, without describing the importance of a specific topic, which is described in Chapter 4.

1. Introduction in IS design

Software engineering, information system, methods, methodic, methodologies, models, design, modeling, IS design life cycle phases and software production, project management, communication and relations with users.

2. Analysis of business and process modeling (ANA)

System analysis, master project, project task, team analysis, data flow chart: data flow, process, data storage, external systems, decomposition, context chart, hierarchical description, process types, balance, recommendations for drawing, constraints, interview, practical representation of structural research, recognition of process and data flow, how to develop the IS in a company, analysis and modeling using students' examples.

3. Conceptual modeling and development of data base (DDB)

Abstraction, method entity-relations (ER), structure of ER: entity, relation, attribute, aggregation, ... Constraints in ER: relationship's and attribute's cardinality,... Data analysis in documents and respective modeling, independent and team modeling, database, 4GL, data dictionary, relational model: relation, attribute, domain, candidate for key, relation key, external key, constraints, relational operators, operator, normalization, translation of ER into a relational model. Object oriented approach, UML/OML, executive project, exercise in modeling.

4. Design practice (DES) and realization of own project

Solving examples, team modeling, case analysis, realization of own project, interview exercise, review and suggestions for models in our work.

5.2. Characteristics of basic training model

Examples of answers given during seminars: Requirements changed during the realization of a project, how to document a project, standardization of projects and methods, planning and management of development, distribution of tasks, etc.

Scope, depth and sequence of curriculum are defined for each seminar. These are three basic didactic dimensions of teaching, aiming to a successful teaching.

Scope of curriculum is the issue of extent of knowledge and skills that designers should acquire and develop during particular seminar.

The depth of curriculum changes over time. Some topics are presented as a mere information, such as normalization, while others become more important.

The sequence of presenting the curriculum is based on the following principles:

- from easier to more difficult
- from known to unknown
- from simple to more complex
- from existing to desired
- from programming practice to abstract modeling of a system.

The sequence of seminars follows the IS development life cycle phases, unlike in the one-year training model.

Success rate in training design could be measured more accurately applying the following criteria:

- quick familiarization with business technology of organization system

- quality of the project
- quality of the developed IS program support.

It is considered that the lessons have been improved over time, which was supported by the fact that lessons have been continuously performed as of 1986 as well as supporting literature, textbooks and presentations.

During teaching, numerous misconceptions were noted. At early stage, results were poorer. Half of the students would not sit for exam or they would not pass the exam. A few students actually became IS designers.

There were quite a great number of students attending first seminars. At the time, a number of various methods for the same task were theoretically studied. The teaching was oriented towards the concepts of method structure. Modeling process was disregarded.

It was teachers who defined and achieved aimed teaching goals. Students showed quite a high degree of knowledge and enthusiasm. But something failed. The number of students that became designers was not as high as desired. Particularly efficient studying did not help so new methods, which would foster creativity, were searched [13].

Teaching can be well elaborated, materials covering topics could be prepared, students could learn everything as planned, so we could speak of a good teaching. Yet, students are not able to independently design IS, and the excellent teaching produced poor results.

Over time, teaching included knowledge from critical-communication didactic. It has been permanently aimed to improve the teaching. In teaching, democratization and humanization is the potential future horizon and any failure in teaching is seen in that light and used to acquire new useful knowledge. Teaching is a communication process. There is interaction between teachers and students. Teaching starts, lasts and ends through an intensive communication. Teacher is a moderator, who is, during teaching, permanently available for communication with students.

Any question to any direction establishes a relation of new quality and provides opportunity for other students to learn. Such flashes of digression are the source of new ideas and they increase the power of designer's thinking - designers are born. In communication, roles are changed, at times teacher assumes a designer's role, and at times he is a user. The same roles are assigned to students in their communication.

Teacher is a communication moderator and he does not allow ineffective communication (wasting time). This is students' role, too, and students are also entitled to join the communication. Person and his/her attitude are respected in communication.

From the communication process, teacher measures and reads expected teaching results (acquisition of designing skill), as shown in students' statements, actions and decision referring to simulated reality. There is an active communication in order to present skill acquired during teaching, while if objectives are not achieved, teaching is intensified in order to reach necessary knowledge and obtain results, i.e. teach a designer. The objective of teaching is to teach students how to design, there is a confidence in students and, at last, difference between the expected and the modeled by students is analyzed, which results in assessment of the quality of a project solution and design speed.

The success of training is controlled immediately during the training. The control can be obvious or hidden, through feed back instructions, advices, desires, opinions, and answers. The control provides and enables the progress of teaching. Otherwise, the progress in not effective.

6. Conclusion

The paper presents the model of basic training for information system designers and the didactics knowledge as a critical theory of teaching communication. The training model has seminars defined for analysis and modeling of business systems before addressing hands-on software production. Minimum required time is six days, i.e., 24 hours of lessons.

Accelerated and abridged presentation of a particular topic can be redesigned as one-day or briefer course for providing information only, where students are not intended to acquire the design skills.

This paper resulted from experience, on the basis of need for making the teaching inventive. Education of designers is a creative communication process.

Further research on effects of teaching could be made through establishing a method for measuring designing potentials after completed education, varying the teaching by sequence, examples, individual effects, effects in group, result relations and emotional moods (confidence and methods, attitudes, believes), measuring changes in attitudes and exploring teaching without and with various level of communication. It is quite clear to the authors that measurements in this field are especially aggravated and that initial states can considerably influence the results and conclusions.

References

1. Brumec, J., Dušak, V., N. Vrček.: Framework for strategic planning of information systems. In Proceedings of the 7th Americas Conference on Information Systems, Boston, USA, 2001., p. 1701-1707
2. Cippico, V.: Faza izvedbe informacijskog sustava, HIZ, Savjetovanje CASE 8, Opatija, 1996., p 19-40
3. Gudjons, H.: Pedagogija, temeljna znanja, Educa, Zagreb, 1994., p. 278
4. Hawryszkiewycz, I., T.: Introduction to Systems Analysis and Design, Prentice - Hall, New York, 1988., p 373
5. Inmon., W. H.: Information Systems Architecture – a System Developers Primer, Prentice-Hall, NJ, 1986., p. 235
6. Kovačić, A., Vintar, M.: Načrtovanje in gradnja informacijskih sistemov, DZS, Ljubljana, 1994., p. 316
7. Mylopoulos, J., Levesque, J. H.: An Overwiev of Knowledge Representation, Proceedings of International Symposium Conference, NY, 1984., p. 3-17
8. Pavlić, M.: Metodika projektiranja informacijskih sustava, IV savjetovanje o informatičkoj djelatnosti u ZO Rijeka, Opatija, 1987., p. 5-1 do 5-16.
9. Pavlić M. i dr.: Školovanje projektanata IS, CASE 5, Opatija, 1993., p. 8_1 do 8-8
10. Pavlić, M.: Tehnologija projektiranja informacijskih sustava, CASE 7, Opatija, 1995., p. 1-13
11. Peters, L.: Advanced Structured Analysis and Design, Prentice Hall, London, 1998., p. 272
12. Simon, J.C.: Introduction to Information Systems, John Wiley & Sons, NY, 2001., p. 327
13. Srića, V.: Upravljanje kreativnošću, Školska knjiga, Zagreb, 1992., p 104
14. Varga, M.: Faze razvoja informacijskog sustava, Infotrend, vol. 43, br. 2., Zagreb, 1996., p. 54-57

APPLICATION OF THE QUEUEING THEORY ON CALCULATION OF THE OPTIMAL NUMBER OF EMPLOYEES OF THE CALL-CENTRE

Marko Potokar[1], Gregor Miklavčič[2]
and Mirjana Rakamarić Šegić[3]
[1]Bankart d.o.o., Celovška 150, 1000 Ljubljana, Slovenia,
email: marko.potokar@bankart.si
[2]Bank of Slovenia, Slovenska 35, 1000 Ljubljana, Slovenia
email: gregor.miklavcic@bsi.si
[3]Politechnic of Rijeka, Vukovarska 58, 51000 Rijeka, Croatia
email: mrakams@veleri.hr

ABSTRACT

The paper deals with the calculation of the optimal number of employees (operators) of the call-centre. We have derived a stochastic model of the call-centre, based on queueing theory. In the first part of the article we defined the objective function (model 1) and calculated its optimum. In the second part of the article we developed advanced objective function (model 2) and performed sensitivity analysis. All the calculations are based on real (life) data which we gathered from the working statistics of the operators of the call-centre.

KEY WORDS

Call-centre, stochastic model, queueing system, Poisson process, stationary regime, sensitivity analysis.

1. INTRODUCTION

The Bankart company offers card and ATM processing services, as well as POS terminals, to all banks and other financial institutions, and to interested non-banking organisations in Slovenia. The company also provides its users with data for effecting payment and debit card payments, authorises transactions and services self-service machines.

One of the most important services provided by Bankart company is to manage ATM and POS network and to provide correct information for its users regarding the processing services of their cards. This services are provided within the so called call-centre which operates in non-stop regime.

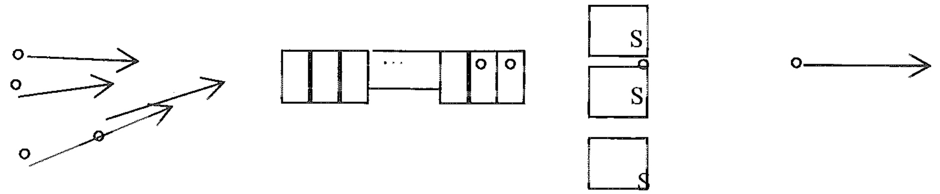
The mathematical models of the call-centre developed to optimize the number of the operators being on duty shown in this paper is based on methods of queueing theory.

2. THE MODEL OF THE CALL-CENTRE

In the call-centre incoming telephone calls are distributed by special server called CTI server. The entrance in the system is represented with unique waiting line (queue). Incoming calls are delivered to operators by the method FIFS (First In First Served). We can interpret the call-centre as the $M/M/m/\infty$ queueing system where:

- telephone calls (jobs) arrive in accordance to a Poisson process with intensity $\lambda>0$
- times of the services are distributed by exponential law

- there are m ≥ 1 identical servers (operators)
- one server (operator) serves one job with intensity $\mu > 0$
- an arriving job goes directly into service if any of the servers are free; otherwise the job joins the waiting line (the queue) with infinite waiting places
- jobs leave the system after service completion



Our statement that the waiting line has infinite waiting places is based on statistical data gathered from the CTI server. It has never happened that waiting line was full.

We have defined the model for time interval from Monday to Friday for the hours between 7 a.m. and 5 p.m. Values of parameters of the objective function (see chapter 3.).are also calculated for this interval. Within this interval the stationary regime exists.

In such a condition the sum $S = 1 + \sum_{i=1}^{\infty} \frac{\lambda_{i-1}\lambda_{i-2}\cdots\lambda_0}{\mu_i\mu_{i-1}\cdots\mu_1}$ has the following form (Potokar et.al., 2004):

$$S = 1 + r\rho + \frac{(r\rho)^2}{2!} + \frac{(r\rho)^3}{3!} + \dots + \frac{(r\rho)^{r-1}}{(r-1)!} + \frac{(r\rho)^r}{r!} + \frac{(r\rho)^r}{r!}\rho + \frac{(r\rho)^r}{r!}\rho^2 + \dots ;\ \rho = \lambda/\mu s$$

From conditions for stationary regime the condition $\rho < 1$ $(s > \lambda/\mu)$ must hold:

$$S = 1 + r\rho + \frac{(r\rho)^2}{2!} + \frac{(r\rho)^3}{3!} + \dots + \frac{(r\rho)^{r-1}}{(r-1)!} + \frac{(r\rho)^r}{r!} \cdot \frac{1}{1-\rho} \quad ;\ \rho = \lambda/\mu s$$

The average waiting time is:

$$W_q = \frac{P(i \geq s)}{s \cdot \mu - \lambda}, \text{ where } P(i \geq s) = \frac{(s \cdot \rho)^s \cdot p_0}{s!(1-\rho)};\rho=\lambda/\mu s \text{ and } p_0 = \frac{1}{\sum_{i=0}^{s-1} \frac{(s \cdot \rho)^i}{i!} + \frac{(s \cdot \rho)^s}{s!(1-\rho)}};\rho = \lambda/\mu s$$

3. THE OBJECTIVE FUNCTION AND ITS PARAMETERS (Model 1)

In our case we are interested in costs of individual costumer being served and in opportunity costs that rise if costumer leaves the system without being served.

3.1 Data gathering

All the calculations are based on real data which we gathered from the working statistics of the operators of the call-centre.

We calculated the parameter a as $a = \frac{\sum_{i=1}^{17} BP_i}{17}$ where BP_i is the operator gross salary in Slovenian tolars, converted in tolars on minute of working (BP_i = (basic gross salary * tariff coefficient + bonus)/number of minuts in one month; ***a* = 21 SIT/min.**).

Parameter b shows the opportunity costs of waiting of each individual customer per minute. We estimated the value b = 3*a (***b* = 63 SIT/min.**). Because it is difficult to estimate the value of this parameter right, we decided to check the quality of models with sensitivity analyses.

For calculation of parameters λ and μ of the objective function we have used the statistical data of the application of the CTI server for the period of one month (March 2005) that has 230 working hours.

$$\lambda = \left[\frac{\sum_{i=1}^{230} n_i}{230 \cdot 60 \text{ min}}\right] ;\ n_i = \textit{the number of calls in one hour } (\boldsymbol{\lambda = 0.427/min.})$$

$$\mu = \left[\frac{\sum_{i=1}^{n} t_i}{n}\right]^{-1} ;\ t_i = \textit{the individual serving time in minutes } (\boldsymbol{\mu = 0{,}291/min.})$$

3.2 The objective function

The objective function is compound of two parts: the cost of serving the customers (a•s) and the cost of customers waiting in line(λ•b•W_q). So we define the objective function as $J = a \cdot s + b \cdot \lambda \cdot W_q(s)$ where:

a = the cost of individual server (operator)
s = the number of servers
b = the cost of individual call
λ = the rate of incoming calls; the rate is calculated as one month average:
μ = service intensity
W_q = the average waiting time (in minutes)

4. ANALYTICAL METHOD OF CALCULATING THE OPTIMAL NUMBER OF THE OPERATORS

We derive the optimal number of the operators by calculating the minimum of the objective function:

$$\min J = \frac{d(a \cdot s + b \cdot \lambda \cdot W_q(s))}{ds} = 0 ;\ \rho = \lambda/\mu s < 1 \Rightarrow s > \lambda/\mu$$

$$\frac{d(a \cdot s + b \cdot \lambda \cdot W_q(s))}{ds} = \frac{d(a \cdot s)}{ds} + \frac{d(b \cdot \lambda \cdot W_q(s))}{ds} = a + b \cdot \lambda \cdot \frac{dW_q(s)}{ds};$$

We can write $W_q(s)$ as:

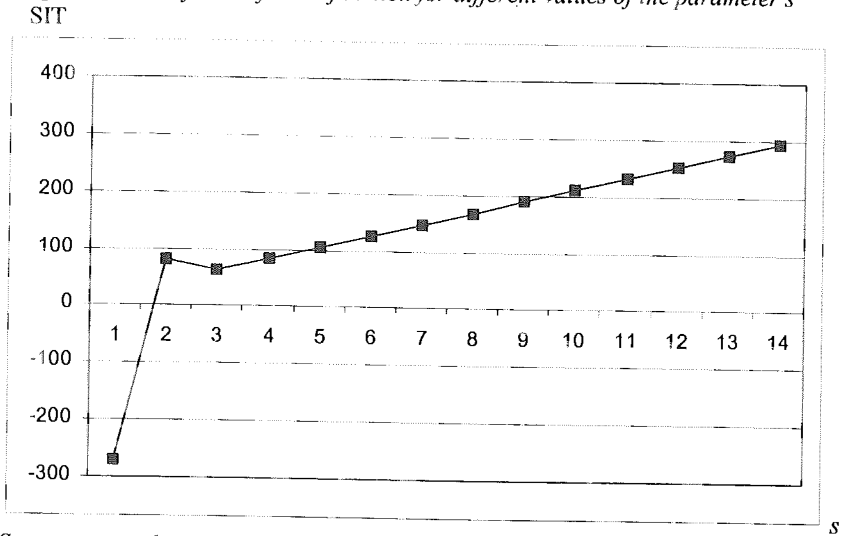
$$W_q(s) = \frac{\left(s \cdot \frac{\lambda}{s \cdot \mu}\right)^s}{(s \cdot \mu - \lambda) \cdot s!\left(1 - \frac{\lambda}{s \cdot \mu}\right) \cdot \left[\sum_{i=0}^{s-1} \frac{\left(s \cdot \frac{\lambda}{s \cdot \mu}\right)^i}{i!} + \frac{\left(s \cdot \frac{\lambda}{s \cdot \mu}\right)^s}{s!\left(1 - \frac{\lambda}{s \cdot \mu}\right)}\right]};$$

The derived function is very difficult to resolve. We rather used the program Excell to determine the optimal number of operators.

5. DETERMINING THE OPTIMAL NUMBER OF OPERATORS WITH EXCELL

We calculated values of the objective function for different values of the parameter s and given values of parameters a, b, λ in μ. Calculated values of the objective function are shown in the *Graph 1* bellow.

Graph 1: values of the objective function for different values of the parameter s



Source: own calculations

We can see that the minimum of the objective function is when s equals 3. So the optimal number of operators is 3.

6. ADVANCED OBJECTIVE FUNCTION (Model 2)

In the first model we presumed that the relation between opportunity costs and average waiting time is linear. But from practical experience we can assume that this relation is more exponential because the customer waiting in the line is more and more unsatisfied and so more likely to get another provider. The customer's opportunity costs rise with time spent in waiting line as well. We write advanced objective function in the form:

$$J_A = as + b\left(e^{\lambda W_q} - 1\right) = as + b\left(e^{L_q} - 1\right);\ \lambda W_q = L_q.$$

7. SENSITIVITY ANALYSIS

In this section we analyse how the optimal number of servers is sensitive to changes in parameters a, b, λ and μ. Only results of the second model will be presented here in a way that one parameter will change and the others will remain fixed. Sensitivity analysis is performed in Excel, where we entered different values for each parameter and then we looked for the optimal solution of the model.

In *Table 1* we presented the results of the sensitivity analysis. By changing parameter a in the model 2 we have calculated that the interval, where $s = 3$, equals $1{,}6 \leq a \leq 51{,}4$. In relation to model 1 the interval has widened, so model 2 is less sensitive to changing a than model 1.The expected costs are increasing very similarly to increment of parameter a, in relative terms.

*Tabel 1: Influence of changing a on s, C and C/C**

a (in SIT/min)	a/a* (in %)	s	C (in SIT/min)	C/C* (in %)
5,25	-75	3	17,33	-73,2
10,50	-50	3	33,08	-48,8
15,75	-25	3	48,83	-24,4
21,00	0	3	64,58	0
26,25	25	3	80,33	24,4
31,50	50	3	96,08	48,8
42,00	100	3	127,58	97,6
52,00	148	2	157,05	143,2

Source: own calculations

Parameter b shows the opportunity costs of waiting of each individual customer per minute and should be inside the interval $25{,}8 \leq b \leq 869$, in case of three operators being on duty. The results of the analysis are similar to the previous one, which means that changes are relatively small (see *Table 2*). The differences between model 1 and 2 are again not significant.

*Tabel 2: Influence of changing b on s, C and C/C**

b (in SIT/min)	b/b* (in %)	s	C (in SIT/min)	C/C* (in %)
15,75	-75	2	55,26	-14,4
31,50	-50	3	63,79	-1,2
47,25	-25	3	64,19	-0,6
63,00	0	3	64,58	0
78,75	25	3	64,98	0,6
94,50	50	3	65,38	1,2
126,00	100	3	66,17	2,5
870	1280	4	84,86	31,4

Source: own calculations

In *Table 3* we presented the results from varying parameter λ (arrival rate). Firstly, there are again no significant changes comparing both models. Secondly, parameter λ must be $0{,}353 \leq \lambda \leq 0{,}796$, if we would like to have $s = 3$.

*Tabel 3: Influence of changing λ on s, C and C/C**

λ (number of calls/min)	λ/λ* (in %)	s	C (in SIT/min)	C/C* (in %)
0,10675	-75	1	35,92	-44,4
0,2135	-50	2	46,62	-27,8
0,32025	-25	2	57,60	-10,8
0,427	0	3	64,58	0
0,53375	25	3	65,74	1,8
0,6405	50	3	67,95	5,2
0,874	100	4	84,33	30,6

Source: own calculations

In the last table we deal with parameter μ (service intensity), where s = 3, if parameter μ is $0{,}156 \leq \mu \leq 0{,}352$. Comparing both models we again realised no significant change.

*Tabel 4: Influence of changing μ on s, C and C/C**

μ (number of calls/min)	μ/μ* (in %)	s	C (in SIT/min)	C/C* (in %)
0,07275	-75	6	126,01	95,1
0,1455	-50	4	84,30	30,5
0,21825	-25	3	66,30	2,7
0,291	0	3	64,58	0
0,36375	25	2	61,65	-4,5
0,4365	50	2	52,52	-18,7
0,582	100	2	46,26	-28,4

Source: own calculations

Hence, from the present analysis we can sum up:

1. From the sensitivity point of view models 1 and 2 do not differ significantly from each other.
2. When we are changing parameters a and b, the influence on the optimal number of operators is only slight, while changing λ and μ, we have more significant effect.

8. CONCLUSION

In this paper the problem of determining the optimal number of the employees of the call-centre was solved by application of the model from the queuing theory. If we compare the results for both models we see they match. This is one indicator that confirms our presumption of their validity.

But we have to be aware that accuracy of the results we get with such models strongly depends on the used data. In our case the most important factor for the validity of the results is the rate of incoming calls (λ) and service intensity (μ) that we calculated. If the data are not observed during a sufficiently long period of time, serious problems might appear regarding their accuracy.

The sensitivity analysis shows that models 1 and 2 do not differ significantly. The influence on the optimal number of operators is only slight when changing parameters a and b, while changing λ and μ, we have more significant effect.

REFERENCES

1. Doob J.L.: Stochastic Processes. New York: Wiley, 1953.
2. European Journal Of Operational Research: Volume 156, Issue 2, 16 July 2004; An analyzing of queuing systems with multi-task servers; Z. G. Zhang and N. Tian
3. Gihman I.I. and Skorohod A.V.: The Theory of Stochastic Processes I-III. Berlin: Springer, 1974.
4. Hudoklin-Božič A.: Stohastični procesi. Kranj: Univerza v Mariboru, 1986. 180 str.
5. Hudoklin A.: Stohastični model, Organizacija, revija za management, informatiko in kadre, Založba Moderna organizacija, letnik 32, št. 10, December 1999, str. 535 – 547.
6. International Abstracts in Opeations Research: Volume 55 Number 1 January/February 2004; Serial Numbers 85496-85945; Analysis of a finite-buffer queue with discrete-Markovian arrival process: D-MAP/$G^{a,b}$/1/N; 85911

7. International Abstracts in Opeations Research: Volume 55 Number 2 March 2004; Serial Numbers 85946-86395; A complete and simple solution for a discrete-time multi-server queue with bulk arrivals and deterministic service times
8. Karlin S. and Taylor H.: A First Course in Stochastic Processes. 2d ed. Orlando, Fla.: Academic Press, 1975.
9. Koechel P.: Stohastic Modelling; Lecture held at the University Ljubljana 1999
10. Legiša P.: Analiza I; zapiski predavanj, FNT 1988
11. Potokar M., Miklavčič G.: Teoretična analiza in empirična preverbal modela za določitev optimalnega števila strežnih mest v klicnem centru. EF Ljubljana, April 2004. 37.
12. QTS software; John Wiley & Sons, Inc; http://www.wiley.com; fftp://ftp.wiley.com/public/sci_tech_med/queueing_theory; http://www.geocities.com/qtsplus
13. SOR'03 Proceedings: The 7th International Symposium on Operational Research in Slovenia; Podčetrtek, Slovenia, September 24-26, 2003
14. Whitt, Ward: Improving Service by Informing Customers About Anticipated Delays; Management Science, Volume 45, Issue 2, 1999, str. 192-207
15. Winston Wayne L.: *Operations Research: Applications and Algorithms*. Belmont: Duxbury Press, 1991. 1025
16. http://www.sigov.si/zrs/publikacije/skd/skd02s1.doc

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Appendix:
Authors' addresses

Addresses of SOR'05 Authors

(The 8th International Symposium on OR in Slovenia, Nova Gorica, SLOVENIA, September 28 – 30, 2005)

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
1.	Zdravka	Aljinović	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	zdravka.aljinovic@efst.hr
2.	Josip	Arnerić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	jarneric@efst.hr
3.	Jan	Babič	Jožef Štefan Institute	Jamova 39	1000	Ljubljana	Slovenia	jan.babic@ijs.si
4.	Zoran	Babić	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	babic@efst.hr
5.	Vlasta	Bahovec	University of Zagreb, Graduate School of Economics and Business	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	bahovec@efzg.hr
6.	Majda	Bastič	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	majda.bastic@uni-mb.si
7.	Tomaž	Berlec	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	tomaz.berlec@fs.uni-lj.si
8.	Ludvik	Bogataj	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	ludvik.bogataj@guest.arnes.si
9.	Marija	Bogataj	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1000	Ljubljana	Slovenia	marija.bogataj@guest.arnes.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
10.	Bernhard	Böhm	Vienna University of Technology,Institute for Mathematical Models and Economics		1040	Vienna	Austria	bernhard.boehm@tuwien.ac.at
11.	Matej	Borovinšek	University of Maribor, Faculty of Mechanical Engineering	Smetanova 17	2000	Maribor	Slovenia	matej.borovinsek@uni-mb.si
12.	Darja	Boršič	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	darja.borsic@uni-mb.si
13.	Marko	Botinčan	University of Zagreb, Department of Mathematics	Bijenička cesta 30	10000	Zagreb	Croatia	mabotinc@math.hr
14.	Valerija	Bublić	Lorana d.o.o.	Zagrebačka 26		Krapina	Croatia	lorana@kr.htnet.hr
15.	Lorenzo	Castelli	DEEI - University of Trieste	Via A. Valerio, 10	34127	Trieste	Italy	castelli@units.it
16.	Katarina	Cechlarova	P. J. Šafarik University, Faculty of Science, Department of Geometry and Algebra	Jasenna 5	04154	Košice	Slovak Republic	cechlarova@ science.upjs.sk
17.	Boris	Cota	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	bcota@efzg.hr
18.	Vesna	Čančer	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	vesna.cancer@uni-mb.si
19.	Sanja	Čandrlić	University of Rijeka, Faculty of Philosophy	Omladinska 14	51000	Rijeka	Croatia	sanjac@ffri.hr
20.	Anton	Čižman	University of Maribor, Faculty of Organizational Sciences Kranj	Kidričeva 55a	4000	Kranj	Slovenia	anton.cizman@fov.uni-mb.si
21.	Mirjana	Čižmešija	University of Zagreb, Graduate School of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	mcizmesija@efzg.hr
22.	Draženka	Čižmić	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	dcizmic@efzg.hr
23.	David	Čreslovnik	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	david.creslovnik@ef.uni-lj.si
24.	Nadja	Damij	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	nadja.damij@ef.uni-lj.si
25.	Samo	Drobne	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	samo.drobne@fgg.uni-lj.si
26.	Ksenija	Dumičić	University of Zagreb, Graduate School of Economics and Business, Department of Statistics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	kdumicic@efzg.hr
27.	Igor	Đukanović	University of Maribor, EPF	Razlagova 14	2000	Maribor	Slovenia	igor.mat@uni-mb.si
28.	Andreas	Ehrenmann	CORE, Universite Catholic de Louvain			Louvain La Neuve	Belgium	ehrenmann@core.ucl.ac.be
29.	Nataša	Erjavec	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	nerjavec@efzg.hr
30.	Daša	Fabjan	University of Ljubljana, Faculty of Maritime and Transport	Pot pomorščakov 4	6320	Portorož	Slovenia	dasa.fabjan@ fpp.edu
31.	Matjaž	Feltrin		Zakraj 1a	5283	Slap ob Idrijci	Slovenia	matjaz_feltrin@email.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
32.	Liljana	Ferbar	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	liljana.ferbar@ef.uni-lj.si
33.	Tamás	Fleiner	Department of Computer Science and Information Theory, Budapest Univ. of Tehnology and Economics	Műegyetem rkp. 3-9	H-1111	Budapest	Hungary	fleiner@cs.bme.hu
34.	Martin	Gavalec	Department of Information Technologies, Faculty of Informatics and Management, University Hradec Králové	Rokitanskehó 62	50003	Hradec Králové	Czech Republic	martin.gavalec@uhk.cz
35.	Janez	Grad	University of Ljubljana, Faculty of Administration	Gosarjeva ulica 5	1000	Ljubljana	Slovenia	janez.grad@fu.uni-lj.si
36.	Jasna	Hrovatin	University of Ljubljana, Biotechnical Faculty	Večna pot 83	1111	Ljubljana	Slovenia	jasna.hrovatin@bf.uni-lj.si
37.	Tihomir	Hunjak	Faculty of Organization and Informatics	Pavlinska 2	42000	Varaždin	Croatia	thunjak@foi.hr
38.	Dušan	Hvalica	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	dusan.hvalica@ef.uni-lj.si
39.	Josef	Jablonsky	Department of Econometrics, University of Economics		130 67	Praha	Czech Republic	jablon@vse.cz
40.	Hubertus Th.	Jongen	RWTH Aachen University (D), Department of Mathematics		D-52056	Aachen	Germany	jongen@rwth-aachen.de
41.	Elza	Jurun	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	elza@efst.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
42.	Vladimir I.	Kalika	Natural Resources and Environmental Research Center, Haifa University	Mount Carmel	31905	Haifa	Israel	kalika@econ.haifa.ac.il
43.	Alenka	Kavkler	University of Maribor, Faculty of Economics and Business Maribor	Razlagova 14	2000	Maribor	Slovenia	alenka.kavkler@uni-mb.si
44.	Peter	Köchel	Chemnitz University of Technology, Faculty of Informatics, Chair of Modelling & Simulation	Straße der Nationen 62	09107	Chemnitz	Germany	peter.koechel@informatik.tu-chemnitz.de
45.	Marek	Kočan	Institute of Mathematics, Faculty of Science, P. J. Šafárik University			Košice	Slovak Republic	kocan@science.upjs.sk
46.	Tomaž	Kramberger	University of Maribor, Faculty of Mechanical Engineering	Smetanova 17	2000	Maribor	Slovenia	tomaz.kramberger@uni-mb.si
47.	Janez	Kušar	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	janez.kusar@fs.uni-lj.si
48.	Lado	Lenart	Jožef Štefan Institute	Jamova 39	1000	Ljubljana	Slovenia	lado.lenart@ijs.si
49.	Igor	Lipušček	University of Ljubljana, Biotechnical Faculty	Večna pot 83	1111	Ljubljana	Slovenia	igor.lipuscek@bf.uni-lj.si
50.	Andrej	Lisec	Post of Slovenia, Ltd., Business Unit Postal Logistics Center Ljubljana	Cesta v Mestni log 81	1002	Ljubljana	Slovenia	andrej.lisec@posta.si
51.	Anka	Lisec	University of Ljubljana, Faculty of Civil and Geodetic Engineering	Jamova 2	1000	Ljubljana	Slovenia	anka.lisec@fgg.uni-lj.si
52.	Zrinka	Lukač	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	zlukac@efzg.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
53.	Špela	Malovrh	University of Ljubljana, Biotechnical Faculty		1000	Ljubljana	Slovenia	spela.malovrh@bf.uni-lj.si
54.	Robert	Manger	University of Zagreb, Department of Mathematics	Bijenička cesta 30	10000	Zagreb	Croatia	manger@math.hr
55.	David	Manlove	Department of Computer Science, University of Glasgow	Glasgow G12 8QQ		Glasgow	UK	davidm@dcs.gla.ac.uk
56.	Branka	Marasović	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	branka.marasovic@efst.hr
57.	Marija	Marinović	University of Rijeka, Faculty of Philosophy	Omladinska 14	51000	Rijeka	Croatia	marinm@ffri.hr
58.	Gregor	Miklavčič	Bank of Slovenia	Slovenska 35	1000	Ljubljana	Slovenia	gregor.miklavcic@bsi.si
59.	Dubravko	Mojsinović	Privredna banka Zagreb d.d.	Račkoga 6	10000	Zagreb	Croatia	dubravko.mojsinovic@pbz.hr
60.	Blaž	Mojškerc	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	blaz.mojskerc@ef.uni-lj.si
61.	Luka	Neralić	Faculty of Economics, University of Zagreb	Trg J. F. Kennedya 6	10000	Zagreb	Croatia	lneralic@efzg.hr
62.	Goranka	Nogo	University of Zagreb , Department of Mathematics		10000	Zagreb	Croatia	nogo@math.hr
63.	Ana	Osvald		Via dei Faiti 7	34170	Gorizia	Italy	ana.osvald@bayern-mail.de
64.	Dejan	Paliska	University of Ljubljana, Faculty of Maritime and Transport	Pot pomorščakov 4	6320	Portorož	Slovenia	dejan.paliska@fpp.edu
65.	Pavels	Patlins	Riga Technical University, Faculty of Engineering Economy, Logistic and Transport Economics Department			Riga	Latvia	pavels.patlins@rtu.lv
66.	Anita	Pavković	University of Zagreb, Graduate School of Economics and Business, Department of Finance	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	apavkovic@efzg.hr
67.	Mile	Pavlić	University of Rijeka, Faculty of Philosophy	Omladinska 14	51000	Rijeka	Croatia	mpavlic@ffri.hr
68.	Bojan	Peček	University of Ljubljana, Faculty of Administration	Gosarjeva 5	1000	Ljubljana	Slovenia	bojan.pecek@fu.uni-lj.si
69.	Jan	Pelikán	Vysokáškola ekonomická v Praze				Czech Republik	pelikan@vse.cz
70.	Raffaele	Pesenti	DINFO – University of Palermo	Viale delle Scienze	I-90128,	Palermo	Italy	pesenti@unipa.it
71.	Raffaele	Pesenti	DIAI – University of Palermo	Viale delle Scienze	90128	Palermo	Italy	pesenti@unipa.it
72.	Snježana	Pivac	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	spivac@efst.hr
73.	Ján	Plavka	Department of Mathematics, Faculty of Electrical Engineering and Informatics, University of Košice	B. Nemcovej 32	04200	Košice	Slovak Republic	jan.plavka@tuke.sk
74.	Marko	Potokar	Bankart d.o.o.	Celovška 150	1000	Ljubljana	Slovenia	marko.potokar@bankart.si

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
75.	Janez	Povh	School of Business and Management	Na Loko 2	8000	Novo mesto	Slovenia	janez.povh@guest.arnes.si
76.	Krunoslav	Puljić	University of Zagreb, Department of Mathematics	Bijenička cesta 30	10000	Zagreb	Croatia	nuno@math.hr
77.	Martin	Rajgelj	University of Ljubljana, Faculty of Economics	Kardeljeva ploščad 17	1001	Ljubljana	Slovenia	martin.rajgelj@ef.uni-lj.si
78.	Mirjana	Rakamarić Šegić	Politechnic of Rijeka	Vukovarska 58	51000	Rijeka	Croatia	mrakams@veleri.hr
79.	Franz	Rendl	Universität Klagenfurt, Institut für Mathematik	Universitätsstrasse 65-67	A - 9020	Klagenfurt	Austria	franz.rendl@uni-klu.ac.at
80.	Lada	Rokov	Faculty of Economics, University of Zagreb	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	rola4@net.hr
81.	Viljem	Rupnik	INTERACTA, LTD, Business Information Processing	Parmova 53	1000	Ljubljana	Slovenia	viljem.rupnik@siol.net
82.	Aleš	Silič	Goriške opekame d.d.	Merljaki 7	5292	Renče	Slovenia	ales.silic@go-opekarne.si
83.	Yves	Smeers	CORE, Universite Catholic de Louvain			Louvain La Neuve	Belgium	smeers@core.ucl.ac.be
84.	Renata	Sotirov	The University of Melbourne, Department of Mathematics and Statistics	Parkville			Australia	rsotirov@ms.unimelb.edu.au
85.	Marko	Starbek	University of Ljubljana, Faculty of Mechanical Engineering	Aškerčeva 6	1000	Ljubljana	Slovenia	marko.starbek@fs.uni-lj.si
86.	Dragomir	Sundać	International Business Consulting Center, Ltd	Bulevar o. 25	51000	Rijeka	Croatia	ibcc@ibcc.hr

ID	First name	Surname	Institution	Street and Number	Post code	Town	Country	E-mail
87.	Ivan	Šmid	Centar za poduzetništvo Krapinsko-Zagorske županije			Krapina	Croatia	i.smid@vip.hr
88.	Kristina	Šorić	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	ksoric@efzg.hr
89.	Neli	Tomić-Plazibat	University of Split, Faculty of Economics	Matice Hrvatske 31	21000	Split	Croatia	neli@efst.hr
90.	Walter	Ukovich	DEEI - University of Trieste	Via A. Valerio, 10	34127	Trieste	Italy	ukovich@units.it
91.	Petra	Ulamec	ARSO	Vojkova 1b,	1000	Ljubljana	Slovenia	petra.ulamec@gov.si
92.	Marko	Urh	University of Maribor, Faculty of Organizational Sciences Kranj	Kidričeva 55a	4000	Kranj	Slovenia	marko.urh@fov.uni-mb.si
93.	Katja	Vadnal	Biotehniška fakulteta	Jamnikarjeva 101	1000	Ljubljana	Slovenia	katja.vadnal@bf.uni-lj.si
94.	Višnja	Vojvodić Rosenzweig	University of Zagreb, Faculty of Economics	Trg J.F. Kennedya 6	10000	Zagreb	Croatia	vvojvodic@efzg.hr
95.	Damir	Vukičević	Universtity of Split, Department of Mathematics	Nikole Tesle 12	21000	Split	Croatia	vukicevic@pmfst.hr
96.	Lidija	Zadnik Stirn	University of Ljubljana, Biotechnical Faculty	Večna pot 83	1111	Ljubljana	Slovenia	lidija.zadnik@uni-lj.si
97.	Lyudmyla	Zahvoyska	Department of Ecological Economics, National University of Forestry and Wood Technology	Gen Chuprynky Str., 103,	79057	Lviv	Ukraine	zld@forest.lviv.ua
98.	Janez	Žerovnik	University of Maribor, Faculty of Mechanical Engineering	Smetanova 17	2000	Maribor	Slovenia	janez.zerovnik@fmf.uni-lj.si

The 8th International Symposium on
Operational Research in Slovenia

SOR '05

Nova Gorica, SLOVENIA
September 28 - 30, 2005

Appendix:
Sponsors' notices

Austrian Science and Research Liaison Office Ljubljana

ASO LJUBLJANA
AUSTRIAN SCIENCE AND RESEARCH LIAISON OFFICE LJUBLJANA

Main tasks of the Austrian Science and Research Liaison Office Ljubljana (ASO Ljubljana)

ASO-Ljubljana aims to facilitate and intensify scientific contacts and cooperation between Austria, Slovenia and South-Eastern European countries. In close thematic cooperation with the new science policy of the Austrian Federal Ministry of Education, Science and Culture (bm:bwk) in South-Eastern Europe, the ASO Ljubljana serves together with ASO Sofia/Bulgaria as focal point in this region.

By organizing and (co-)financing of international seminars, workshops, conferences, as well as training seminars for researchers from SEE region the ASO Ljubljana provides and promotes opportunities for collaborative work, scientific networking and cooperation with institutions of higher education and research facilities in Austria, Slovenia and the SEE Countries.

Special attention is paid to joint research projects in EU-programs, where ASO Ljubljana acts as project partner as well as mediators between potential project partners.

Main fields of activities of ASO Ljubljana

▪ Contribution of science to EU-integration of the Western Balkan Countries ▪ Governance ▪ Civil Society – Knowledge Society ▪ Science and Technology Policy ▪ Technological and Regional development in SEE ▪ Entrepreneurship development in SEE ▪ Information and Communication Technologies (ICT) ▪ Ethnic Minorities in Science

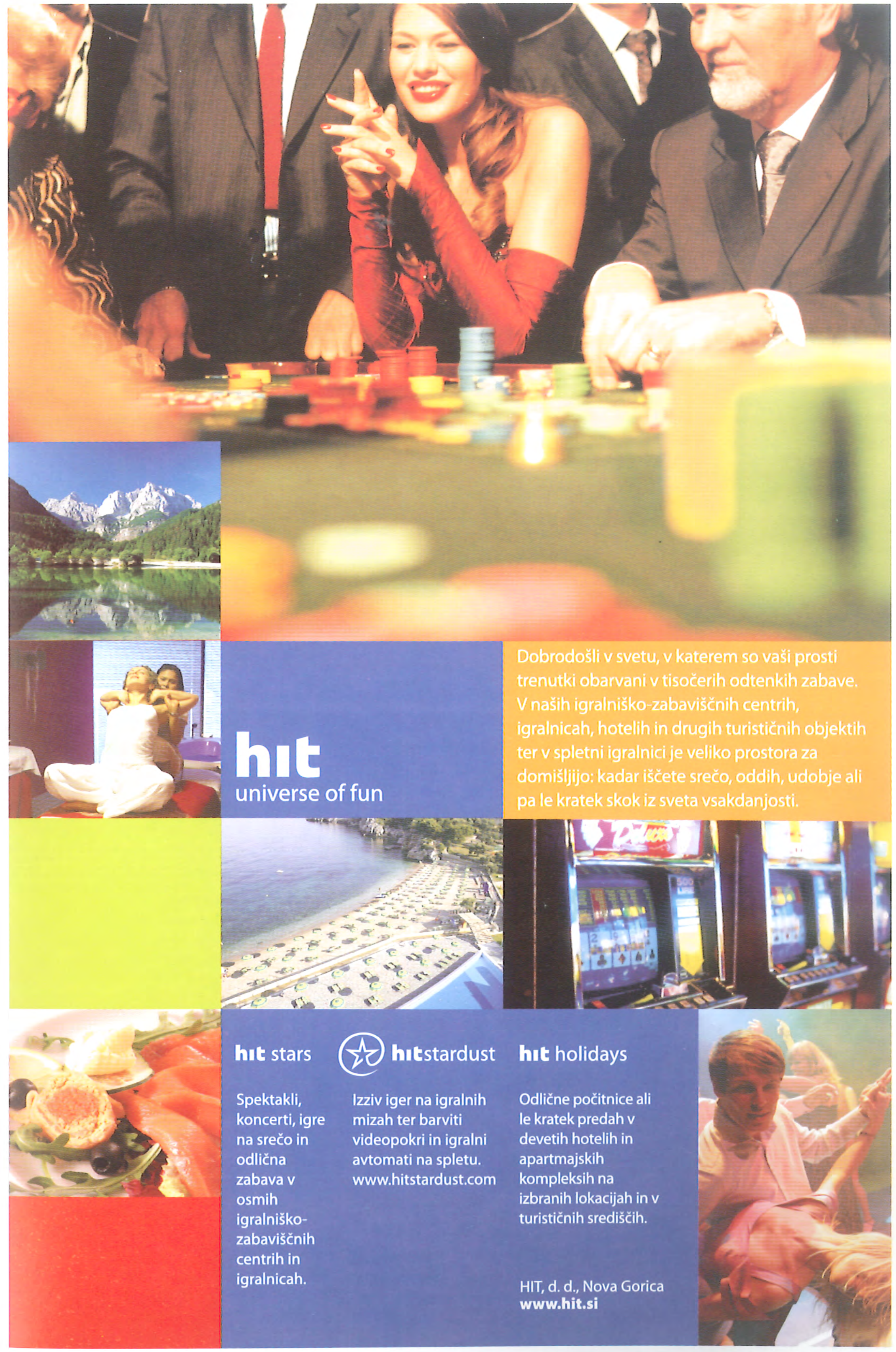
Austrian Science and Research Liaison Office Ljubljana

director:	**dr. Miroslav Friedrich Polzer** (economist and environmental engineer)
office employees:	Gorazd Weiss, mag. Marija Malle
address:	Avstrijski znanstveni institut v Ljubljani Dunjaska 104; SI-1000 Ljubljana tel.: +386 1 5684168; fax: +386 1 5684169 e.mail: polzer@zsi.at homepage: www.aso.zsi.at

Head office/Coordination in Vienna:

Centre for Social Innovation – Zentrum für Soziale Innovation

Linke Wienzeile 246; A-1150 Vienna
tel. +43-1-4950442; fax: +43-1-495044240
e-mail: institut@zsi.at


Dobrodošli v svetu, v katerem so vaši prosti trenutki obarvani v tisočerih odtenkih zabave. V naših igralniško-zabaviščnih centrih, igralnicah, hotelih in drugih turističnih objektih ter v spletni igralnici je veliko prostora za domišljijo: kadar iščete srečo, oddih, udobje ali pa le kratek skok iz sveta vsakdanjosti.
hit
universe of fun
hit stars
Spektakli, koncerti, igre na srečo in odlična zabava v osmih igralniško-zabaviščnih centrih in igralnicah.
hitstardust
Izziv iger na igralnih mizah ter barviti videopokri in igralni avtomati na spletu.
www.hitstardust.com
hit holidays
Odlične počitnice ali le kratek predah v devetih hotelih in apartmajskih kompleksih na izbranih lokacijah in v turističnih središčih.
HIT, d. d., Nova Gorica
www.hit.si

Proceedings SOR

Rupnik V. and L. Bogataj (Editors): The 1st
Research, SOR'93. Proceedings. Ljubljana:
Section for Operational Research, 1993, 3

Rupnik V. and M. Bogataj (Editors): The 2
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.

ISBN 961616520-8
9 789616 165204